

Les effets des indices non-verbaux sur les activités de communication à distance. Deux études expérimentale sur le dialogue tutoriel

Federico Tajariol

▶ **To cite this version:**

Federico Tajariol. Les effets des indices non-verbaux sur les activités de communication à distance. Deux études expérimentale sur le dialogue tutoriel. domain_other. Université Joseph-Fourier - Grenoble I, 2006. Français. NNT : . tel-00128077

HAL Id: tel-00128077
https://theses.hal.science/tel-00128077

Submitted on 30 Jan 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Joseph Fourier – Grenoble I

Laboratoire Clips – IMAG

Ecole Doctorale «Ingénierie pour la Santé, la Cognition et l'Environnement»

THESE

Pour l'obtention du grade de Docteur

en Sciences Cognitives

Présentée et soutenue publiquement le 30 Octobre 2006 par

Federico Tajariol

Les effets des indices non-verbaux sur les activités de communication à distance.

Deux études expérimentales sur le dialogue tutoriel.

Rapporteurs

Sylvie Lainé-Cruzel, Professeur, Université Jean Moulin, Lyon

Ioan Roxin, Professeur, Université de Franche-Comté

Examinateurs

Anne Guérin, Professeur, Université Joseph Fourier, Grenoble

Dongo Rémi Kouabenan, Professeur, Université Mendès-France, Grenoble

Directeurs de thèse

Jean-Pierre Peyrin, Professeur, Université Joseph Fourier, Grenoble

Jean-Michel Adam, Maître de Conférences, Université Mendès-France, Grenoble

Michel Dubois, Maître de Conférences, Université Mendès-France, Grenoble

RESUME

Notre travail s'insère dans le domaine de la communication audio-vidéo médiatisée. Les recherches de ce domaine ont l'objectif de comprendre l'impact des technologies de la communication sur les différentes activités de communication et de réaliser des dispositifs adéquats pour ces activités. Quelle que soit la situation de communication, l'interface n'est plus un simple objet d'adressage et de traduction pratique des commandes de l'opérateur, comme dans la vision classique de l'interaction homme-machine, mais devient plutôt un espace de communication interpersonnelle

L'action de communiquer avec autrui ne se limite ni à une simple adéquation de codes verbaux et non-verbaux, ni à un processus de production d'inférences et d'interprétation d'indices contextuels. Les interlocuteurs agissent selon une logique coopérative (Clark, 1996) : cela signifie que les interlocuteurs doivent réciproquement accepter chacune de leurs contributions pendant un échange, afin de pouvoir partager des connaissances ou des faits relatifs à leur activité. Une communication réussie demande que les interlocuteurs co-construisent mutuellement leurs référentiels, s'appuyant sur les indices verbaux et non-verbaux.

Parmi ces derniers, on se doit de distinguer les indices non-verbaux kinésiques (e.g., les expressions faciales, les regards, les gestes du corps, etc.) et les indices non-verbaux ostensifs-inférentiels (e.g. les gestes déictiques, les actions physiques). Nous avançons deux hypothèses de travail principales : A) les indices non-verbaux kinésiques facilitent surtout le processus de communication et favorisent un sentiment de proximité sociale entre les partenaires distants ; B) les indices non-verbaux ostensif-inférentiels facilitent surtout l'intercompréhension et la réalisation de la tâche commune.

Pour vérifier ces hypothèses, nous avons construit une tâche expérimentale dans laquelle un tuteur se doit d'assister deux apprenants distants impliqués dans la réalisation d'un travail pratique (la réalisation d'une page en notation html) et qui ne peuvent pas communiquer entre eux. Dans la première expérimentation (n tutor=12, N étudiants = 48), nous étudions les effets des indices kinésiques, véhiculés par l'image-vidéo montrant les partenaires distants. Les résultats indiquent que les indices kinésiques facilitent le processus de compréhension mutuelle et l'activité de suivi du tuteur, sans pour autant améliorer la qualité du travail des apprenants. Dans la deuxième expérimentation (n tutors = 12, étudiants = 72), nous nous intéressons aux indices ostensifs-inférentiels, véhiculés par l'image-vidéo relative aux actions effectuées par les partenaires distants sur les objets appartenant à leur espace de travail. Les indices ostensifs-inférentiels favorisent la construction des référentiels communs, facilitent le suivi du tuteur et induisent une amélioration du travail des apprenants. Nous discutons les résultats de deux expérimentations et nous avançons des propositions pour la conception des dispositifs consacrés à la communication tutorielle.

MOTS CLES : communication médiatisée, vidéo-présence, vidéo-activité, interaction tuteur-apprenant.

REMERCIEMENTS

J’adresse mes sincères remerciements à Mme Sylvie Lainé-Cruzel, M. Ioan Roxin et M. Dongo Rémi Kouabénan pour avoir accepté d’évaluer ce travail de recherche : leurs questions et leurs remarques m’ont beaucoup aidé à expliciter des idées et, je l’espère, à améliorer globalement le manuscrit. Je remercie aussi Mme Anne Guérin d’avoir accepté de participer au jury de thèse.

Ma profonde gratitude va principalement à mes directeurs de thèse, pour avoir eu confiance en moi et avoir accepté de suivre mon travail. Un grand merci à Jean-Michel Adam et à Michel Dubois pour m’avoir initié à la rigueur scientifique : j’espère avoir mis en œuvre au moins une partie de leurs enseignements. Merci infiniment pour leur soutien constant, pour leurs conseils avisés indispensables pour l’élaboration de ce manuscrit et leur grande disponibilité. Merci à Jean-Pierre Peyrin de m’avoir offert les meilleures conditions scientifiques qu’un doctorant peut imaginer pour concrétiser ses propres recherches.

Mes profonds remerciements aux membres de l’équipe de recherche ARCADE (Clips-Imag) pour leur disponibilité et leur aide tout le long de l’élaboration de cette recherche. Un remerciement particulier à Dominique, pour sa relecture attentive de mon texte.

Je remercie vivement Brigitte Meillon et Martine Harsch (équipe Multicom, Clips-Imag) pour leur disponibilité et pour avoir permis la mise en place des expérimentations rapportées dans cette thèse. Merci aussi à tous les sujets qui ont accepté d’y participer.

J’exprime ma reconnaissance aux membres du laboratoire Laseldi (Université Franche-Comté, site de Montbéliard) pour m’avoir offert les conditions matérielles pour finaliser ce manuscrit.

Enfin, merci à toutes les personnes qui m’ont soutenu moralement, de près et de loin.

