

Conception d'un système de communication multiéchelle

Sofiane Gueddana
Laboratoire de Recherche en Informatique
Université Paris-Sud
91405 Orsay Cedex France
+33 6 99 45 04 09

sofiane.gueddana@laposte.net

ABSTRACT

De nos jours, la plupart des systèmes de communication sont basés sur plusieurs services séparés, correspondants à différents niveaux de détail. Cette séparation impose de changer de service pour changer de niveau de détail. Nous proposons une approche nouvelle de services unifiés pour des systèmes de communication qui peuvent être qualifiés de multi-échelles. Un système de communication multi-échelles est capable de transmettre un niveau d'information variable et permet des transitions fluides entre ces niveaux, pour s'adapter aux contextes de son utilisation.

L'objectif de mon stage a été de concevoir et de réaliser un système de communication multi-échelles, exploitant principalement la vidéo et dont l'utilisation est destinée aux membres du cercle familial. Après avoir étudié et présenté les aspects théoriques liés à la conception d'un tel système, nous présenterons le prototype conçu et réalisé. Enfin, nous verrons les réponses qu'il a apportées aux problématiques initialement posées.

Categories and Subject Descriptors

H.4.3 [Information System Application]: Communication Applications; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces

General Terms: Design.

Keywords

Communication médiatisée, vidéo, système multi-échelles, technologie domestique, système interactif, interaction homme-machine, sonde technologique.

1. INTRODUCTION

Les technologies de la communication évoluent vers une diversification des usages et des services proposés, et la plupart des systèmes de communication se basent sur plusieurs services, correspondants à différents niveaux de détail. Pourtant, ils imposent la plupart du temps de changer de service pour changer de niveau d'information. Par exemple, les téléphones portables les plus récents combinent un ensemble de services correspondants à différents niveaux de détail (e.g. identification de l'appelant, SMS, MMS, messagerie vocale, téléphonie et éventuellement visiophonie). Cet éventail de possibilités permet de contrebalancer l'accessibilité permanente liée à la possession du téléphone. Cependant, les transitions entre niveaux sont généralement difficiles, voire impossible.

L'approche nouvelle que nous proposons surmonte cette difficulté en introduisant la notion de système de communication multi-

échelles : Un système de communication peut être qualifié de multi-échelles s'il est capable de transmettre un niveau d'information variable et permet des transitions fluides entre ces niveaux. Cette définition est à rapprocher de la notion d'espace géométrique 2D zoomable à l'infini [44].

Le stage de recherche que j'ai effectué, avait pour but de concevoir et de réaliser un système multi-échelles qui utilise les images fixes et la vidéo pour la communication dans un cadre familial et domestique. Les usages auxquels nous nous sommes intéressés couvrent à la fois la simple coordination entre individus, la communication informelle ainsi que le partage d'objets physiques ou informatiques.

Dans ce rapport, nous étudierons les aspects théoriques liés à la conception d'un tel système et nous présenterons le prototype conçu et développé. Dans une première partie, nous commençons par un aperçu du contexte scientifique spécifique à notre problème. Nous présenterons d'abord, le contexte général des outils en interaction Homme-Machine puis nous décrirons la stratégie de conception participative que nous avons adoptée. Ensuite, nous explorerons quelques systèmes communiquant utilisant l'image, nous dégagerons les difficultés que la littérature nous permet de prévoir. Puis, nous présenterons une synthèse bibliographique sur les rôles pour notre contexte familial, de la vision dans la communication et sur la capacité de la vidéo à les véhiculer. Ensuite, la revue de quelques logiciels de communication nous permettra de situer relativement notre approche. Dans une seconde partie, nous expliciterons les services que doit fournir le système que nous nous proposons de réaliser, et les contraintes qu'il doit prendre en compte. Ensuite, nous détaillerons la conception et le développement de ce système, autour de la notion de système de communication multi-échelles, et au regard des requis précédents. Puis nous présenterons un scénario d'utilisation mettant en relief quelques fonctions que notre système réalise. Nous terminerons ce rapport en concluant sur les éléments de réponses qu'il a apporté et les améliorations possibles qui restent à réaliser.

2. CONTEXTE GENERAL

Dans cette partie, nous présentons le contexte scientifique de notre étude. D'abord, nous définirons notre stratégie de recherche et les paradigmes d'interaction dans lesquels les nouveaux outils sont conçus. Ensuite nous décrirons la méthodologie de conception participative que nous avons choisi et en particulier les *sondes technologiques* utilisées dans le contexte familial. Puis, nous étudierons l'utilisation de la vidéo dans le contexte de la communication informelle. Ensuite, nous verrons les rôles importants pour la communication familiale de la vision et nous évaluerons la capacité de la vidéo à les véhiculer. Enfin, nous ferons la revue de quelques logiciels de communication pour situer notre approche.

2.1 Stratégie de recherche en Interaction Homme-Machine

Les modèles utilisés par les chercheurs en Interaction Homme-Machine sont des simplifications abstraites de la réalité choisies en fonction de leur capacité à représenter les problèmes posés, à prédire le réel et à produire des solutions utiles. Ces modèles qui permettent de guider la conception de systèmes interactifs sont souvent issus des sciences humaines (e.g. psychologie, ergonomie, sociologie).

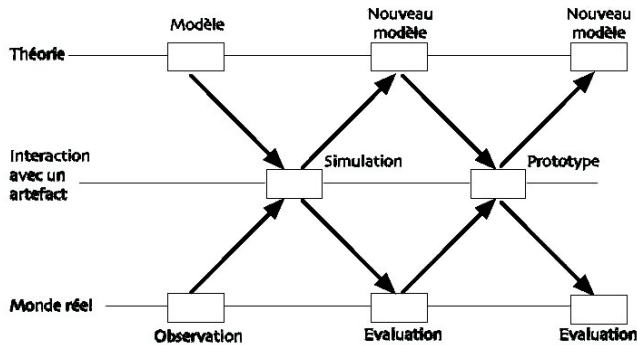


Figure 1. Modèles et conception participative.

Les théories qui guident la conception des systèmes interactifs viennent souvent des sciences humaines, en particulier de la psychologie (par exemple la perception pré-attentive [54]), de l'ergonomie et de la sociologie. Elles peuvent se résumer dans des principes généraux (Gestalt perception), peuvent constituer des théories à part entières (Théorie écologique de la perception [28]) ou être des techniques descriptives plus proches de l'implémentation (Machines à état, GOMS [8]). On utilise aussi des lois empiriques qui sont issues d'observations contrôlées (par exemple, loi de Fitts [25]).

L'observation de l'interaction entre des utilisateurs et le système étudié, plus ou moins réalisé ou simulé, fournit également de nombreuses informations utiles à sa conception, permettant de corroborer ou invalider les hypothèses issues des modèles théoriques. La conception d'un système interactif doit donc reposer sur des itérations répétées entre une approche théorique et la réalisation d'artefacts permettant l'évaluation de cette approche (Figure 1).

Norman en 1990 [42] a mis en évidence l'importance de la correspondance entre les modèles mentaux des utilisateurs et le modèle conceptuel utilisé dans la conception : L'utilisation de métaphores appropriées et l'exploitation des *affordances*¹ naturelles conduit à des systèmes exploitant nos réflexes naturels, plus intuitifs et demandant peu d'apprentissage. Les métaphores sont un outil efficace pour la conception d'interface : en calquant le fonctionnement d'une application sur celui d'un élément courant du monde réel, on décharge l'utilisateur de la phase d'apprentissage [21].

Les objets du monde physiques sont pour la plupart facile à utiliser, tout le monde sait comment interagir avec eux. Ils sont faciles à adapter à des situations différentes, et portent en eux leur mode d'emploi. Selon l'approche écologique de la perception de Gibson [28], la perception des objets constituant notre

environnement est directement la perception de possibilités d'interaction avec ceux-ci. Ceci notamment par le fait qu'ils obéissent tous aux mêmes lois universelles et familières de la physique. Les ordinateurs eux, sont pour la plupart rigides et sont rarement d'usage intuitifs. Ils se comportent selon les règles définies par leurs concepteurs et programmeurs et nécessitent un apprentissage.

Aux débuts de l'informatique, plusieurs personnes devaient partager un seul ordinateur. Avec l'avènement des ordinateurs personnels, chaque personne pouvait avoir une machine. Aujourd'hui, chacun peut avoir à sa disposition de nombreux systèmes informatiques interactifs. Weiser en 93 [55], voyant la convergence de plusieurs facteurs dans cette évolution (une miniaturisation plus importante, l'accélération et la baisse des prix des processeurs, le nombre plus important d'ordinateurs par personne et le développement des réseaux non filaires) a prédit l'arrivée de l'*ubiquitous computing*² : « invisible everywhere computing » : Invisibles car petits, embarqués, dédiés à une tâche. Plusieurs par personne, de différentes tailles, accessibles à distance sans fils, configurables dynamiquement donc et répartis dans l'environnement. Les badges actifs développés par Olivetti sont un bon exemple d'applications ubiquitaires : ils permettent de recevoir des messages, et de rappeler les rendez-vous et les réunions, ils sont munis de cadrans alphanumériques et réagissent à des balises disséminées dans l'environnement qui les connectent au réseau informatique et localisent leur propriétaire. Ils réagissent à la lumière et sont désactivé dans l'obscurité [33].

C'est aussi dans cette optique que s'est développée la réalité augmentée. La *réalité augmentée* (ou *réalité mixte*) vise à "augmenter" les propriétés des objets de notre environnement de capacités de traitement d'information, par exemple en superposant des informations à l'image du monde réel, comme l'application de maintenance d'imprimantes (Karma de Feiner & Macintyre en 91 [22]). D'autres applications existent : donner à voir les anciens bâtiments d'une ville (tourisme augmenté), pour les archéologues voir les annotations en même temps que le site de fouille archéologique, et en particulier en médecine, où les données 3D provenant de l'imagerie médicale et l'image normale sont fusionnées. La *réalité augmentée* inclut aussi les *interfaces tangibles* définies par Ishii [36] comme l'incarnation d'informations virtuelles dans des objets physiques, comme dans la Marble Answering Machine de Bishop [16], et permettent ainsi de se dispenser des interfaces usuelles comme les écrans et les boutons. Les interfaces tangibles permettent aussi à l'utilisateur de contrôler l'ordinateur à travers des objets physiques (en utilisant les objets habituels comme c'est le cas du Digital Desk de Wellner [56]). Les interfaces tangibles peuvent aussi servir à communiquer directement de façon tactile entre deux personnes distantes, comme dans le cas de « In Touch » [7].

De manière plus générale, la réalité augmentée cherche à établir des ponts entre le monde physique et le monde informatique à travers de nouveaux paradigmes d'interaction. Pour réaliser la réalité mixte, il faut augmenter l'utilisateur, augmenter l'objet ou augmenter l'environnement. Augmenter l'utilisateur c'est lui faire porter l'interface (« wearable computing » [39]). Augmenter l'objet c'est utiliser un objet ou un ensemble d'objets comme interface, comme dans le projet « Table probe » ou « story table

¹ Opportunités d'interaction offertes par le système

² En français, *informatique ubiquitaire* ou *disséminée*

». Pour augmenter l'environnement, il s'agit d'augmenter plusieurs éléments de l'environnement, et aboutit à des systèmes immersifs dans lesquels l'interface est autour de l'utilisateur (AmbientRoom de Ishii & al [36]), par exemple, cela peut consister à détecter l'utilisateur et utiliser ses gestes pour piloter le système.

Nous avons vu les nouveaux paradigmes d'interaction dans lesquels notre système va s'inscrire, nous présenterons à présent les nouvelles méthodologies de conceptions que l'on utilisera.

2.2 Conception participative

Le processus traditionnel de production de logiciel se déroule de la façon suivante (appelée en management *cycle de vie en V*): d'abord les spécifications du besoin des utilisateurs sont produites et fixées, puis on conçoit la solution logicielle répondant à ces besoins, ensuite on développe cette solution. Enfin on livre à l'utilisateur final un logiciel définitif et fonctionnel.

La conception participative a un déroulement différent (appelé en management *cycle de vie en étoile ou en spirale*): Il s'agit de partir aussi des spécifications mais pour concevoir et produire d'abord un prototype, ensuite ce prototype est évalué auprès de l'utilisateur final et son usage observé, puis l'évaluation et l'observation est analysée afin de spécifier les nouveaux besoins et enfin relancer le cycle de conception et de développement.

La spécificité de la conception participative tient surtout dans la participation de l'utilisateur final dans le processus de conception.

Dans la conception des interfaces des systèmes interactifs, il s'agit de trouver l'équilibre entre les moyens technologiques, l'utilisabilité des systèmes pour la tâche à laquelle ils sont dédiés, et les interactions nécessaires de l'utilisateur. La conception participative est centrée sur l'utilisateur et allie l'ingénierie aux connaissances des sciences humaines et aux méthodes du design et de l'innovation. Elle en emprunte et en dérive les techniques. Parmi ces techniques on peut citer: l'incident critique, les scénarios, le brainstorming et le prototypage.

Cette méthode de conception engendre un phénomène de co-adaptation empêchant de prédire a priori la définition de l'outil idéal. En effet, au cours de la conception de ce dernier, les concepteurs corrigent et améliorent les dispositifs qu'ils développent en tenant compte des remarques des utilisateurs. Chaque itération du cycle de conception, va modifier la perception de ces derniers qui devront se réadapter pour soumettre à nouveau leurs commentaires.

La production de prototype fait intervenir plusieurs techniques qui visent à des améliorations justifiées du design, en considérant plusieurs alternatives de conception et en s'assurant de leurs utilisabilité. Les prototypes réalisés doivent être rapides à faire, construits de façon modulaire pour supporter une évolution incrémentale, afin de pouvoir incorporer facilement des changements dans la conception et peut-être aboutir à un produit final. Les techniques de prototypage peuvent être de différents niveaux: simulation papier ou vidéo, réalisation plus ou moins complète du système.

Les techniques d'évaluation utilisées peuvent être soit subjectives (l'introspection, les interviews, les questionnaires, les notes d'observation, etc) soit objectives par observation directe, enregistrement et codage de données (études d'utilisabilité, enregistrement audio ou vidéo, capture clavier et souris, etc). Elles peuvent être formelles et ciblées comme les

expérimentations contrôlées, ou informelles et de portée plus large. Souvent, la combinaison de plusieurs techniques est nécessaire afin de ne pas négliger des aspects importants.

La conception est aussi une activité créative génératrice d'idées. Elle s'inspire des systèmes existants tout en restant critique et tente de comprendre les utilisateurs. Par exemple les ateliers (workshops) de conception participative rassemblent les utilisateurs, les concepteurs et les chercheurs, et organisent les idées à travers des scénarios, des brainstorming et des analyses de tâches pour les présenter par des prototypes et des simulations.

La conception doit aussi tenir compte des spécificités des supports utilisés : par exemple le point d'attention des utilisateurs lors de la consultation de vidéos est au centre de l'écran.

Utilisées en interaction homme machine, ces techniques ont pour objectifs de confronter au plus tôt les prototypes produits à partir des modèles théoriques au besoin réel, de corriger le modèle théorique et de concentrer le développement sur les vrais problèmes. Le résultat de cette façon de procéder est d'obtenir une meilleure adéquation du modèle théorique à la réalité et entre les logiciels produits, l'activité à laquelle ils répondent et l'utilisateur qui réalise cette tâche.

Si le concepteur est expert dans la production de solutions logicielles à des besoins qu'on lui soumet, l'utilisateur est expert de la tâche qu'il doit accomplir et même s'il n'est pas toujours conscient de ses besoins et ne peut pas les expliciter, sa participation aux produits qu'il utilise est néanmoins très utile, voire indispensable.

