

# Espaces conceptuels pour l'interaction multimédia et multimodale

**Laurence Nigay, Joëlle Coutaz**

Laboratoire CLIPS-IMAG  
B.P. 53  
38 041 Grenoble Cedex 9

---

*RESUME. L'innovation en interaction multimédia et multimodale tient à la disponibilité et à la combinaison de mécanismes sophistiqués tels que la parole, la vision par ordinateur, la génération automatique de textes et d'images réalistes, la vidéo. Ces avancées brutales se manifestent aussi par une prolifération de termes employés le plus souvent sans rigueur, en général les uns pour les autres. Cet article fait le point sur la terminologie et sur les référentiels applicables à la conception d'interfaces multimédia et multimodales. Ces espaces conceptuels visent à forcer les "bonnes questions" mais adoptent des perspectives différentes : les uns centrés sur l'utilisateur, d'autres sur la technique, les uns raisonnant à bas niveau d'abstraction, les autres à haut niveau, les uns se limitant aux entrées des échanges, les autres aux sorties. Ces choix sont validés par une finalité précise. Après avoir introduit les référentiels les plus significatifs, nous présentons notre propre contribution motivée, pour l'essentiel, par les problèmes de la conception logicielle.*

*ABSTRACT. The combined use of multiple interaction techniques such as speech and computer vision as well as video and the automatic generation of text and realistic images, has opened the way to a new world of experience: that of multimedia and multimodal interaction. The unexpected and sudden success of this new area of research has resulted in a profusion of confusing terms. This article aims at clarifying the relationships between the most significant taxonomic spaces that support reasoning about multimedia and multimodal interaction. All of these conceptual spaces trigger the "right design questions" but they do so from different perspectives: some models are motivated by user-centered concerns, others are system driven ; some of them enforce reasoning at a low level of abstraction, others at a high level ; some are concerned with input others with output. Following the description of the relevant conceptual spaces, we describe our own contribution motivated by the software engineering aspects of multimedia and multimodal interaction.*

*MOTS-CLES: Modalité, média, multimodalité, multimédia, espace de conception.*

*KEYWORDS: Modality, media, multimodality, multimedia, design space.*

---

## 1. Introduction

L'exploitation optimisée des ressources matérielles s'est imposée d'emblée comme un facteur central en conception des systèmes informatiques. La baisse du coût des composants électroniques, leur miniaturisation et l'amélioration des performances ont tempéré depuis peu cette attitude obsessionnelle pour enfin rendre à l'utilisateur la place qui lui revient. Aujourd'hui, l'efficacité recherchée n'est plus celle de la technologie prise isolément mais celle du couple "machine-humain". Il ne s'agit plus seulement d'exploiter au mieux les ressources techniques, mais de développer une technologie qui soit conforme aux attributs de la communication humaine. Et l'être humain voit, entend, parle, gesticule, manipule [COU 96].

Les recherches en reconnaissance et synthèse de la parole, la vision par ordinateur et la synthèse d'image, les nouveaux dispositifs d'interaction comme le visiocasque et les transducteurs gestuels rétroactifs visent l'équivalent technique de la perception et de l'action humaines. Mais dès l'origine, ces études sont menées de manière cloisonnée faisant l'objet de disciplines distinctes. En 1983, N. Negroponte fonde le MediaLab au MIT avec l'objectif d'étudier l'intégration de ces nouvelles technologies. En France, la prise de conscience a lieu plus tard avec la création en 1990 d'un groupe d'étude au sein du GDR-PRC Communication Homme-Machine. Le concept d'interaction multimodale est né. Parallèlement, le multimédia gagne le milieu industriel.

Nous assistons aujourd'hui à une explosion de la technologie communicante aiguillonnée par les autoroutes de l'information [THE 94], vecteur incontournable de l'ubiquité instantanée et universelle de l'information. Cette avancée brutale se manifeste notamment par une prolifération de termes employés le plus souvent sans rigueur, en général les uns pour les autres. Il en va ainsi pour le multimodal et le multimédia.

Cet article vise à faire le point sur la terminologie et sur les cadres d'étude liés au multimédia et à la multimodalité. Après une revue terminologique sur les notions de média et de modalité, nous présentons les cadres taxonomiques les plus pertinents, les uns centrés sur les techniques d'entrée, les autres sur les systèmes de représentation en sortie. Un bilan analytique nous conduit à présenter notre propre contribution : l'espace MSM.

## 2. Terminologie

Parce que les acteurs de la communication, l'homme et la machine, posent des problèmes différents, les notions de multimédia et de multimodalité sont appréhendées sous différentes perspectives, les unes centrées sur l'individu, les autres sur la technologie. L'ambiguïté des expressions repose pour l'essentiel sur cette dualité.

Toutefois, la diversité des points de vue ne justifie que partiellement l'absence de consensus. En vérité, et notre analyse le démontre, les désaccords qui se manifestent au sein d'une même perspective traduisent soit une mauvaise maîtrise des concepts soit des finalités distinctes. Notre première contribution est de choisir et proposer un ensemble de termes qui couvrent raisonnablement le domaine de réflexion pour une

finalité donnée. Avant de présenter les taxonomies relevées dans la littérature, il convient de considérer les définitions du sens commun.

### 2.1. Terminologie et le sens commun

Sur le plan linguistique, les termes “multimédia” et “multimodalité” ont en commun le préfixe “multi” qui traduit le nombre. Les radicaux “média” et “modalité” les distinguent. Si l’on se réfère aux dictionnaires de langue française (Hachette 1991, Larousse 1986) :

- un *média* est un “procédé technique permettant la distribution, la diffusion ou la communication des œuvres de l’esprit écrites, sonores, ou visuelles (depuis la presse imprimée jusqu’à l’ordinateur ....)”. Appliquée à notre domaine, cette définition couvre le système informatique dans son tout. Même si la conception d’un tel système nécessite une analyse plus fine de la notion de média, il convient de relever dans cette définition deux notions clés : celle de dispositif (“procédé technique”) en relation étroite avec les qualités sensorielles et cognitives humaines (“œuvres ... écrites, sonores, ou visuelles”).

- la *modalité* a plusieurs acceptions selon le domaine de référence. Nous retenons deux domaines en relation directe avec la communication homme-machine : la psychologie et la linguistique. En psychologie, une modalité désigne l’une des “grandes catégories de qualités sensorielles (vision, audition, olfaction, etc.)”. En linguistique, la modalité permet “d’exprimer les diverses manières dont le contenu est envisagé” par le locuteur. Ces manières se manifestent par des marqueurs linguistiques (mode grammatical, adverbe, interjection) auxquels il faut ajouter le rôle important de l’intonation de la voix ou l’attitude corporelle. Ainsi pour un même contenu (neige-tomber), les expressions “il faut que la neige tombe”, “Ah, si la neige tombait!”, “la neige tombera”, traduisent, par leurs modalités, des engagements différents du locuteur vis-à-vis du contenu informationnel. Nous constatons que ces deux définitions, les catégories sensorielles et les marqueurs de statut linguistiques, sont deux ingrédients valides de la communication homme-machine. Il convient de les distinguer sans confusion ou de les encapsuler en une définition commune.

Corrélé à *modalité*, nous trouvons le *mode*. En philosophie, le mode est la manière d’être d’une substance (un objet, un être, une entité) et la modalité est la propriété que possède la substance d’avoir des modes. En limitant la discussion à l’action, l’un des constituants essentiels de la communication homme-machine, le mode est la manière dont se fait l’action et la modalité en désigne la forme. Notons que la manière de faire influence la forme et que la forme permet de déterminer la manière ! Il n’est donc pas surprenant que les termes mode et modalité soient l’un et l’autre employés pour qualifier une action. Enfin, l’adjectif modal relatif à la fois aux substantifs mode et modalité souligne l’ambiguïté du terme multimodal : lorsque nous caractérisons une interface de multimodale, s’agit-il de multi-*modalité* ou de multi-*mode* ?

Ayant rappelé les définitions de référence académique, nous sommes en mesure de les infléchir au cas de la communication homme-machine.

## 2.2. Terminologie et communication homme-machine

Une revue de la littérature nous a permis de relever trois groupes de termes directeurs : média, modalité, multimédia et multimodal. Dans chacun de ces groupes, les auteurs se démarquent par le point de vue adopté (utilisateur versus technologie) et le niveau d'abstraction. Lorsque c'est possible, nous explicitons ces différences de perspective puis nous prenons position.

### 2.2.1. Média

La plupart des auteurs s'accorde à penser qu'un *média* est un support technique de l'information. Exprimée ainsi, cette définition est cohérente avec l'acception du sens usuel. Mais chaque auteur affine ce pivot commun de précisions personnelles qui montrent la diversité des points de vue ou des finalités. Nous avons relevé trois perspectives essentielles : point de vue purement technique, point de vue purement centré sur le sujet humain, et un modèle hybride qui concilie technologie et facultés humaines.

Pour certains, un média est un dispositif physique capteur ou effecteur d'un système informatique. Une souris est un capteur, l'écran est un effecteur. Ainsi défini, un média est une spécialisation de la classe "support technique". Mais pour Blattner, un média n'a pas nécessairement une incarnation physique [BLA 90]. En tant que véhicule d'information, il peut se voir réalisé par du logiciel. Dans ces conditions, un message électronique est un média. On observe ici une distinction qui tient à la différence des niveaux d'abstraction. Les uns s'en tiennent au support matériel, les autres voient dans le média toute forme logique, mais purement technique, capable de véhiculer de l'information. Nous verrons au paragraphe 4.2. comment la théorie des dispositifs de Mackinlay, Card et Robertson concilie ces deux points de vue. Si elles diffèrent par le degré de conceptualisation, toutes ces définitions relèvent du même parti : celui de la technologie.

D'autres auteurs occultent la composante matérielle et technique du média pour y voir un lien prédominant avec les capacités perceptuelles et/ou cognitives humaines :

- pour Bernsen, un média est un ensemble de qualités perceptives couplé au dispositif sensoriel humain nécessaire à la perception de ces qualités [BER 94]. Par exemple, les qualités visuelles et graphiques nécessitent la vision, les qualités du son, l'audition. Ce couplage définit un média. Le concept de média, au sens de Bernsen, peut s'assimiler à la notion de modalité (sensorielle) telle qu'on l'entend en psychologie.
- pour Alty, un média est un système représentationnel [ALT 91, FRO 91]. Un texte en langage naturel, un graphe sont des systèmes représentationnels qui font sens. Ils font donc appel à l'interprétation et à la compréhension. Arens dans son système de présentation automatique d'information, utilise également média dans ce sens [ARE 93].

Un troisième modèle de la notion de média concilie perception et support matériel. Dans ces conditions, un média est un dispositif qui a la capacité matérielle de véhiculer des informations d'une certaine "substance" organisée en une "forme" [BOU 92]. Chez Hjelmslev [HJE 47], la substance d'une expression désigne sa réalité physique observable en tant que matériau non analysé (par exemple, le son) tandis que la forme lui donne une structure (par exemple, une pièce musicale). "La

substance et la forme ... sont un peu à l'exemple du bloc de granit à l'état brut (= substance) dont le sculpteur tire une statue (= forme)" [BEN 95]. Le média est associé à substance et forme mais n'a pas de capacités d'interprétation. L'extraction du contenu sémantique, par exemple le message conceptuel d'une œuvre, échappe au média.