TABLE DES MATIERES

INTRODUCTION 1

CHAPITRE 1 - COMMUNIQUER : UNE ACTIVITE ORIENTEE VERS AUTRUI 9

1.1 Communiquer : agir *sur* autrui et construire une entente *avec* autrui. 9

1.1.1 « *Communiquer pour agir sur la réalité* » 10

1.1.2 « *Communiquer c'est* [aussi] *produire et interpréter des indices* » 11

1.1.3 « *Communiquer c'est construire une entente commune* » 14

Synthèse 17

1.2 Les indices dédiés à l'activité de communication 18

1.2.1 Les indices non-verbaux 19

1.2.2 Les indices verbaux 23

1.2.3 Les indices verbaux dans le dialogue tutoriel 25

Synthèse 30

Conclusion 31

CHAPITRE 2 - VOIR LE PARTENAIRE DISTANT : UN AVANTAGE OU UNE ENTRAVE POUR L'ACTIVITE DE COMMUNICATION ? 33

2.1 Faut-il se voir pour bien coopérer ? 37

Synthèse 44

2.2 Peut-on ne voir que le visage du partenaire distant pour bien coopérer ? 48

Synthèse 52

2.3 Faut-il se regarder dans les yeux pour se comprendre ? 54

2.3.1 Le regard réciproque 55

2.3.2 Connaître la direction du regard du partenaire distant 60

Synthèse 64

Conclusion 67

Problématique de recherche et hypothèses de travail 70

CHAPITRE 3 - PARTAGER UN ESPACE COMMUN AVEC LE PARTENAIRE DISTANT : UN AVANTAGE OU UNE ENTRAVE POUR COMMUNIQUER ? 75

3.1 Le partage d'un espace visuel commun. 76

Synthèse 89

3.2 Le partage d'un espace de travail commun. 92

Synthèse 96

Conclusion 98

Problématique et hypothèses de recherche 99

CHAPITRE 4 - ETUDE I : LES EFFETS DE L'IMAGE-VIDEO DU PARTENAIRE SUR L'INTERACTION TUTEUR-APPRENANT 105

4.1 Méthodologie 106

4.1.1 La population 106

4.1.2 Le matériel 107

4.1.3 La procédure 109

4.1.4 Les conditions expérimentales 110

4.1.5 Justifications des choix méthodologiques 110

4.1.6 Méthodologie d'analyse du corpus 111

4.1.7 Opérationnalisation des variables dépendantes 120

4.2 Hypothèses opérationnelles 124

4.3 Les résultats 127

4.3.1 Les effets de l'image-vidéo sur le volume des dialogues (H1) 127

4.3.2 Les effets de l'image-vidéo sur les interventions du tuteur (H2). 128

4.3.3 Les effets de l'image-vidéo sur le processus de compréhension mutuelle (H3) 129

4.3.4 Les effets de l'image-vidéo sur le contenu du dialogue tutoriel (H4) 130

4.3.5 Les effets de l'image-vidéo sur le score de performance (H5) 133

4.3.6 Les effets de l'image-vidéo sur l'évaluation subjective des contributions (H6). 134

4.4 Discussions des résultats. 137

CHAPITRE 5 - ETUDE II : LES INDICES NON-VERBAUX RELATIFS A L'ESPACE VISUEL COMMUN. 143

5.1 Méthodologie 143
5.1.1 La population 144
5.1.2 Le dispositif technique 144
5.1.3 Les conditions expérimentales 144
5.1.4 La grille de codage 146
5.1.5 Les variables dépendantes 146

5.2 Hypothèses opérationnelles 146

5.3 Les résultats 150
5.3.1 Les effets de l'espace visuel commun sur le volume des échanges verbaux. 150
5.3.2 Les effets de l'espace visuel commun sur les interventions spontanées du tuteur 150
5.3.3 Les effets de l'espace visuel commun sur le processus de compréhension mutuelle 151
5.3.4 Les effets de l'espace visuel commun sur la construction du dialogue tutoriel 153
5.3.5 Les effets de l'espace visuel commun sur la performance de réalisation de la tâche 157
5.3.6 Les effets de l'espace visuel commun sur l'évaluation des contributions des participants ... 159

5.4 Discussion des résultats 162

CONCLUSIONS 167

BIBLIOGRAPHIE GENERALE 175

LISTE DES TABLEAUX 187

LISTE DES FIGURES 189

Introduction

Notre travail concerne le domaine de la communication audio-vidéo médiatisée (Finn, Sellen & Wilbur, 1997). Ce domaine de recherche est issu des études consacrées aux environnements informatiques pour le travail coopératif, dont l'objectif est de réaliser des collecticiels, c'est-à-dire des plates-formes logicielles permettant à plusieurs acteurs de travailler ensemble (Carstensen & Schmidt, 1999). Dans ce sens, un collecticiel se doit d'optimiser la coordination entre les différents acteurs (espace de coordination), de favoriser les actions sur des entités (espace de production) et de rendre efficace l'échange d'informations entre les acteurs (espace de communication) (Salber, 1995). Les recherches de ce domaine ont une double finalité : *a) comprendre l'impact des technologies de la communication sur les différentes activités de communication* : de l'échange informel entre des amis aux activités professionnelles spécifiques (e.g. le contrôle aérien, la chirurgie, l'enseignement, etc.) ; *b) concevoir et développer des environnements logiciels efficaces pour favoriser au mieux les échanges entre les partenaires distants*.

Pour rendre efficace les espaces de communication, malgré les nombreuses recherches conduites, des questions restent encore irrésolues, à la fois sur le plan théorique et applicatif. Sur le plan théorique, comme nous le montrons, il manque encore un cadre théorique qui permet de généraliser et de prédire l'effet des modalités de communication sur les activités humaines à distance. Sur le plan applicatif, l'ergonomie de l'interaction homme-machine appliquée au domaine de la communication n'a pas encore réussi à décliner la notion d'utilisabilité en celle de « communicabilité » (Prates, de Souza & Barbosa, 2000). En effet,

quelle que soit la situation de communication, l'interface n'est plus un simple objet d'adressage et de traduction pratique des commandes de l'opérateur dans le domaine d'intervention, tel que le domaine de l'interaction homme-machine la voyait, mais devient plutôt un « espace de communication interpersonnelle » (Navarro, 2001). Notre recherche expérimentale veut donc contribuer à enrichir les connaissances de ce domaine d'étude, s'intéressant aux *communications synchrones médiatisées* et notamment *au dialogue tutoriel synchrone*.

La problématique et l'objectif de recherche

Du point de vue théorique nous nous accordons avec la vision coopérative de Clark (1996), pour qui toute activité de communication est à la fois cognitive et sociale. Dans ce sens, les co-locuteurs reconnaissent réciproquement leurs intentions communicatives (selon le *principe de la responsabilité mutuelle*) et visent à produire leurs propres énoncés de manière suffisamment claire afin de ne pas devoir les réitérer (selon le *principe du moindre effort collaboratif)* (Clark, 1996). Chaque locuteur se construit un modèle hypothétique de son interlocuteur (Cahour & Falzon, 1992), dans le cadre d'un *contrat de communication* (Ghiglione, 1986) stipulé et reconnu par eux-mêmes, c'est-à-dire en rapport avec les enjeux et les objectifs de l'activité de communication dans laquelle les locuteurs sont engagés. Grâce à ce modèle, contenant l'ensemble des représentations et des hypothèses que le locuteur se construit sur son interlocuteur (e.g. son identité sociale, ses intentions, les connaissances supposées partagées, etc.), les co-locuteurs construisent leurs interactions communicatives.

Nous définissons le dialogue tutoriel comme une activité de communication coopérative (Fox, 1993). Dans ce sens, le tuteur a besoin de comprendre les objectifs et les difficultés de l'apprenant afin de l'aider de manière adéquate. Pour ce faire, comme toute autre activité de communication, le tuteur et l'apprenant communiquent grâce à des indices verbaux et non-verbaux. Parmi ces derniers, on se doit de distinguer les indices non-verbaux kinésiques de l'apprenant (e.g., ses expressions faciales, ses regards, ses gestes du corps, etc.) et les indices non-verbaux ostensifs-inférentiels (e.g. les gestes déictiques, les actions physiques). Lorsque le tuteur (et l'apprenant) les perçoit ou les interprète de manière non correcte, son diagnostic peut se révéler inadéquat et ne pas bien comprendre ce dont l'apprenant a besoin ; d'autre part, l'apprenant peut mal comprendre ce que le tuteur veut lui expliquer et l'aide peut se révéler non seulement inefficace, mais aussi désavantageuse pour l'apprentissage. Ces

considérations amènent à s'interroger sur l'impact des modalités de communication et sur les indices non-verbaux par elles-mêmes véhiculés : Est-il nécessaire que le tuteur et l'apprenant distants se voient pour assurer la compréhension réciproque ? Est-il nécessaire que le tuteur observe ce que l'apprenant fait pour lui fournir l'aide la plus adéquate à ses besoins ?

L'objectif de ce travail de recherche est donc de comprendre d'une part l'influence de l'image-vidéo du tuteur et de l'apprenant (les indices non-verbaux kinésiques), et d'autre part l'influence de l'image-vidéo relative à un espace visuel commun (les indices non-verbaux ostensifs-inférentiels) sur les interactions communicatives, les impressions subjectives et le niveau de réalisation d'une tâche effectuée par l'apprenant. Nous avançons *deux hypothèses de travail principales*. Selon la *première*, les indices non-verbaux kinésiques facilitent le processus de communication et favorisent un sentiment de proximité sociale entre les partenaires distants. Selon la *deuxième* hypothèse, les indices non-verbaux ostensif-inférentiels facilitent principalement la compréhension réciproque et la réalisation de la tâche coopérative.