2.3 InterLiving et le concept de sondes technologiques

InterLiving est un projet de recherche européen qui a récemment impliqué le LRI et l'INRIA (projet IST Disappearing Computer initiative, jan 2001 - déc 2003). L'objectif de ce projet était la conception de nouvelles technologies de communication pour l'environnement familial. L'une des originalités de ce projet était qu'à cette équipe de chercheurs pluridisciplinaire (ethnologie, psychologie, design industriel et informatique) ont été associées trois familles suédoises et trois familles françaises. Une part de la coopération entre les chercheurs et les familles s'était effectuée par le biais d'interviews et de rencontres-ateliers classiques, mais cette coopération s'était également effectuée par l'installation de sondes technologiques chez les familles.

Dans le projet InterLiving, les méthodes classiques de conception participative ont été complétées par la réalisation de *sondes technologiques*. Les contraintes liées au contexte familial ne permettaient pas aux chercheurs d'analyser in situ l'utilisation de ces technologies. Les sondes avaient permis à la fois de tester de nouvelles technologies, de collecter des données d'utilisation dans un contexte réel et d'encourager les familles à réfléchir avec les chercheurs à leurs besoins et aux solutions qui pouvaient être mises en oeuvre pour y répondre. L'utilisation des sondes au sein des familles avait aussi pour but de développer des technologies de la communication influençant le mode de vie des gens les utilisant. L'utilisation des sondes permettait de comprendre comment l'informatique et les nouvelles technologies pouvaient s'insérer dans la vie quotidienne de la famille. Il a été montré par Dewsbury 2001 [17], qu'une utilisation appropriée de ces technologies pouvait améliorer la qualité de vie des personnes qui les utilisent.

Le concept de *sonde technologique* diffère de celui de prototype sur plusieurs points. Une sonde est très simple et propose très peu de fonctionnalités, alors qu'un prototype peut en offrir plusieurs. De plus, une sonde est malléable et est destinée à être abandonnée : elle doit inspirer des idées, et les utilisations non envisagées sont encouragées. D'ailleurs, puisqu'elle est destinée à être jetée, l'utilisabilité d'une sonde n'est pas une caractéristique aussi importante qu'elle puisse l'être pour un prototype. Et surtout, une sonde doit être instrumentée afin de fournir des données d'utilisation.

Au cours du projet InterLiving, plusieurs sondes avaient été élaborées, telles que le VideoProbe [12] et MirrorSpace [48].

2.3.1 VideoProbe



Figure 2. Le VidéoProbe installé chez une famille.

VidéoProbe [12] est un dispositif qui permet de prendre automatiquement des images de la vie familiale et de les partager entre les foyers d'une même famille (Figure 2). Il est constitué d'un écran et d'une caméra, et se trouve connecté à des vidéoProbes situés dans d'autres foyers de la même famille. Les photos qui sont prises par un videoProbe sont instantanément partagées et peuvent être consultées sur les videoProbes distants. Afin d'éviter aux utilisateurs du VidéoProbe d'être submergés par un nombre d'images grandissant dû aux nombreuses photos qu'ils prennent, VidéoProbe propose une solution à ce problème en incluant un mécanisme de vieillissement des images analogue à celui des photos réelles : au fur et à mesure que le temps passe, les couleurs, ainsi que le contraste des images disparaissent. Au bout d'un certain temps, si les images n'ont pas été explicitement retenues, elles sont éliminées (Figure 3).



Figure 3. Mécanisme de vieillissement des images : les couleurs et les contrastes disparaissent progressivement

Les résultats du VideoProbe ont été intéressants. En effet, les scènes échangées entre les familles, intentionnellement ou automatiquement ont été utiles et représentent un usage nouveau de la photographie, qui a été intégré par les familles au même titre qu'un répondeur téléphonique. La présence du VideoProbe a augmenté la fréquence des appels téléphoniques et a contribué à rapprocher les familles.

Un autre projet appelé « Digital Family Portrait » mené par Mynatt & al en 2001, visait aussi à surmonter le problème de la distance géographique dans une famille. Toutefois, ce projet s'adressait aux personnes âgées dont l'éloignement physique rend difficile leur surveillance par les autres membres de la famille. Dans ce projet, un système combine des informations disparates pour fournir aux autres membres de la famille une information résumée dans un portrait. Contrairement au VidéoProbe, le Digital Family Portrait n'est pas centré sur la communication. Son objectif est la surveillance d'un proche à distance.

2.3.2 MirrorSpace

MirrorSpace [48] est un système de communication vidéo original reposant sur la métaphore du miroir qui prend en compte la notion de distance. Tandis que les systèmes traditionnels se contentent de créer un espace partagé correspondant à une distance interpersonnelle particulière, mirrorSpace est à l'inverse conçu pour offrir un continuum de distances permettant l'expression d'une grande variété de relations entre individus (Roussel & Al 2003).



Figure 4. MirrorSpace

Les flux vidéo des lieux reliés par mirrorSpace sont affichés sur un écran unique, fusionnant par transparence l'image des participants locaux et distants. Afin de permettre des formes de communication intimes où le regard joue un rôle très important, la caméra est placée au centre de l'écran. Un utilisateur peut ainsi se placer très près de celle-ci tout en étant toujours capable de voir la personne distante et de communiquer avec elle (Figure 4). Le dispositif comporte également un capteur de proximité qui mesure en continu la distance à la personne ou l'objet le plus proche. Les distances mesurées sont utilisées pour appliquer un effet de flou sur chacune des images affichées. Cet effet a pour but de situer de façon intuitive les personnes ou objets perçus à travers mirrorSpace dans un espace virtuel partagé. Le flou permet de percevoir le mouvement d'une personne éloignée avec un

minimum d'implication. Il offre également un moyen naturel et intuitif pour initier ou éviter une transition vers un mode de communication plus engagé en se déplaçant simplement vers le dispositif (Figure 5) ou au contraire en s'en éloignant.



Figure 5. Diminution de l'effet de flou accompagnant l'approche d'une personne

Le logiciel créé pour ces installations a été conçu selon le principe des sondes technologiques évoqué précédemment. À l'occasion des différentes expositions, il a ainsi enregistré de nombreuses données d'utilisation qui ont permis de mieux comprendre comment les utilisateurs perçoivent le système et d'en améliorer la conception matérielle, logicielle et esthétique.

Nous avons présenté la méthode de conception participative, et nous avons expliqué un concept important pour notre projet, celui de sonde technologique dont nous avons décrit plusieurs applications. Le chapitre suivant est consacré à d'autres systèmes de communication centrés sur la notion de communication informelle que nous allons expliquer.

2.4 Communication informelle et MediaSpaces

La présence physique dans un même environnement commun permet la connexion non planifiée entre les personnes pour des communications informelles. Contrairement aux communications formelles qui sont prévues à l'avance, avec des participants fixés et un ordre du jour, les communications informelles sont imprévues, spontanées et opportunes, et s'effectuent dans un langage informel. Cette distinction que nous faisons entre communication informelle et formelle est théorique. Dans sa pratique, la frontière est floue et les situations de communication mêlent formel et informel.

Ces communications informelles sont particulièrement importantes et sont essentielles pour la survie du groupe qu'il soit professionnel ou familial. Au quotidien et particulièrement au travail, la plupart des communications qui se produisent sont informelles et il est important de pouvoir « tomber » sur quelqu'un à certains moments, tel que les apartés au cours d'une réunion ou les rencontres dans les couloirs, et c'est parfois de cette façon que des éléments importants sont échangés.

Les dispositifs de communication traditionnels utilisant la vidéo, tels que les systèmes de vidéoconférences sont adaptés à un usage formel. Cet usage de la vidéo semble inadapté, d'où peut-être une raison de son échec. Les dispositifs orientés vers la communication informelle ont connu plus de succès, malgré les difficultés inhérentes à ce type de communication. Parmi ces derniers, on peut citer MirrorSpace mais aussi les MediaSpaces.

Un MediaSpace est un dispositif permettant à un groupe éventuellement géographiquement dispersé, de communiquer par l'intermédiaire de moyens audiovisuels et informatiques. Techniquement, un MediaSpace repose sur un réseau audio/vidéo

piloté par des moyens informatiques ; chaque utilisateur dispose d'une station de travail, ainsi que d'un moniteur vidéo, d'une caméra, de haut-parleurs et d'un microphone. À l'aide de logiciels adaptés, tout utilisateur peut établir une connexion audio/vidéo avec n'importe quel autre membre du groupe.

Les principales installations MediaSpace sont au nombre de quatre : le premier est le MediaSpace du Xerox PARC en Californie qui a donné son nom au concept [52] ; RAVE de Xerox EuroPARC en Grande-Bretagne est son descendant direct [27], et CAVECAT de l'Université de Toronto repose sur une partie de la technologie de RAVE [40] ; enfin Cruiser de Bellcore est un système MediaSpace assez différent des trois précédents [23]. Depuis ces premières expériences, plusieurs autres laboratoires ont développé des MediaSpaces à titre expérimental, comme le CoMedi au CLIPS-IMAG à Grenoble [15].

Les MediaSpaces, permettent de s'affranchir de la distance entre les correspondants, pourtant ils sont principalement utilisés par des personnes présentes dans un même bâtiment. Ils ont pour objectif de renforcer la communication et la collaboration au sein d'un groupe. Cet aspect est complètement absent dans le vidéophone. Les MediaSpaces sont aussi un bon outil pour faciliter la communication informelle entre les membres d'un groupe. Et cela à travers les opportunités de rencontre non planifiées, similaire à croiser une personne dans le couloir ou jeter un coup d'œil dans un bureau ouvert, mais aussi en renforçant la perception du groupe, permettant par cela une contextualisation de l'action personnelle.

Les services offerts par un MediaSpace sont d'établir des connexions audio/vidéo avec tout autre utilisateur du MediaSpace. Ces connexions peuvent être de différents types :

- une connexion réciproque de type vidéophone: un utilisateur initie une connexion avec un autre utilisateur, et tous deux sont alors reliés par une connexion audio/vidéo qui peut être interrompue à l'initiative de l'un quelconque des deux utilisateurs.
- une connexion de courte durée, qui permet de "jeter un coup d'oeil" (glance) chez un autre utilisateur.
- une connexion de longue durée, l'"office-share", qui peut durer de quelques heures à plusieurs années: cette connexion est établie de façon permanente entre deux utilisateurs, qui leur permet ainsi de partager un bureau virtuel. Cette connexion permet une communication informelle sur une longue durée [18].
- une connexion "background", qui est une connexion de longue durée, unidirectionnelle et uniquement vidéo, avec une partie commune, typiquement la cafétéria, ou bien un équipement vidéo tel un magnétoscope ou un tuner TV.

Ces modes de communication soulèvent de nombreuses questions. La plus importante est sûrement celle de la protection de l'espace privé des utilisateurs du système: les utilisateurs doivent clairement avoir des moyens de contrôler l'accès à leur espace privé. Un MediaSpace doit inclure des mécanismes de protection de l'espace privé de l'utilisateur. Mais on peut aussi s'interroger sur l'utilisation même du MediaSpace. Le MediaSpace ne vise pas à remplacer la communication humaine directe ; il offre des possibilités de communication supplémentaires [5]. Pour protéger l'espace privé des utilisateurs, les MediaSpaces ont été dotés de mécanismes de contrôle. Pour que la connexion s'établisse, il faut que l'utilisateur accepte explicitement. Toutefois, pour un service

comme le glance, cette négociation explicite est trop intrusive [13], aussi il fallait associer au contrôle la notion de notification.

Le système d'origine a été progressivement modifié et enrichi. De ce point de vue, un MediaSpace peut être un bon terrain d'expérimentations sur le fonctionnement d'un groupe. L'apport de la psychologie sociale et de l'anthropologie est capital pour la compréhension des interactions au sein d'un groupe [31]. Le MediaSpace favorise la communication, mais certaines particularités de la communication humaine ne peuvent être transmises par le système. Le contact visuel direct en est l'exemple le plus flagrant : pour croiser le regard de son interlocuteur, il faudrait regarder l'objectif de la caméra, ce qui empêche de regarder en même temps l'écran du moniteur.

La vision par ordinateur peut apporter des avantages certains à ce type de systèmes de communication. Elle offre la possibilité de capter le comportement de l'utilisateur dans son milieu naturel sans adjonction d'artifices contraignants comme les cordons de connexion du gant numérique. L'une des applications de la vision par ordinateur est le suivi de mouvement. Les techniques nécessaires existent déjà isolément, mais elles présentent des faiblesses. Coutaz [14] propose un processus de coopération entre ces techniques jouant à la fois sur la redondance d'information et la complémentarité fonctionnelle pour doter les interfaces utilisant la vision par ordinateur de robustesse et d'autonomie.

Les problèmes centraux qui déterminent le succès d'un mediaspace sont intégration dans les habitudes des utilisateurs, sa flexibilité par rapport aux besoins de ses usagers et les mécanismes permettant la préservation de l'intimité des utilisateurs. Lorsque ces exigences sont atteintes, l'espace partagé par le mediaspace devient un lieu social d'échange et de communication.

Ces conclusions nous donnent un aperçu des points essentiels qu'il faut prendre en compte pour la conception d'un système de communication utilisant la vidéo. Dans le chapitre suivant nous avons choisi d'analyser le rôle de la vision dans la communication et la capacité de la vidéo à véhiculer ce rôle. Ce chapitre servira à déterminer les comportements visuels que notre système utilisant la vidéo, permettra aux membres d'une famille d'utiliser pour communiquer, et dont il faudra tenir compte pour sa conception.

2.5 Le rôle de la vision dans la communication.

Dans la communication naturelle, la vision a un rôle important, qui peut-être partiellement véhiculé par l'utilisation de la vidéo dans la médiation de la communication à distance.

Exception faite des études sur les coups d'œil (glance), la plupart des études sur l'usage de la vidéo la considèrent comme un complément à la communication audio, contrairement aux projets MirrorSpace et VideoProbe qui se basent sur l'image seule comme support de la communication. Cependant, on peut souligner l'importance du canal oral puisque sa suppression a un impact énorme sur la communication [9], probablement parce que l'oral est le support langagier naturel. Allant dans le sens de cette explication, parmi les usages intentionnels faits par les familles utilisant le videoProbe, à plusieurs reprises les utilisateurs ont simplement laissé un message écrit posé devant le videoProbe. Cela montre qu'il reste difficile de se dispenser totalement du langage pour communiquer.

La plupart des études prenaient comme référence la communication face à face, car c'est là que l'on constate les mécanismes naturels de comportement visuels dans la communication. La communication face à face (Face-to-Face) est un processus multimodal, qui engage une interaction complexe entre des comportements verbaux et visuels. Malgré la nature multimodale de la communication présente, la technologie de communication distante la plus persuasive et la plus populaire reste le téléphone, qui n'est le support que de la modalité vocale. Les tentatives d'enrichissement de la modalité vocale par l'ajout d'informations supplémentaires n'ont pas conduit aux améliorations attendues dans la communication à distance. Les études faites en laboratoires pour montrer les bénéfices de la modalité visuelle dans la communication, ont montré peu d'améliorations objectives [9]. Les technologies qui ont ajouté l'image à la voix, comme les vidéophones ou les dispositifs de vidéoconférence n'ont toujours pas prouvé leur réussite sur le marché [19].

Aussi, pour mieux situer le rôle que peut jouer la vidéo dans la communication, nous allons d'abord définir les aspects importants de la communication, voir comment les comportements visuels se manifestent et évaluer l'efficacité de la vidéo pour véhiculer ces comportements.

2.5.1 La communication

Il y a plusieurs aspects fondamentaux dans la communication qui ont besoin d'être pris en compte, quelque soit le mode de communication utilisé. D'après Clark & Brennan [11], la communication est une activité qui nécessite la coordination conjointe du processus et du contenu par les participants.

