

Armchair travellers

Louis-José Lestocart

Fin années 1970, divers systèmes d'imageries interactives avec mini-ordinateurs, télévision numérique couleur et vidéodisques sont mis en œuvre à l'Architecture Machine Group (Arch-Mac, pré-Media Lab) du MIT (Cambridge, Massachussets) par Nicholas Negroponte et d'autres informaticiens placés sous l'égide de Richard A. Bolt : Spatial Data Management System, Put-That-There System, World of Windows, Aspen Movie Map. Ces environnements à processus d'affichage (cartes, menus détaillés, indications de direction, réception de programmes de télévision ou de films) où sens de l'espace compte dans l'organisation et le traitement des données, ancrent un langage de Réalité augmentée proche des canons actuels de communication (tablette, téléphones mobiles, Internet). Inspirés de l'ars memoriae (art de la mémoire) ancien, ces dispositifs hypermédiats de gestion visuelle de données, financés par le Cybernetics Technology Office du DARPA (Defense Advanced Research Projects Agency), montrent la richesse cognitive potentielle d'un ordinateur via capacité d'apprentissage et interactions diverses. Assis sur un fauteuil de commande face à un écran, non plus spectateur d'images, on agit sur elles, les change, les modèle à sa guise. Armchair traveller (voyageur en fauteuil) ; on « entre » proprement en une réalité augmentée.

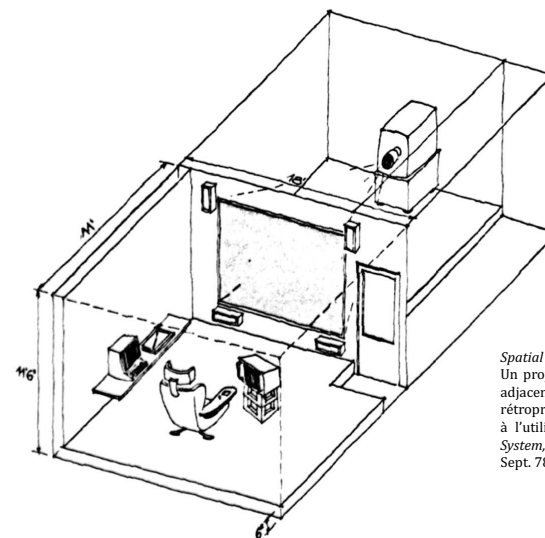


Le Spatial Data Management System (SDMS, 1976-79)

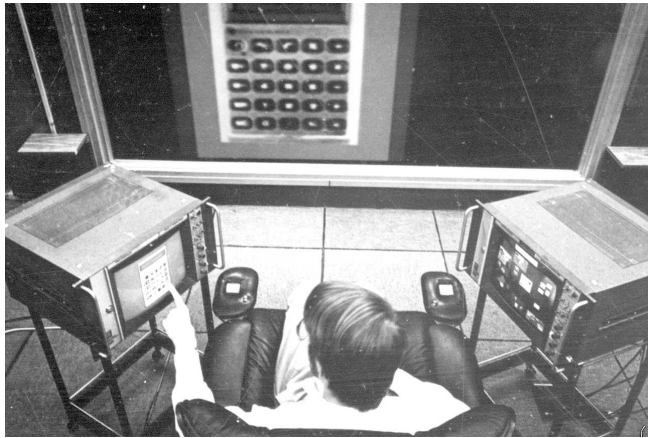
Dès le printemps 1976, l'Arch-Mach bâtit un système multimédia interactif de gestion de données graphiques avec affichage interactif : le Spatial Data Management System. Cet outil de coopération multi-sensorielle homme/ordinateur (Homme-Machine Interaction, HMI), simule, via perception et mémoire, l'organisation d'informations dans l'espace naturel. Les codages de celles-ci (sons, images, mots, nombres, données) y sont numériques, et les supports de communication (diffusion hertzienne, réseau câblé, disques vidéos et services télématiques) liés entre eux¹. Ce « méta-média » numérico-optico-audiovisuel, mêlant pouvoir d'absorption du cinéma et traitement informatique, s'appelle « effet Simonide », du nom de Simonide de Céos (VI^{ème}-V^{ème} siècles av. J-C), père de la mnémotechnique. Procédé où l'orateur mémorise son discours via des images/signes (*imagines*) situées en des architectures imaginaires – des palais (ou temples) de la mémoire avec pièces distinctes et ornements architectoniques – , « lieux » (*loci*) peuplés d'objets codés identifiables. Chaque