Le contexte scientifique et le domaine d'application

Ce travail de recherche s'intègre dans le projet FORMID (FORMation Interactive à Distance), mené par les équipes ARCADE et IIHM (Laboratoire Clips-Imag). L'objectif du projet est de proposer une plate-forme logicielle permettant à des formateurs d'interagir de manière synchrone avec des apprenants distants impliqués dans des travaux pratiques (Pernin, 2000). Les plates-formes pour l'enseignement synchrone à distance (e.g., WebCT, TopClass, LearningSpace, etc.) disposent généralement de fonctionnalités pour la communication asynchrone (e.g. mail, forum, chat), parfois pour la communication synchrone (e.g. audio) et plus rarement d'un espace de communication audio-vidéo ou d'un espace visuel ou un espace de travail commun (e.g. tableau blanc). Notre recherche a donc pour objet de *comprendre les avantages et les entraves d'un espace de communication interpersonnelle audio-vidéo et d'un espace visuel commun dans une situation tutorielle synchrone*, où un formateur doit assister des apprenants distants. L'objectif, sur le plan applicatif, est de *fournir des recommandations sur la conception des dispositifs dédiés à la communication entre le formateur et les apprenants distants*.

Ce dont cette thèse ne traite pas

Cette thèse ne s'insère pas dans le débat sur la formation à distance du point de vue pédagogique ou de l'ingénierie pédagogique : notre objectif n'est ni d'améliorer les performances des situations d'enseignement à distance (cf. Cavanaugh, 2001 pour une meta-analyse), ni de comprendre comment renouveler les méthodes d'enseignement (cf. Laurillard, 1993). Nous ne proposons donc pas une comparaison des pratiques d'ingénierie pédagogique associées à la formation à distance, mais plutôt une *réflexion sur les dispositifs permettant la communication médiatisée*.

Par ailleurs, cette thèse ne se propose pas d'étudier l'activité tutorielle (Fox, 1993 ; Wagemann & Percier, 1995) ni le préceptorat (Soury-Lavergne, 1998). Notre objectif est d'apporter *des éléments de compréhension des mécanismes socio-cognitifs impliqués dans un dialogue tutoriel* (aspect cognitif) *qui se construit dans un environnement de communication médiatisée* (aspect ergonomique).

La structure du manuscrit

Le manuscrit est organisé en une partie théorique et une partie expérimentale. La première partie, composée de trois chapitres, présente le cadre théorique de notre recherche.

Dans le *premier* chapitre, nous exposons la notion de communication, ses composants et les cadres conceptuels, issus principalement de la psychologie cognitive et de la pragmatique cognitive. Nous portons une attention particulière dans la description des indices, verbaux et non-verbaux, *intrinsèques au dialogue tutoriel*.

Le *deuxième* chapitre est consacré aux *études expérimentales concernant l'impact des indices non-verbaux kinésiques*, véhiculés par l*'image-vidéo du partenaire distant*. Nous montrons que des facteurs différents sont à considérer lorsque l'on veut comprendre la fonction de l'image du partenaire distant sur l'activité de communication.

Le *troisième* chapitre s'intéresse aux effets des *indices non-verbaux ostensifs-inférentiels*. Nous présentons des études concernant l'apport des logiciels qui mettent à disposition des utilisateurs un « espace visuel commun » et « un espace de travail commun ».

La partie expérimentale se compose de deux chapitres. Dans le *quatrième chapitre*, décrivant la première étude expérimentale, nous étudions les *effets des indices verbaux kinésiques sur l'interaction tutorielle médiatisée*. Dans notre situation expérimentale, un tuteur

doit assister deux apprenants impliqués dans la réalisation d'une page html. Deux conditions de communication, « audio seul » et « audio & image-vidéo » sont comparées, afin d'évaluer l'apport de l'image du visage du partenaire distant sur l'interaction tuteur-apprenant.

Dans le *cinquième chapitre*, nous utilisons la même situation expérimentale pour étudier l'impact des *indices non-verbaux ostensifs-inférentiels*, notamment *les indices relatifs à l'activité de l'apprenant distant*, sur le dialogue tutoriel.

Dans les *conclusions*, les résultats de deux expérimentations sont résumés et discutés à la lumière des recherches relatives au domaine et des élements théoriques évoqués dans la première partie. Des perspectives sont proposées, à la fois pour la recherche dans le domaine de la communication médiatisée et pour la conception de dispositifs de communication dédiés aux activités de communication tutorielle.

PARTIE THEORIQUE

Chapitre 1 - Communiquer : une activité orientée vers autrui

Nous communiquons avec quelqu'un pour accomplir différents types d'activités comme « … *planifier, accomplir, modifier, coordonner, négocier, discuter, évaluer* » (Lacoste, 1991 : 205). Nous n'avons pas un accès direct au contenu de ses pensées, de ses émotions et de ses intentions et nous ne pouvons faire que des suppositions à partir de l'ensemble d'indices verbaux et non-verbaux qu'il (elle) produit (Argyle & Graham, 1977). Qu'est-ce donc que l'activité de communication ? Quels sont ses mécanismes intrinsèques ? L'objectif de ce chapitre est de présenter les positions théoriques principales relatives à l'activité de communication (1.1) et d'en comprendre les mécanismes spécifiques (1.2).

1.1 Communiquer : agir *sur* autrui et construire une entente *avec* autrui.

Sans vouloir prétendre à l'exhaustivité, nous nous limitons à décrire trois positions théoriques relatives à l'activité de communication. La première conception (1.1.1), issue de la philosophie analytique du langage, décrit la communication comme un *processus d'action d'un émetteur sur un récepteur*. La deuxième (1.1.2), issue des recherches en pragmatique et en psychologie cognitive, considère que la communication est un *processus de construction d'inférences* entre les locuteurs. Enfin, la troisième position théorique (1.1.3), issue des travaux

en psychologie de la communication, définit l'activité de communication comme une *action sociale et cognitive de coopération avec autrui*.

1.1.1 « *Communiquer pour agir sur la réalité* »

D'après Austin (1962), le langage est un outil dédié non seulement à *énoncer* (l'acte physique du dire) et à *dire en disant* (l'acte locutoire ou du produire des mots avec un sens), mais il est surtout un outil pour *faire en disant* (l'acte illocutoire ou de l'agir sur le monde). Searle (1972) ajoute que le langage est aussi un outil pour *faire croire en disant* (l'acte perlocutoire ou de l'agir sur l'interlocuteur). Par exemple, lorsque le locuteur T dit « *L'icone bleue change la couleur de ton texte* », il s'attend à ce que l'allocutaire A comprenne que, dans un contexte d'usage d'un logiciel pour l'écriture électronique, un certain objet dont le référent linguistique est le mot « *icone bleu* » modifie la couleur d'une partie du texte. Le locuteur T s'attend donc à ce que l'allocutaire A identifie et reconnaisse la chaîne de sons phonétiques, qu'il se représente le contenu propositionnel de cet énoncé et modifie ainsi l'état du texte. La compréhension de A se concrétise par le fait qu'il utilise l'objet « *icone bleue* » pour modifier la couleur de son texte. Ainsi faisant, l'allocutaire A reconnaît « l'intention communicative » (Grice, 1975) de T de vouloir l'informer sur comment changer la couleur de son texte. Or, le contenu d'un message n'est reconnu que si l'allocutaire comprend l'intention communicative du locuteur ou, en d'autres termes, s'il se représente ce que le locuteur attend de lui (Bange, 1992). En d'autres termes, si A ne change pas la couleur du texte avec l'icone bleue, l'acte de parole de T échoue. Les causes peuvent être les suivantes : a) A n'a pas compris *la signification de l'énoncé* de T (ce que T voulait dire) ; b) A n'a pas entendu *l'énonciation* (ce que le locuteur T a dit) ; c) A ne veut pas réaliser cet acte.