Dans la coordination du processus, il y a deux aspects importants : la prise de parole et la disponibilité. Pour la communication utilisant l'image seule, seul le second aspect nous intéresse. La plupart des communications ne sont pas planifiées, elles requièrent que les participants puissent savoir précisément quand les autres personnes sont disponibles et la pertinence de commencer une interaction spontanée, et cette connaissance est basée sur la perception du mouvement et des activités des autres [32].

Dans la coordination du contenu, il s'agit de comment les participants atteignent et maintiennent une compréhension commune dans une conversation [11]. Un aspect important de la coordination du contenu est la référence. La référence permet aux participants d'identifier conjointement les objets de la communication [10]. Un autre aspect de la coordination du contenu concerne l'état affectif des participants et l'attitude interpersonnelle. C'est une information sociale sur les sentiments des participants, leurs émotions, et l'attitude à l'égard des autres participants et envers l'objet de la communication.

2.5.2 Le rôle et la fonction de l'information visuelle dans la communication

Les premiers travaux sur le rôle de l'information visuelle dans la communication ont montré qu'il est subtil et complexe. D'un point de vue théorique, nous avons besoin de comprendre en détail la fonction que l'information visuelle joue dans la communication. D'un point de vue pratique, il faut comprendre

quand et comment l'information visuelle est utilisée pour la communication.

Dans la communication visuelle, il y a deux types d'informations visibles. Le premier est l'information à propos des comportements des autres participants, c'est-à-dire le regard, l'expression du visage et la posture. Le second ensemble d'informations visuelles est l'environnement que les participants partagent.

Le regard est la façon dont on extrait des informations visibles de l'environnement. La direction vers laquelle une personne regarde, la durée du regard dans une certaine direction, et la façon de regarder sont des aspects importants du comportement visible.

Le regard est en général un indicateur de l'attention et peut être dirigé vers les autres participants, aussi bien que vers des éléments physiques de l'environnement. Les gens sont très bons pour déterminer l'endroit où les autres regardent. Cela facilite l'attention conjointe, et permet une grande flexibilité dans la référenciation des objets. Le regard est aussi un indicateur de l'attitude interpersonnelle ou affective. Les gens évaluent les autres d'après leur façon de regarder : Ceux qui regardent peu leurs interlocuteurs sont jugés « défensifs » ou « évasifs », ceux qui regardent beaucoup sont jugés « amicaux » et « sincères » [38].

Pour préserver le contact oculaire dans les systèmes vidéo, les participants doivent regarder directement la caméra, mais comme la personne doit regarder l'image de l'autre personne, ou un écran d'ordinateur, le contact oculaire n'est pas possible. Le compromis typique dans les vidéophone ainsi que pour les ordinateurs est de placer la caméra au dessus de l'écran. Il y a des nouvelles techniques pour sauvegarder le contact oculaire : Le ClearBoard [35] surimpose la vidéo des autres participants sur un tableau commun, et la caméra préserve le contact visuel en utilisant des miroirs semi-réfléchissants (tunnel vidéo). Sellen [49] décrit un dispositif combinant écran et caméra où la caméra est intégrée à un écran suffisamment petit pour que le contact visuel soit possible. Ott [43] propose un système dans lequel l'image de l'utilisateur vu de face est calculée numériquement à partir des images de deux caméras situées l'une en dessous et l'autre au-dessus de l'écran et pointées vers l'utilisateur. Cette dernière approche est la plus intéressante car elle ne nécessite pas de matériel spécifique.

L'expression faciale est portée par les yeux, les sourcils, le nez, la bouche et le front [20]. Le visage est aussi une source d'information riche sur l'état émotionnel des participants. Les yeux, la bouche et les sourcils sont hautement expressifs. Ekman et Friesen [20] ont montré que les personnes quelque soit leur culture sont capable de reconnaître sept expressions faciales distincte d'après des photographies (la joie, la tristesse, la surprise, la colère, le dégoût, la peur et l'intérêt).

La posture est l'information fournie par l'inclinaison du corps et l'orientation du corps d'un participant, en particulier son tronc et le haut de son corps. La posture est un autre indicateur de l'intérêt et de l'engagement d'un participant [6]. La position corporelle et l'orientation peuvent aussi être utilisées pour inclure ou exclure une personne de la communication [30].

Le fait que les participants aient accès à un espace physique partagé signifie que d'autres types d'informations visibles sont disponibles. Les interactions sur le lieu de travail ne sont généralement pas planifiées [58] et l'information visible

représente un mécanisme pour initier de telles communications. Les participants peuvent faire des inférences sur la disponibilité des autres pour la communication en se basant sur l'information visible. La disponibilité de cette information aide dans le processus d'initiation et de terminaison d'une communication. L'environnement visible inclut des informations à propos des objets et des événements dans l'environnement partagé aussi bien que leur configuration spatiale. L'environnement visible fournit une information contextuelle cruciale [59].

2.5.3 L'usage de la vidéo pour véhiculer des informations visuelles

Trois hypothèses sur l'efficacité de la vidéo pour ce rôle sont étudiées et présentées : (a) La vidéo comme support de comportements visibles et de communication non verbale. (b) La vidéo fournit une information visuelle sur la disponibilité des personnes et encourage la communication non prévue et spontanée. (c) La vidéo est aussi un support à l'information proprement visuelle à propos d'objets ou d'événements qui ont une importance pour les tâches collaboratives (la vidéo comme donnée).

L'hypothèse de communication non verbale est que les comportements visibles comme le regard, les gestes, l'expression faciale et la posture peuvent être véhiculés par la vidéo. Il y a trois versions de cette hypothèse : (a) la vidéo fournit des indices cognitifs facilitant le partage de la compréhension ; (b) la vidéo offre des éléments de coordination de la communication ; (c) la vidéo offre des indices sociaux et permet l'accès à l'information émotionnelle. La première hypothèse suppose un usage de la vidéo utilisée comme complément de la parole ne nous intéresse pas, dans le contexte de ce travail. En ce qui concerne l'hypothèse de la coordination de la communication, les systèmes de communication vidéo en général, ne reproduisent pas les processus du face à face. Ils ont tendance à présenter l'image depuis un seul écran, ce qui compromet la direction de la tête et la direction du regard. Le dispositif utilisé pour le MirrorSpace [48], est l'un des plus intéressants de ce point de vue, car non seulement il permet la conservation de la direction du regard (on peut plonger dans le regard de l'autre) mais il introduit la distance au dispositif comme moyen de communication. L'usage de la vidéo change surtout l'issue et le caractère des communications qui requièrent l'accès à l'affect et aux facteurs émotionnels. Les groupes utilisant vidéo ont tendance à s'aimer les uns les autres [60]. Les contraintes technologiques qui limitent la perception du regard et l'utilisation du contact oculaire, contribuent à la difficulté pour la vidéo à transmettre les signaux non verbaux [60].

La seconde hypothèse est que la vidéo fournit une information sur la disponibilité, le mouvement et l'interruptibilité des autres personnes. L'information dans l'environnement visible facilite la connexion pour des communications non planifiées. Deux classes d'applications vidéos ont testées, les hypothèses suivantes : (a) le coup d'œil (glance) qui permet à l'utilisateur de regarder rapidement dans le bureau d'un collègue, pour s'assurer de sa disponibilité à la communication. Et (b) des liens permanents dans lesquels un canal vidéo persistant est maintenu entre deux endroits séparés. Fish & al [23] ont testé différents types de coups d'œil et l'efficacité de chaque type à établir une interaction spontanée. Les résultats ont montré que les participants veulent

contrôler directement quand et avec qui ils se connectent et utilisent le coup d'oeil comme préparation à la communication. La vidéo peut aussi être utilisée de façon continue entre les bureaux de deux collaborateurs distants [23]. Ce lien est sensé approximer le partage physique d'un même bureau. Toutefois, Fish & al [23] ont rapporté que l'usage de ces dispositifs a donné lieu à des interactions brèves plutôt que des communications longues. Les études sur terrain reliant des lieux publics, rapportent une utilisation fréquente de ces liens pour saluer [1]. Ces résultats montrent qu'il manque des preuves que les coups d'oeil et les liens permanents puissent être utiles à établir la connexion. Ces défaillances semblent dues à des facteurs de l'évaluation, ou à des problèmes d'implémentation comme l'impossibilité d'interrompre le lien permanent ou de l'utiliser pour autre chose [57]. En l'absence de mécanismes de contrôle et de notification, l'utilisation de la vidéo pour fournir des informations sur la disponibilité est compromise par les besoins sociaux de préservation de l'intimité, ce résultat rappelle les conclusions que nous avons tirés de l'étude des médiaspace.

Une hypothèse alternative est que le bénéfice majeur de la vidéo est lié à sa capacité à transmettre des informations complexes et dynamiques sur les objets 3D partagés, plutôt que sur les participants eux-mêmes. Cette approche est motivée par le fait que les participants passent la plupart du temps à regarder les objets du travail plutôt que les autres personnes. Ainsi, cette information transmise en temps réel peut être utilisée pour la coordination du contenu entre les équipes distribuées et constitue un contexte physique partagé.

2.5.4 Conclusion sur l'usage de la vidéo pour la communication visuelle

Nous avons présenté les fonctions des informations visibles dans la communication, et les arguments des trois hypothèses à propos du rôle de la vidéo dans la communication interpersonnelle.

A l'exception des communications qui nécessitent l'accès à l'information affective, peu de résultats soutiennent l'hypothèse de la communication non verbale. L'information visuelle change l'issue des tâches dépendant de l'affect et de l'émotion, soutenant l'hypothèse des indices sociaux. Cette hypothèse est très importante dans le cadre présent de la communication familiale, car les liens entre les membres d'une famille sont de nature affective. Les éléments de coordination du processus de communication relativement restent mal transmis par la vidéo. Une explication possible est que les systèmes actuels ne simulent pas avec assez de précision certains aspects de la communication face à face, comme la spatialisation de la vidéo [50].

Malgré l'importance montrée par les études de l'aspect opportuniste des communications, le rôle de la vidéo dans l'initiation de telles communications n'est pas très clair. D'autres facteurs de conception ont aussi besoin d'être pris en compte comme le temps mis à établir la communication, le style de l'initiation et les problèmes d'intimité [57]. Il serait aussi intéressant de voir quelles alternatives technologiques pourraient fournir l'information sur la disponibilité et être un substitut pour l'information visuelle, comme les badges actifs [45]. Il est aussi intéressant de voir comment les technologies de communication asynchrones pourraient substituer partiellement les rencontres spontanées.

Finalement, la vidéo comme donnée est un secteur prometteur. Les travaux récents sur l'hypothèse de la communication non verbale offrent indirectement un support pour les objets et l'environnement partagés. Toutefois, comme pour la connexion opportune, il y a des problèmes sociaux liés à l'intimité et à l'accès qui doivent être résolus pour la vidéo comme donnée.

Globalement, ces travaux suggèrent que le bénéfice de la vidéo est spécifique aux tâches et aux situations, et dépend des dispositifs utilisés. La vidéo est donc utile pour initier des communications opportunistes, pour partager des objets et pour communiquer le rapport affectif et l'état émotionnel. Notre système de communication prendra en compte ces résultats en fournissant des fonctionnalités pour supporter ces types de communications. Pour situer notre projet relativement aux logiciels de communication existant nous allons présenter quelques un de ces logiciels.

2.6 Revue de quelques dispositifs de communication interpersonnelle.

Les dispositifs de communication matériels (téléphones) ou logiciels (messagerie instantanée, vidéoconférence) actuels offrent souvent différents services pour communiquer. Ces services permettent une communication synchrone ou asynchrone, avec différentes quantités d'informations et sont plus ou moins engageants et intrusifs. Toutefois, ils sont organisés comme des collections de services séparés, sans réelle cohérence d'ensemble ni transition entre chaque mode.

Du côté matériel, les téléphones mobiles les plus récents combinent un ensemble de services correspondants à différents niveaux de détail (e.g. identification de l'appelant, SMS, MMS, messagerie vocale, téléphonie et éventuellement visiofonie). Cet éventail de possibilités permet de contrebalancer l'accessibilité permanente liée à la possession du téléphone. Cependant, les transitions entre niveaux sont généralement difficiles, voire impossibles. On ne peut pas prolonger une conversation téléphonique par une conversation textuelle, lorsqu'on entre dans un lieu silencieux, et faire la transition entre téléphonie et SMS. Ou envoyer sur la messagerie tous les appelants excepté ceux dont on attends un coup de fil lorsqu'on est occupé, en liant l'identification de l'appelant et messagerie.

Du côté logiciel, il existe beaucoup de produits qui permettent la communication interindividuelle. Certains sont spécifiquement conçus pour la communication vidéo, et d'autres recouvrent des modalités diverses (textuelle, audio, téléphonie via ip, vidéo, etc)

En ce qui concerne la vidéo, on peut citer iChat [3]. C'est le logiciel d'Apple qui permet la vidéoconférence, il est orienté vers la communication vidéo comme complément de la communication audio. Le mode vidéo permet de se voir et de choisir la position et la taille de cette vue miroir sur l'image de l'autre. Les images des utilisateurs distants, lorsque ces derniers sont plusieurs, sont tournées sur les cotés de façon donner l'impression d'un petit espace 3D.

Les logiciels plus génériques et plus populaires comme MSN Messenger [41] ou AIM [2] (AOL Instant Messenger) sont structurés autour d'une liste de contact. Ils permettent de savoir qui est connecté, quand il se connecte, et qui est disponible à la communication. Leur fonction par défaut (double click sur un élément de la liste) est le chat, qui se transforme en envoi d'e-mail pour les personnes non connectées. MSN remplit aussi la fonction

de connexion vidéo ou audio, permettant de faire de la vidéoconférence, seulement ces fonctions sont présentées comme des extensions du chat. On peut signaler le logiciel Skype [29], bien qu'il ne véhicule que l'audio, il a l'avantage d'être opérant avec des téléphones classiques.

Ces différents logiciels réalisent certes une communication multimodale ou vidéo seule, mais ils n'offrent aucune transition entre les modes. Aucun de ces logiciels ne permet de modifier à sa guise la quantité d'information transmise ou reçue (à part l'image miroir dans iChat), comme l'intelligibilité du son ou la taille de l'image vidéo. Au mieux on peut modifier le volume du son du correspondant, ou diviser par deux la taille de l'image sur son écran. Pour MSN qui supporte des modes multiples de communication et donc des niveaux différents de détail, il n'existe aucune continuité entre les différents niveaux, et il est toujours un peu étrange d'expliquer formellement à l'autre personne qu'on veut couper le lien vidéo et continuer avec le texte. De plus, aucune réciprocité n'est exigée, le correspondant doit lui aussi couper son image de son côté, ce qu'il fait en général.

Ces systèmes ne peuvent s'utiliser pour des connexions longues ou permanentes. Ce type de communication en tout ou rien combiné à la taille fixe de l'image vidéo limite l'usage que l'on peut en faire à une vague simulation d'une communication face à face, c'est-à-dire à seconder la communication en audio. Il est difficile d'imaginer utiliser ce type de connexion de façon permanente, par exemple pour faire du « office share » ou pour partager son salon. Un autre inconvénient à établir une liaison permanente est l'encombrement de l'écran de l'ordinateur par l'image et par l'interface du logiciel, inévitable à moins de consacrer sa machine à cette application. Un autre problème d'intégration lié plutôt au système qu'aux applications elles-mêmes est le contrôle exclusif du périphérique par l'application, qui empêche que l'on puisse utiliser la même image dans plusieurs logiciels.

Enfin, MSN présente certains problèmes de protection de la vie privée. Il permet même de modifier intentionnellement l'état affiché aux autres utilisateurs, donc de se connecter de façon furtive et de pouvoir observer l'état des autres personnes sans être vu. Toutefois, il n'est pas possible de filtrer sélectivement les demandes de connexions et de permettre à certaines personnes et pas à d'autres de demander la connexion, ce qui restreint l'usage à des groupes homogènes.