période du discours se lie à ces topiques qui, « activés » par l'esprit, se font *imagines agentes* (images agissantes). Leur montage génère un espace mental organisé, enregistré, puis restitué.

Ici, plus qu'un palais, ce qui est à portée : la bibliothèque et le bureau. Trouver un livre requiert à la fois mémoire spatiale et motrice, et perception (couleur, texture, forme). Livre sur l'étagère, saisi, ouvert, pages tournées : l'activité physique, également mentale, qui ressort du comportement et de l'apprentissage, est principe d'organisation et d'orientation spatiales. De même, connaître l'agencement d'objets sur un bureau, et leur fonction, le fait de les prendre, de s'en servir, répond à un codage mental et moteur. Dans le SDMS, cela se traduit par un *Spatial Data Objects* ou *Data-Land* : un système virtuel fait d'images digitalisées et colorées, de films, de sons et d'icônes affichées sur un écran. Ce champ de données graphiques (*Graphical Data Space*), riches en images et en sons, sorte de « tableau noir » virtuel spatialement défini, compte diverses informations à organiser et à traiter.



Spatial Data Management System.
Un projecteur de télévision dans la pièce adjacente à la Media Room opère une rétroprojection sur le grand écran face à l'utilisateur. *Spatial Data Management System, Final technical report, 1 Oct. 76-30 Sept. 78, 12.*



Spatial Data Management System, Final technical report, 1 Oct. 76-30 Sept 78, 47.

Les « images agissantes » potentielles de l'*ars memoriae*, issues, là, de données numérisées et stockées sur des vidéodisques optiques MCA DiscoVision², se retrouvent au sein d'un monde implicite ordonné tel un bureau (*desktop metaphor*³).

Un monde d'« intensité d'informations⁴ » qui engage la cognition pour choisir des objets, agir comme en un environnement naturel. L'immersion, la navigation, l'interaction se font par capteurs de position, outils de poursuite du regard (*eye tracking*), systèmes de reconnaissance de la parole, etc.⁵

Dans la Media Room du MIT, un ensemble de mini-ordinateurs gère écrans de visualisation – moniteurs TV couleur – et outils de commande vocale et de reconnaissance des gestes. Le DataLand s'affiche par rétroprojection sur une plaque de verre dépoli servant d'écran mural. En face et au centre, à 2,50 m de l'écran, un fauteuil instrumenté. Chaque bras a un joystick sensible à la pression et à la direction, et deux petites manettes carrées, tactiles. Le joystick de droite, poussé ou tiré, guide le mouvement horizontal et vertical, ainsi que le mouvement latéral, lors du voyage dans le DataLand ; le gauche sert à zoomer. À cela s'ajoute une table traçante avec stylo optique sélectionnant diverses fonctions dans un menu, table ayant un petit microphone sur son bord supérieur, recueillant des messages vocaux, traités en simultané par système de reconnaissance de