Cependant, cette perspective théorique présente deux aspects réducteurs :

a) toute activité de communication est considérée selon une « logique d'échange », où un seul acteur possède à la fois le droit de parole de manière exclusive (Cutler & Pearson, 1986). L' « agir » est considéré dans une seule direction : du locuteur (le sujet actif) à l'allocutaire (le sujet passif) de l'échange (Brennan, 1990 ; Clark, 1996 ; O'Connell, Kowal & Kaltenbacher, 1990) ;

b) la compréhension réciproque est considérée comme une réponse adéquate à un contenu sémantique spécifique : lorsque le locuteur émet un énoncé, il s'attend à ce que

l'allocutaire réagisse selon le contenu de cet énoncé. Par exemple, lorsque le locuteur énonce « *fermez la fenêtre* », il s'attend à ce que son auditoire réalise l'action de fermer la fenêtre, montrant ainsi qu'il a compris le « contenu sémantique » de l'énoncé.

Or, comme nous le montrons par la suite, *a)* l'action de communication est régie par des mécanismes conversationnels dynamiques (§ 1.1.3) et *b)* la compréhension réciproque est le résultat d'un ajustement des représentations des interlocuteurs actifs plutôt que le résultat d'un décodage (Navarro, 1999), car un énoncé du locuteur fournit plutôt des indices à partir desquels on peut inférer à la fois le sens de l'énoncé et l'intention du locuteur.

1.1.2 « *Communiquer c'est* [aussi] *produire et interpréter des indices* »

D'après Sperber & Wilson (1989), la communication ne se limite pas à un simple processus de codage et décodage sémantique (ou d'association signifiant-signifié issue des travaux de Saussure). Les auteurs, s'appuyant sur le *principe conversationnel* de Grice[1] (1975), suggèrent que l'acte de communication est régi par un processus de production d'inférences et un processus d'interprétation d'indices (Sperber & Wilson, 1989 : 14). Ces indices sont les manifestations de l'intention communicative du locuteur plutôt que de son intention informative. En effet, Sperber & Wilson (1989) distinguent l'*intention informative*, c'est-à-dire l'intention du locuteur de fournir des éléments informatifs à son interlocuteur, de l'*intention communicative*, c'est-à-dire l'intention explicite du locuteur de renseigner l'interlocuteur sur quelque chose. La notion d'intention communicative est l'occurrence sur le plan linguistique du concept de *communication ostensive-inférentielle* (Reboul & Moeschler, 1998), selon laquelle un locuteur produit des indices (verbaux et non-verbaux) afin de faire comprendre à son interlocuteur qu'il a l'intention de lui faire connaître une certaine information.

[1] Grice (1975) suppose que les participants à une conversation respectent le principe conversationnel suivant : *chacun s'attend à ce que l'autre participe à la conversation de manière coopérative*, dans le sens de faciliter l'interprétation des énoncés. Ce principe est décliné par Grice en quatre maximes : de quantité (chaque locuteur fournit seulement la quantité d'information nécessaire à la situation de communication), de qualité (chaque locuteur est tenu d'affirmer la vérité), la relation (le contenu d'un énoncé de chaque locuteur est pertinent avec les énoncés précédents) et de manière (la présentation du contenu doit être claire et non ambiguë).

Comment l'interlocuteur peut-il attribuer une valeur d'intentionnalité à un indice produit par le locuteur ?

La valeur d'intentionnalité est déterminée par le *principe de pertinence*, que nous illustrons avec un exemple entre un tuteur et un apprenant placés en condition de communication « face à face ». Ce dernier doit réaliser une page html à partir d'un texte brut. Le tuteur lui demande de mettre en italique un mot du texte brut. Comme l'apprenant hésite, le tuteur lui indique sur l'écran (avec son doigt) la balise html « I » et deux crochets, sans ajouter rien d'autre. Il s'agit de la part du tuteur d'un acte *ostensif* de communication (il agit pour montrer) qui induit l'apprenant à produire une inférence (utiliser la balise et les crochets). Bien que l'apprenant ait sous ses yeux toutes les données du problème, il peut attribuer à la balise « I » et aux crochets un certain degré de *pertinence* seulement grâce à l'acte ostensif du tuteur, car il comprend qu'il doit construire la solution à partir de ces éléments. Comme Reboul & Moeschler (1998) le soulignent, « *…l'acte de communication ostensive-inférentielle provoque l'attente de pertinence* » (1998 :76).

Pour Sperber & Wilson (1989), le principe de pertinence, sur lequel l'interlocuteur s'appuie pour donner une signification au contenu des énoncés du locuteur, n'est pas normatif mais plutôt interprétatif. Ce principe s'articule autour de deux critères :

a) le *critère du moindre effort* : plus le coût d'interprétation d'une information est faible, plus cette information est pertinente dans un contexte donné ;
b) le *critère des effets contextuels* : plus une information engendre des informations nouvelles, plus cette information est pertinente dans un contexte donné.

Reprenons l'exemple du tuteur et de l'apprenant. Lorsque l'apprenant focalise son attention sur la balise et les crochets, il produit des raisonnements pour rendre ces informations cohérentes par rapport à son objectif (i.e., mettre en italique le mot du texte). Cette activité inférentielle est produite à partir de connaissances procédurales (i.e., « toute balise doit être accompagné par des crochets » ; « il y a un ordre spécifique à respecter » ; « si on insère le mot à mettre en italique à l'intérieur de deux crochets, alors on obtient le résultat attendu ») et comporte un *effort* sur le plan cognitif. Le coût de cet effort d'interprétation est moindre lorsque la pertinence de l'information est élevée. Ainsi, l'apprenant produira plus facilement des inférences lorsque le tuteur lui montre la balise html « I », car cette information est hautement pertinente.

Est-il suffisant de produire des inférences pour comprendre son interlocuteur ?

Pour se comprendre, les interlocuteurs doivent accéder à des connaissances «*manifestement partagées* »[2], c'est-à-dire accessibles et perceptibles à des individus en interaction. Elles constituent un « *environnement cognitif partagé* », dans lequel se situent tous les faits mutuellement manifestes aux individus et les faits supposés partagés[3]. Cet environnement cognitif permet aux interlocuteurs d'accéder aux mêmes références, mais non nécessairement de produire les mêmes inférences. Cette différence est fondamentale, car elle permet de rendre compte des échecs de la compréhension réciproque. En effet, l'échec de compréhension entre les interlocuteurs est provoqué par la spécificité des mécanismes inférentiels qui «… *assurent à la communication un succès tout au plus probable, mais non pas certain* » (Sperber & Wilson, 1989 : 33). Ainsi, la compréhension est la résultante d'une activité inférentielle ancrée dans un contexte : on adopte telle ou telle hypothèse relative au contenu sémantique d'un énoncé si cette hypothèse est adéquate au contexte cognitif, qui contient « *l'ensemble de prémisses utilisées pour l'interprétation d'un énoncé* » (Sperber & Wilson, 1989 : 31). Par ailleurs, pour résoudre l'incompréhension, l'interlocuteur accepte l'hypothèse la plus pertinente parmi les autres afin d'attribuer une interprétation plausible au discours du locuteur.

En conclusion, le modèle de communication de Sperber & Wilson (1989) n'est pas innovateur par rapport à la théorie des actes de langage car, comme Brassac (2000) souligne :

---

[2] D'après Sperber & Wilson (1989) un fait est considéré « *manifeste à un individu à un moment donné si et seulement si cet individu est capable à ce moment-là de se représenter mentalement ce fait et d'accepter sa représentation comme vraie ou probablement vraie* » (Sperber & Wilson, 1989 : 65).