Bien entendu, tous ces logiciels ne supportent pas des fonctions comme le glances utilisé dans les médiascares, ce qui restreint les initiatives d'établissement de la connexion vidéo, bien qu'ils fournissent une information sur la disponibilité des autres.

Enfin, le fait que ces logiciels soient construits à partir de listes de contacts ou de carnets d'adresse, ils ne tiennent pas compte de l'existence des groupes et se basent sur la pré-existence de liens dans la réalité, prolongés par le logiciel. Par exemple, grâce au logiciel, il est possible de faire la connaissance de quelqu'un par l'intermédiaire d'un autre, comme cela se produit incidemment dans la réalité lorsqu'on arrive chez un ami et qu'on le trouve avec quelqu'un. Mais il faut pour cela que l'intermédiaire invite explicitement les deux personnes à partager la même discussion, ce qui ne se produit quasiment jamais. Un autre aspect qui n'est pas supporté est qu'on ne sait jamais qui connaît qui et qui parle à qui. Parfois, la motivation à se joindre à une discussion tient

justement à la perception d'une discussion préexistante. Grâce à ces logiciels, on peut arriver à la situation étrange où A parle avec B et séparément avec C, et B et C qui parlent aussi séparément, sans qu'aucun ne sache que les trois personnes parlent ensemble. Un dernier problème lié à cette organisation autour d'une liste est la gestion de la liste elle-même, quand on ne désire plus être vu par quelqu'un. Alors que dans la réalité physique les liens qui se desserrent entre les individus se passent progressivement et les personnes disparaissent l'un de la vue de l'autre, dans ce type de logiciel il faut supprimer explicitement la personne de sa liste de contacts, si on ne veut plus qu'elle voie les informations que l'on destine aux autres membres de la liste, tels que les images et les noms dont on se dote, ainsi que les moments où l'on se connecte.

Globalement, c'est principalement sur le modèle du téléphone que sont calqués les communications, et qu'ils soient matériels ou logiciels ces systèmes de communication interpersonnelle proposent des modalités différentes de communication sans transition entre chaque modalité, ni contrôle de la quantité d'information offerte dans chaque modalité. Ils présentent des lacunes dans la protection de l'intimité et ne soutiennent pas la notion de groupe social.

Après ce tour d'horizon théorique et technique, nous allons maintenant présenter le système que nous avons réalisé, tenant compte des différentes conclusions tirées dans cette première partie.

3. SYSTEME DE COMMUNICATION MULTI-ECHELLES

Nous expliciterons les contraintes que respecte le système que nous avons réalisé. Ensuite, nous détaillerons la conception et le développement de ce système. Enfin nous présenterons un scénario d'utilisation mettant en relief quelques fonctions que notre système prend en charge.

Contraintes du projet: multiéchelles, communication proche et sonde technologique

Dans la problématique du sujet de mon stage, nous avons identifié trois points principaux. La conception d'un système multiéchelles, dédié à la communication dans le cadre domestique et familial, ayant les caractéristiques des sondes technologiques

Multi-échelles

Un système de communication peut être qualifié de multi-échelles s'il est capable de transmettre un niveau d'information variable et permet des transitions fluides entre ces niveaux. Cette définition est à rapprocher de la notion d'espace géométrique 2D zoomable à l'infini (Perlin & Fox 93). En particulier, les notions de zoom sémantique et de zoom continu dans un espace 2D correspondent exactement aux concepts de niveau d'information (niveaux d'ordre sémantique) et de transitions fluides.

Pour tenir compte de ces caractéristiques, la communication entre deux foyers doit être envisagée de manière synchrone (i.e. directe, en temps réel) ou asynchrone. Elle doit s'adapter au contexte, et notamment selon leur disponibilité, les utilisateurs préféreront une communication plus ou moins directe et plus ou moins riche. Dans certains cas, l'utilisation du système de communication doit être considérée comme une activité de premier plan. Dans d'autres, elle doit être une activité secondaire parmi d'autres.

Pour faire varier le niveau de détail associé à un flux d'images, on peut faire varier la taille mais aussi le taux de rafraîchissement des images transmises et/ou affichées. Pour enrichir les images on peut aussi utiliser des procédés de composition spatiale ou temporelle des images. Des modes de communication légers peuvent s'appuyer sur un principe de rendu stylisé. D'autres techniques pourront au contraire enrichir le message transmis, en combinant par exemple des images plus anciennes à la dernière image capturée. Ces différents procédés d'enrichissement ou d'allègement de l'image transmise doivent être articulés de façon à passer d'un niveau à l'autre d'une façon fluide et à constituer une adaptation contextuelle.

Enfin, le système peut combiner différents média (e.g. image, son, texte) et utiliser différents capteurs (e.g. détecteur de mouvement, capteurs de pression) pour déterminer le contexte d'utilisation, et s'affranchir des périphériques traditionnellement associés aux systèmes informatiques (e.g. clavier, souris). On peut également utiliser la caméra comme périphérique d'entrée, en s'appuyant sur des techniques classiques de vision par ordinateur.

Communication dans le cadre domestique et familial

La famille est d'abord un groupe d'individus qui est lié par une connaissance et une confiance mutuelle. Pour tenir compte de cette caractéristique, le système doit être capable de gérer une communication multisite, et créer un réseau entre tous les membres.

Du point de vue de l'usage du système, un des problèmes à résoudre concerne le choix de la métaphore et des techniques de visualisation des images provenant des différents sites. La plupart des systèmes reposent sur la métaphore de la fenêtre ou du miroir et ne sont pas adaptés à des communications mettant en jeu plus de deux sites.

La communication entre les membres d'une même famille est quasi exclusivement informelle. Aussi l'information sur la disponibilité des autres est cruciale, et le besoin de perception périphérique de l'activité de chacun des membres doit être supportée par le système, afin de coordonner leurs échanges. Cependant, si une liaison vidéo permanente paraît la meilleure solution pour assurer la communication de l'information sur la disponibilité des personnes, il est évident que la vie privée de chacun doit être préservée. L'un des inconvénients pour les familles du VideoProbe a été qu'une fois qu'une photo était prise, ils n'avaient aucun moyen d'empêcher sa diffusion. Il est essentiel pour assurer l'adoption du dispositif de communication dans le contexte familial d'assurer à chacun des moyens de protéger son intimité tout en lui permettant de la partager. Un autre aspect de la coordination des échanges est l'appel explicite. En effet, si on peut se contenter de partager incidemment sa vie avec ses proches, parfois on a besoin de les solliciter pour des communications de premier plan, de les appeler ou au moins de leur laisser un message.

Enfin, le lien familial est aussi souvent le lieu d'échanges autres que de communication ou d'activités; on partage souvent ses photos, ou bien on montre aux autres membres des films ou des vidéos enregistrées. En général, des objets de toutes sortes circulent entre les membres, intentionnellement ou incidemment et conduisent à la constitution d'un contexte partagé. Le système doit pouvoir servir à ce type de circulation.

Sonde technologique

Le processus de conception de ce système de communication doit être alimenté par l'utilisation de sondes technologiques. Pour cela, le système doit être instrumenté afin de récolter des données utiles sur l'usage qui en est fait. Ces données doivent d'abord mesurer l'impact que son introduction dans le contexte domestique a sur les participants, et rendre compte de l'efficacité du système à rapprocher les différents membres. Ensuite, les données peuvent servir à comprendre comment le système est utilisé afin de pouvoir l'améliorer. Enfin, les données récoltées peuvent être étudiées sous leur aspect social, anthropologique et psychologique pour mieux comprendre le fonctionnement des groupes familiaux distants à travers leurs interactions.

Pour tenir compte de ces aspects, le système peut enregistrer trois types de données différentes : Le résultat final que l'utilisateur voit sur son écran peut servir à analyser comment se passe l'interaction. Capturer les sources d'images utilisées et les valeurs des capteurs permet de rejouer les séquences enregistrées tout en changeant les algorithmes qui régissent l'interaction. On peut aussi enregistrer des données quantitatives et qualitatives, comme le nombre et la durée des contacts ou la proportion de contacts qui aboutissent à un échange, afin d'analyser plus globalement les résultats des interactions.

3.1 Conception et développement du prototype :

La conception du prototype, que nous avons appelé Pêlemêle, devait donc tenir compte de la communication domestique et familiale par une approche multiéchelle, et dans une moindre mesure des caractéristiques de sonde technologique.

Pour la communication intime et informelle, s'inspirant du vidéoProbe, l'objectif de la conception était de tenir compte de deux aspects : Le premier est le partage accidentel de vidéo des personnes, d'objets et de situations de tous les jours. En effet, le vidéoProbe a révélé l'importance pour les familles des images illustrant les petits moments de la vie quotidienne que l'on ne pense pas à capturer intentionnellement. Ce type de partage similaire à l'« office share » soutient l'aspect opportuniste de la communication, et doit pouvoir se transformer en une communication intentionnelle. Le second, le partage intentionnel de vidéos ou d'images, comme on le fait physiquement en montrant les photos des vacances aux autres membres de la famille. Pour cela, nous avons pensé que le panneau sur lequel on accroche les photos, que l'on trouve parfois chez les familles est un bon modèle pour le fonctionnement automatique du système : Lorsqu'il n'est pas utilisé pour la communication intentionnelle, l'appareil peut présenter soit des photos ou des vidéos intentionnellement partagées (par exemple, des photos de vacances) soit des vidéos enregistrées automatiquement dans les différents foyers.

Bien que ce soit à l'intérieur d'une même famille, cette forme de partage d'espace privé pose des problèmes de protection de l'intimité. En fait, il n'est pas nécessaire d'avoir une liaison vidéo continue, et il est plus utile de détecter un certain type d'évènement intéressant à partager. En dehors de ces moments, on peut protéger l'intimité des familles en réduisant la quantité d'information transmise.

Ce changement de quantité d'information relativement au contexte correspond à la communication multiéchelles. Nous avons pensé à utiliser une forme de tableau de photos « réactives », où certaines photos ou vidéo sont des objets partagés, tandis que d'autres représentent des lieux partagés. En fonction du contexte, les images pourraient changer de taille et de position, mais surtout de quantité d'information. Pour effectuer ces changements de niveaux d'informations, nous avons décidé d'utiliser des techniques de composition temporelles ou des rendus stylisés, comme la peinture à huile ou le flou.

Il est important étant donné le contexte de communication familiale, c'est-à-dire assez proche voire intime de pouvoir se regarder dans les yeux. Aussi au niveau matériel, ce prototype était destiné à fonctionner sur un dispositif qui permet de supporter la communication du regard, similaire à celui du MirrorSpace, présent actuellement au LRI. La prise de vue du système MirrorSpace est assurée par une caméra USB démontée dont la partie optique a été placée au centre d'une plaque de verre elle-même placée devant l'écran. Cette disposition permet aux utilisateurs du système d'être très proches du dispositif tout en étant à la fois vus par la caméra et capables de voir les images affichées.

La première question que pose la conception est la présentation des images sur l'écran.

3.1.1 Présenter et partager des photos réactives

L'idée du tableau de photo de laquelle on est parti pose le problème de la présentation. En effet, comment doit-on disposer et présenter les images et les vidéos partagées ou enregistrées et les images venant des autres participants ? Nous avons envisagé et essayé plusieurs modèles avant de choisir.

Nous sommes partis d'une disposition aléatoire ayant une évolution lente et aléatoire. Cependant, la disposition et l'évolution aléatoires ont comme inconvénient principal l'impossibilité de prévoir quelle est la place de chaque image, et à fortiori la position de l'image de la caméra locale et des lieux distants. Un autre problème lié à cette méthode est que les images locales et distantes peuvent être cachées et donc l'information sur les autres personnes indisponible. Ce type d'évolution est aussi problématique pour l'aspect multiéchelle du système, changer le niveau d'information et la taille des images entre en conflit avec la disposition aléatoire, étant donné que si une image grandit elle peut en cacher d'autres. De cette première disposition c'est surtout l'aspect évolutif qui est intéressant et a été retenu pour la suite.

Comme autre modèle, nous avons pensé à un modèle en spirale. Ce modèle est intéressant car il permet une représentation du temps, ainsi une vidéo qui vient d'arriver ou une connexion vidéo live peut apparaître au centre de la spirale et pousser les autres images vers la périphérie et les images trop anciennes atteignant le bord de l'écran pourront être supprimées. Son inconvénient principal est qu'il semble peu adapté aux possibilités de communication synchrone et en particulier multisite. De plus, changer la taille des éléments affichés était assez compliqué dans ce modèle. De cette seconde disposition nous avons retenu la nécessité d'offrir une bonne représentation du temps.

Dans la littérature scientifique, l'exemple du médiaspace CoMedi [15] utilise un « porthole », une mosaïque présentant les différents

endroits, en fisheye, c'est-à-dire de forme hyperbolique, pour supporter l'information sur la disponibilité du groupe. Cette présentation est intéressante parce qu'elle permet de modifier le niveau de granularité en élargissant l'image de la personne qui nous intéresse sans perdre complètement l'information sur les autres personnes. Le modèle hyperbolique du « porthole » du médiaspace CoMedi évoqué précédemment est aussi une solution intéressante pour disposer les images sur l'écran. Cependant, il est fait pour se concentrer sur une seule image et n'est pas adapté pour augmenter la granularité de deux images simultanément, en particulier si ces images sont situées aux deux extrêmes de la grille.

Pour la disposition des images sur l'écran, l'utilisation d'un modèle géométrique s'est imposée. En effet, le système était plus logique et plus intuitif à partir du moment où des informations implicites étaient représentées géométriquement. Notre choix s'est porté sur une disposition circulaire et dynamique (Figure 5a). Cette disposition permet aussi directement d'avoir une notion de périphérie et de centre, véhiculant naturellement les fonctions de perception périphérique et d'activité centrale de la façon suivante : Au repos, les images sont placées en périphérie et lorsqu'une image est activée, c'est-à-dire quand une personne veut utiliser le système ou quand le système choisit une vidéo à montrer, l'image se déplace au centre et grandit. Ce choix était justifié sur le fait que le point d'attention d'une personne consultant une vidéo est porté au centre de l'écran. Cette disposition permet de changer de niveau d'information et d'agrandir certaines images sans qu'elles interfèrent avec les autres. De plus, elles associent une position sur le cercle à un lieu ou à une image. Toutefois, ce choix avait plusieurs inconvénients. Le plus important est la faible possibilité d'évolution de la disposition : En réservant le centre à l'image active, on ne peut faire évoluer les images que le long du cercle, ce qui rompt l'association entre lieu et position. L'autre inconvénient de cette disposition est la mauvaise exploitation des coins de l'écran, puisque le modèle est circulaire, il aurait fallu un écran circulaire. De plus la vidéo étant carrée, il en résulte aussi une mauvaise exploitation de l'espace central. Ce problème a été contourné en utilisant une disposition en carré arrondi à la place de la disposition circulaire. Nous avons pensé aussi à utiliser un modèle semi-circulaire (Figure 5b), de la forme d'un arc en ciel, car sur la machine qui a servi au développement l'écran est de format 16/9, la disposition en cercle ou en carré arrondi étant plus adaptée aux écrans de format 4/3, comme celui des vidéos utilisées. Un dernier problème est lié à l'utilisation dans le temps. En effet, il est important de savoir le moment où une vidéo arrive et l'ordre d'arrivée de plusieurs vidéos, et le moment où une personne était devant l'appareil et le temps écoulé depuis lequel elle n'y est plus. Nous avons amélioré ce modèle de présentation en introduisant une troisième dimension, utilisant la perspective : les vidéos qui arrivent du même site sont placées sur une ligne de fuite les plus récentes devant les plus anciennes en fonction de leur heure d'arrivée. Avec le temps, ces images dérivent et rétrécissent pour finir par atteindre le point de fuite, au bout d'une période plus ou moins longue en fonction de l'échelle temporelle utilisée (Figure 5c). De plus, l'image distante des différents sites se met aussi à dériver à partir du moment où il n'y a plus de présence détectée sur ces sites. Quand les vidéos capturées ou échangées sont assez fréquentes, les images affichées successivement font apparaître les lignes de fuites de la perspective, et on voit les moments où les interactions ont eu lieu (par contraste aux vides des moments sans

interactions). Cette amélioration au modèle de disposition permet d'une part de savoir quand chaque vidéo est arrivée, les moments où des interactions ont lieu, mais aussi approximativement depuis combien de temps personne n'est présent sur chaque site. Le modèle résultant évolue avec le temps tout en donnant une assez bonne représentation des durées.