En résumé, la notion de média désigne un dispositif servant de support à l'information. Cette définition générale est affinée pour répondre à des finalités distinctes :

- finalité purement matérielle : média vu comme capteur ou effecteur d'un système informatique. Dans cette acception, un écran, une souris sont des médias. Nous les appelons communément *dispositifs d'entrée/sortie* ;
- finalité purement technique : média en tant que procédé physique et/ou logiciel utilisable comme véhicule d'information (cf. Blattner et Makinlay) ;
- finalité technico-humaine à deux niveaux d'abstraction : média comme couplage d'un dispositif physique aux qualités sensorielles humaines (cf. Bensen), et média comme système représentationnel (cf. Alty et Arens).

### 2.2.2. Modalité

Au sens large, une modalité est assimilée à une technique d'interaction, une manière d'interagir avec le système. Comme le concept de média, le terme recouvre différents niveaux d'abstraction : les expressions "modalité clavier", "modalité langue naturelle", ou encore "modalité parole", se côtoient fréquemment dans la littérature courante. En conception d'interface, il convient cependant d'observer davantage de rigueur.

Les concepteurs d'interfaces (et notamment les ergonomes et les psychologues) associent la modalité aux capacités de perception du sujet humain mais aussi à ses facultés de compréhension. Bensen, on l'a vu, utilise le concept de média pour traduire le niveau perceptuel. Dans sa théorie des modalités pures (se référer au paragraphe 5), Bensen voit dans le concept de modalité, un système représentationnel de l'information : un graphe, un texte, etc. (Notons que cette définition est équivalente à la notion de média chez Alty et Arens !)

Pour les concepteurs de logiciel, la modalité est liée au contenu et à la nature des informations que le système est capable de traiter [BLA 90, BOU 92]. Par exemple, Martin [MAR 95] définit une modalité comme un processus (informatique) d'analyse ou de synthèse défini sur des ensembles de données d'entrée et de sortie. Nous voyons dans cette définition une traduction informatique des concepts de contenu et de nature de l'information : le processus d'interprétation est utile à l'obtention d'un contenu (analyse) ou à la production d'un contenu (synthèse) tandis que la nature de l'information véhiculée se définit par le type des ensembles de données d'entrée et de sortie.

Le modèle "syndésique" de Duke vise, comme son nom l'indique, à articuler les deux perspectives système et utilisateur au moyen d'une théorie axiomatique [DUK 94]. Dans ce contexte, une modalité se réfère aux capacités sensorielles mais aussi aux dispositifs physiques ou logiques engagés dans l'interaction.

En résumé, une modalité est assimilée au sens large à une technique d'interaction. Comme pour le concept de média, une modalité recouvre différents niveaux d'abstraction et s'envisage selon plusieurs perspectives :

- approche purement utilisateur et à haut niveau d'abstraction (comme chez Bernsen),
- point de vue purement système et à haut niveau d'abstraction (Martin [MAR 95], Bourguet [BOU 92, Nigay [NIG 94]),
- ou encore une vision conjointe du système et de l'utilisateur recouvrant différents niveaux d'abstraction, du physique au sens, du perceptif au cognitif et vice versa (Cf. Duke).

### 2.2.3. *Multimédia et multimodal*

La diversité des définitions des termes média et modalité voire leur recouvrement, laisse entrevoir la difficulté d'identifier une définition consensuelle pour multimédia et multimodal. Nous retiendrons que les objectifs justifient les différences. A l'évidence, ces termes sont tirillés entre d'une part, les préoccupations des psychologues et des ergonomes qui visent à développer des théories explicatives ou prédictives sur l'utilisation des interfaces nouvelles, et d'autre part, les considérations des informaticiens qui visent à définir des modèles d'architecture logicielle et des outils pour la réalisation de telles interfaces. Nous tentons ici une présentation synthétique de la littérature puis nous précisons notre point de vue.

Pour les psychologues et les ergonomes, la multimodalité d'un système tient au fait que la machine sollicite les capacités multi-sensorielles et cognitives de l'utilisateur. Multimédia couvre la diversité des supports techniques de ce système.

Pour la plupart des informaticiens, multimodal est lié à la capacité du système à extraire du sens des informations de "forme" et de "substance" diversifiées tandis que multimédia dénote l'incapacité à le faire. Nous verrons avec l'espace MSM présenté au paragraphe 7, une définition plus précise de cette distinction.

Pour d'autres informaticiens, un système multimédia se caractérise en entrée, par l'utilisation simultanée de différents dispositifs physiques, par la gestion en parallèle des événements qui en découlent, et par la construction d'événements de haut niveau d'abstraction à partir d'événements de plus bas niveau. Le multimédia se caractérise en sortie, par différents canaux en parallèle (texte, vidéo etc.) [DUC 90]. Pour d'autres encore, tel Buxton [BUX 90], multimédia ou multimodal impliquent : plusieurs modalités sensorielles, plusieurs canaux appartenant à une ou plusieurs modalités (par exemple les deux oreilles sont deux canaux appartenant à la même modalité : l'ouïe), plusieurs tâches en parallèle voire plusieurs utilisateurs effectuant la même tâche en parallèle !

### 2.2.4. *Notre point de vue*

Toutes ces définitions sont valides, chacune s'inscrivant dans un système de réflexion cohérent. Il suffit que les termes soient situés dans leur contexte d'utilisation. Par exemple, en phase de conception, nous recommandons pour le concept de média la perspective technico-humaine qui associe le dispositif physique aux qualités sensori-motrices de l'utilisateur. En revanche, pour les phases de mise en œuvre, nous assimilons un média à un dispositif d'entrée/sortie (vision purement matérielle) assorti d'un type qui caractérise l'information brute acquise ou produite (par exemple, une chaîne de caractères, vision logique de la perspective matérielle). Ce type est la représentation informatique de l'association "substance-forme"

évoquée plus haut. Dans les deux cas, conception et mise en œuvre, nous considérons que le concept de média couvre un bas niveau d'abstraction seulement : le niveau physique matériel pour la technologie et le niveau perceptuel du côté humain. Dès qu'il y a interprétation pour extraire du sens, dès que l'on pénètre dans le représentationnel, nous parlons de modalité. Mais, à l'évidence, le matériel et le représentationnel sont intimement liés.

Par conséquent, nous considérons qu'il convient, tant en conception qu'en mise en œuvre, de distinguer deux niveaux d'abstraction pertinents et de les mettre en étroite relation tant du côté humain que du côté machine. Nous distinguons le niveau physique (ou matériel) qui fait appel à la perception et le niveau représentationnel qui fait appel à la cognition. Un ensemble de couples  $\langle p, r \rangle$  où  $p$  désigne un dispositif physique et  $r$ , un système représentationnel, est une façon de caractériser les échanges entre un système et son utilisateur.

Par exemple, dans nos travaux de recherche sur la mise en œuvre logicielle, nous avons choisi pour  $p$ , l'ensemble des *dispositifs d'entrée/sortie* d'une machine (écran, clavier, etc.) et pour  $r$ , la notion de langage d'interaction. Nous appelons *langage d'interaction* un système conventionnel structuré de signes qui assure une fonction de communication. De manière plus formelle, un langage d'interaction se définit par un vocabulaire d'éléments terminaux et une grammaire. Les éléments terminaux sont produits ou captés par les dispositifs d'entrée/sortie. Par exemple, dans le système MATIS [NIG 94], un utilisateur peut formuler des requêtes de renseignement sur des horaires d'avion en saisissant au clavier une phrase en langage naturel comme "Flights from Pittsburgh to Boston", ou bien énoncer oralement cette même phrase ou encore remplir un formulaire par manipulation directe. Les couples  $\langle$ clavier, langage naturel $\rangle$ ,  $\langle$ microphone, langage naturel $\rangle$  et  $\langle$ souris, formulaire graphique $\rangle$  décrivent ces trois possibilités.

Cette distinction entre les mondes physique et représentationnel permet de structurer l'espace de conception en deux espaces de préoccupations tant du point de vue conception (la perception et la cognition) que du point de vue de la mise en œuvre (traitements de bas niveau et traitements sémantiques). Nous avons exploité cette distinction pour l'analyse de systèmes existants et notamment pour définir une taxonomie [NIG 94] ; nous l'avons utilisée comme point de départ à la définition d'un modèle d'architecture [NIG 95], de même pour l'étude de relations entre techniques d'interaction [COU 95]. Une *technique d'interaction* se définit soit comme un dispositif physique d'entrée/sortie, soit comme un langage d'interaction, soit comme le couplage d'un dispositif et d'un langage. Le choix entre ces trois perspectives dépend des besoins de l'analyse. L'important est de choisir la perspective qui convient aux objectifs et de s'y tenir.

Dans nos études, nous assimilons le concept de modalité à la notion de technique d'interaction et laissons ouvert le choix du niveau d'abstraction (dispositif physique d'entrée/sortie, langage d'interaction, couplage des deux). Nous n'utilisons pas le terme "média" mais celui de dispositif physique d'entrée/sortie. Nous considérons que le multimodal, tout comme le multimédia, sollicite les multiples capacités sensori-motrices naturelles de l'homme. Mais un système multimodal a la faculté d'abstraire donc de comprendre ce qui est reçu ou émis via ses dispositifs d'entrée/sortie. Un système multimédia n'a pas ces capacités. Du point de vue technique, le multimodal et le multimédia se distinguent par la capacité à modéliser la sémantique des informations échangées avec l'utilisateur.

Nous venons de proposer une revue terminologique du sujet et avons précisé notre point de vue. Il convient maintenant de situer les termes évoqués dans leur espace taxonomique d'origine. Dans cet objectif, nous analysons les propositions de la littérature en adoptant une démarche que nous précisons maintenant.

### 3. Notre démarche d'analyse : interface d'entrée et interface de sortie

La plupart des espaces de conception dissocient l'interface d'entrée de l'interface de sortie. D'autres ne considèrent qu'une seule des deux dimensions. L'interface d'entrée est constituée des moyens de communication artificiels mis à la disposition de l'utilisateur pour modifier l'état du système. L'interface de sortie couvre les moyens de communication qui rendent observable l'état du système.

Pour Card [CAR 86], ces deux interfaces, qui permettent un échange bidirectionnel d'information, sont les ingrédients incontournables de l'interaction. L'information spécifiée par l'utilisateur traverse l'interface d'entrée en subissant des transformations pour aboutir à une représentation interne au système. En sens inverse, l'information interne traverse l'interface de sortie en subissant des transformations pour aboutir à une représentation observable. L'information perçue est alors traitée par l'utilisateur selon un processus similaire d'acquisition-traitement-action que Card, Newell et Moran ont modélisé schématiquement avec leur Modèle du Processeur Humain [CAR 86]. La distinction entre interface d'entrée et interface de sortie se retrouve dans la théorie de l'action de Norman [NOR 86]. Cette théorie, qui centre l'analyse sur l'utilisateur, met en évidence les efforts mentaux et physiques mis en jeu par un sujet en situation d'interaction avec un système informatique.

Cette distinction entre interface d'entrée et interface de sortie se conçoit à des fins pédagogiques ou analytiques. Le processus cyclique :

- 1) interface d'entrée : acquisition de données,
- 2) traitement interne : changement d'état,
- 3) interface de sortie : expression perceptible du changement d'état,
- 4) retour à l'étape 1,

est une vue réductrice du fonctionnement souhaitable d'un système. Les interfaces d'entrée et de sortie ne peuvent être conçues de manière indépendante car les données reçues par l'une peuvent influencer les résultats de l'autre avant même que la boucle ait été exécutée dans sa totalité. Le tracé des formes élastiques dans les éditeurs graphiques illustre ce cas. De même, plusieurs formes de coopération entre les modalités de l'interface d'entrée et de sortie peuvent être intéressantes à analyser [MAR 95]. On ne peut davantage se satisfaire d'un schéma pipe-line séquentiel : le principe du "retour d'information immédiat" prôné en ergonomie, nécessite souvent le parallélisme ou tout du moins l'entrelacement des traitements de l'interface d'entrée avec ceux de l'interface de sortie.