[3] Dans le domaine de pragmatique cognitive, la notion d'*environnement cognitif partagé* est proche de la notion de *savoir commun* (Lewis, 1969) et de *savoir mutuel* (Schiffer, 1972). Selon ces deux notions, la connaissance est une forme de savoir où le locuteur (L) sait que l'allocutaire (A) sait que L sait que A sait etc. (le savoir en forme de régression). Si A veut être sûr d'interpréter l'énoncé de L de manière adéquate aux intentions de L, il faut que L et A partagent un certain nombre de connaissances. Cependant, lorsque L et A dialoguent, ils n'ont pas le temps de vérifier leur savoir mutuel. Il semble donc que la notion de *savoir mutuel* (et *savoir commun*) soit à considérer comme un niveau idéal que L et A devraient atteindre pour éviter autant que possible les malentendus, plutôt qu'un mécanisme pragmatique fonctionnel à la compréhension réciproque.

a) la position théorique de Sperber & Wilson (1989) est encore *monologiste*, où la construction du sens est le résultat de l'adéquation entre le contenu sémantique de l'énoncé émis par le locuteur et les inférences produites par son interlocuteur ;

b) l'aspect dynamique et conversationnel n'est pas intégré dans leur perspective théorique.

En regard de ces caractéristiques, leur modèle semble adéquat à l'étude des situations de communication asynchrone (e.g., en modalité textuelle) plutôt qu'à l'étude de l'interaction dialogique, où des mécanismes conversationnels et socio-cognitifs sont mis en place pendant l'énonciation des locuteurs (Argyle, Lalljee & Cook, 1968 ; Sacks, Schegloff & Jefferson, 1974).

1.1.3 « *Communiquer c'est construire une entente commune* »

Les deux positions théoriques présentées précédemment décrivent la communication comme un acte plutôt individuel. En revanche, l'approche de Clark (1996), s'insérant dans le courant interactionniste de la psychologie sociale, considère la communication comme un processus coopératif visant la compréhension mutuelle entre les locuteurs. Pour Clark (1996), l'*interaction* désigne toute action individuelle qui implique la participation active d'autrui, comme lors d'une vente aux enchères ou lors d'un vis-à-vis entre deux voyageurs dans un train. Dans toute activité de communication, les contributions dialogiques des participants s'ajustent pendant toute l'interaction selon l'axe du processus et du contenu.

1.1.3.1 Le processus et le contenu de communication

D'après Clark & Brennan (1991), deux personnes qui communiquent sont comme deux pianistes qui jouent à quatre mains : ils doivent se mettre d'accord sur quelle partition jouer (le *contenu*) et synchroniser la dynamique et l'intensité de leurs gestes (le *processus*).

Le *contenu* d'une activité de communication est constitué de l'ensemble des *connaissances et croyances* que les locuteurs *partagent* ou supposent partager entre eux (Clark & Marshall, 1981 ; Lewis, 1969). Cet ensemble de contenus, dont les locuteurs parlent, est un terrain commun d'entente co-construit par eux-mêmes (*common ground*) (Clark, 1996). Pour la co-construction du terrain commun d'entente, les locuteurs produisent des inférences qui s'appuient à la fois sur le contenu sémantique de l'énoncé et sur les connaissances qu'ils partagent ensemble (« …*A knows that A and B mutually know* » (Clark & Marshall, 1981:18).

Il s'agit d'une activité coopérative, qui se déroule pendant que les co-locuteurs[4] interagissent, parlent et se comprennent, ce qui constitue le *processus* de l'activité de communication.

Le *processus* est le plan de la *dynamique conversationnelle* (Schlegoff, 1987), de l'alternance de la parole entre celui qui parle et qui écoute (Sacks & al., 1974). Dans le processus de communication, les co-locuteurs s'efforcent afin que le passage de parole se réalise de la manière la plus explicite possible et ils veillent à maximiser le niveau de *compréhension réciproque* (Clark, 1996). En d'autres termes, les interlocuteurs *coopèrent* de façon à s'assurer que ce qui est dit soit compris (Schober & Clark, 1989). Par ce processus assurant la compréhension réciproque, les co-locuteurs assurent le terrain commun d'entente (« …*they try to ground what has been said – that is, make it part of their common ground*. » (Clark & Brennan, 1991 : 128), c'est-à-dire que, pendant le déroulement de l'interaction, ils mettent à jour leurs connaissances et croyances communes. Mais pour le faire, leur compréhension réciproque doit être assurée : comment ?

Les processus de compréhension réciproque

Ce processus coopératif se compose de deux phases et donne forme à une *contribution* (Clark & Schaefer, 1989 ; Clark & Wilkes-Gibbs, 1986). Les deux phases sont :

a) *la présentation* : le locuteur A adresse un énoncé *e1* au co-locuteur B et A fait l'hypothèse que si B produit un indice évident de sa compréhension, alors B a compris ce que A voulait dire ;
b) *l'acceptation* : le co-locuteur B *accepte* l'énoncé *e1* par l'intermédiaire d'un indice qui montre qu'il a compris (ou qu'il croit avoir compris) ce que le locuteur A veut dire avec l'énoncé *e1* (Clark & Wilkes-Gibbs, 1986 : 151).

Toute contribution peut s'imbriquer. Par exemple, lorsque le locuteur A présente un énoncé au co-locuteur B, ce dernier peut soit accepter la présentation de A soit la refuser. Dans le premier cas, les indices que B produit pour accepter la présentation de A doivent être

[4] Dans le modèle de Clark (1996), la distinction entre le locuteur et l'allocutaire est très nuancée : l'allocutaire est plutôt un co-locuteur, il a la même responsabilité que le locuteur, à la fois sur le plan du processus et du contenu.

reconnus aussi par A. Ces indices sont de nature verbale (e.g. les acquiescements de type « *OK* », « *oui* », « *mhm mhm* » etc. ; § 1.2.3.1) et non-verbale (e.g. les hochements de tête, les gestes, etc. ; § 1.2.3.2). Ainsi la confirmation de B est un nouvel énoncé, une nouvelle présentation soumise à l'acceptation de A, et ainsi le cycle se répète[5]. Lorsque A et B ont atteint un niveau de compréhension mutuelle satisfaisante, A et B reconnaissent réciproquement[6] que B a compris ce que A voulait exprimer lors de sa dernière énonciation (cf. le *principe de responsabilité mutuelle* ; Clark, 1996). Mais la compréhension réciproque entre A et B est le résultat d'un ajustement continu, car elle n'est ni immédiate ni parfaite.

1.1.3.2 Les stades de compréhension et le principe du moindre effort collaboratif

D'après Clark & Wilkes-Gibbs (1986), la compréhension réciproque se construit par des stades différents. Au contraire de Sperber & Wilson (1989), pour qui le résultat de la compréhension est discret (réussie *vs.* échouée), Clark & Wilkes-Gibbs (1986) décrivent quatre stades de compréhension mutuelle :

- stade 0 : le locuteur A n'est pas conscient que le co-locuteur B a produit un énoncé (e.g. il ne l'a pas entendu) ;
- stade 1 : le locuteur A perçoit physiquement l'énoncé de B, mais il ne l'a pas entendu de manière correcte (détection imparfaite) ;
- stade 2 : le locuteur A entend correctement l'énoncé de B, mais ne le comprend pas (identification syntaxique imparfaite, comme pour un nouveau mot) ;
- stade 3 : le locuteur A comprend ce que B veut dire (e.g. la compréhension réussie).

Pour passer d'un stade à l'autre, le locuteur doit reformuler sa présentation (ou son acceptation) afin de se faire comprendre par son allocutaire, ce que Schlegoff (1987) désigne « *organization of repair* » (1987 :157). La réparation[7] est un ensemble de pratiques linguistiques que les

[5] Traum (1998) souligne que dans le modèle de Clark (1996) aucune limite temporelle n'est indiquée, avec le risque d'une itération infinie des présentations et des acceptations.

[6] Pour mieux cerner l'aspect de réciprocité, Brennan (1998) définit toute contribution dialogique un *acte collectif*, car cela implique la participation d'au moins deux sujets.