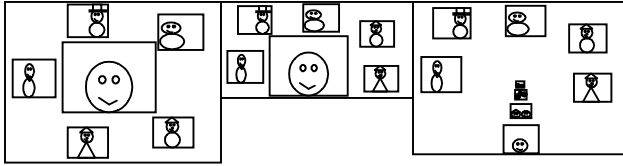


Figure 5. (a) Présentation circulaire, (b) semi-circulaire et (c) Dérive temporelle

Une autre question se pose du fait de la propriété communicante du dispositif, à savoir quelle correspondance existe entre les différents pèlemèles connectés ensemble ?

Nous avons vu que la direction du regard joue un rôle dans l'attention conjointe dans la communication, et que les individus sont très bons dans la prédiction de la direction du regard des autres. Il paraît important alors que les pèlemèles aient des présentations identiques pour permettre aux participants de deviner quelle est l'image ou le lieu que le correspondant regarde en ce moment. Un autre problème est relatif au partage de vidéos et de photos. Lorsque le système est en veille il nous paraît important que la lecture d'images ou de vidéos soit synchronisé sur tous les pèlemèles afin de constituer un partage d'objet et devenir potentiellement le sujet d'une communication. Une consultation automatique de cette sorte permet en regardant un poste de savoir ce qui se passe sur les autres postes, et si une personne y est présente, ce qu'elle peut voir actuellement.

3.1.2 Visualisation et communication multi-sites

A part la disposition globale des images sur l'écran, plusieurs questions se sont posées sur la métaphore et les techniques de visualisation adaptée au fait d'avoir plusieurs sites.

La méthode la plus couramment choisie pour le multisites est la mosaïque dans laquelle les différents intervenants sont placés cote à cote sur une grille.

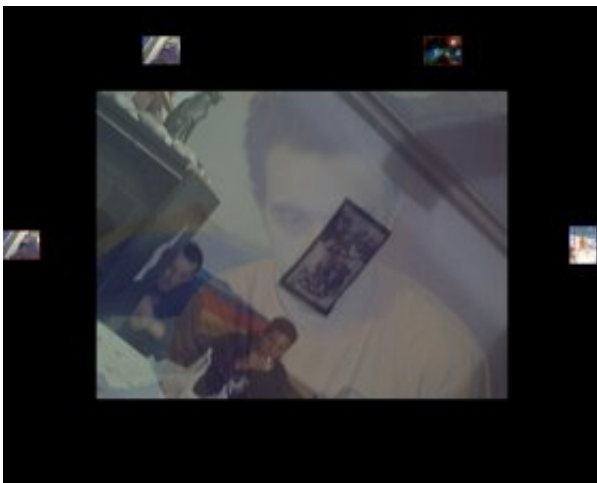


Figure 6. Utilisation de la superposition par transparence

MirrorSpace résout ce problème de façon originale, en superposant les images des différents sites par transparence. C'est la méthode que nous avons choisi et implémenté (Figure 6). Cette méthode est intéressante bien qu'elle pose un problème quand les images à superposer sont plus de quatre, il devient difficile de distinguer les visages et surtout qui est dans quel contexte. Nous avons pensé à ce stade à plusieurs modes de superposition, comme l'incrustation utilisée dans l'hypermirror ou la multiplication, ou encore à des méthodes qui améliorent le contraste du résultat final, comme le truchement d'histogramme (Vernier & al). Nous avons pensé aussi à une superposition partielle, cette solution semble intéressante mais elle est difficile à mettre en œuvre.

Nous avons aussi testé une solution inspirée par le multiblending expliqué par Baudisch [4]. L'image de la caméra locale lorsqu'elle était superposée à l'image distante était préalablement transformée en image « en verre ». Pour faire cela, nous avons simplement utilisé un filtre passe haut (emboss) pour ne garder que les contours, puis superposé cette image en utilisant un mode « linear light » proche de l'addition. Plusieurs utilisateurs ont trouvé que le résultat était assez esthétique. Cette approche est justifiée par le fait que le système visuel humain traite de façon différenciée les hautes et les basses fréquences (voies visuelles dorsale et ventrale).

Sur ce point, nous avons remarqué que la plupart des dispositifs utilisant la vidéo pour la communication interpersonnelle donnent systématiquement plus de détail pour le correspondant que pour l'image locale (cela est vrai pour les logiciels MSN et iChat, ainsi que les systèmes de vidéophonie). La justification est que l'image du correspondant est le lieu de beaucoup plus de prélèvement d'informations (affectives, sociales, etc) tandis que l'on connaît mieux ces informations pour soi-même et l'image locale sert surtout à savoir ce qu'on donne à voir, et est utile par exemple pour se recadrer. Dans le système que nous avons conçu, l'application de cette remarque peut passer par l'utilisation de l'effet « verre » présenté juste précédemment ou tout simplement par l'utilisation d'un alpha plus important pour le correspondant que pour l'image locale pour la superposition par transparence.

3.1.3 Détecter des situations d'usages

Un des aspects des systèmes de communication multiéchelles est leur capacité à s'adapter à différentes situations d'utilisation, pour cela il faut pouvoir les détecter.

Bien que nous devions au départ utiliser des capteurs divers, par exemple des capteurs de distance, tel que ceux utilisés pour le MirrorSpace. Notre choix s'était porté sur l'utilisation de l'image en tant périphérique d'entrée car l'information qu'elle porte est très riche, et cette utilisation nous dispense d'un périphérique supplémentaire. Seulement utiliser le périphérique de capture principal, c'est-à-dire la webcam, est aussi très complexe et très coûteuse en temps de calcul.

Etant donnée l'expérience du VidéoProbe pour lequel était utilisée la détection du changement de scène, nous avons pensé à détecter les mêmes événements, c'est-à-dire que quelque chose de nouveau est devant la caméra, pour déclencher le début de la capture. Cependant, cette méthode a quelques inconvénients. En effet, il y a trois événements qui peuvent résulter en une différence entre les images actuelle et de référence : la présence d'une personne devant la caméra, le déplacement de la caméra ou

un changement de la luminosité. En utilisant le dispositif prévu, dans lequel la caméra est fixée au centre de l'écran, le deuxième cas est éliminé. Reste le dernier cas, qui n'est pas totalement résolu. Toutefois, il y a une début de solution mais que nous n'avons pas testé : dans la plupart des situations, un changement de luminosité affecte toute l'image, alors que la présence d'une personne affecte une partie seulement de l'image.

Nous avons choisi et implémenté la méthode utilisant la différence entre l'image courante et l'image de référence permettait cependant de détecter la présence dans la plupart des cas. En utilisant la différence entre deux images successives il est aussi possible de détecter le mouvement.

Au départ, la détection de mouvement était utilisée pour capturer la vidéo : la quantité de mouvement était accumulée et quand un seuil était dépassé, la capture se déclenchait et en l'absence de mouvement, l'accumulateur se vidait progressivement. Le problème de cette méthode est qu'elle n'enregistre que les scènes de mouvement, et un utilisateur qui fait peu de mouvements devant la caméra était rapidement ignoré. De plus, la quantité de mouvement (la surface de l'image modifiée) n'est pas une information très pertinente pour l'utilisateur, aussi nous avons introduit l'utilisation du temps, plus intuitive, et nous avons fini par combiner les deux informations de présence et de mouvement de façon un peu similaire au VidéoProbe. La combinaison que nous avons utilisée diffère cependant sur plusieurs points, étant donné que nous souhaitons capturer de la vidéo et non des images fixes : la présence est détectée à partir du moment où la différence entre image de référence et image actuelle persiste plus de 2 secondes. De façon similaire le mouvement est détecté si la différence entre les images successives persiste pendant plus de 2 secondes. Si cette dernière est nulle pendant plus de 10 secondes, l'image de référence est mise à jour. Ces choix sont plus justifiés par l'utilisation qui est faite des événements détectés que par la détection d'une « vraie » présence ou d'un « vrai » mouvement.

Par ailleurs, nous avons rencontré aussi un problème technologique dans l'utilisation de la différence d'image pour la détection. Ce problème est lié à la sensibilité de la caméra aux conditions lumineuses : pour la lumière naturelle et celle des ampoules à incandescence quand elle est suffisamment forte l'image qui en résulte est assez stable tandis que pour la lumière des néons et des ampoules économiques l'image est légèrement instable. Mais quand la lumière est faible l'image est très parasitée. De plus la stabilité de l'image dépend de la caméra utilisée. Il en résulte pour la détection utilisant la différence la nécessité d'un ajustement des seuils de réactions. Cet ajustement peut-être fait manuellement, en fonction de la caméra et mis assez haut par défaut pour tenir compte des mauvaises conditions de luminosité, il en résulte que le système de détection « voit mal » au sens où la modification de l'image doit être importantes pour être prise en compte.

Vers la fin de mon stage, nous avons exploré et implémenté un autre type d'évènements : grâce à la librairie OpenCV [34], nous avons pu utiliser la détection et le suivi de visage pour extraire trois événements différents. Les deux premiers sont la présence et le mouvement d'un visage face à la caméra. Le troisième est la distance de ce visage : étant donné la faible variation de taille des visages humains adultes, en particulier de la largeur, nous avons pu en déduire la distance du visage à l'appareil. Cette dernière mesure est à prendre avec précaution et serait faussée si un enfant

se sert de l'appareil. Cette difficulté pourrait être contournée en utilisant la reconnaissance de visage fournie par OpenCV pour corriger la mesure.

La reconnaissance des visages est loin d'avoir pour seule application la correction de la mesure de distance, et bien que nous n'ayons pas conçu des applications pour ce type d'information, nous avons cherché à la tester pour en comprendre les limites. Cette reconnaissance est assez efficace pour un petit nombre de visages, comme c'est le cas dans le contexte de l'usage domestique, et serait très utile à notre système de communication.

Les événements auxquelles nous nous sommes intéressé et ces méthodes pour les détecter, sont principalement orientés pour supporter une interaction intentionnelle de l'utilisateur avec le système de communication. Notre système ne se limite pas aux usages intentionnels et doit permettre des usages « accidentels ». Malheureusement, les méthodes utilisant la détection du visage est particulièrement sensible à la luminosité et à l'orientation du visage. Et même dans de bonnes conditions de luminosité, ce type de détection n'est pas approprié pour la présence d'une personne qui n'interagit pas avec le système, et qui a peu de chance de se placer assez près et assez en face de l'appareil. Cependant, pour la détection utilisant la différence d'image, elle est bien plus robuste, même si son utilisation en situation réelle n'a pas été évaluée. Utilisée pour enregistrer des vidéos, les enregistrements sont assez fréquents. Mais il n'est pas sûr que les vidéos résultantes soient suffisamment intéressantes. En fait, il est difficile de dire à priori ce qu'est un événement intéressant à partager et encore plus difficile de le capturer complètement afin d'avoir une vidéo intéressante que l'on voudrait garder. Ce problème nécessite d'avantage d'investigations qui amélioreraient nettement l'intérêt de l'enregistrement et du partage automatique.

Pour partager la vidéo comme donnée dans la communication, l'utilisation de la détection de visage pose certains problèmes. Par exemple, on ne peut se contenter de mettre un message écrit sur un papier devant le système pour que celui-ci transmette le message. L'utilisation de la différence par contre peut servir plus facilement à faire cela, il suffirait d'agiter le papier puis de le placer devant la caméra et le système se charge de transmettre l'image aux autres utilisateurs.

Un problème très important que nous n'avons pas résolu au cours de ce stage est la possibilité d'interaction directe avec le système pour la consultation des vidéos enregistrées. En effet, il est évident que si l'on reçoit une vidéo et que l'on a envie de la regarder il faut fournir un moyen de la déclencher manuellement, sans cela il faudrait attendre que le système la sélectionne de lui-même, ce qui peut prendre beaucoup de temps. Nous avons pensé pour cela à deux solutions : la première solution inspirée de l'usage classique de la souris était de disposer d'un écran tactile et de toucher l'élément que l'on veut visualiser. Cette solution est assez intuitive, mais elle a l'inconvénient d'un équipement spécialisé incompatible avec le dispositif que nous avions prévu. Elle est plus facilement réalisable en combinant à l'écran tactile des techniques de vision par ordinateur pour simuler un point de vue virtuel (e.g. le centre de l'écran) à partir des images de plusieurs caméras.

Nous avons aussi tenté d'utiliser des techniques rudimentaires d'interaction par l'image comme activer la zone où il y a le plus de mouvement ou celle qui correspond à la moyenne des mouvements. Nous avons même développé ces solutions par

l'intermédiaire d'un spot que l'on est censé pouvoir placer sur une zone particulière. Malheureusement, le spot restait difficile à diriger, ces tentatives n'ont pas été très concluantes, et nous ont découragé à aller plus loin.

3.1.4 Préserver l'intimité

Un autre aspect important du multiéchelles que nous avons exploré est la réduction d'information. Cet aspect est essentiel pour la protection de la vie privée des individus.

En plus de la réduction de la taille de l'image ou de sa fréquences d'affichage, il existe plusieurs façon de réduire la quantité d'information portée par une image. Notons que du fait de nature dynamique des images vidéo, celles-ci sont perçues comme plus détaillées que les images fixes de la même taille et de la même résolution, car elles permettent de faire des inférences à partir des images successives pour extraire un plus grand nombre d'informations. On peut signaler aussi que les différents filtres n'éliminent pas tous les mêmes types d'informations, par exemple le filtre de publication utilisé dans le médiaspace CoMedi [15] filtre uniquement les informations jugées socialement indésirables, en se basant sur une base d'image socialement correctes.

Les filtres que nous avons utilisés avaient pour objectif de diminuer le détail des images et peuvent être considérés comme des filtres passe bas.

D'abord, nous avons utilisé le flou gaussien pour effectuer cette réduction d'information. En appliquant un flou sur la vidéo, on peut distinguer qu'il y a quelqu'un ou quelque chose, sans savoir qui il est ou quelle est l'activité filmée en fonction du niveau de réduction utilisé. Le flou peut-être considéré comme équivalent à une réduction de la résolution ou une pixellisation. Cependant le flou a l'avantage esthétique de préserver la résolution d'origine et de donner une image plus. Toutefois, le flou gaussien est assez gourmand en temps de calcul comparé à la réduction de résolution.