Le cycle ci-dessus a donc valeur pédagogique. Il ne peut servir de cadre de référence à l'étude de tous les phénomènes et notamment de l'interaction qui, précisément, est le fruit du couplage étroit entre action et réaction, entre entrée et sortie. Dans les sections qui suivent, nous présentons les espaces de conception selon leur axe de portée (entrée ou sortie, entrée et sortie) et le niveau d'abstraction envisagé (dispositifs physiques ou traitements de haut niveau). Au vu de nos remarques, nous pouvons d'emblée retenir le caractère réducteur des espaces de conception dont la portée est limitée à une seule direction (entrée ou sortie).

### 4. Les espaces de conception des interfaces d'entrée

Les premières tentatives de taxonomies concernent les dispositifs physiques d'entrée. Dans cette lignée, nous trouvons les espaces de Buxton [BUX 83], de Card, Mackinlay et Robertson [CAR 90, MAC 90] et de Foley [FOL 84] qui correspondent chacun à des niveaux d'abstraction ou d'affinement différents.

#### **4.1. La taxonomie des dispositifs physiques selon Buxton**

Selon Buxton, un *dispositif d'entrée* est un transducteur de propriétés physiques que l'on peut caractériser dans un espace à trois dimensions :

- le premier axe de l'espace de classification dénote les propriétés que le dispositif est capable de capter : la position, la pression, le mouvement ;
- le second axe correspond au nombre de dimensions captées pour chaque propriété. Par exemple, pour la souris, la position est déterminée par deux valeurs prises dans un système de coordonnées. Pour le manche à balai, la position est un triplet ;
- le troisième axe distingue le type direct ou indirect de capture. Pour un écran tactile, la mesure est directe alors que la souris implique un intermédiaire mécanique.

Les deux premiers axes de cette taxonomie expriment la capacité du dispositif à traduire les actions de l'utilisateur en une forme exploitable par le système informatique, c'est-à-dire un type de donnée. La richesse du dispositif est alors fonction du nombre de propriétés physiques captées et pour chaque propriété, le nombre de dimensions spatiales. Le type (direct ou indirect) de la capture, qui traduit le degré d'engagement physique de l'utilisateur, est une notion intéressante qu'il convient d'associer aux propriétés des interfaces à manipulation directe [HUT 86].

La taxonomie de Buxton ne concerne que les dispositifs physiques continus. En particulier, les dispositifs à fonctionnement discret comme les boutons crantés ne sont pas représentés. De plus, ne sont considérés que les dispositifs actionnables par le geste instrumental : le microphone et la caméra ne sont pas évoqués. Ces derniers constituent cependant des cas intéressants : une caméra peut servir à reconnaître un geste, à suivre les déplacements d'un utilisateur, à le reconnaître [CRO 96]. De même, le microphone peut numériser le signal sonore d'une phrase, fournir un mot ou une suite de mots, modéliser ou non la prosodie. La théorie de Mackinkay et de ses co-auteurs répond à ce problème de modélisation d'un dispositif à différents niveaux d'abstraction.

#### **4.2. L'espace de conception des dispositifs physiques de Mackinlay, Card et Robertson**

Mackinlay, Card et Robertson observent que l'assemblage de dispositifs élémentaires en unités de contrôle plus complexes est une dimension importante en conception. L'objectif est alors de définir une théorie qui systématise la connaissance et le raisonnement sur les dispositifs d'entrée.

Le domaine de cette théorie est l'ensemble des dispositifs d'entrée. Chaque *dispositif* est défini par un sextuplet  $\langle M, In, Out, R, S, W \rangle$ . Les opérateurs de la théorie expriment la possibilité d'assembler plusieurs dispositifs :

- **M** est un opérateur de manipulation ("Manipulation operator") appliqué par l'utilisateur. Il désigne une propriété physique captable par le dispositif (par exemple, une position).
- **In** est le domaine des valeurs d'entrée possibles de M. Par exemple, l'ensemble des réels compris entre 0 et 180 degrés pour un bouton rotatif.
- **Out** est le domaine des valeurs de sortie possibles du dispositif. Si le bouton comporte 5 positions (cas d'un bouton cranté), Out contient alors les 5 valeurs angulaires correspondantes.
- **R**, ou fonction de résolution ("Resolution function"), définit la correspondance entre les domaines In et Out. Dans le cas du bouton, cette fonction exprime les conditions de passage entre les angles du mouvement effectué par l'utilisateur et les crans (angles en sortie) imposés par le dispositif.
- **S** dénote l'état actuel du dispositif ("state"). Il comprend la valeur *In* de l'entrée prise dans In à l'instant présent t, la valeur *Out* prise dans Out et un état interne.
- **W** désigne le fonctionnement externe et interne du dispositif physique ("Works"). Il permet de modéliser des relations intéressantes entre In et Out en fonction de l'état. Par exemple, certains dispositifs captent des valeurs sans percevoir les valeurs intermédiaires entre deux valeurs captées successives. La tablette est dans ce cas : il y a des gestes humains que le dispositif ne peut percevoir d'où un risque de perte d'informations potentiellement utiles à la qualité de la communication.

Les opérateurs de composition, la fusion ("merge"), la connexion et la disposition ("layout"), permettent de modéliser des assemblages de dispositifs. Ces opérateurs sont complétés par un mécanisme d'encapsulation qui permet de définir des "dispositifs génériques" c'est-à-dire des classes. Formellement, la fusion est le produit cartésien de deux dispositifs sur les domaines d'entrée des dispositifs. Par exemple, la position d'une souris est le produit cartésien des deux dispositifs élémentaires captant respectivement l'ordonnée et l'abscisse. La connexion traduit la composition en cascade de dispositifs : la sortie de l'un sert d'entrée à d'autres. Par exemple, un bouton de ventilateur est connecté au dispositif qui agit sur la vitesse de rotation du ventilateur. La disposition traduit une composition spatiale comme les trois boutons de la souris sont liés spatialement aux dispositifs de position (x,y).

En synthèse, Mackinlay et ses coauteurs systématisent la description des dispositifs d'entrée au moyen d'une algèbre simplifiée. L'avantage est double : élargir le champ de conception à la prospection de nouveaux dispositifs et ouvrir la voie à l'évaluation du pouvoir d'expression des dispositifs en relation avec la tâche. Cette taxonomie comme celle de Buxton, prend le dispositif comme élément directeur. Toutes deux permettent, en considérant les caractéristiques d'un dispositif donné de s'interroger sur son adéquation à une tâche donnée. La taxonomie de Foley prend le point de vue inverse : étant donné une tâche, elle vise à proposer un ensemble de dispositifs idoines.

### 4.3. La taxonomie des tâches graphiques selon J. Foley

A l'origine de nombreux travaux comme le modèle d'interaction de Myers [MYE 90], la taxonomie de Foley adopte la tâche comme point d'ancrage. Pour classer un dispositif, Foley propose de le mettre en relation avec les tâches graphiques qu'il permet d'accomplir : sélection, positionnement, rotation, cheminement,

quantification et saisie de texte. Selon le dispositif, ces tâches peuvent être accomplies de manière directe ou indirecte. On retrouve ici l'idée d'engagement de l'utilisateur dans l'action.

Cette taxonomie dirigée par la tâche a l'avantage d'offrir des recommandations prescriptives sur le choix des dispositifs en fonction des tâches à accomplir. Toutefois l'espace des tâches envisagées concerne le domaine graphique uniquement et le modèle n'a pas le pouvoir d'expression de la théorie générative de Mackinlay.

#### **4.4. Notre point de vue**

Les opérateurs de composition de la théorie de Mackinlay assortie du mécanisme d'encapsulation offre les moyens de modéliser des dispositifs abstraits et de passer progressivement et élégamment du monde physique au monde logiciel. Par exemple, pour la caméra, un niveau d'abstraction pertinent est la détection des contours comme dans le "videoplace" de Krueger [KRU 91], ou mieux encore, la position (x, y, z) de chaque doigt comme il faudrait le faire pour le "bureau digital" [WEL 93]. (Le bureau digital est l'un des premiers démonstrateurs du concept de réalité augmentée. Il permet par exemple de sélectionner avec ses doigts un texte imprimé.) J. Crowley traduit l'existence de niveaux d'abstraction avec la notion de "capteur logique" [CRO 94]. Plus généralement, un système interactif peut être vu comme un dispositif composé d'une société de dispositifs plus simples. Les modèles d'architecture logiciels multiagent répondent au même mécanisme. Comme les dispositifs de Mackinlay, ces agents ont un état, un fonctionnement, des domaines d'entrée et de sortie [DUK 92, NIG 93].

Pour résumer, et en adoptant l'éclairage de notre propos central sur le multimédia et le multimodal, ces trois taxonomies permettent à l'évidence de raisonner au niveau du média en relation avec la tâche : média comme dispositif matériel chez Buxton et Foley, média comme dispositif technique chez Mackinlay et al. qui recouvrent les niveaux physique et logique. On peut noter le mécanisme de composition de la théorie de Mackinlay pour raisonner sur le facteur *multi* des notions de multimédia et de multimodalité. De plus, son mécanisme d'encapsulation offre une base de raisonnement sur le franchissement entre le multi-*média* (bas niveau d'abstraction) et le multi-*modal* (haut niveau d'abstraction). Ces trois modèles sont centrés sur l'interface d'entrée. Une piste intéressante, nous semble-t-il, serait d'étudier la composition de dispositifs d'entrée et de sortie à différents niveaux d'abstraction. C'est ce que nous avons fait pour la modélisation d'architecture logicielle [NIG 94, NIG 95].

A l'inverse des interfaces d'entrée, peu d'efforts se sont portés sur la classification des interfaces de sortie. La taxonomie sur les modalités pures de Bernsen est une exception. Nous la présentons dans la section qui suit.

### **5. Un espace de conception des interfaces de sortie**

En accord avec la métaphore des distances sémantiques et articulatoires de la théorie de Norman, Bernsen s'interroge sur le problème de la représentation optimale des informations que le système et l'utilisateur sont amenés à échanger au cours de

leur tâche commune. Dans cet espace problème, Bernsen centre son étude sur la représentation des expressions de sortie qu'il appelle *modalités représentationnelles* [BER 94]. Un graphe, un texte, une icône sont de telles modalités. Il s'agit de représentations produites par le système à destination de l'utilisateur.

### 5.1. Les principes

Chez Bernsen, nous l'avons vu, une modalité est un système représentationnel. Il ne s'agit pas des capacités sensorielles humaines telles qu'on l'entend en psychologie. L'objectif visé est de permettre au concepteur d'identifier les modalités qui conviennent à la représentation des concepts d'un système donné. Le schéma ci-dessous en décrit les principes.

"Ce qui doit être représenté" -> "principes de correspondance" -> "représentations"

Tout comme la taxonomie de Mackinlay, Card et Robertson, la classification de Bernsen se veut générative : elle s'appuie sur une classe de modalités élémentaires dites "modalités pures" combinables en représentations plus complexes. Bernsen définit une modalité représentationnelle par le couple <média, profil> :

- le *média* désigne le support d'expression en relation avec les capacités sensorielles humaines. Il prend sa valeur dans l'ensemble {graphique, son, toucher}. Chacune de ces valeurs - graphique, son, toucher - est caractérisée par un ensemble de qualités perceptuelles particulières (respectivement, visuelles, auditives, et tactiles) que Bernsen appelle *canaux d'information*. Par exemple, la couleur et la texture sont des canaux. Dans un tableau, les lignes et les colonnes sont aussi des canaux. En informatique, nous parlerions d'attributs.
- le *profil* regroupe quatre propriétés : statique ou dynamique, linguistique ou non-linguistique, analogique ou non-analogique, arbitraire ou non-arbitraire.