parole. De chaque côté du fauteuil, un écran TV couleur couvert d'une pellicule plastique transparente, tactile, livre au système les coordonnées du point touché et met soudain dans le monde interactif⁶. L'écran de droite montre le DataLand en 2D (World View). Fenêtre visuelle (ou « carte ») sur les données disponibles sous forme d'icônes (modélisations informatiques d'objets réels) identifiables : visages, téléviseurs, couvertures de livres, fac simile d'une lettre du DARPA, animaux, photo satellite miniature LANDSAT d'une partie de la Nouvelle-Angleterre, téléphone, agenda, calculatrice, etc. « Images sur les choses » (*memoria rerum*) de la mnémotechnique antique, se sentant presque observées. Rangées par genre et par associations sur des arrière-plans colorés, leur aspect leur donne une identité⁷. Outre le joystick, la désignation agit sur elles. Sur l'écran TV de droite, un carré en surbrillance *you-are-here* touché du doigt va aussitôt dans le DataLand et signale la zone requise qui, par rétroaction visuelle, surgit, immense, en HD sur le grand écran. « Téléportation » qui avertit qu'une partie du paysage informationnel s'« ouvre », prête à servir sur le grand écran et les deux petits écrans des moniteurs. L'« Ouverture » est l'interaction.

Interface, avatar et téléprésence

L'interface, d'une identité quasi insaisissable, sonde la limite entre réel, vivant et machinique. Activé, l'avatar de l'objet sur l'écran a la puis-

sance de ce qu'il symbolise. Appuyer dessus, lui parler, le désigner est tel un acte performatif : un dire c'est faire (John Austin)⁸. Lors de la désignation d'un emplacement dans le DataLand du SDMS, une interface, faite de deux petits cubes, effectue une mesure par champ magnétique. Ces cubes ROPAMS (Remote Object Position Attitude Measurement) de la société Polhemus Navigation Science à Junction (Vermont), montés au poignet, sont des capteurs de distance et de direction. Chacun a trois bobines électriques parcourues par des courants créant un champ magnétique dipolaire toroïdal autour d'eux. Par la juxtaposition des deux champs, l'ordinateur définit l'orientation dans l'espace du cube transmetteur sur le poignet par rapport au cube receveur placé sur un pied du fauteuil, et la distance qui les sépare ; il sait ce qui est désigné⁹.

Les visées du SDMS semblent banales face aux menus d'Internet, des tablettes et des portables de maintenant, mais ce n'est pas le cas alors. La couverture du livre virtuel paraît en grand sur l'écran mural. Sur le moniteur de gauche, la table des matières. Quand le chapitre choisi paraît sur le grand écran, on a la sensation d'un livre réel : pages tournées (bruit en feedback) via les manettes tactiles du siège, voix commentant le texte et les illustrations, lecture... Sur toute la surface de l'écran mural, la TV virtuelle diffuse son programme préenregistré sur vidéodisque. La visualisation se contrôle sur un des moniteurs où l'image digitale incrustée d'une horloge indique le temps de diffusion. Des curseurs virtuels l'accélèrent ou la ralentissent, et l'on choisit la langue désirée. Pour le téléphone, la communication se fait par appui sur ses touches virtuelles, sur le petit écran du moniteur. Le destinataire, trouvé en une liste dans un frame sur l'image du téléphone, s'obtient par la voix. Enfoncer les touches virtuelles de la calculatrice ou la commander vocalement pour des opérations suffit pour que celle-ci, élargie sur grand écran, inscrive le résultat¹⁰. Dire « Menez-moi à la calculatrice », exauce aussitôt le souhait. Il y a aussi l'*eye tracking* qui mesure le lieu de fixation du regard par repérage du déplacement des yeux (oculomotricité). Si ce procédé sert plus à l'ori-

gine aux laboratoires d'études comportementales et psychologiques, son irruption dans la RV forge des voies rapides d'interaction¹¹. Un des facteurs premiers de la communication étant la direction du regard, en suivant les mouvements oculaires, l'ordinateur peut saisir l'intention poursuivie.