Nous avons exploré aussi d'autre façon de filtrer l'information visuelle, par un rendu stylisé. Nous avons pensé à deux types de rendu : le premier basé sur la distorsion de l'image est similaire aux vitres de salles de bain. Et le deuxième est semblable à l'effet « peinture à huile » que l'on trouve dans les logiciels de retouche d'image du type de Photoshop. Le premier effet n'a pas été développé car nous n'avons pas trouvé suffisamment d'informations sur comment le réaliser. Le dernier filtre a fait l'objet de développements et a été intégré dans le système. En fait, pour réaliser cet effet de peinture à huile, il faut remplacer la couleur de chaque point de l'image par la couleur la plus fréquente dans une zone circulaire autour de ce point. Nous avons développé deux versions de ce filtre, la première basée sur l'information de couleur et la seconde sur l'information de luminance. La seconde version du filtre a été retenue pour être intégrée car elle avait un résultat plus lisse. Lors de l'intégration de ce filtre nous avons rencontré un problème dû au fait que ce filtre est extrêmement coûteux en temps de calcul. Pour que ce filtre cache suffisamment d'information et préserve l'intimité il fallait que l'effet soit très prononcé et le coût en temps de calcul devenait extrêmement élevé pour générer le type de filtrage qui nous intéressait. Nous avons d'abord pensé à exporter le calcul sur une autre machine, comme c'est le cas dans le mediaspace CoMedi [15] pour le filtre de publication utilisé. Mais étant donné

l'objectif de concevoir des technologies ad hoc, nous avons préféré ne pas retenir cette possibilité, d'autant plus que cela pouvait encombrer le réseau et introduire du délai dans les communications synchrones. Pour résoudre ce problème nous avons décidé d'utiliser une solution adaptée à notre usage : nous utilisons ce filtre pour cacher une image qui variait peu et correspondait à l'absence de toute personne. Aussi notre solution était de découper l'image en petits morceaux et de faire un morceau du calcul à chaque fois. Cette solution a pourtant l'inconvénient de donner des résultats peu esthétiques lors des transitions. En effet lorsqu'une personne vient se placer devant le système, la présence est détectée et le filtre est progressivement supprimé, mais ce découpage de l'image devient visible car la mise à jour de chaque carré résulte en un carré discontinu par rapport à ses voisins.

Enfin, Les résultats du filtre peinture à huile ne sont pas de la même qualité esthétique que celui nommé « Aquarelle » dans Photoshop. Nous avons cherché à un moment à améliorer le filtre produit et cela était possible de plusieurs façons. Nous avons pensé à l'idée d'un filtre entre la peinture à huile et le blur, ou plus exactement à une classe de filtres combinant ces deux filtres. Cependant la conception du filtre peinture à huile est seulement un des éléments du système de communication.

3.1.5 Enrichir l'image par le temps

Si il est possible de réduire grâce à filtres précédents la quantité d'informations transmises par les images sans modifier leur taille ni leur résolution, il est possible aussi de l'augmenter.

On peut enrichir le message transmis, en combinant par exemple des images plus anciennes à la dernière image capturée. De cette façon, le passage d'une personne laisserait des traces. Ce principe de composition temporelle amène à se poser deux questions : d'abord quel type de combinaison il faut utiliser, ensuite quelles images plus anciennes faut-il garder. Nous avons développé et testé différentes solutions pour ces deux problèmes. Pour le premier problème, nous avons essayé deux possibilités de combinaisons basées sur la superposition appelée alpha blend. La première possibilité était de superposer les n dernières images de façon récursive : à l'image la plus ancienne on superpose l'image suivante avec un alpha égal à 50% pour chaque image, puis au résultat on superpose celle d'après et ainsi de suite jusqu'à l'image actuelle. Le résultat de cette façon de procéder est que dans l'image finale, la dernière image a comme importance 50%, la précédente 25%, la suivante 12,5% et celle de rang n a $100/2^n$ %. En pratique, le résultat ressemble à la rémanence vidéo (appelée aussi motion blur), présente une traînée lorsqu'il y a des mouvements, si on prend un intervalle temporel assez faible (moins de 0.5 sec) entre chaque image. Ce qui nous intéressait c'est sa capacité à enrichir l'image en donnant une perception des instants précédents, aussi son utilisation devait être faite avec un intervalle temporel assez important (de l'ordre de 2-3 secondes). Le problème avec cette méthode d'enrichissement est qu'elle ne permet d'avoir plus de trois images mélangées, car au-delà les images ont une participation de moins de 7%, c'est-à-dire qu'ils ne sont plus perceptibles. Son avantage est de donner une certaine perception de l'ordre temporel des images, puisque la plus actuelle sera la plus visible et plus on recule dans le temps moins l'image est visible. Nous avons aussi exploré une autre possibilité, celle de superposer les n dernières images mais de façon à obtenir à la fin une contribution de $100/n$ % pour chacune. Pour cela nous

avons compensé l'effet d'atténuation du aux compositions successives : le premier alpha blend se fait avec $\alpha = 1/2$, le deuxième $1/3$, le troisième $1/4$, etc. Le résultat de ce type de composition est que toutes les images ont la même participation à l'image finale. L'avantage de procéder de cette façon est qu'on peut voir plus de trois images et en distinguer jusqu'à dix différentes. L'inconvénient de cette méthode est qu'elle ne permet pas de savoir quelle est l'image la plus récente et laquelle est la plus ancienne. Un autre problème révélé à l'usage est qu'elle favorise les positions statiques. En effet, si dans un mouvement sur dix images successives trois correspondent à la position de départ, quatre correspondent aux différentes positions du mouvement et trois correspondent à la position d'arrivée, on verra principalement les positions d'arrivée et de départ car leurs occurrences vont se renforcer les unes les autres et occulter les images du mouvement. Cette méthode semble pourtant intéressante lorsqu'on veut superposer des instants qui ne succèdent pas, par exemple toutes les quelques minutes. En général, l'utilisation des deux méthodes précédentes nous suggère que le mieux serait d'avoir une composition entre les deux.

Un dernier problème relatif à la méthode de superposition tient à l'utilisation du alpha blending pour la composition des images. Ce type de superposition donne à l'image finale un aspect terne et gris, qui ne facilite pas la distinction entre les images initiales. Mais il existe d'autres types de superpositions qui sont plus adaptés au contexte de notre utilisation. Nous avons pensé en particulier à la superposition par incrustation, car si le dispositif ne change pas de place, cela signifie que le fond est toujours le même et qu'on pourrait superposer uniquement les parties nouvelles de l'image. Cette méthode n'a pas été développée même si des modules allant dans ce sens ont été élaborés, car l'extraction du fond et du premier plan restent des techniques de vision par ordinateur trop complexes pour entrer dans le cadre de ce stage, étant donné que ce n'est pas le sujet central.

La deuxième question posée par l'enrichissement temporel des images est moins technique mais plus complexe. Dans le traitement de la première question, nous avons supposé que les images alimentant les procédés expliqués étaient prises à intervalles réguliers. Cette supposition limite l'utilisation de ce procédé soit à une fenêtre temporelle assez petite, de l'ordre de quelques secondes, soit à une fenêtre trop vaste ou les quelques images prises au hasard et superposées ne sont pas toujours représentatives de l'activité qui a lieu pendant ce temps là. Savoir quelles sont les images intéressantes à superposer dépend aussi de l'usage que l'on veut en faire. Dans notre cas, c'est principalement représenter l'activité et la disponibilité de l'autre personne. C'est l'idée du résumé généré automatique qui se pose et qui n'a pas été résolue dans le contexte de ce stage.

Cependant, nous avons pensé à une autre façon d'utiliser la superposition temporelle, mais que nous n'avons pas testé ni évalué, inspirée par l'installation de Vincent Levy appelée Le panneau du temps qui passe. Dans cette installation, neuf images sont affichées en mosaïque : celle de l'instant actuel, une seconde avant, une minute avant, une heure avant, etc. Pour la composition temporelle nous pouvons mélanger à l'instant actuel les images provenant de quelques secondes avant, quelques minutes avant, etc. La différence avec cette méthode est qu'on ne va pas supprimer la plus ancienne des images pour rajouter une plus récente, mais on aura plutôt une modification de toutes les images

(puisque l'instant actuel a changé, les autres qui y sont relatifs aussi ont changé).

3.1.6 Lecture automatique et communication différée

Pour remplir la fonction de consultation automatique, nous avons exploré deux points :

Le premier est la lecture à vitesse variable de vidéos enregistrées. Nous nous sommes intéressés à ce point car contrairement aux images fixes où on peut jeter un simple coup d'œil ou passer plusieurs minutes à les regarder, voire à les afficher physiquement ou en arrière plan d'un bureau d'ordinateur, la consultation des vidéos pose un problème de temps et de durée, particulièrement lorsqu'il s'agit de vidéos brutes et non montées. La lecture en accéléré permet d'avoir un aperçu de qu'on ne veut pas voir, tandis que le ralenti nous donne le temps de prendre encore plus d'informations dans un flux qui passe. L'un des composants développés avait donc cet objectif.

Si nous nous sommes intéressés à ce point c'est pour deux raisons. D'abord pour explorer le problème du résumé automatique : en se basant sur certains paramètres nous avons voulu faire un système qui va lire la vidéo en fonction de l'intérêt du moment visualisé : vite lorsque rien ne se passe, lentement lorsque quelque chose de rapide se passe. Ce qui était particulier dans ce composant c'est le fait qu'il passait de façon fluide d'une vitesse à l'autre.

Le développement de ce composant a eu une continuation imprévue : nous avons alors pensé à l'application de ce procédé de lecture à vitesse variable à la vidéo capturée par la caméra. Ce procédé du temps réel à vitesse variable a été appliqué pour obtenir des transitions fluides entre la diffusion différée et la diffusion directe de l'image. On pouvait alors ralentir le flux de diffusion direct et passer d'une façon fluide à une diffusion différée et réciproquement. Dans notre système de communication, lors de la présence d'une personne, nous avons choisi de diffuser son image en différé. Cela permet principalement d'éviter qu'une personne ne diffuse son image alors qu'elle la juge gênante. Elle a un délai d'une dizaine de secondes pour se rendre compte qu'elle est filmée et de sortir du champ de la caméra. A partir du moment où elle n'est plus visible par la caméra la diffusion est arrêtée, et comme la diffusion est faite en retard les dix dernières secondes ne sont pas diffusées. Quand la personne désire passer de la simple présence à la communication utilisant l'appareil, le retard est progressivement comblé en accélérant le flux pour passer à une diffusion en direct. Réciproquement, lorsque la personne veut arrêter la communication et revenir à une simple présence, le flux est retardé en le ralentissant pour repasser à une diffusion retardée.

Le deuxième point est l'importance du feedback temporel lors de la consultation d'une vidéo. Nous avons pensé à implémenter un trait qui se déplacerait de droite à gauche sur l'image, et qui permettrait de situer l'instant actuel dans la durée du flux : à gauche le début et à droite la fin.

Pour assurer une communication asynchrone, nous avons conçu et développé dans notre prototype un système qui enregistre automatiquement la vidéo à partir du moment où on interagit avec l'appareil. L'enregistrement est arrêté lorsque l'absence est constatée. Nous avons prévu mais nous n'avons pas développé, un échange des vidéos enregistrées ainsi que des images et des vidéos partagées entre les pèlemèles connectés ensemble, lorsque

l'absence est détectée. Lors de l'absence, la connexion vidéo n'est plus établie et la bande passante réseau est libre pour ces échanges. Le résultat prévu est que les différents pélemèles vont avoir les mêmes fichiers images et vidéo, de cette façon on pourra déclencher la lecture automatique simultanément des mêmes fichiers sur tous les pélemèles, créant un contexte partagé.

3.1.7 Extensions du système communicant

L'un des besoins importants identifié dans notre système de communication est l'utilisation de l'audio pour compléter la communication vidéo. Nous pensons que les capacités communicantes du système seraient étendues par le support de telles fonctionnalités. L'utilisation de l'audio ne se limite pas à la seule communication orale et une perception périphérique de l'audio pourrait avoir une place importante dans notre système de communication, en permettant le partage d'un contexte audio.

Pour supporter l'établissement du canal audio deux questions se sont posées : la voie que le canal oral allait emprunter et le moment où ce canal allait être établi. Nous avons examiné et tenté de développer deux possibilités basées sur les deux outils Skype [29] et les téléphones BlueTooth pour répondre à la première question. Récemment, le logiciel de téléphonie via ip appelé Skype a mis à la disposition des utilisateurs une API permettant de le contrôler de façon externe, depuis un autre programme. Etant donné la popularité de Skype, cette API en faisait le candidat idéal à remplir la fonction que l'on avait besoin d'implanter. Malheureusement, les tentatives de l'utiliser se sont heurtées à de nombreux problèmes d'intégration dans l'environnement Mac OSX, liés au programme qui servait d'interface à cette API, nommé « dbus » [26]. A l'heure de la rédaction de ce rapport, l'API a été améliorée et adaptée à OSX, et nous pensons que les problèmes que nous avons rencontrés sont dépassés. On peut signaler que ces problèmes ne se posaient pas dans l'environnement Linux dans lequel notre système était aussi testé. Pour le bluetooth, son intérêt dépasse simplement l'établissement d'un lien audio. Cependant, vu les difficultés de la première solution, il remplissait la fonction que nous avions besoin d'utiliser. En ce qui concerne la deuxième question relative au moment d'établissement du canal audio. Nous avons pensé d'abord à cliquer sur un élément qui apparaît lors de l'interaction de deux personnes avec le système. Nous avons pensé aussi à faire un geste ou un mouvement de la tête, mais nos méthodes de reconnaissance des gestes et de la tête n'étaient pas au point pour faire cela.

Finalement, nous nous sommes inspirés des données récoltées par le MirrorSpace et en particulier du jeu spontané de ses utilisateurs de superposer leurs visages pour en faire une commande qui établit le lien audio. L'interruption de ce lien peut-être faite en s'éloignant de l'appareil, c'est-à-dire en repassant de l'interaction à la présence simple. L'intérêt de cette façon d'établir un canal audio réside d'abord dans sa spontanéité et par la réciprocité qu'elle exige. En effet, il n'est pas besoin de décrocher, il suffit de se laisser superposer le visage. L'inconvénient de cette façon d'appeler un correspondant réside dans l'intimité qu'elle suppose, généralement on n'a pas envie de superposer son visage avec n'importe qui, mais dans notre contexte l'idée reste tout de même acceptable.

Une autre utilisation de l'audio à laquelle nous avons pensé et qui nous semble primordiale est la notification par des sons qu'une personne demande à communiquer ou qu'elle est en train

d'interagir avec l'appareil. En effet, la notification nous paraît très importante car si on voit qu'une personne est présente mais que cette personne est occupée par autre chose, on peut faire des signes devant la caméra, mais si la personne ne regarde pas dans la bonne direction, on a aucune chance d'attirer son attention. Nous avons identifié d'autres besoins de notification. Comme la prise en compte par l'appareil de la personne qui se place devant. Si cette personne n'y fait pas attention, son image sera diffusée sans qu'elle se rende compte. En général, il nous paraît utile de notifier les changements d'état de l'appareil comme l'établissement d'un lien audio, le passage de l'appareil en état de veille, etc.

Nous avons évoqué précédemment l'utilisation de Skype et du téléphone bluetooth. Cette possibilité d'intégrer des outils extérieurs nous semble importante dans le cadre de la communication familiale. Nous pensons que dans ce contexte, il faut que notre système soit ouvert au sens où il est interopérant avec d'autres systèmes déjà existant dans les familles. Par exemple, les téléphones bluetooth pourraient être utilisés pour identifier la personne qui est dans la pièce ou encore en approchant un téléphone du système passer des images du téléphone vers le pélemèle. Une autre utilisation possible de l'interopérabilité, c'est la diffusion sur une télévision d'une vidéo ou d'une image dans le cadre de la consultation automatique ou intentionnelle. En effet l'écran d'une télévision est généralement assez grand et on peut y afficher sur tout l'écran sans craindre de cacher des parties du pélemèle, on disposerait alors de meilleures possibilités de visualisation.