Par exemple, la modalité langue naturelle écrite a pour média "graphique", et pour profil le quadruplet <linguistique, non analogique, non arbitraire, statique>. Le langage parlé a pour média d'expression le son et présente le même profil que la langue naturelle écrite à l'exception de la dimension temporelle. Nous allons décrire ce que recouvre les qualificatifs d'un profil.

### 5.2. Le profil d'une modalité représentationnelle

Le profil définit une propriété discriminante entre modalités. Le caractère *dynamique* ou *statique* d'une modalité repose sur la présence de la composante temporelle dans la représentation. Une *modalité linguistique* désigne un langage, c'est-à-dire un système fondé sur des relations conventionnelles entre des éléments lexicaux, syntaxiques et pragmatiques pour faire sens. Par exemple, la langue parlée et écrite, le langage gestuel, sont des modalités représentationnelles linguistiques : tous s'appuient sur l'acquis d'un lexique et d'une syntaxe et tous, pour remplir leur fonction de communication, recouvrent une sémantique en situation (ou pragmatique).

Une *modalité analogique* entretient un rapport de ressemblance avec la réalité : elle fonctionne par ressemblance au concept représenté. Par exemple, le dessin d'une maison est une représentation analogique par rapport à la maison qu'il représente. On parle aussi de représentation iconique ou isomorphique. Notons que les icônes que l'on conçoit en interface graphique ne sont pas toujours des représentations analogiques. Une icône est qualifiée d'*abstraite* lorsque le dessin qui la constitue n'est pas analogique mais contient un indice sémantique (par exemple, une flèche pour désigner une direction) ; elle est arbitraire lorsqu'elle ne présente aucun indice sémantique sur le signifié. Une icône qui opère par analogie est nommée *icône représentationnelle*. Les *icônes semi-abstraites* résultent de la composition d'éléments abstraits et représentationnels

Une *modalité arbitraire* fonctionne en dehors d'un système conventionnel. A l'inverse, les *modalités non arbitraires*, tels les langages, reposent, pour remplir leur rôle, sur l'existence d'un système sémantique appris. Cependant, à un niveau fin, le signe linguistique, unité constitutive du langage, se définit par l'association arbitraire d'une forme et d'un contenu. Par exemple, l'association entre la forme graphique "loup" ou le son [lu] et l'animal sauvage qu'ils évoquent n'est pas logiquement motivée. Néanmoins il semble que la part d'arbitraire dans le lexique ne soit pas aussi systématique que cela : les onomatopées lexicalisées comme "clic, vlan" ou les noms d'oiseaux constitués par imitation de leurs cris n'ont pas été formés arbitrairement. A notre sens, ces termes ne sont pas arbitraires et revêtent en plus un caractère analogique.

En combinant les caractéristiques du profil d'une modalité et les valeurs possibles pour le média de restitution, on obtient 48 modalités pures de base : 2 (linguistique, non linguistique) x 2 (analogique, non analogique) x 2 (arbitraire, non arbitraire) x 2 (dynamique, statique) x 3 (graphique, son, toucher). De ces combinaisons, il faut éliminer les incohérences : les modalités analogiques et linguistiques ne sont pas arbitraires et les médias son et toucher ne peuvent être statiques. Bernsen obtient ainsi 28 modalités pures qu'il illustre par des types représentationnels connus. Par exemple, la modalité pure "langage écrit analogue statique" est illustrée par les hiéroglyphes, la modalité "son réel", tel le bruit d'une porte qui s'ouvre, est une modalité à profil <non linguistique, analogique, non arbitraire, dynamique>.

Les dimensions du profil d'une modalité ont des propriétés que l'on peut relier aux capacités perceptuelles et cognitives humaines.

### **5.3. Modalités représentationnelles et facultés humaines**

La "dynamicité" confère un caractère éphémère qu'il convient d'exploiter avec prudence en fonction du contexte situationnel. Par exemple, Palmiter constate que les représentations graphiques animées, admises pour faciliter l'apprentissage initial, ne promeuvent pas nécessairement la rétention de l'information à plus long terme [PAL 91] : des sujets ayant reçu un enseignement par démonstration animée des services "auteur" d'HyperCard se sont révélés plus performants (en terme de temps d'exécution de tâche) que les sujets instruits par écrit. Huit jours plus tard, la supériorité des performances est inversée résultant, pense-t-on, d'un traitement superficiel de l'information dans le cas de l'apprentissage par illustration animée.

Au caractère éphémère des modalités dynamiques, il convient de relier les notions d'inévitable et de prolongation introduites par Sellen et ses coauteurs. Sellen [SEL 92] caractérise les retours d'information selon trois critères : éphémère/non-éphémère, évitable/inévitable, maintenu/non-maintenu ("sustained/unsustained"). Par exemple, un bip émis à l'affichage d'une fenêtre est éphémère (car dynamique), inévitable (dans des conditions normales) et non-prolongé (une fois émis, il n'est pas répété). Il est maintenu par le système s'il est répété jusqu'à ce que l'action attendue soit réalisée par l'utilisateur (cas de l'imprimante qui signale l'absence de papier). Ces trois dimensions, sont, à notre sens, d'utiles affinements de la dimension statique/dynamique.

Les représentations analogiques, qui imitent ou reproduisent la réalité, induisent certainement moins d'efforts cognitifs que les représentations non analogiques. De même, les représentations arbitraires impliquent de la part du récepteur l'effort d'apprendre une nouvelle convention. Il faut insister sur le fait que le concepteur peut choisir d'utiliser une modalité analogique mais adopter une correspondance arbitraire avec le concept à représenter. Par exemple, il pourrait décider de représenter l'état de transfert d'un fichier par le grincement d'une porte ! A l'inverse, Gaver note à propos du SonicFinder, que l'association d'un son analogique ne convient pas nécessairement à décrire le changement d'état d'objets métaphoriques [GAV 89]. Par exemple, l'une des premières versions du SonicFinder associait des bruits de fenêtre du monde réel aux changements d'état des fenêtres de l'écran. L'expérience démontre que des sons analogiques mais impropres aux fenêtres physiques (par exemple, "schussss"), étaient mieux appropriés.

Les représentations linguistiques sont focalisées mais manquent de spécificité [BER 94]. A l'inverse, les représentations analogiques sont spécifiques mais manquent de focalisation. On entend par *spécificité* la capacité d'une représentation à limiter le champ interprétatif. La *focalisation* propre aux modalités linguistiques met en lumière le sujet du discours, l'intention communicationnelle de l'auteur. Par exemple, le texte "un angle de 60°" est plus focalisé que la représentation graphique d'un angle de 60°. L'observateur du dessin est amené à s'interroger sur le trait important qu'il convient de relever : est-ce l'existence de l'angle qui compte ou bien est-ce la valeur 60°? La modalité linguistique "un angle de 60°" ne laisse aucun doute. Elle est focalisée mais elle est moins spécifique : elle laisse de côté les détails telles que l'épaisseur du trait, l'orientation de l'angle dans l'espace, etc. La complémentarité "focalisation-spécificité" des modalités linguistique et analogique conduit à les combiner. Les cartes annotées de texte ou les textes illustrés de représentations analogiques sont des exemples courants. On trouvera dans [NIE 93] une revue intéressante d'études sur la combinaison de la langue naturelle écrite et des représentations graphiques.

#### 5.4. En synthèse

Pour résumer, la résolution du problème de la correspondance entre concepts et représentation perceptible exige l'existence d'une taxonomie solide qui offre des possibilités d'analyse. La théorie de Mackinlay, Card et Robertson vise cet objectif pour les dispositifs d'entrée. Bernsen poursuit un but similaire pour les sorties. Si la théorie des premiers est dirigée par la technologie (les dispositifs d'entrée sont

modélisés comme des transducteurs de propriétés physiques), la taxonomie de Bernsen est davantage liée aux capacités cognitives et sensorielles humaines avec les notions de profil et de média.

De fait, la multimodalité chez Bernsen est la composition de plusieurs systèmes représentationnels de sortie mais la théorie reste silencieuse sur la nature du profil d'une représentation multimodale. Est-ce, par exemple, un produit cartésien ? Alors que la théorie de Mackinlay permet de modéliser informatiquement le processus d'abstraction des informations acquises, la théorie de Bernsen n'a pas cette finalité pour les informations restituées. Que ces représentations soient engendrées par le système à partir d'une représentation interne de haut niveau ou qu'elles soient rejouées telles quelles comme dans un système hypermédia, la théorie n'a pas la prétention de faire la différence. Pour le développeur, la différence est pourtant pertinente.

Nous venons de présenter des taxonomies qui concernent soit les entrées soit les sorties. Nous considérons maintenant des espaces de conception qui couvrent à la fois l'interface d'entrée et l'interface de sortie. L'espace de Frohlich, centré sur l'utilisateur, est présenté en section 6. Les classifications de la section 7 proposent une vue complémentaire orientée système.

## **6. L'espace de conception de Frohlich**

Frohlich adopte comme point de départ le canevas du Modèle du Processeur Humain [CAR 86] introduit brièvement au paragraphe 3. Pour le besoin de notre analyse, ce modèle se résume ainsi : 1) l'interaction est le fruit d'un échange bidirectionnel d'informations à travers deux interfaces distinctes, l'une d'entrée, l'autre de sortie ; 2) du côté humain comme du côté informatique, l'information subit des transformations selon le cycle "acquisition-traitement-action". Partant de ce canevas, Frohlich introduit les éléments de son espace de classification [FRO 91].

### **6.1. Présentation des dimensions de l'espace**

Dans son analyse, Frohlich adopte la distinction entre entrée et sortie. Il en résulte deux espaces de conception. L'espace de conception en entrée est illustré par le schéma de la figure 1. On y distingue quatre éléments structurants : les notions de mode, de canal, de média et de style.