Put-That-There System (1980), World-of-Windows (1981)

Ces deux installations sont à la fois des prototypes et des déclinaisons du SDMS. Le *Put-That-There System* de Christopher Schmandt et Eric Hulthen, supervisé par Bolt, s'attache plus précisément aux interactions gestes/paroles et au dialogue homme/machine en environnement immersif. L'union d'un écran vidéo mural, d'un périphérique de saisie gestuelle (par pointage du doigt) et d'un système de commande vocale CSRS (*Connected Speech Recognition System*) de la NEC (Nippon Electric Company), crée un outil de commande verbo-gestuelle¹². Après des analyses syntaxiques et sémantiques, et des recherches sur l'intonation et l'articulation, il devient possible de donner des instructions à un ordinateur (doté d'un dictionnaire de 120 mots dans sa mémoire active), selon une formule de cinq mots maximum, qu'il déchiffre par reconnaissance de parole en 300 millisecondes.

Le système compte des outils logiciels et des matériels informatiques divers décrivant de façon réaliste cette interaction sur un vaste affichage graphique virtuel. Pour que l'ordinateur reçoive, comprenne et réponde, des capteurs magnétiques de position de main – deux petits cubes ROPAMS Polhemus spatio-sensitifs (un au poignet, l'autre près du siège de l'utilisateur) –, et un casque-micro. Assis face à un vaste écran vide bleu, « océan sans rivage »¹³ via une reconnaissance vocale et gestuelle simultanées, on crée, déplace, modifie forme et couleur, copie, nomme, annote (ou détruit), des carrés, des cercles, des triangles, grands, moyens, petits et en couleur (rouge, jaune, orange, vert ou bleu). « Put a blue triangle (pause) there » : on indique, à l'aide du capteur, un point sur l'écran, (un curseur blanc y apparaît). Une « voix » réagit : « Where ? » Si l'on répond « There », la machine obéit,

matérialise le triangle. Ensuite, on produit de la même façon un carré vert et, montrant le triangle bleu, on demande à le placer à droite du carré, en bougeant le doigt au lieu souhaité. Le geste s'accompagne du pronom démonstratif « ça » (*that*). Tel un petit enfant signalant un objet par « ça », on dit : « *Put that... there.* ». Seuls les pronoms démonstratifs (ceci, cela) et les prépositions (au dessus de, à droite de...) jouent pour l'ordinateur dont le vocabulaire s'enrichit peu à peu.

De multiples « fenêtres » animées ou fixes, sur grand écran à présent : *World-of-Windows* (Monde de fenêtres). Brièveté, fragmentation, diversité d'informations simultanées sans lien logique, caractérisent ce *Gaze-Orchestrated Dynamic Windows* chaotique (actualités d'agences de presse diffusées par satellite en temps réel, conversations, vidéoconférences, événements militaires ou sportifs en direct, graphiques, animations vidéo, diapositives, textes, images fixes s'animant en diaporama). Cette surcharge cognitive, perception comme entendement ne peut l'englober. La sélection d'une partie correspond au « filtre » d'attention sélective dont se sert le cerveau face aux stimuli sensoriels excédant sa capacité de traitement¹⁴. Eye tracking, commandes vocales et manuelles, capteurs ROPAMS ont ici ce rôle, interagissant sur les fenêtres, contrôlant les images et orchestrant les sons. L'exploitation des données devient « feuilletage » de flots structurés d'informations¹⁵.

The Aspen Movie Map (1978-79)

Cette carte-film interactive 2D et 3D (première de l'Histoire) réalisée par Andrew Lippman et d'autres chercheurs tels Michael Naimark (qui sera ensuite artiste multi-média) et Scott Fisher, permet l'immersion dans la ville d'Aspen (Colorado)¹⁶. Automne 1978, quatre caméras 16 mm, à l'arrière d'un camion, filment ses rues par déclenchement automatique, tous les trois mètres, à 360° (parfois avec une lentille anamorphique) et l'on prend, sous les mêmes angles, des photos historiques et l'intérieur de bâtiments (hôtels, restaurants, magasins). Ces séquences et d'autres (interviews, cinéma-vérité, sons synthétisés ou enregistrés, animation

générée par ordinateur), montées en simulation de promenade, sont transférées et digitalisées sur vidéodisque avec base de données 3D complète de la ville. On mixe un menu en surimpression graphique couleur et légende (flèches) pour s'orienter. Avec 30 images/s (taux de télévision standard) et une vitesse de 200 km/h, on doit ralentir à 10 images/s, soit 110 km/h.