3.1.8 Aspects généraux de conception

Le prototype a été conçu de façon incrémentale à l'aide de brainstorming avec mon encadrant de stage. A chaque étape de la conception, les choix de conception ont abouti au développement de composants autonomes accomplissant une partie des fonctions du système. Ces composants ont été ensuite évalués sur le plan de la performance et sur celui de l'interaction. Quand le composant développé était jugé satisfaisant, il était intégré aux autres composants pour constituer par agrégation le prototype qui sera présenté dans les parties suivantes du rapport. Il est important de signaler que la plupart des composants conçus et développés ne sont pas présents dans le prototype final car soit ils n'ont pas franchi l'étape de la première évaluation, soit ils ont été intégrés puis remplacés plus tard par d'autres composants plus appropriés.

Nous avons essayé quand c'était possible d'utiliser le principe de redondance informationnelle qui facilite la compréhension pour l'utilisateur du fonctionnement de l'appareil. Par exemple, la taille et la position sont deux indicateurs différents qui reproduisent la même information (au centre et grand signifie l'interaction avec l'appareil, un peu périphérique et petit de taille signifie la présence, etc). Nous avons aussi essayé d'avoir dans la mesure du possible une vitesse d'interaction fixe pour donner à l'utilisateur une interaction fluide. Par exemple, toutes les évolutions sont fonction du temps, et si dans un déplacement la machine est plus lente parce qu'elle a plus de calculs à faire, ce n'est pas la vitesse du mouvement qui est affectée mais la vitesse de rafraîchissement. Nous avons aussi lissé les mouvements en utilisant les algorithmes de easing quadratique afin de rendre les mouvements et les évolutions plus confortables et prédictibles, par exemple pour les déplacements l'utilisateur a surtout besoin de voir de quelle position on part et à laquelle on arrive, les déplacements

linéaires ont l'inconvénient d'un déplacement continu dont on ne peut prévoir l'arrêt.

Nous avons aussi constaté dans la programmation comme dans la conception du prototype que sans même le vouloir intentionnellement, nous avons introduit à beaucoup d'endroit des comportements similaires à celui des systèmes perceptifs et cognitifs humains. Que cela soit au niveau de la vision par ordinateur, c'est peut-être évident, mais c'est aussi à d'autres niveaux moins évidents, par exemple l'introduction des seuils pour passer des variables continues à des états discrets évoque la capacité à catégoriser ou le comportement d'un neurone, de même l'interprétation d'une présence sans mouvement trop longue rappelle la capacité des récepteurs sensoriels qui soumis à un signal continu finissent par s'adapter et considérer que ce n'est pas un signal.

Enfin, l'aspect sonde technologique est celui qui a été le moins développé dans ce stage. Bien que l'enregistrement des sources vidéo est daté de façon à pouvoir les utiliser plus tard pour voir comment l'appareil a été utilisé, l'organisation de toutes ces fonctionnalités et le développement du prototype ont été privilégiés sur ce dernier aspect.

3.1.9 Aspects techniques et survol du code développé

Le prototype du Pêlemêle qui a été programmé est développé en c++ sous les deux plateformes Mac OSX et Linux. Il s'appuie principalement sur la librairie Nucleo [47] qui permet d'explorer les nouveaux usages de la vidéo et des nouvelles techniques d'interaction homme-machine. Il utilise cette librairie aussi pour la communication réseau pour l'échange de vidéos en TCP et UDP, et utilise le module de rendez-vous Howl [46] pour trouver automatiquement d'autres Pêlemêles sur réseau local. Pêlemêle utilise aussi sur OpenGL [51] pour ce qui concerne l'affichage et OpenCV [34] la librairie de vision par ordinateur, pour la détection et le suivi de visages.

Pêlemêle a été testé sur différentes machines, il tourne sans problèmes sur un PowerPC G4 ou un Pentium4 1,6GHz, mais de façon plus fluide sur un Mac Mini ou un Mac Titanium. On peut signaler que même lors de la communication synchrone sur un réseau local une latence de 1 à 2 secondes peut-être observée. Cette latence fait qu'on n'observe pas exactement le même résultat sur les différents Pêlemêles

Pêlemêle prends en charge un fichier de configuration dans lequel on peut spécifier les sources (images, vidéo et caméra) qu'il doit charger, l'adresse ip ou le nom de service de autres Pêlemêles auxquels il doit se connecter, et les modes de communications (UDP, TCP ou les deux) qu'il doit prendre en charge. Il utilise aussi un fichier de description qu'il fournit à OpenCV lequel l'utilise pour la détection des visages de face.

Le code du Pêlemêle contient 10 fichiers principaux, consacrés à la gestion des éléments sur l'écran, aux filtres utilisés, à la détection d'évènements, au positionnement des éléments, à la composition temporelle, et aux fonctions de communication réseau.

Nous avons rencontré lors du développement de ce programme deux principales difficultés. D'une part, l'utilisation de la programmation réactive basée sur les évènements dont nous connaissions les principes mais dont la pratique grâce à la librairie Nucleo nous a permis de maîtriser en détail le fonctionnement.

D'autre part, l'évolution incrémentale du projet nous a empêché de pouvoir prévoir son évolution et bien que nous avons utilisé une programmation modulaire, l'évolution anarchique du code nous a amené à plusieurs à plusieurs moments à le restructurer complètement afin de pouvoir continuer de le développer. Nous pensons à présent que la définition de règles de programmation au départ nous aurait probablement permis de mieux maîtriser cette évolution. Enfin, nous avons approfondi notre connaissance du langage c++ au contact de la librairie Nucleo, car certains aspects les plus avancés nous étaient jusqu'alors inconnus.

3.2 Scénario d'utilisation : Un évènement dans un salon partagé

3.2.1 Mode d'emploi du système

Dans ce scénario, on s'intéressera pour simplifier uniquement au mode utilisant la détection de visage.

Lorsqu'on est absent ou loin du dispositif, le Pêlemêle se met en veille, il va lire des vidéos enregistrées. Les images correspondant à chaque site distant ou local où il y a absence, se mettent à dériver avec le temps le long d'une ligne de fuite. Ces images figées sont en plus brouillées par l'application d'un filtre « peinture à huile ».

Lorsqu'on est présent mais assez loin du dispositif, mais détecté. Le Pêlemêle est toujours en veille, et continue de lire des vidéos enregistrées ou partagées, mais les images correspondant à chaque site distant ou local où il y a présence ne dérive plus et reviennent à la périphérie de l'écran. L'application du filtre disparaît et la composition temporelle est utilisée, elle mélange les images des instants passés à l'image de l'instant présent différé. L'image la plus récente que l'on voit date déjà de plusieurs secondes.

Lorsqu'on s'approche de face du dispositif. Le Pêlemêle interrompt l'activité de veille, les vidéos en train d'être lues reprennent leur place en périphérie, et les images correspondant à chaque site distant ou local où un utilisateur s'est approché se déplacent vers la position centrale et accélèrent. Lorsqu'elle atteint le centre le différé n'existe plus et l'image est en direct. Si plusieurs images sont en position centrale, les plus récentes deviennent transparentes et on continue de voir toutes les images. Si pendant cette phase deux visages sur les images centrales ont la même position et la même taille alors un lien audio est établi entre les deux sites.

Lorsqu'on s'éloigne du dispositif, l'image repart en périphérie en ralentissant et en fonction de si on reste présent ou si on s'absente, soit la composition temporelle et le différé reprennent, soit l'image s'arrête et le filtre « peinture à huile » et la dérive temporelle reprennent. Si aucune image de site distant n'est au centre de l'écran, l'activité de lecture automatique reprends. Une petite image correspondant à l'image capturée pendant l'interaction apparaît alors derrière l'image locale.

3.2.2 Who, Where & What for

Utilisateurs : famille, groupes proches

Le système que nous avons conçu et développé est destiné à la communication dans un groupe proche comme une famille. Dans ce scénario nous prenons le cas (réel) de plusieurs membres de la même famille qui vivent à des distances importantes : Moi Sofiane vivant à Paris, mon frère Slim habitant Londre, ma sœur Héla à Montréal et mes parents Nébiha et Béchir habitant à

Hammamet en Tunisie. Pour plus de réalisme, nous situons ce scénario dans le futur, quand le réseau Internet pourra transporter la vidéo rapidement sur ces distances importantes.

A quel endroit: Salon, LivingRoom

Le dispositif est destiné à être utilisé dans un cadre domestique, dans la maison. En général, c'est le salon qui est considéré comme le lieu ouvert où les échanges entre les membres se passent. Pour ce scénario, les pêle-mêles sont placés dans le salon de chacun des participant, ou dans ce qui fait office de salon.

Pour faire quoi : Partager sa vie de tout les jours, des images et des vidéo mais aussi pour communiquer ensemble

Le PêleMêle sert à rapprocher les familles distantes géographiquement. Pour cela, il permet de partager sa vie et la vie des autres, à travers des images de leur espace et de ce qu'il s'y passe mais aussi à communiquer intentionnellement avec eux.

3.2.3 Démonstration

En rentrant d'une journée de travail, Sofiane arrive chez lui et s'installe devant sa télé afin de se détendre. A côté de sa télé, son PêleMêle tourne depuis plusieurs jours et il s'est débrouillé pour que chacun des membres de sa famille en ai un dans son salon.

(1) Dérive automatique : Il y jette un coup d'œil et il voit que l'image qui lui vient du PêleMêle de sa sœur Héla à Montréal a presque disparu dans l'horizon de l'appareil ; depuis la veille elle n'est pas revenue dans son salon. Par contre l'image qui lui vient de son frère Slim à Londres dérive depuis ce matin ; il ne doit pas encore être rentré du travail. Il voit aussi que quelqu'un est venu dans la journée face à l'appareil dans le salon de ses parents en Tunisie car l'image a l'air de dériver seulement depuis quelques heures, mais le rendu stylisé l'empêche de voir de qui il s'agit.

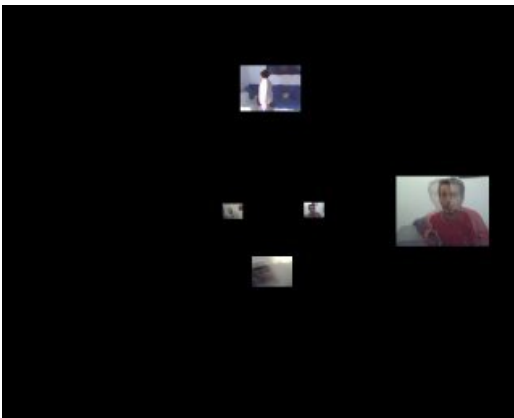


Figure 6. Dérive et perception du temps

(2) Consultation et enregistrement automatique de vidéos : Chacun de ses correspondant a un axe dans son écran, et il voit au centre des vidéos de sa discussion la veille avec son frère : Il avait l'air amusé.



Figure 6. Lecture automatique de vidéos

(3) Communication synchrone : Interrompant ses réflexions, l'image de Héla se met à grandir et chasser la vidéo en lecture. Sur son visage un grand sourire, et quand elle atteint le centre de l'écran et la taille maximale elle affiche le relevé de ses résultats d'examens pour que Sofiane le voie. Il s'avance pour mieux regarder, il voit alors qu'elle lui montre son relevé de notes : elle a réussi sa révision comptable. Son image se met aussi à grandir et se superpose à l'image de Héla.

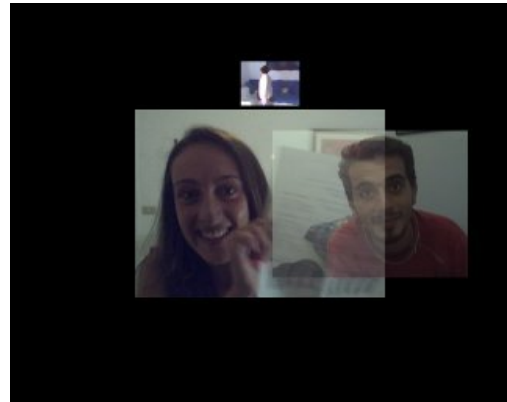


Figure 6. Communication video synchrone

(4) Etablissement d'un contact audio : Comme leurs deux visages étaient proches de la caméra et au centre de l'écran, Sofiane entend le son du papier qu'elle agite devant lui, il lui dit fait alors félicitations et la discussion s'engage d'elle-même.

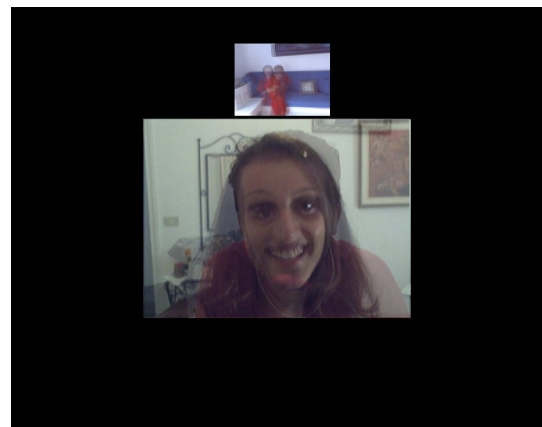


Figure 6. Etablissement d'un contact audio

(5) Perception périphérique (peripheral awareness): Peu de temps après l'image venant de Tunisie cesse de dériver et revient vers le bord de l'écran, Sofiane voit sa mère Nébiha dans différentes images superposées. Héla se met alors à faire des signes de la main pour attirer son attention. Nébiha l'aperçoit et viens voir ce qu'il se passe, elle se superposent le visage et se mettent à bavarder intensément, Sofiane s'éloigne un peu de l'écran et son image se replace en périphérie.



Figure 6. Perception périphérique

4. Solutions apportées et problèmes restant:

La conception et le développement du système de communication multiéchelles centré sur l'usage de la vidéo sont une tâche difficile qui implique des problèmes nombreux et d'autant plus complexes quand ce système regroupe plusieurs services et se destine à plusieurs usages. Nous avons essayé en respect du principe de conception en interaction homme-machine « less is more » et en s'inspirant des systèmes assez simples du VideoProbe et du MirrorSpace, de cacher au mieux la complexité à l'utilisateur et d'aboutir à un système simple, homogène et complet, mais nous nous n'avons pu apporter de réponse à tous les problèmes soulevés par un tel système dans les six mois de stages qui nous étaient impartis. Ce stage nous a tout de même permis d'apporter des solutions logicielles ou des éléments de conception répondant à la plupart des contraintes que nous avons définies au départ du projet.

Ces contraintes sont au nombre de trois, la caractéristique multiéchelle, le contexte familial et l'aspect sondes technologiques.

En ce qui concerne l'aspect multiéchelles, le PèleMèle permet de communiquer de façon synchrone par le partage des images des différents sites en temps réel. Il était prévu pour supporter la communication asynchrone en échangeant vidéos capturée intentionnellement ou accidentellement et des images et des vidéos intentionnellement partagées. Nous avons conçu ce dernier aspect mais nous n'en avons pas achevé le développement. Le PèleMèle s'adapte au contexte des utilisateurs en détectant leur présence et leur besoin de communiquer, utilisant leur distance à l'appareil et des procédés basés sur la différence entre images ou sur la détection de visages, et en fonction de ces contextes, va offrir une communication directe ou différée, de premier plan ou secondaire, en faisant varier le niveau de détail associé au flux d'images transmises ou affichées. Pour cela, il utilise des procédés

de composition spatiale en jouant sur la position et la taille des flux affichés, et temporelles en fusionnant des images successives dans le temps. Il utilise aussi des filtres stylisés de type « peinture à huile » et des mécanismes de diffusion différée pour protéger l'intimité de ses utilisateurs et leur permettre de communiquer de façon plus légères. Ces adaptations contextuelles offrant des niveaux variés d'informations sont articulées de façon continue propre à donner des transitions fluide d'un niveau à l'autre.