[7] Le locuteur répare son énoncé en présentant un autre indice verbal (cas de l'*autoréparation*, e.g. : D « *Well that's a little different from last week* », L « *Yeah we were in hysterics last week* », D «

locuteurs utilisent pour résoudre des difficultés d'énonciation, d'écoute ou de compréhension lors d'un échange verbal (Hirst, McRoy, Heeman, Edmonds, & Horton, 1994). Toute réparation engendre un coût cognitif, que les locuteurs essayent de réduire, appliquant le *principe du moindre effort collaboratif* (Clark, 1996 ; Clark & Wilkes-Gibbs, 1986). Selon ce principe, les locuteurs essayent de produire leur présentation (ou acceptation) de manière suffisamment claire afin de ne pas devoir la réitérer (sous une même forme ou sous une autre, verbale ou non-verbale). Cependant, ce principe n'est pas toujours appliqué par les locuteurs et notamment lorsque (Clark & Brennan, 1991 : 134) :

a) *la référence linguistique est trop complexe* pour être acceptée par le co-locuteur : ainsi, il utilise plusieurs présentations partielles qui demandent des acceptations partielles de la part du co-locuteur[8] ;
b) *le temps d'énonciation est trop court* : le locuteur est ainsi obligé de produire une première proposition provisoire pour ensuite l'améliorer après que le co-locuteur a produit son acceptation.

Synthèse

L'apport principal de la position théorique « *langage comme action* » est de mettre en valeur la dimension performative du langage. Cependant, cette perspective réduit la communication à un simple jeu de codage et décodage entre un locuteur actif et un allocutaire passif. La position *ostensive-inférentielle* met en valeur l'aspect intentionnel du locuteur, qui veut montrer à son interlocuteur son intention de lui faire connaître une information par l'intermédiaire d'indices communicationnels. L'interlocuteur interprète ces indices selon un critère de pertinence. En d'autres termes il produit des inférences, qui n'assurent cependant pas la compréhension réciproque. En effet, elles doivent être ancrées dans un contexte, c'est-à-dire dans un environnement cognitif partagé entre les participants. Bien que la notion de contexte

No I mean Al », L « *Oh ...* », extrait de Hirst & al. (1994) ou bien c'est son co-locuteur qui apporte d'autres indices (*réparation par autrui*).

[8] Voici un exemple issu de notre corpus : A : « *tu vois l'objet rouge qui se trouve en haut à droite de l'écran ?* », B « *oui* », A « *eh bien il s'agit d'une icone* », B « *d'accord* », A « *et tu dois la cliquer deux fois pour lancer le programme* ».

soit importante, la perspective ostensive-inférentielle considère la compréhension entre les locuteurs comme le résultat d'une adéquation entre le contenu intentionnel des énoncés du locuteur (la partie active) et le contenu de l'inférence produite par l'allocutaire (la partie passive).

Dans une perspective interactionniste, le modèle *coopératif* de Clark (1996) propose que l'activité de communication se structure selon des formes conversationnelles (côté social) et s'articule autour de mécanismes cognitifs inférentiels (côté individuel). Selon ce modèle, le locuteur et le co-locuteur partagent la responsabilité de leur compréhension mutuelle, c'est-à-dire qu'ils coopèrent ensemble pour choisir, modifier et reformuler leurs énoncés, jusqu'à atteindre un niveau d'acceptation réciproque satisfaisant, « … *les locuteurs* [étant en même temps] *émetteurs et récepteurs* » (Abric, 1999 : 7). Pour construire leur compréhension réciproque, les locuteurs suivent le principe du moindre effort collaboratif et produisent (et interprètent) des indices. Quels sont ces indices ? Quelle est leur nature ? Quelles sont leurs fonctions ?

1.2 Les indices dédiés à l'activité de communication

Lorsque des individus communiquent en condition « face à face », ils sont impliqués dans un processus multimodal (McGurk & MacDonald, 1976) : lorsqu'ils parlent, ils se regardent, ils observent leur environnement et produisent des gestes tout en bougeant leur corps. En d'autres termes, ils produisent et interprètent des indices de nature différente, à savoir les indices[9] non-verbaux (1.2.1) et verbaux[10] (1.2.2).

[9] Nous ne présentons ici que ces deux types d'indices. Cependant, la littérature propose aussi les indices para-verbaux, dont Kerbrat-Orecchioni (1998) identifie quatre types principaux : a) les pauses et les hésitations, b) le rythme et l'intonation de la voix, c) l'intensité vocale et d) les productions vocales particulières (e.g. raclements de gorge, grognements, etc.). A différents degrés, ces indices sont responsables de l'alternance de parole (Sacks & al., 1974). Par exemple, lors d'une conversation, l'intonation de la voix du locuteur permet de faire comprendre à l'allocutaire qu'un énoncé est bien une question plutôt qu'une description (Amado & Guittet, 2003). Dans la situation tutorielle, la durée de l'intervalle de temps entre la question du tuteur et la réponse de l'apprenant peut inciter le tuteur à apporter son aide ou bien à attendre (Fox, 1993).

[10] Nous rappelons que, dans la littérature issue des recherches sur la communication médiatisée, les termes « *verbal* » et *« linguistique* » ont le même référent conceptuel, désignant une forme de

1.2.1 Les indices non-verbaux

Lors d'une interaction dialogique, nous avons accès aux expressions faciales, aux gestes et aux actions de notre partenaire. Ainsi, nous produisons et nous percevons plusieurs types d'indices non-verbaux, que nous présentons selon deux catégories : les indices *kinésiques* et les indices *ostensifs-inférentiels*[11] (Kerbrat-Orecchioni, 1998).

1.2.1.1 Les indices non-verbaux kinésiques

Les indices *non-verbaux kinésiques* comprennent les gestes corporels, les regards entre les interlocuteurs, les expressions faciales et la posture (Kendon, 1967 ; Goodwin, 1981). Globalement, leur fonction est d'assurer la régulation des échanges pendant l'interaction, ainsi que d'informer les interlocuteurs sur les intentions et l'état émotionnel d'autrui (Amado & Guittet, 2003). Nous allons décrire leurs fonctions dans la construction du processus de compréhension réciproque et du contenu de la communication.

* *Les gestes corporels*. Certains gestes corporels ont une fonction sur le plan du *processus de communication*, notamment les gestes qui contribuent à maintenir le contact avec les partenaires, ce que Le Breton (1998) nomme les *gestes à dimension phatique*. Ce sont surtout des gestes définis par des conventions culturelles et qui signalent le début et la fin de la prise de parole, facilitant la gestion des tours de paroles. Par exemple, Goodwin (1981) observe que le geste de la main du locuteur en fin de tour de parole a la fonction d'inviter le co-locuteur à prendre la parole (la paume de la main ouverte selon l'axe horizontal) ou bien de l'empêcher de parler (la paume de la main ouverte selon l'axe vertical). Bavelas, Chovil, Coates & Roe (1995) nomment ces gestes « *interactifs* » et ils apportent une preuve empirique à l'hypothèse que ces gestes servent à coordonner le processus de prise de parole entre les interlocuteurs. Ils

communication dialogique (cf. Whittaker, 2002), alors que le terme « *textuel* » est employé surtout pour désigner une forme de communication linguistique en format écrit. Pour des exigences de clarté par rapport aux recherches relatives au format de présentation de l'information (cf. Paivio, 1986) nous jugeons plus pertinent d'employer « *verbal* et *non-verbal* » à la place de « *linguistique* et *visuel* ».

comparent deux conditions expérimentales : dans la condition « collaborative », un individu est chargé de résumer un récit face à un auditoire avec un partenaire, alors que dans la condition « monologique » le même individu doit résumer ce récit tout seul. Les résultats principaux montrent que le nombre des gestes interactifs est plus important en condition collaborative (lorsque les partenaires doivent coordonner leur processus) qu'en condition monologique.

D'autres gestes exercent une fonction sur le *plan du contenu*. Par exemple, des gestes « *emblèmes* », comme la lettre « V » formée avec le deuxième et le troisième doigt de la main, ont une signification conventionnelle partagée par les membres d'une même communauté (McNeill, 1992), suggérant un événement positif. Egalement, les hochements de la tête (approbation *vs*. refus) accompagnent ou parfois se substituent aux énoncés linguistiques (Goodwin, 1981).