#### **6.1.1. Modes**

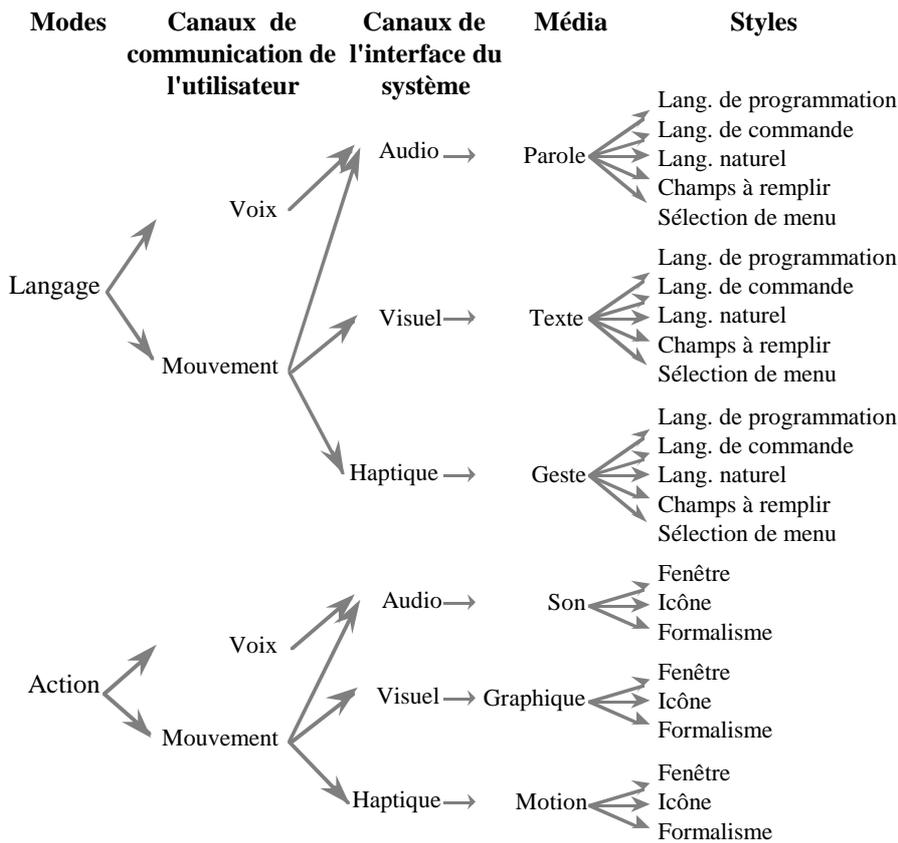
D'après Frohlich, les *modes* d'une interface se définissent comme "des états dans lesquels différentes actions de l'utilisateur peuvent avoir le même effet." [FRO 91, p. 58]. Un exemple nous paraît indispensable pour éclairer cette définition quelque peu allusive : la destruction d'un fichier peut s'exprimer par la saisie de la commande "Détruis <nom du fichier>" ou se faire en glissant l'icône du fichier dans la corbeille. Le premier procédé relève d'une activité langagière, la seconde d'un engagement dans l'action physique. Comme le montre la figure 1, action et langage définissent les deux modes possibles de l'espace de conception de Frohlich.

Cette notion de mode correspond aux deux paradigmes de l'interaction homme-machine présentées par Hutchins et ses coauteurs dans [HUT 86] : la métaphore du monde réel (“model world metaphor”) et la métaphore conversationnelle (“conversation metaphor”). Elle contraste avec la définition communément admise en informatique qui assimile la notion de mode à un contexte interactionnel. Avec cette acception, une même suite d'actions utilisateur peut, au contraire de la définition de Frohlich, conduire à des effets différents pour des modes différents [THI 90].

### 6.1.2. *Canaux*

Frohlich définit un *canal* “comme une interface qui opère une transformation d'énergie” [FRO 91, p. 58]. Il adopte une position anthropomorphique en mettant en relation les capteurs et effecteurs qui définissent les canaux de communication humains telles l'ouïe ou la vue, avec les dispositifs physiques qui correspondent aux canaux de l'interface du système. A chaque canal correspond un dispositif chargé de la transformation d'énergie. Par exemple, pour le canal “voix humaine”, les dispositifs sont les cordes vocales. Pour le canal de l'interface “audio”, le dispositif est un microphone. Dans le premier cas, l'énergie est de nature électrochimique. Elle est électrique pour le microphone.

D'après la figure 1, les modes langage et action en entrée font appel aux canaux de communication humains voix et mouvement. La voix humaine est captée par le canal audio du système tandis que le mouvement peut l'être par les trois canaux système audio, visuel et haptique. En sortie, il y a correspondance biunivoque entre canal système et canal humain.



**Figure 1.** L'espace de conception des interfaces en entrée de Frohlich

### 6.1.3. Média

Un *média* est “un système représentationnel qui permet l'échange d'information” [FRO 91, p. 58]. Frohlich distingue six médias : la parole, le texte, le geste qu'il rattache au mode langage ; le son, le graphique et la motion attribués au mode action. La parole recouvre toute forme de vocalisation à fonction communicationnelle y compris les silences. Le texte recouvre la trace visible d'un mouvement communicationnel tandis que le geste désigne un mouvement communicationnel sans trace visible. Le son est toute forme audible autre que la parole, le graphique toute forme de trace visible autre que le texte et la motion tout mouvement autre que le geste.

Les définitions que Frohlich adopte pour ses six catégories de médias sont certainement discutables. L'intérêt qui se dégage de son espace vient de la mise en relation entre les notions de mode, de canal et de média. En particulier, le média pris comme système représentationnel, est fixé par le couple <canal, mode>. Par exemple, la distinction entre un geste et une motion ne réside pas tant dans la nature des muscles mis en jeu, mais dans la fonction intentionnelle, c'est-à-dire, pour revenir au

modèle de Frohlich, au mode langage ou action. De même, ayant choisi le canal système visuel, le seul média d'entrée possible en mode langage est le texte alors que pour le mode action, ce média ne peut être que graphique. Notons qu'en sortie, le choix du média implique le choix du canal système.

#### 6.1.4. *Styles d'interface*

Dans son espace de conception, Frohlich définit un *style d'interface* comme "une classe reconnue de méthodes qui rend possible l'activité d'interaction" [FRO 91, p. 59]. Il distingue deux classes de méthodes correspondant chacune aux modes langage et action et qu'il note respectivement "style langage" et "style action". Les styles langage permettent de produire des expressions alors que les styles action permettent la production d'événements. L'auteur ne précise pas ce qu'il entend par expression et événement.

Les styles langage recouvrent les langages de programmation, les langages de commande, la langue naturelle, les champs de formulaires que l'on remplit et les menus. Les styles action incluent les environnements fenêtrés qui structurent l'espace de travail par groupes fonctionnels, les interfaces iconiques qui s'appuient sur une métaphore du monde réel, les formalismes graphiques qui correspondent à une interaction non-métaphorique et où le sens de l'activité est défini par le logiciel d'application.

Cette classification est reprise de celle de Baecker et Buxton [BAE 87]. Par rapport aux taxonomies de Mackinlay et de Bernsen, son manque de généralité nous donne le droit de nous interroger sur sa complétude.

#### 6.1.5. *Synthèse*

Pour résumer, nous explicitons les liens entre les cinq dimensions de l'espace de Frohlich : mode, canaux humain et système, média, et style.

- 1) Un média est défini par le couplage d'un mode et d'un canal système.
- 2) A un média correspond un style adéquat. Par exemple, au média parole est associé le langage naturel comme style approprié.
- 3) A un mode d'interaction correspond une et une seule classe de styles.

Jusqu'ici, nous avons décrit les éléments structurants de l'espace de Frohlich et émis quelques jugements de valeur sur chacune de ces dimensions. Dans la discussion qui suit, nous analysons l'apport et les faiblesses de cet espace en le considérant dans son ensemble.

## 6.2. *Discussion*

### 6.2.1. *Intérêt*

L'espace de Frohlich, et c'est là l'intérêt d'un espace de conception, peut être utilisé a posteriori pour analyser le résultat d'une conception ou a priori pour guider les choix de conception. Pour tout système existant, il est possible d'instancier le modèle général appliqué au cas particulier du système à étudier. Par exemple, l'interface du Finder d'Apple, qui s'appuie sur la métaphore du bureau, recouvre les deux modes action et langage en entrée comme en sortie. En entrée, les actions de

manipulation directe sur les icônes de fichier se caractérisent par le quintuplet : <mode = action, canal humain = mouvement, canal système = haptique, média = motion, style = iconique>. Certaines fonctions, telle l'ouverture de fichier, sont disponibles de manière équivalente dans le mode langage via les menus, les formulaires et les raccourcis clavier. Dans le modèle de Frohlich, un raccourci clavier se caractérise par : <mode = langage, canal humain = mouvement, canal système = haptique, média = geste (le mouvement ne laisse pas de trace), style = langage de commande>. Un raisonnement similaire vaut pour l'interface de sortie.

Cet exercice démontre qu'il convient d'être précis lorsque l'on vise à évaluer l'utilisabilité d'un système en fonction de ses caractéristiques ou à effectuer une étude comparative. Comme le note Frohlich, on avance peut être un peu hâtivement que la manipulation directe, comparée au mode langage, augmente l'utilisabilité du système. Dans une interface à manipulation directe, il n'y a pas d'intermédiaire entre l'utilisateur et le monde présenté. B. Laurel qualifie d'interfaces "à la première personne" ces systèmes qui offrent le sentiment d'engagement direct dans l'action sans relais intermédiaire [LAU 86] par opposition aux interfaces langage à "la deuxième personne" qui instituent une barrière linguistique entre l'utilisateur et les concepts internes du système. Si l'on reprend la terminologie des actes de langage d'Austin, le mode langage conduit l'utilisateur à "faire-savoir" au système ou à "faire-faire" via l'interface [AUS 62] alors que dans le mode action, l'utilisateur "fait".

La plupart des systèmes informatiques offrent, en vérité, un curieux mélange de ces interfaces à la première et à la deuxième personne. Par exemple, les éditeurs de texte adoptent tous l'action pour ce qui est de la saisie de texte mais ils imposent le conversationnel pour l'activation des commandes. Certes, les menus, les formulaires et les boutons, font l'objet de manipulations directes mais ils servent d'intermédiaires entre l'utilisateur et les concepts du domaine qui sont alors manipulés par indirection [KAR 94] et selon des conventions langagières. Il convient donc d'être prudent sur la qualification "interface à manipulation directe". Il faut savoir que la présence d'objets de présentation comme les menus et les formulaires, ne suffit pas à garantir l'action directe sur les concepts du domaine. Le Finder d'Apple, que l'on s'accorde à dire "facile d'emploi parce que relevant de la manipulation directe", montre à l'analyse, que son interface accorde une large part au mode langage. Son utilisabilité ne provient donc pas uniquement du mode action mais d'une combinaison adaptée de plusieurs modes.

Le second apport de l'espace de Frohlich est son utilisation en conception. L'espace définit des niveaux d'abstraction auxquels correspondent des décisions de conception. Les liens entre les niveaux définissent des implications entre les décisions. Par exemple, dans une approche descendante, le choix d'un mode implique un nombre limité de styles. Inversement, dans une approche ascendante, le choix d'un style implique un mode. Cette structuration de l'espace de conception en niveaux d'abstraction n'est pas sans rappeler l'approche en couches dont CLG s'est fait le pionnier [MOR 81].

En résumé, l'espace de Frohlich permet de caractériser de façon concise aussi bien les interfaces d'entrée que les interfaces de sortie de systèmes existants. En conception, il permet de situer différentes décisions et d'explicitier leurs liens. De manière générale, il fixe une terminologie organisée dans un cadre. Ainsi,

introduisant les termes mode et média, il permet de distinguer un système multimode d'un système multimédia. Mais à ces avantages correspondent aussi des limitations.

### 6.2.2. Limitations

L'espace de Frohlich ne couvre pas les dispositifs d'entrée/sortie dont les propriétés, nous l'avons vu, ont des incidences sur l'utilisabilité du système. Centré sur l'utilisateur, cet espace ne permet pas de considérer les problèmes liés à la mise en œuvre des systèmes.

La notion de canal encapsule, chez Frohlich, un ensemble de capteurs et d'effecteurs sans préciser leur nature. Puisque chaque dimension du modèle de Frohlich peut être vue comme un niveau d'abstraction, nous suggérons d'injecter dans l'interface d'entrée les taxonomies des dispositifs physiques de Buxton ou de Mackinlay comme une dimension intermédiaire entre les canaux de communication de l'utilisateur et les canaux système.

L'espace de Frohlich ne permet pas de relier les décisions de conception à leurs conséquences en terme de réalisation. Il n'explique pas davantage le parallélisme qui peut exister dans l'utilisation des canaux de communication qu'ils soient humains ou système. Cette dimension nous semble importante à considérer aussi bien pour l'utilisateur du système que pour le réalisateur :

- Pour l'utilisateur du système : peut-il parler tout en sélectionnant des objets avec la souris ? Comment les modes langage et action sont-ils disponibles ? Le sont-ils simultanément ou de manière exclusive ? Autant de questions qui ont trait à l'usage temporel des canaux humains et à la résolution de tâche.
- Pour la réalisation du système : faut-il considérer des traitements de données en parallèle ? Le faut-il en entrée uniquement ou bien en sortie ? Autant de questions en relation avec les traitements temporels du système depuis les canaux système jusqu'aux services internes de plus haut niveau.