Pour tenir compte de la communication dans le cadre familial, le PèleMèle supporte une communication multi-sites. Pour cela, la technique de visualisation qu'il utilise est la fusion des images des correspondants par transparence. Afin de rendre possible une communication informelle entre les membres de la même famille, le PèleMèle permet le partage accidentel ou intentionnel de scènes de la vie quotidienne, la perception périphérique des autres membres, et offre une information sur leur disponibilité, leur permettant de coordonner leurs échanges, sans utiliser des connexions vidéos permanentes et ininterrompibles, et en préservant leur intimité par l'utilisation de filtres et la diffusion différée. L'échange de messages, d'images ou de vidéo ainsi que l'appel explicite, ont fait l'objet d'une conception, mais n'ont pas été développés dans le PèleMèle faute de temps. C'est l'exploration de ce dernier point a d'ailleurs mis en évidence l'importance pour les systèmes communicant pour l'usage familial d'être des dispositifs ouverts et interopérants avec les outils déjà utilisés par les familles. Allant dans ce sens, PèleMèle permet d'enrichir la communication vidéo l'établissement d'un canal audio entre les sites utilisant les logiciels de téléphonie via ip.

Le dernier aspect, concernant les sondes technologiques, a été moins pris en compte dans ce projet. Nous avons prévu au départ la capture du résultat final de l'interaction, des valeurs des capteurs, de données quantitatives et qualitatives sur l'usage et les sources vidéo échangées et utilisées pour la communication. PèleMèle tient compte uniquement du dernier point car il enregistre et il date les sources vidéo des interactions avec le système, qui pourront être analysées pour mieux comprendre l'usage qui en est fait et pouvoir l'améliorer

5. CONCLUSION & PERSPECTIVES

Dans ce rapport, nous avons présenté un système de communication basé sur l'image, pour le cadre domestique et familial. La solution conçue et développée, fondée sur la notion de communication multi-échelles, s'adapte aux contextes d'usages des familles, leur permettant de partager des images de la vie quotidienne tout en préservant leur intimité. Notre système couvre des situations de communication exigeant des niveaux d'informations et d'engagement variable, tout en permettant des transitions fluides entre ces niveaux.

D'un point de vue théorique, notre étude du contexte général nous a permis de définir la méthodologie de conception et d'organiser les problèmes liés à l'utilisation de la vidéo pour la communication entre les membres distants d'une même famille. Nous avons vu que la méthodologie de la conception participative est appropriée à notre problème et que le concept de sondes technologiques permet de l'appliquer dans le contexte familial. Nous avons délimité les problèmes qu'il faut prendre en compte pour qu'un système de communication utilisant la vidéo soit adopté par ses usagers: la flexibilité des usages, l'intégrabilité des outils et la préservation de la vie privée. Ensuite, nous avons

présenté le rôle de la vidéo comme véhicule de la communication des indices visuels et nous avons dégagé les aspects importants pour notre contexte : d'une part, la communication informelle et la coordination des membres d'une même famille, d'autre part, les liens et les indices affectifs visuels et le partage de contexte physique.

Du point de vue pratique, nous avons développé autour de la notion de communication multi-échelles les solutions conceptuelles et logicielles permettant de prendre en compte tous ces aspects. Le programme développé, appelé Pêlemêle, permet de détecter et d'adapter le niveau de détail à des situations d'usages différentes. Il rend possible de communiquer de façon synchrone en direct ou en différé, d'établir un lien audio ou de supporter la communication périphérique, le partage de contexte et la coordination, tout en fournissant des transitions fluides entre tous ces modes. Nous avons présenté ces fonctions sous la forme d'un scénario d'utilisation.

Malgré ces avancées, le problème des outils de communication multi-échelles dans le cadre familial est loin d'être entièrement résolu. La prochaine étape est d'abord de compléter le développement de notre logiciel, le Pêlemêle, pour répondre aux besoins que nous avons identifiés mais qu'il ne supporte pas encore, c'est-à-dire le partage d'images pour la communication asynchrone et la notification. Ensuite, il s'agit d'instrumentaliser correctement le dispositif afin d'améliorer sa capacité à récolter des données in situ. Enfin, il s'agit d'explorer l'interopérabilité du système avec d'autres systèmes existants, les possibilités d'utilisation de capteurs et de configurations matériels différentes, et surtout d'expérimenter le système dans des conditions réelles chez des familles.

6. REMERCIEMENTS

Tout d'abord, je tiens à remercier toute l'équipe du projet inSitu pour son accueil chaleureux et en particulier Emmanuel Nars qui m'a aidé à chaque fois que j'en avais besoin et Jacob Eisenstein qui m'a passé un morceau de code qui m'a permis d'utiliser facilement OpenCV.

Je remercie surtout M. Nicolas Roussel, mon encadrant de stage, dont l'aide et l'éclairage m'ont été très précieux et qui m'a formé, soutenu et encouragé jusqu'à la fin. J'aimerais aussi remercier Michel Beaudouin-Lafon et Wendy Mackay dont les cours m'ont été très utiles durant ce stage.

Je tiens aussi à remercier les membres de ma famille qui du fait de notre dispersion géographique m'a rendu plus sensible à la situation que j'ai étudiée. En particulier, je remercie ma sœur Héla et ma mère Nébiha qui ont contribué à mon travail en jouant dans mon scénario, et mon père qui a tenu à venir à Paris pour me voir soutenir ce projet.

Je voudrais aussi remercier mon ancienne colocataire Aurélie Vandeginste pour sa contribution en me prêtant sa machine pour tester mon système et en acceptant de jouer les utilisateurs naïfs (de moins en moins naïve d'ailleurs). Je remercierais aussi volontiers tous mes amis qui ont été intéressés par mon système et qui en l'essayant et en discutant de certains aspects m'ont inspiré des nouvelles idées (à savoir Fly, Dom, Dim, Cedbou, etc).

J'aimerais enfin remercier Lina qui, par l'enthousiasme qu'elle & montré pour ce travail, m'a beaucoup encouragé à m'y investir et m'a persuadé de son intérêt.

7. BIBLIOGRAPHIE

- [1] Abel, M.J. Experiences in an exploratory distributed organization, In J. Galegher, R. Kraut & C. Egidio, (Eds.), *Intellectual Teamwork*, Hillsdale, NJ: Lawrence Erlbaum, 489-510, 1990.
- [2] AOL, AIM (AOL Instant Messenger) <http://www.aim.com/>
- [3] Apple, iChat AV <http://www.apple.com/macosex/features/ichat/>
- [4] Baudisch, P. and Gutwin, C. Multiblending: displaying overlapping windows simultaneously without the drawbacks of alpha blending. In *Proceeding of CHI 2004*, Vienna Austria, pp. 367-374, April 2004.
- [5] Bly, S.A., Harrison, S.R. & Irwin S. Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment, *Communications of the ACM*, Janvier 1993.
- [6] Bull, P.E. The interpretation of posture through an alternative methodology to role play. *British Journal of Social and Clinical Psychology* 17, 1-6, 1978.
- [7] Buxton, W. Hill, R. & Rowley, P. Issues and Techniques in Touch-Sensitive Tablet Input, *Computer Graphics*, 19(3), 215-224. 1985.
- [8] Card, S., Moran, T. & Newell, A. *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Erlbaum, 1983.
- [9] Chapanis, A. *Interactive Human Communication*. Scientific American, Vol. 232, pp 36-42, 1975.
- [10] Clark, H. H., & Marshall, C. R. Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge: Cambridge University Press, 1981.
- [11] Clark, H.H., & Brennan S.E. Grounding in Communication. In L. Resnick, J. Levine & S. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (127-149). Hyattsville, MD: American Psychological Association, 1991.
- [12] Conversy, S., Roussel, N., Hansen, Evans, H. H., Beaudouin-Lafon, M. and Mackay, W. Partager les images de la vie quotidienne et familiale avec videoProbe. In *Proceedings of IHM 2003*, pages 228-231, ACM, International Conference Proceedings Series. Novembre 2003.
- [13] Cool, C., Fish, R.S., Kraut R.E., and Lowery, C.M. Iterative Design of Video Communication Systems. In *Proceedings of ACM CSCW'92 Conference on Computer-Supported Cooperative Work*, Toronto, Ontario, pages 25-32. ACM, New York, November 1992.
- [14] Coutaz, J. Bérard, F. and Crowley, J.L. Coordination of perceptual processes for Computer Mediated Communication, in *Procs. of Second International Conference on Automatic Face and Gesture Recognition*, Killington, Vermont, 1996.
- [15] Coutaz, J., Bérard, F., Carraux, E. & Crowley, L. Early Experience with the Mediaspace CoMedi. *EHCI 1998*: 57-72, 1998.
- [16] Crampton Smith, G. The Hand That Rocks the Cradle. *I.D. May/June*: 60-65. 1995.
- [17] Dewsbury, G and Edge, M. "Designing the home to meet the needs of tomorrow ... today", *Open House International*,

- Vol. 26 (2). (2001)
www.rgu.ac.uk/files/DewsburyEdge2001.pdf
- [18] Dourish, P. & Bly, S.A. Portholes : Supporting Awareness in a Distributed Work Group, Proceedings of the CHI'92 Conference on Human Factors in Computing Systems, Monterey, p. 541, 1992.
- [19] Edigo, C. Teleconferencing as a technology to support co-operative work: Its possibilities and limitations. In J. Gallegher, R. E. Kraut, & C. Edigo (Eds.) Intellectual teamwork: Social and technological foundations of cooperative work (pp. 351-371). Hillsdale, NJ. Erlbaum Associates, 1990.
- [20] Ekman, P., & Friesen, W. V. Unmasking the face. Englewood Cliffs, N. J.: Prentice-Hall, 1975.
- [21] Erickson, T.D. Working with Interface Metaphors, in The Art of Human-Computer Interface Design, Brenda Laurel, ed., Addison- Wesley, 1990.
- [22] Feiner, S., MacIntyre, B. and Seligmann, D. Karma (knowledge-based augmented reality for maintenance assistance), 1993.
<http://www.cs.columbia.edu/graphics/projects/karma/karma.html>
- [23] Fish, R., Kraut, R., Root, R. & Rice, R. Video as a technology for informal communication. Communications of the ACM, 36:1, 48-61, 1993.
- [24] Fish, R.S., Kraut, R.E., Root, R.W. & Rice R.E. Evaluating Video as a Technology for Informal Communications, Proceedings of the CHI'92 Conference on Human Factors in Computing Systems, Monterey, p. 37, 1992.
- [25] Fitts, P. The information capacity of the human motor system in controlling amplitude and movement. In: Journal of Experimental Psychology. 47 - p. 381-391, 1954.
- [26] Freedesktop.org, dbus
<http://www.freedesktop.org/Software/dbus>
- [27] Gaver, W., Moran, T., MacLean, A., Löfstrand, L., Dourish, P. Carter, K. & Buxton, W. Realizing a Video Environment: EuroPARC's RAVE System, Proceedings of the CHI'92 Conference 1992.
- [28] Gibson, J. The Ecological Approach to Visual Perception, 1979.
- [29] Global P2P Telephony Company, Skype
<http://www.skype.com/>
- [30] Goodwin, C. Conversational Organization : interaction between speakers and hearers, Academic Press, New York and London, 1981.
- [31] Grudin, J. Why groupware applications fail: problems in design and evaluation, in Office: Technology and People, Elsevier Science Publishers, p. 245, 1989.
- [32] Heath, C. C. & Luff, P. K. Disembodied Conduct: asymmetries in video mediated interaction in an office environment, CHI'91: Reaching Through Technology. New Orleans. pp. 92-106, 1991.
- [33] Hopper, A., Harter A. and Blackie, T. The Active Badge System. In Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems, pages 335-341. ACM, New York, 1993.
- [34] Intel, L de la librairie OpenCV (Open Source Computer Vision Library).
<http://www.intel.com/research/mrl/research/opencv/>
- [35] Ishii, H. and Kobayashi, M., "ClearBoard: A Seamless Media for Shared Drawing and Conversation with Eye-Contact," Proceedings of Conference on Human Factors in Computing Systems (CHI '92), ACM SIGCHI, Monterey, pp. 525-532, 3-7 May 1992.
- [36] Ishii, H. and Ullmer, B. Tangible bits: Towards seamless interfaces between people, bits and atoms. In Proceedings of ACM CHI 97 Conference on Human Factors in Computing Systems, volume 1 of PAPERS: Beyond the Desktop, pages 234-241, 1997.
- [37] Ishii, H., et al., ambientROOM: integrating ambient media with architectural space, CHI'98,173-174, 1998.
- [38] Kleck, R. E., & Nuessle, W. Congruence between the indicative and communicative function of eye contact in interpersonal relations. British Journal of Social and Clinical Psychology, 7, 241-246, 1968.
- [39] Mann, S. Smart clothing: The shift to wearable computing. Communications of the ACM, pages 23-24, August 1996.
- [40] Mantei, M., Backer, R.M., Sellen, A., Buxton, W. Milligan, T., Wellman B.: Experiences in the use of a Media Space, Proceedings of the CHI'91 Conference on Human Factors in Computing Systems, Nouvelle-Orléans, p. 203, 1991.
- [41] Microsoft, MSN Messenger <http://messenger.msn.com/>
- [42] Norman, D. Psychology of Everyday Things. Basic Books, 1988.
- [43] Ott, M., Lewis, J.P., Cox, I. Teleconferencing Eye Contact Using a Virtual Camera, Adjunct Proceedings of InterCHI'93, Amsterdam, p. 109, 1993
- [44] Perlin, K. and Fox, D. Pad: An alternative approach to the computer interface. In Proc. of ACM SIGGRAPH, pages 57-64. ACM Press, 1993.
- [45] Pier, K. (Ed.) Active Badge Panel. Proceedings,Conference on Organizational Computing Systems,November 5-8, Atlanta, Georgia, 1991.
- [46] Porchdog Software, Howl
<http://www.porchdogsoft.com/products/howl/>
- [47] Roussel, N. de la librairie Nucleo
<http://insitu.lri.fr/~roussel/projects/nucleo/>
- [48] Roussel, N., Evans, H. and Hansen, H. Proximity as an interface for video communication. *IEEE Multimedia*, 11(3):12-16, July-September 2004.
- [49] Sellen A., Buxton B. Using Spatial Cues to Improve Videoconferencing, Proceedings of the CHI'92 Conference on Human Factors in Computing Systems (Video), Monterey, p. 651, 1992.
- [50] Sellen, A. Remote Conversations: Theeffects of mediating talk with technology. HumanComputer Interaction, Vol. 10, No. 4, pp.401-444, 1995.
- [51] Silicon Graphics, OpenGL <http://www.opengl.org/>

- [52] Stults, R. MediaSpace, rapport technique Xerox PARC, 1986.
- [53] Tang, John C., Ellen A. Isaacs, and Monica Rua, "Supporting Distributed Groups with a Montage of Lightweight Interactions", Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW) '94, Chapel Hill, NC, pp. 23-34, October 1994.
- [54] Triesman, A. Preattentive Processing in Vision. Computer Vision, Graphics, and Image Processing 31, 156-177, 1985.
- [55] Weiser, M. Some computer science issues in ubiquitous computing. Communications of the ACM, 36(7):75--83, July 1993.
- [56] Wellner, P. Interacting with paper on the digitaldesk. Communications of the ACM, 36(7):87--96, July 1993.
- [57] Whittaker, S. Rethinking Video as a Technology for Interpersonal Communications: Theory and Design Implications. In International Journal of Human-Computer Studies, 42 (5) p. 501-529, 1995.
- [58] Whittaker, S., Frohlich, D. and Daly-Jones, O., "Informal workplace communication: what is it like and how might we support it?" in the Proceedings of the ACM 1994 conference on Human factors in computing systems (CHI '94), pp. 131-137, 1994.
- [59] Whittaker, S., Geelhoed, E. and Robinson, E. 'Shared workspaces: how do they work and when are they useful?' mt. T Man-Machine Studies 39, 813442, 1993.
- [60] Williams, E. Experimental comparisons of face-to-face and mediated communication: A review. Psychological Bulletin, 84(5), 963-976, 1977.