* *Les regards*. Sur le plan du *processus de communication*, les regards échangés entre les locuteurs peuvent influer sur l'alternance de parole. Notamment, Kendon (1967) montre que lorsque le locuteur regarde son allocutaire en fin de tour de parole, le temps de transition est plus court que lorsqu'il ne le regarde pas. En d'autres termes, le regard favoriserait la fluidité des échanges entre les interlocuteurs. Cependant, d'autres recherches ont nuancé l'influence du regard sur la gestion des tours de parole (Duncan & Fiske, 1977 ; Beattie, 1981). Par exemple, Beattie (1981) observe que lorsque le locuteur adresse son regard au co-locuteur en fin de tour de parole, l'intervalle de la prise de parole est du même ordre de grandeur que lorsqu'il ne le lui adresse pas.

Sur le *plan du contenu*, pendant que le locuteur parle, le co-locuteur lui manifeste son degré d'attention par l'intermédiaire du regard (Goodwin, 1981). Le locuteur peut ainsi supposer si le contenu de ses énoncés est compris par le co-locuteur et modifier sa contribution en conséquence. Par exemple, Fox (1993) observe que le tuteur arrête son explication dès qu'il voit un regard moins attentif de la part de l'apprenant. Le regard mutuel aurait aussi une

[11] Un troisième type d'indices non-verbaux, les indices proxémiques, concernent la distance spatiale entre les interlocuteurs co-présents (Hall, 1971). Notre domaine d'étude étant les activités de communication médiatisée à distance, nous préférons approfondir surtout les deux autres indices.

fonction d'*activation cognitive*, permettant l'augmentation de l'attention et le rappel d'information. Par exemple, des études expérimentales (Fry & Smith, 1975 ; Otteson & Otteson, 1980, cités in Beall, Bailenson, Loomis, Blascovich & Rex, 2003) suggèrent que le regard peut aussi influer sur la performance de la tâche des apprenants : plus la fréquence des regards adressés par l'enseignant aux étudiants est élevée, plus leur score de rappel est élevé.

En condition de « face à face », les locuteurs ont accès au même environnement physique et peuvent observer les actions que chacun d'eux réalise sur les objets de cet environnement. Un autre type de paramètre est donc disponible, notamment la direction du regard. Sa fonction est de faciliter la compréhension réciproque (Sanford, Anderson & Mullin, 2004), notamment en aidant le co-locuteur et le locuteur à construire leurs « référentiels communs »[12].

* *Les expressions faciales et la posture* influent surtout sur le *plan du contenu* de la communication. Les expressions faciales du co-locuteur (e.g., les plis du front, la direction des lèvres de la bouche, les sourcils, etc.) expriment son intérêt, son niveau de compréhension ainsi que son état émotionnel au regard du locuteur (Ekman, 2003). Concernant la posture, les bras et la partie supérieure du corps indiquent l'engagement de la personne dans l'interaction (Amado & Guittet, 2003). Par exemple, l'ennui de l'interlocuteur lors d'une interaction est exprimé par sa tête appuyée sur son coude (Goodwin, 1981).

1.2.1.2 Les indices non verbaux ostensifs-inférentiels

Les indices *ostensifs-inférentiels* comprennent les gestes déictiques et les actions que les interlocuteurs réalisent sur les objets de leur environnement. Grâce à ces indices, les

[12] La notion de « *référentiel commun* » se prête à de nombreuses interprétations (cf. Giboin, 2004 pour une revue des questions). Par exemple, il désigne une représentation interne d'un support ou d'un contexte externe ou bien une « *représentation fonctionnelle commune aux opérateurs, qui oriente et contrôle l'activité que ceux-ci exécutent collectivement* » (Leplat, 1991). Par ailleurs, Karsenty (1999) emploie la notion de « *contexte partagé* » qui implique l'idée d'un cadre inséré dans la communication, alors que Clark (1996) emploie la notion de « *terrain commun d'entente* » plutôt que celle de « *contexte* » ou « *connaissances mutuelles* » et Sperber & Wilson (1986) préfèrent l'expression « *évidence mutuelle* » plutôt que « *connaissances mutuelles* » ou « *contexte* », car ces expressions ne seraient pas plausibles sur le plan psychologique.

partenaires montrent leurs intentions (Argyle & Graham, 1977 ; Clark & Marshall, 1981) et peuvent ainsi construire immédiatement leurs référentiels communs (Navarro, 2001).

Cela est illustré dans une recherche conduite par Nardi, Kuchinsky, Whittaker, Leichner & Schwarz (1997) sur la fonction des indices ostensifs-inférentiels en milieu professionnel. Pendant trois mois, ils ont observé le travail d'une équipe de neurochirurgiens en salle opératoire. La salle est équipée d'un système vidéo, avec une micro-caméra montée sur un microscope utilisé par le chirurgien. L'image-vidéo est affichée en temps réel sur des écrans situés dans la salle et aussi sur des écrans placés dans des bureaux proches, dans lesquels d'autres médecins peuvent observer les phases de l'intervention. D'après leurs observations, les indices ostensifs-inférentiels ont plusieurs fonctions : a) la réduction des ambiguïtés entre les participants ; b) l'anticipation des exigences et des nécessités du chirurgien, sans qu'il ait besoin de les expliciter verbalement ; c) la coordination des actions des participants ; d) le maintien de l'attention du groupe sur le déroulement de l'activité.

Cette dernière fonction est confirmée par la recherche conduite par Fox (1993) dans les interactions tutorielles en face-à-face. Fox (1993) observe que le tuteur utilise des indices ostensifs-inférentiels pour attirer l'attention de l'apprenant sur l'élément d'un problème (e.g., indiquer avec un stylo une partie d'une équation), de sorte que l'apprenant puisse le prendre en compte dans son raisonnement personnel. De la part de l'apprenant, le geste ostensif adressé au tuteur a une fonction équivalente à une réponse linguistique ou bien souvent l'accompagne, comme lorsque l'apprenant répond au tuteur.

En résumé, les indices non-verbaux, kinésiques et ostensifs-inférentiels (Tableau 1-1), ont principalement pour fonction de gérer l'alternance de parole, de faciliter la compréhension réciproque et de permettre aux co-locuteurs de rendre plus clair le contenu de leur interaction.

	Indices non-verbaux					
	Kinésiques				**Ostensifs-Inférentiels**	
	Gestes corporels	*Regards*	*Expressions faciales*	*Posture*	*Gestes déictiques*	*Actions et objets*
Processus	Alternance et prise de parole	Alternance et prise de parole	Monitorage/expression de l'attention	Monitorage de l'attention	Réduction des ambiguïtés	Usage de pronoms déictiques à la place du nom des objets
Contenu	Construction de référentiels communs	Monitorage et expression attention pour le dialogue ou pour un objet; référentiel commun	Etat émotionnel du locuteur (accord, désaccord)	Intérêt et implication dans l'interaction	Construction de référentiels communs	Construction de référentiels communs

Tableau 1-1 : Les fonctions principales des indices non-verbaux en situation de communication face-à-face.

1.2.2 Les indices verbaux

Nous présentons les indices verbaux qui participent au processus de compréhension réciproque (*plan du processus*) et à la co-construction du contenu (*plan du contenu*).

* Les indices verbaux dédiés au *processus de compréhension réciproque*.

Lorsqu'un locuteur A s'adresse à l'interlocuteur B, comment A sait si B a compris ? Clark & Brennan (1991) signalent que A ne peut pas le savoir à moins que l'interlocuteur B ne lui *manifeste* avoir compris par l'intermédiaire d'un indice évident. Cet indice est donc l'*acceptation* que A fait de la *présentation* de B, une preuve de compréhension (Cohen & Levesque, 1995). Cette manifestation de compréhension peut être positive ou négative (Clark & Brennan, 1991) :

- *la manifestation positive* se concrétise dans la production de *marqueurs verbaux de validation* (ou d'*acceptation*) qui expriment l'agrément du locuteur B par rapport à ce que le locuteur A veut dire. Les formes de manifestation positive sont les suivantes :

l'acquiescement[13], qui est la forme positive d'un accord réciproque et qui clôture une entente commune (e.g. un simple « *ok* » qui peut être interprété comme « je confirme avoir compris ce que tu as dit » ou bien « je comprends ce que tu as dit » ou encore « je partage ton avis »), *la répétition* de l'énoncé du locuteur (e.g. L : « *to the right* » I : « *to the right* » ; Clark & Brennan, 1991), *les répliques courtes* (e.g. « *uhuh, mhm, right* ») (O'Conaill, Whittaker & Wilbur, 1993 ; Yngve, 1970) et les *régulateurs* (e.g. « *uh huh, yeah* ») que le co-locuteur utilise pour montrer son accord au locuteur sans l'interrompre (Schlegoff, 1987);

- *La manifestation négative,* par laquelle l'interlocuteur B manifeste au locuteur A ne pas avoir entendu (stade 1 du modèle de compréhension de Clark) ou avoir mal compris ce que A veut dire (stade 2 du modèle de compréhension de Clark). La manifestation négative se concrétise par la production de *marqueurs de vérification,* comme les *queues de phrase interrogatives* (e.g. « *you changed your ticket, didn't you* ? », Haviland & Clark, 1974) et les *questions fermées* (e.g. « *do you see it in the right corner ?* », Isaacs & Clark, 1987).