En résumé, cet espace a le mérite d'être le premier, à notre connaissance, à avoir défini une taxonomie qui recouvre à la fois les interfaces d'entrée et de sortie. S'il constitue un cadre de pensée général, il ne permet pas de différencier les systèmes en fonction des possibilités d'utilisation offertes à l'utilisateur et en fonction des problèmes de réalisation. Ces aspects vont être abordés avec l'espace MSM.

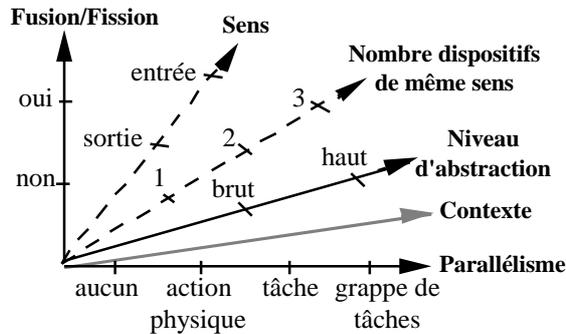
## 7. L'espace MSM

Avec le concept de *système Multi-Sensori-Moteur*, nous proposons un espace de référence qui caractérise un système en des termes utiles au concepteur de logiciel [COU 93, NIG 93] : les concepts auxquels il se réfère ont des implications directes sur les solutions logicielles aussi bien sur le plan architecture que sur la réalisation de mécanismes dédiés. On trouvera dans [NIG 95, MAR 95, BRI 94, BOU 92] des exemples de telles solutions.

Après l'exposé des dimensions de l'espace MSM, nous mettons en évidence ses points contributifs et notamment sa capacité à distinguer le multimédia de la multimodalité. Nous le comparons ensuite avec le cadre d'étude TYCOON [MAR 95].

## 7.1. Les dimensions de MSM

Comme le montre la figure 2, l'espace MSM comprend 6 dimensions. Les deux dimensions "Sens" et "Nombre de dispositifs de même sens" ont trait à la notion de dispositif d'entrée et de sortie. Les autres dimensions caractérisent le degré de sophistication computationnelle des fonctions d'acquisition et de restitution du système : niveaux d'abstraction, contexte, fusion et fission, granularité du parallélisme.



**Figure 2.** Le référentiel MSM : un espace de référence pour système multi-sensorimoteur

### 7.1.1. Dispositif d'entrée/sortie

Dans MSM, un dispositif d'entrée/sortie doit s'entendre au sens du paragraphe 2.2 : un capteur et/ou un effecteur. Notons que certains dispositifs actionnables par le geste instrumental présentent à la fois des caractéristiques d'entrée et de sortie contrôlables par programme. Par exemple, un manche à balai avec retour d'effort est présent sur l'axe des entrées mais aussi sur celui des sorties. Inversement, la souris est, dans un cadre usuel d'emploi, considérée comme un dispositif d'entrée seulement : ses boutons (et son poids) ont des effets sur la perception tactile humaine mais leur comportement n'est pas contrôlable par programme. S'il en est besoin, comme dans l'étude de F. Jambon sur les effets d'une souris sans clic [JAM 94], on affinera la décomposition d'un dispositif physique en autant de composantes sur l'axe des entrées et des sorties. Il suffit, une fois encore, de choisir la granularité qui convient et de s'y tenir.

### 7.1.2. Fonctions d'acquisition et de restitution

L'information acquise par les dispositifs d'entrée est transformée par les processus internes du système. Cette suite d'opérations modélise la *fonction d'acquisition* (ou d'interprétation). Dans l'autre sens, l'information interne (c'est-à-dire l'état interne du système) est transformée jusqu'à ce qu'elle soit acceptable par les dispositifs de sortie du système. Cette suite d'opérations constitue la *fonction de restitution* (ou de synthèse).

A chaque dispositif d'entrée, nous associons une fonction d'acquisition qui prolonge au cœur du système, les capacités de calcul du dispositif. (Sur ce point,

nous rejoignons le mécanisme d'abstraction de la théorie de Mackinlay.) De même, à chaque dispositif de sortie, nous associons une fonction de restitution. Acquisition et restitution font intervenir quatre ingrédients en intime relation : niveaux d'abstraction, contexte, fusion et fission, parallélisme.

#### 7.1.3. Niveaux d'abstraction

La notion de niveau d'abstraction exprime le degré de transformations subies par les informations reçues ou émises par les dispositifs. Elle couvre également l'éventail des représentations que gère le système depuis les informations brutes (les signaux) jusqu'aux représentations symboliques (le sens). Une fonction d'acquisition se caractérise par son *pouvoir d'abstraction* des événements reçus du dispositif d'entrée auquel elle est associée. Une fonction de restitution se qualifie par son *pouvoir de matérialisation* depuis les représentations internes jusqu'aux représentations de bas niveau de son dispositif de sortie.

#### 7.1.4. Contexte

La capacité d'une fonction à abstraire ou à matérialiser peut dépendre de *variables contextuelles* ou *contexte*. Nous définissons de manière plus complète cette notion dans [NIG 94, chapitre 3]. Un contexte comprend un ensemble de variables d'état utilisées par les processus internes pour contrôler l'acquisition ou la restitution. Par exemple, dans les systèmes actuels, nous observons deux contextes d'acquisition pertinents : les commandes et les concepts du domaine. Les informations liées aux commandes font l'objet d'une acquisition à un haut niveau d'abstraction tandis que les informations ayant trait aux concepts du domaine, sont souvent laissées inchangées. Par exemple, dans les systèmes de traitement de texte, les saisies au clavier ou à la souris qui correspondent aux commandes sont "comprises" par le système tandis que ces mêmes saisies à destination du contenu des documents ne sont pas interprétées.

Ainsi, le pouvoir d'abstraction ou de matérialisation d'un système est propre à chaque dispositif, mais pour un dispositif donné, ce pouvoir dépend du contexte.

#### 7.1.5. Fusion et fission d'information

La *fusion* est la combinaison de plusieurs unités d'information pour former de nouvelles unités. La *fission* correspond au processus inverse. L'une et l'autre traduisent deux activités importantes des processus d'acquisition et de restitution.

En acquisition, la *fusion* peut intervenir à un bas niveau d'abstraction entre des informations provenant de plusieurs dispositifs d'entrée. Elle peut aussi s'effectuer à un plus haut niveau d'abstraction pour des informations issues de différents contextes. Par exemple, l'expression orale "mets ça ici" nécessite la fusion de l'événement parole reçu via le microphone avec les événements souris qui désignent l'objet et le lieu dont il est question. Plus fréquente est la fusion de clics souris répartis sur une palette et une zone de dessin pour construire une figure géométrique. Ces informations issues du même dispositif (la souris) transitent selon des contextes différents pour être regroupés à un haut niveau d'abstraction [NIG 94].

En restitution, la *fusion* intervient également à plusieurs niveaux d'abstraction. Au niveau le plus haut, elle a lieu dans l'adaptation des informations du noyau

fonctionnel aux besoins conceptuels de l'interface [COU 91]. Au niveau le plus bas, elle se manifeste par exemple sous forme d'incrustations (vidéo et graphique).

La *fission* en acquisition traduit le besoin de décomposer une information issue d'un dispositif ou d'un contexte pour franchir un niveau d'abstraction. Par exemple, l'acte de parole "dessine un cercle dans une nouvelle fenêtre" fait référence à deux domaines de discours : les figures géométriques ("dessine un cercle") et l'interface homme-machine ("nouvelle fenêtre"). Cette phrase dont le sens a pu être identifié à partir de données captées par un seul dispositif (le microphone) doit être décomposée en deux primitives de haut niveau du système : "créer fenêtre" à destination du gestionnaire de fenêtre et "créer cercle" dans la nouvelle fenêtre à destination de l'éditeur graphique.

La *fission* en restitution revêt plusieurs formes. La plus courante est la représentation multiple d'un même concept sur un dispositif donné. Par exemple, le concept de température est restitué sous forme d'un thermomètre gradué ou d'un nombre réel. On dit que les deux représentations sont redondantes. La représentation multiple peut aussi s'effectuer en coréférence sur des dispositifs distincts tel le message oral "attention à cette température" accompagné de l'affichage en rouge du thermomètre à surveiller. Dans ce cas, nous parlons de complémentarité des informations. Nous reviendrons au paragraphe 7.1.7 sur ces notions de redondance et de complémentarité en liaison avec le parallélisme.

#### 7.1.6. *Parallélisme*

Nous limitons la discussion au parallélisme perceptible à l'interface. Au niveau de l'interface, le parallélisme se manifeste à trois niveaux de granularité : action physique, tâche élémentaire, grappe de tâches.

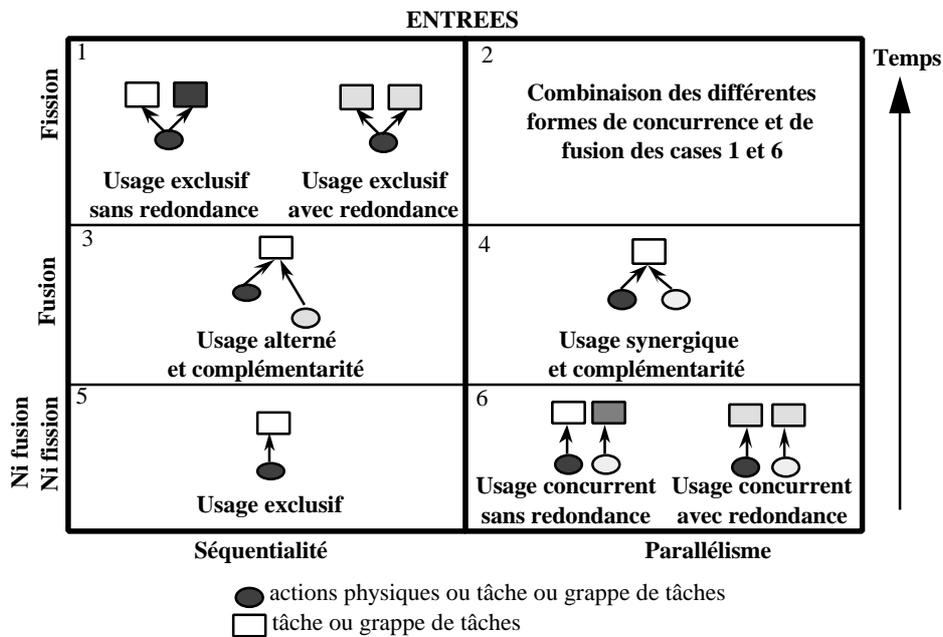
Au niveau physique, le parallélisme en entrée autorise l'utilisateur à agir simultanément sur plusieurs dispositifs d'entrée comme dans l'exemple "mets ça ici". En sortie, le parallélisme signifie que plusieurs primitives de sortie peuvent être exécutées simultanément par un même dispositif (comme en animation graphique) ou par plusieurs dispositifs (comme pour l'exemple "attention à cette température"). Notons que la fusion et la fission à bas niveau d'abstraction dépendent de l'existence du parallélisme au niveau physique.