Au fauteuil, face au grand écran de projection 3D, on ajoute un joystick de vitesse et de direction et, de part et d'autre, deux TV couleur avec écran tactile et pictogrammes interactifs¹⁷. À droite, l'affichage graphique de la carte de la ville. À gauche, sa vue aérienne en 2D. Pointer le doigt sur l'écran met aussitôt au lieu choisi. Débute le voyage en temps réel. Le dialogue continu avec la carte le réorganise sans cesse (liberté de choix de la vitesse et de points de vue). Tourner aux carrefours, pénétrer dans les maisons « vides » en cette sorte de Panavision mobile illimitée, est proche d'un « regard de Dieu ». Mais ce regard a ses limites, car il instaure un champ visuel restreint où le bord des choses reste caché, contrairement au hors-champ cinématographique, lui (presque) toujours signifiant. Des dispositifs plus innovants et performants veulent offrir aujourd'hui une totale implication corporelle et mentale, comme le montre des outils déjà ou bientôt démocratisés tels l'Oculus Rift, le HTC Vive ou le fleur de la mix reality, l'HoloLens (technologie holographique), succédant aux CAVEs des années 1990 toujours à l'œuvre, et même des combinaisons possibles entre ces divers éléments.

Une autre limite dans ce genre de dispositif est peut-être celle de la sensation. Cette dernière, reconstruite artificiellement (« *je suis en train d'être dans un univers virtuel où je peux agir, déclencher tel ou tel phénomène, ce qui m'étonne ou me ravit* »), ne restitue ni une vraie sensation, ni la mémoire et le plaisir lié à elles deux. La redécouverte d'un livre qu'on a lu étant enfant ou plus jeune, ou encore de tel objet à l'histoire particulière dans notre vécu, est, lui, porteur de cette sensation que Marcel Proust a désigné dans *À la Recherche*

du temps perdu (1913-1927) par le terme de « mémoire involontaire » : c'est-à-dire une mémoire qui opère une conjonction, une identité de la sensation présente avec une sensation passée. « *Telle sensation, perçue par hasard, me rappelle malgré moi tel évènement oublié, produisant en moi un autre évènement puisant, à la fois cognitif et affectif* ». En somme, une sorte de mémoire affective du corps et de l'esprit, à son tour porteuse de tant de riches sensations mêlées indissolublement à elle. Car les choses ne sont pas qu'utilitaires, au service de nos habitudes, dédiées à notre seul plaisir de consommateur, ou juste conçues pour nos loisirs. Elles engendrent des impressions qui concourent au développement de notre organisme (via le cerveau), et à notre apprentissage lors la reconnaissance de notre environnement. Et cela ne se passe pas de façon factuelle et formelle, mais au cours de multiples essais/erreurs. Ces impressions nous font. C'est ainsi que les nourrissons se développent et que les jeunes enfants apprennent de leur environnement, de manière aléatoire, et que les adultes agissent, encore et toujours. Elles, qui nous bâtissent et nous fondent, semblent trop en désaccord avec la pensée répandue vis-à-vis du virtuel (augmentation des sens, des sensations). Car aucun de ces « objets » que nous manipulons dans cette « culture du virtuel » ne contient en lui la sensation de la lumière qu'il y avait ce jour-là, ou du temps qu'il faisait lors de la découverte d'un « objet » particulier – livre ou paysage –, ni du sentiment que l'on avait alors, de joie, de bonheur ou de tristesse mélancolique. Toutes ces impressions pouvant nous être d'un coup rendues, selon la mémoire involontaire proustienne, aucun des artefacts virtuels, pourtant bâtis sur la mémoire, n'en renferme a priori.