On appelle *tâche élémentaire*, toute tâche que l'utilisateur peut accomplir avec le système via une et une seule commande. Au niveau tâche élémentaire, le parallélisme en entrée autorise l'utilisateur à construire plusieurs commandes de front : en parallélisme vrai si le parallélisme est géré au niveau physique ou, en l'absence de parallélisme physique, de manière entrelacée (dialogue à plusieurs fils). En sortie, une tâche élémentaire recouvre l'ensemble des primitives de sortie pour exprimer un changement d'état du système.

Nous appelons *grappe de tâches*, un ensemble de tâches élémentaires permettant à l'utilisateur d'accomplir une tâche composée. Au niveau grappe de tâches, le parallélisme exprime les possibilités d'entrelacement entre des espaces de travail. Par exemple, pour un robot mobile de surveillance, les tâches de l'utilisateur peuvent être structurées en trois espaces : description de la géographie, spécification de missions, et contrôle d'exécution de mission.

### 7.1.7. Parallélisme/séquentialité et fusion/fission

Le parallélisme (ou la séquentialité) couplé(e) au phénomène de fusion (ou de fission) offre un champ intéressant de réflexion. Nous limitons ici la discussion à l'usage en entrée des modalités ; nos résultats sont néanmoins directement applicables pour caractériser l'usage en sortie des modalités. Comme le montre la ligne du centre de la figure 3, la production séquentielle (case 3) ou simultanée (case 4) de deux (ou plusieurs) informations peut donner lieu à une fusion pour faire sens. Dans le premier cas, on parlera d'*utilisation alternée* de modalité [IHM 92] : les modalités sont utilisées de manière séquentielle ou bien, le système impose une utilisation séquentielle. Par exemple dans *Munix*, l'utilisateur doit compléter les commandes vocales de manipulation de fichiers Unix en saisissant leur nom au clavier [LEF 92]. Dans le second cas, la fusion avec production simultanée (ou quasi simultanée) d'informations correspond à un *usage synergique* de plusieurs modalités. Par exemple, dans *MATIS* [NIG 94], l'expression orale "flights to this city" accompagnée de la désignation simultanée à la souris du nom de la ville est un usage synergique de plusieurs modalités. *TAPAGE*, qui permet l'édition de tableaux par la parole et le geste au moyen d'un stylo, est un autre exemple de système capable de synergie [JUL 93]. Que l'usage soit synergique ou alterné, il y a émergence de sens par suite d'une fusion d'informations. Ce phénomène se produit lorsque des modalités entretiennent une relation de *complémentarité* : chaque modalité véhicule une partie d'information qui, prise isolément, est insuffisante pour faire sens.



**Figure 3.** Les différents cas d'usage en entrée de modalités selon que cet usage par l'utilisateur se fasse de manière séquentielle ou en parallèle et selon que cet usage

*conduise le système à pratiquer ou non une fusion ou une fission d'information pour faire sens*

La ligne du bas de la figure 3, correspond à l'absence de fusion et de fission. On parle d'*usage exclusif* lorsque l'utilisateur emploie (ou lorsque le système impose) au plus une modalité à la fois. Dans la case 5, on n'observe ni fusion, ni parallélisme. Par exemple, un utilisateur de MATIS ferait un usage exclusif des modalités parole et manipulation directe s'il utilisait, à un instant donné, une seule des deux possibilités. Il énoncerait la phrase "flights to Pittsburgh" ou bien il saisirait au clavier la ville de Pittsburgh dans le menu des villes de destination. Mais comme le montre la case 1, l'usage exclusif d'une modalité peut amener le système à pratiquer une fission. Le résultat de cette opération conduit à plusieurs informations distinctes comme dans l'exemple cité en 7.1.5 ("Dessine un cercle dans une nouvelle fenêtre"), ou bien produire plusieurs informations redondantes (cas des pléonasmes, par exemple).

La production simultanée de plusieurs informations ne conduit pas nécessairement à une fusion. Chaque unité d'information peut faire sens. Il s'agit alors d'un *usage concurrent* (voir case 6). Par exemple, dans MATIS, un utilisateur peut énoncer "flights to Pittsburgh" tout en remplissant la ville de destination d'un formulaire avec la ville de Boston. Pour MATIS, il s'agit de deux requêtes menées concurremment. La *redondance* est un cas particulier de concurrence lorsque les deux informations produites en parallèle couvrent le même sens. Par exemple, dans MATIS, l'utilisateur qui énonce "flights to Pittsburgh" tout en choisissant le mot "Pittsburgh" comme ville de destination avec la souris, se comporte de manière redondante. Le système doit éliminer le superflu ou au contraire en milieu bruité, il doit en tirer partie comme information de complémentarité.

La complémentarité et la redondance caractérisent l'usage de plusieurs modalités. Orthogonale à la notion de multi-modalité, vient la notion de choix. Le choix ne désigne pas ici uniquement celui effectué par l'utilisateur mais aussi celui du système. Par exemple, le système expert de Gargan [GAR 88] fait intervenir les préférences de l'utilisateur pour produire une présentation multimodale à partir du contenu sémantique de l'information. La possibilité de choix entre plusieurs modalités pour atteindre un but (ou exprimer un changement d'état) s'appelle *l'équivalence*. Dans MATIS, l'utilisateur dispose de plusieurs modalités pour spécifier une destination. L'absence de choix est une *assignation*. Complémentarité, Assignation, Redondance, Equivalence constituent les propriétés CARE. Dans [COU 95], on trouvera une définition formelle de ces propriétés, de même qu'une réflexion sur leur utilisation en évaluation d'interface et en conception logicielle. [MIG 93] et [CAT 95] rapportent des résultats expérimentaux sur l'usage des modalités pour des tâches de conception et de contrôle de processus.

## **7.2. Contributions de MSM**

Les dimensions de MSM fournissent une définition pour les termes multimédia et multimodal. Noter que MSM centré sur les caractéristiques des systèmes, offre une définition dirigée par la technique.

Les interfaces multimédia et multimodale requièrent l'une et l'autre au moins deux dispositifs, soit d'entrée, soit de sortie. En ce sens, ces interfaces sont toutes

deux multi-sensori-motrices. Une interface multimodale se caractérise par la capacité supplémentaire à disposer d'au moins deux dispositifs (soit d'entrée, soit de sortie) qui, fonctionnant de manière combinée ou non, sont doués d'un haut pouvoir soit d'abstraction, soit de matérialisation.

La distinction technique entre multimédia et multimodalité repose pour l'essentiel sur les capacités d'abstraction des dispositifs d'entrée et /ou de concrétisation vers les dispositifs de sortie. Un haut niveau d'abstraction ou de concrétisation implique l'existence d'un langage d'interaction en entrée et respectivement en sortie. Ainsi un système multimodal doit offrir au moins deux dispositifs physiques de même sens couplés à un ou plusieurs langages d'interaction. Nous retrouvons ici notre définition d'une modalité, introduit au paragraphe 2.2.4, comme le couple <dispositif, langage d'interaction>.

Un système hypermédia, qui permet de jouer des séquences audio-vidéo, est multimédia en sortie (le système ignore la signification des séquences) et monomodal en entrée (manipulation directe des liens hypertextes par la souris). Si l'on ajoutait le contrôle du système à la voix, il deviendrait multimodal en entrée (manipulation directe et parole) tout en restant multimédia en sortie.

### **7.3. Lien avec l'espace TYCOON**

Le cadre d'étude TYCOON [MAR 95] repose sur la notion de modalité définie comme un processus informatique d'analyse ou de synthèse (paragraphe 2.2.2.). Chaque modalité est définie par un domaine de valeurs d'entrée D, par un domaine de valeurs de sortie R, par un programme P et par deux domaines de valeurs de contrôles d'entrée et de sortie de P, notés respectivement CE et CS. La multimodalité est alors définie par la coopération entre plusieurs modalités ou processus. Le cadre d'étude TYCOON identifie cinq types de coopération entre processus et met en relation ces types avec des buts de coopération tels que la reconnaissance, l'apprentissage ou le gain de temps. Les types de coopération permettent de composer des modalités élémentaires. Les cinq types de coopération entre modalités sont le transfert, l'équivalence, la spécialisation, la redondance et la complémentarité.

Nous constatons que l'espace TYCOON repose sur des concepts et un mécanisme de combinaison similaires à ceux de l'espace de conception des dispositifs de Mackinlay, Card et Robertson présenté au paragraphe 4.2. Un dispositif physique comme une modalité dans TYCOON est défini sur des domaines de valeurs en entrée et en sortie (respectivement noté In/Out dans l'espace de Mackinlay et D/R dans TYCOON). De plus un dispositif est défini par une fonction de résolution noté R, qui correspond au concept de programme (P) dans TYCOON. Enfin ces deux espaces proposent de composer des éléments élémentaires pour décrire des éléments plus complexes. Nous pouvons mettre en relation l'opérateur de fusion de l'espace de Mackinlay avec celui de complémentarité de TYCOON. Cependant TYCOON distingue différents opérateurs de complémentarité. L'un des "complémentarité-données" ou complémentarité sur les données, correspond à l'opérateur de fusion de Mackinlay. De même nous pouvons mettre en relation l'opérateur "transfert-données" de TYCOON avec celui de connexion de Mackinlay.

Par rapport à l'espace MSM, nous constatons que les types de coopération sauf le transfert s'obtiennent par projection en ayant fixé la direction et le nombre de

dispositifs. Par exemple, l'assignation se définit par le point dont les coordonnées sont : (plusieurs dispositifs de même sens, haut niveau d'abstraction, pas de choix). Les niveaux d'abstraction, le parallélisme et le contexte dans les fonctions d'abstraction et de concrétisation ne sont pas pris en compte dans l'espace TYCOON.

Néanmoins par ses définitions formelles des types de coopération pour composer les modalités, le cadre d'étude TYCOON définit sans ambiguïté des points dans l'espace MSM.

## 8. Conclusion

Nous avons présenté une revue de la terminologie et les cadres susceptibles de nous aider à structurer l'espace problème des nouvelles interfaces. Ces espaces conceptuels visent à forcer les "bonnes questions" mais adoptent des perspectives différentes : les uns centrés sur l'utilisateur, d'autres sur la technique, les uns raisonnant à bas niveau d'abstraction, les autres à haut niveau, les uns se limitant aux entrées des échanges, les autres aux sorties. Ces choix sont validés par une finalité précise. Notre propre contribution est motivée, pour l'essentiel, par les problèmes de la conception logicielle. Il nous revient maintenant d'étudier les relations de dépendances entre les modalités d'entrée et de sortie puisque c'est de leur couplage qu'émerge l'essence même de l'interaction.

## 9. Remerciements

Les auteurs remercient les membres du groupe de travail GT10 du GDR-PRC Communication Homme-Machine pour les échanges d'idées fructueuses sur l'interaction multimodale ainsi que leurs partenaires du projet ESPRIT AMODEUS. La présentation et les rapports de ce projet sont disponibles : <http://www.mrc-apu.cam.ac.uk/amodeus/amodeus.html> ou <ftp://ftp.mrc-apu.cam.ac.uk/pub/amodeus>.

## 10. Bibliographie

- [ALT 91] ALTY J., "Multimedia -- What is It and How do we Exploit It?", *Actes de la conférence Human Computer Interaction'91*, People and Computers VI, édités par D. Diaper et N. Hammond, 20-23 Août 1991, p. 31-44, 1991.
- [ARE 93] ARENS Y., HOVY E., VAN MULKEN S., "Structure and Rules in Automated Multimedia Presentation Planning", *Actes de la conférence IJCAI'93*, Chambéry, France, p. 1253-1259, 1993.
- [AUS 62] AUSTIN J., *How to do thing with words*, Oxford : Clarendon Press, 1962.
- [BAE 87] BAECKER R., BUXTON W., *Readings in human computer interaction*, Los Altos, California: Morgan Kaufmann Inc, 1987.
- [BEN 95] BENOIT D. et al., *Introduction aux Sciences de l'Information et de la Communication*, Les Editions d'organisation, 414 pages, 1995.
- [BER 94] BERSEN O., "Foundations of multimodal representations. A taxonomy of representational modalities", *Interacting with Computers*, Vol. 6, No. 4, 1994, p. 347-371, 1994.