Recherches Xerox PARC (Palo Alto Research Center) développant l'Alto, ordinateur de la taille d'un grand bureau avec une interface graphique.

⁴ M. R. HEIM, *Virtual Realism*, Oxford, Oxford University Press, 1998.

⁵ R. A. BOLT, *Spatial Data Management System, Final technical report*, 1 Oct. 76-30 Sept. 78. En ligne : <https://apps.dtic.mil/dtic/tr/fulltext/u2/a070243.pdf>

⁶ Huit hauts parleurs diffusent du son octaphonique.

⁷ R. A. BOLT, *The Human Interface : When People and Computers Meet*, New York, John Wiley & Sons, 1984.

⁸ Dire « je vous marie » ou « je vous baptise », rend le mariage ou le baptême effectif par son énonciation.

⁹ R. A. BOLT, « Les images interactives », *La Recherche, Spécial Images* 14(144), mai 1983.

¹⁰ R. A. BOLT, *Spatial Data Management System, Final technical report*, op. cit.

¹¹ Des lunettes conçues par le Denver Research Institute (Colorado) sont dotées d'une source de lumière (diode) infrarouge et d'une petite caméra TV pour la détecter. La caméra localise la position de l'œil par rapport au verre et un cube de captation spatiale sur les lunettes donne leur position dans l'espace. R. A. BOLT, « Eyes at the interface », *Proceedings of the 1982 conference on Human factors in computing systems*, mars 1982.

¹² R. A. BOLT, « 'Put-that-there' : Voice and gesture at the graphics interface », in *SIGGRAPH '80 Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, Seattle, Washington, USA — July 14-18, ACM SIGGRAPH Computer Graphics 14(3), 1980, 262-270.

¹³ Plus tard une carte de la région des Caraïbes.

¹⁴ Phénomène de « limitation » ou « canal à capacité limitée » observé dans la perception par le psychologue Donald Broadbent (hypothèse du goulot d'étranglement central ou de l'entonnoir).

¹⁵ R. A. BOLT, « Gaze-Orchestrated Dynamic Windows », *SIGGRAPH '81 Proceedings of the 8th annual conference on Computer graphics and interactive techniques*, Dallas, Texas, USA — August 03-07, 1981, ACM SIGGRAPH Computer Graphics 15(3), 1981, 109-119. D'un regard, une fenêtre s'agrandit, diffuse sa bande-son, les autres restant en fond sonore.

¹⁶ A. LIPPMAN, « Movie-maps : An application of the optical videodisc to computer graphics », *SIGGRAPH '80 Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, op. cit.

¹⁷ Au dispositif s'ajoutent deux lecteurs de disques.

¹⁸ G. HENROT SOSTERO, « La réminiscence comme évènement : saillance, argentivité, transformation », in E. BALLARDINI, R. PEDERZOLI, S. REBOUL-TOURÉ et G. TREGULIER-FELTEN (eds.), *mediAzioni* 15, « Les facettes de l'évènement : des formes aux signes », 2013. Texte en ligne. http://www.sitlec.unibo.it/images/stories/PDF_folder/document-pdf/15-2013/henrot.pdf

¹ S. BRAND, *The Media Lab : Inventing the Future at MIT*, New York, Viking-Penguin, 1987.

² 1977, MCA Corporation envoie à des chercheurs dont ceux de l'Arch Mach, 25 premiers prototypes de disques lasers DiscoVision contenant 54000 images par face, soit une demi-heure de film.

³ R. A. BOLT, « Touch Sensitive Displays », *DARPA Report, MIT Architecture Machine Group*, March 1977. Le SDMS est en opposition avec le GUI (Graphical User Interface, 1970) d'Alan Kay et d'un groupe de Stanford au Centre de