- [BLA 90] BLATTNER M., DANNENBERG R., "CHI'90 Workshop on multimedia and multimodal interface design", *SIGCHI Bulletin*, ACM, Volume 22, Number 2, p. 54-58, 1990.
- [BOU 92] BOURGUET M-L., Conception et réalisation d'une interface de dialogue personne-machine multimodale, thèse de l'Institut National Polytechnique de Grenoble, 1992, 206 pages.
- [BRI 94] BRISON E., VIGOUROUX N., "Interprétation des événements dans l'interaction multimodale", *Actes de la conférence Ingénierie des IHM'94*, Lille, p. 23-34, 1994.
- [BUX 83] BUXTON W., "Lexical and pragmatic considerations of input structures", *Computer Graphics*, 17 (1), p. 31-37, 1983.
- [BUX 90] BUXTON W., "Smoke and Mirrors", *Byte*, p. 202-251, 1990.
- [CAR 86] CARD S., MORAN T., NEWELL A., *The Psychology of Human-Computer Interaction*, Chapitre 2, The Human Information-Processor, Hillsdale, New Jersey : Lawrence Erlbaum Associates, p. 23-97, 1986.
- [CAR 90] CARD S., MACKINLAY J., ROBERTSON G., "The Design Space of Input Devices", *Actes de la conférence Computer Human Interaction'90*, Seattle, p. 117-124, 1990.
- [CAT 95] CATINIS L., CAELEN J. , "Analyse du comportement multimodal de l'utilisateur humain dans une tâche de dessin", *actes de la conférence Ingénierie des IHM'95*, Cepadues Eds: Toulouse, p. 123-129, 1995.
- [COU 91] COUTAZ J., BALBO S., "Applications: A Dimension Space for User Interface Management Systems", *Actes de la conférence Computer Human Interaction'91*, ACM, New Orleans, p. 27-32, 1991.
- [COU 93] COUTAZ J., NIGAY L., SALBER D., "The MSM framework: A Design Space for Multi-Sensori-Motor Systems", *actes Human Computer Interaction, 3rd International Conference EWHCI'93, East/West Human Computer Interaction*, Moscow. L. Bass, J. Gornostaev, C. Unger Eds. Springer Verlag Publ., Lecture notes in Computer Science, Vol. 753, p. 231-241, 1993.
- [COU 95] COUTAZ J., NIGAY L., SALBER D., BLANDFORD A., MAY J., YOUNG R., "Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The CARE Properties", *actes de la conférence INTERACT'95*, p. 115-120, 1995.
- [COU 96] COUTAZ J., "L'art de communiquer à plusieurs voies", *Spécial La Recherche, L'Ordinateur au doigt et à l'œil*, La Recherche, no 285, p.38-43, 1995.
- [CRO 94] CROWLEY J., BEDRUNE J., "Integration and Control of Reactive Visual Processes", *1994 European Conference on Computer Vision, (ECCV-94)*, Stockholm, mai 1994.
- [CRO 96] CROWLEY J., COUTAZ J. "Vision for Man Machine Interaction", *Engineering for computer Human Interaction EHCI'95*, L.J. Bass & C. Unger Eds., Chapman & Hall Publ., p. 28-45, 1996.
- [DUC 90] DUCE D., GOMES M., HOPGOOD F., LEE J., *User Interface Management and Design, Workshop on User Interface Management Systems and Environments*, Springer, 4-6 Juin 1990.
- [DUK 92] DUKE D., HARRISON M., Abstract Models for Interaction Objects, SM/WP1, System Modelling, Working Paper 1, Amodeus Project, Esprit Basic research Action 7040, Novembre 1992.
- [DUK 94] DUKE D., BARNARD P., DUCE D., MAY J., Syndetic model for human-computer interaction. TR ID/WP35, ESPRIT BRA 7040 Amodeus-2, 1994.
- [FOL 84] FOLEY J., WALLACE V., CHAN P., *The Human Factors of computer Graphics interaction techniques*, IEEE computer Graphics and Applications, 4(11), p. 13-48, 1984.
- [FRO 91] FROHLICH D., "The Design Space of Interfaces, Multimedia Systems", *Interaction and Applications, 1st Eurographics Workshop*, Stockholm, Suède, Springer Verlag, p. 53-69, 1991.
- [GAR 88] GARGAN R., SULLIVAN J., TYLER S., "Multimodal response planning: an adaptive rule based approach", *actes de la conférence Computer Human Interaction'88*, New York, p. 229-234, 1988.

- [GAV 89] GAVER W., "The SonicFinder, An Interface That Uses Auditory Icons", *Human Computer Interaction*, 4(1), Lawrence Erlbaum Ed., 1989.
- [HJE 47] HJELMSLEV L. "Strutural Analysis of Language", *Studia Phonetica*, Vol. 1, p. 69-79, 1947.
- [HUT 86] HUTCHINS E., HOLLAN J. ET NORMAN D., "Direct Manipulation Interfaces", *User Centered System Design, New Perspectives on Computer Interaction*, édité par Norman D., Draper S., Hillsdale, New Jersey : Lawrence Erlbaum Associates, p. 87-124, 1986.
- [IHM 92] Atelier IHM'92, ENST, Interfaces Multimodales et Architecture Logicielle, 30 Novembre-2 Décembre 1992.
- [JAM 94] JAMBON F., COUTAZ J. "Retours d'information sensori-moteurs des dispositifs physiques : Tâche de sélection avec une souris sans 'clic'", *actes 12th Triennial Congress of the International Ergonomics Association (IEA'94)*, International Ergonomics Association Publ., Toronto, vol. 4, p. 464-464, 1994.
- [JUL 93] JULIA L., FAURE C. "A Multimodal Interface for Incremental Graphic Document Design", *actes Human Computer Interaction International'93*, Orlando, 1993.
- [KAR 94] A. KARSENTY, GroupDesign : un collecticiel synchrone pour l'édition partagée de documents, Thèse de doctorat, Université d'Orsay Paris-Sud, 1994.
- [KRU 91] KRUEGER M., *Artificial Reality II*, Addison Wesley, 1991.
- [LAU 86] LAUREL B., "Interface as Mimesis", *User Centered System Design, New Perspectives on Computer Interaction*, édité par Norman D., Draper S., Hillsdale, New Jersey : Lawrence Erlbaum Associates, p. 67-85, 1986.
- [LEF 92] LEFEBVRE P., POIRIER F., DUNCAN G., "A Multimodal Approach to Man-Machine Dialogue with the Unix Operating System", *Actes IFIP TC2/WG2.7 Working Conference on Engineering for Human Computer Interaction*, Larson J. & Unger C. Eds, North Holland publ., p. 285-294, 1992.
- [MAC 90] MACKINLAY J., CARD S., ROBERTSON G., "A Semantic Analysis of the Design Space of Input Devices", *Human Computer Interaction*, Lawrence Erlbaum, Vol. 5, No 2 & 3, p. 145-190, 1990.
- [MAR 95] MARTIN J-C., Coopérations entre modalités et liage par synchronie dans les interfaces multimodales, Thèse Ecole Nationale Supérieure des Télécommunications, Paris, Mars 1995.
- [MIG 93] MIGNOT C., VALOT C., N. CARBONELL N., "An Experimental Study of Future 'Natural' Multimodal Human-Computer Interaction", *actes de la conférence InterCHI'93 (Adjunct Proceedings)*, Amsterdam, Avril, p. 67-68, 1993.
- [MOR 81] MORAN T., "The Command Language Grammar: a representation for the user interface of interactive computer systems", *International Journal of Man-Machine Studies*, No 15, p. 3-50, 1981.
- [MYE 90] MYERS B., "A New Model for handling Input", *ACM Transactions on Information Systems*, Vol. 8. No. 3, p. 289-320, 1990.
- [NIE 93] NIELSEN J., "Noncommand user interfaces", *Communications of the ACM*, Vol. 36, No 4, p. 83-99, 1993.
- [NIG 93] NIGAY L., COUTAZ J., "Espace problème, fusion et parallélisme dans les interfaces multimodales", *actes de la conférence InforMatique'93*, Montpellier, p.67-76, 1993.
- [NIG 94] NIGAY L., Conception et modélisation logicielles des systèmes interactifs : application aux interfaces multimodales, Thèse de l'Université de Grenoble I, 1994, 315 pages.
- [NIG 95] NIGAY L., COUTAZ J., "A Generic Platform for Addressing the Multimodal Challenge", *Actes de la conférence Computer Human Interaction'95*, ACM Press, Denver, Mai, p. 98-105, 1995.
- [NOR 86] NORMAN D., *Cognitive Engineering, User Centered System Design, New Perspectives on Computer Interaction*, édité par Norman D., Draper S., Hillsdale, New Jersey : Lawrence Erlbaum Associates, p. 31-61, 1986.

- [PAL 91] PALMITER S., ELKERTON J, BAGGETT P., "Animated demonstrations versus written instructions for learning procedural tasks", *International Journal of Man-Machine Studies*, 34, p. 687-701, 1991.
- [SEL 92] SELLEN A., KURTENBACH G., BUXTON W., "The Prevention of Mode Errors Through Sensory Feedback", *Human Computer Interaction*, Lawrence Erlbaum, Vol.7, No 2, p. 141-164, 1992.
- [THE 94] G. THERY : *Les autoroutes de l'information*, Rapport au Premier Ministre, La documentation Française Ed., Paris, 1994.
- [THI 90] THIMBLEBY H., *User Interface Design*, ACM Press, Frontier Series, Addison-Wesley Publ., 1990, 470 pages.
- [WEL 93] WELLNER P., "Interacting with paper on the Digital Desk", *Communications of the ACM*, Vol. 36, No. 7, p. 87-97, 1993.

## 11. Biographie

Laurence Nigay est Maître de Conférence à l'Université Joseph Fourier et membre de l'équipe "Ingénierie de l'Interaction Homme-Machine" du laboratoire Communication Langagière et Interaction Personne-Système de la Fédération IMAG (CLIPS-IMAG). Ses travaux de thèse présentés en 1994 ont trait à l'interaction multimodale et à ses aspects logiciels. Elle a effectué plusieurs séjours à l'Université de Carnegie-Mellon (USA). Elle est membre du groupe de travail WG 2.7(13.4) de l'IFIP, membre du groupe de travail sur l'interaction multimodale du GDR-PRC Communication Homme-Machine et a participé aux travaux du projet ESPRIT Basic Research AMODEUS (1989-1995).

Joëlle Coutaz est professeur à l'Université Joseph Fourier et responsable de l'équipe Ingénierie de l'Interaction Homme-Machine du laboratoire CLIPS-IMAG. Elle est vice-présidente du groupe de travail WG2.7(13.4) de l'IFIP et partenaire du projet ESPRIT Basic Research AMODEUS (1989-1995). En France, Joëlle Coutaz est membre du GDR-PRC Programmation. Dans le cadre du GDR-PRC Communication Homme-Machine, elle a fondé le pôle Interface Homme-Machine Multimodale dont elle a partagé la responsabilité de 1990 à 1993. Elle est membre fondateur du groupe de travail SCOOP sur les collecticiels au GDR-PRC CHM et du colloque annuel "Ingénierie des IHM".