* Les indices verbaux dédiés à *la co-construction du contenu*.

Grâce aux indices dédiés à la compréhension réciproque, les locuteurs s'assurent que *ce qui est dit soit compris*. Chaque locuteur utilise le langage pour établir une entente commune adéquate pour satisfaire ses intentions communicatives (Clark, 1996). Cette activité se concrétise dans la production d'énoncés, contenant des *actes de parole*[14] et des *référents* de ces actes. Par exemple, lorsque A *demande* à B une information précise (A : « *what time does your train leave* ? »), A veut savoir l'heure de départ et A *suppose* que B la connaisse et A *veut* que

[13] Des ambiguïtés sont présentes dans la littérature, car un même indice est classé dans des catégories différentes. Par exemple, « *mhm* » est une réplique courte pour O'Conaill & al. (1993) alors que pour Daly-Jones, Monk & Watts (1998) il s'agit d'un acquiescement.

[14] Il nous semble important de distinguer la *notion d'acte de parole* et d'*acte de dialogue* : la première met en jeu l'interaction comme « *une dynamique de co-construction de sens* » (Brassac, 2000) et comme un support « *à une visée intentionnelle d'action* » (Chabrol & Bromberg, 1999), alors que la notion d'acte de dialogue repose sur le concept d'action individuelle (Searle, 1972). Nous situant dans la perspective coopérative de Clark (1996), nous adoptons la notion d'*acte de parole*.

B lui *donne* cette information (B : « *it leaves at two o'clock* »). Si A *accepte* ce que B veut dire grâce à un marqueur verbal (A : « *ok* »), alors cette information devient une *connaissance partagée* entre A et B (leur *common ground*). Ainsi, A manifeste à B avoir compris et valide son énoncé. Les locuteurs sélectionnent les indices verbaux par rapport à *la situation de communication* dans laquelle ils sont impliqués (encadré n.1). Par exemple, si A rencontre B qui porte des valises dans la rue, A entame une *conversation* pour savoir ce que B va faire ; ou bien, si A et B sont en train de travailler ensemble, A dirige le *dialogue* avec B afin de savoir de combien de temps ils disposent.

Encadré n.1 : contrat de communication, conversation & dialogue

Toute action de communication s'insère dans une situation sociale précise, qui est définie par un *contexte externe*, c'est-à-dire par l'ensemble des caractéristiques physiques (le lieu, les modalités de communication, etc.) et par un *contexte interne*, c'est-à-dire par les représentations que les locuteurs se construisent sur les *enjeux* de la situation, ainsi que sur les *compétences et les connaissances* des co-locuteurs (Pavard & Karsenty, 1997). En d'autres termes, une interaction n'est communicative que lorsque les interlocuteurs reconnaissent implicitement ou explicitement un *contrat de communication* (Ghiglione, 1986). Cette notion, issue de la psychologie sociale, propose un cadre interprétatif qui permet au co-locuteur de savoir quelles croyances et quelles représentations sont pertinentes pour rendre intelligibles les énoncés du locuteur (Bromberg, 2004). Par exemple, une personne entre dans un café et demande au serveur s'il vend des cigarettes : les deux interlocuteurs construisent un contrat de communication « marchand » qui est différent du contrat de communication qui s'établirait si la même personne avait posé la même question à un client du café. Ainsi, des stratégies discursives sont activées plutôt que d'autres, jusqu'à ce qu'un nouveau contrat soit mis en place. Selon cette notion, un *dialogue* se distingue d'une *conversation* par le fait que, dans le dialogue, les participants acceptent de partager un certain nombre de présuppositions et de suivre une certaine argumentation pour accomplir une *activité commune* (Karsenty & Falzon, 1993).

Nous présentons maintenant les indices verbaux intrinsèques à la situation de communication qui concerne notre recherche, notamment le dialogue tutoriel.

1.2.3 Les indices verbaux dans le dialogue tutoriel

Des champs de recherche divers s'intéressent au dialogue tutoriel : de la psychologie de l'apprentissage (Wynnikamen, 1998) à la psycholinguistique (Chi, 1997 ; Fox, 1993), de l'ergonomie cognitive (Navarro, 2001) à l'intelligence artificielle (Graesser, Person &

Magliano, 1995 ; Wenger, 1987). En accord avec la la notion de contrat de communication (cf. encadré n.1), nous définissons le *dialogue tutoriel* comme une activité de communication où un tuteur, sans agir à la place d'un apprenant (ce qui est connu comme *effet Topaze* ; Bruner, 1983), l'assiste, le guide et le motive, afin que ce dernier construise des représentations cognitives correctes. Le dialogue tutoriel implique :

- de la part du tuteur, qu'il *observe* l'activité de l'apprenant afin d'en *comprendre* les difficultés et qu'il *décide* le type d'aide à lui offrir (Wenger, 1987)[15];
- de la part de l'apprenant, qu'il *comprenne* l'aide du tuteur et *puisse l'intégrer* dans des représentations circonstancielles (Richard, 1990) pour construire des modèles mentaux corrects (Gentner & Stevens, 1983).

Comme toute activité de communication, le dialogue tutoriel comporte un processus de compréhension réciproque et un contenu co-construit par les participants lors de l'interaction.

1.2.3.1 Les marqueurs dédiés au processus de compréhension réciproque

Nous décrivons les marqueurs utilisés par l'apprenant (i-ii) et par le tuteur (iii-iv).

i) Apprenant : Les marqueurs de vérification de compréhension

Afin de s'assurer d'avoir bien compris, l'apprenant invite le tuteur à répéter son énoncé (e.g., « *what ?* », Fox, 1993 ; « *pardon ?* » Wynnikamen, 1998). D'après Graesser & al. (1995), 54% des questions que l'apprenant adresse au tuteur sont consacrées à obtenir une confirmation de compréhension de la part du tuteur.

ii) Apprenant : Les marqueurs de validation de compréhension

Les répliques courtes, les acquiescements et les régulateurs utilisés par l'apprenant offrent au tuteur un indice sur le niveau de compréhension atteint par l'apprenant (Chi,

[15] Concernant le tuteur, ces opérations cognitives sont à rapprocher d'une part de l'activité de diagnostic et d'autre part de l'activité d'assistance. Le *diagnostic*, typique du domaine médical (Clancey, 1985 ; Sebillotte, 1984) et de la gestion des processus dynamiques (Hoc & Amalberti, 1994 ; Hoc, 1996), se compose de deux phases : une phase analytique, dans laquelle le sujet se doit d'estimer les informations recueillies sur un processus en cours, et une phase synthétique, où le sujet se doit de produire des hypothèses relatives à l'état de processus (Hoc, 1996). L'*assistance* est typique des situations de communication où les participants ne possèdent ni le même vocabulaire ni les mêmes connaissances (Falzon, 1994).

1997). Ils sont souvent produits par l'apprenant pendant que le tuteur lui fournit une explication (Graesser & al., 1995).

iii) Tuteur : Les marqueurs de vérification de compréhension

Pour vérifier que l'apprenant a compris, le tuteur utilise généralement des queues de phrases interrogatives (e.g., «*… this is the value of the normal function, isn't it ?* » Fox, 1993) ou des questions fermées (e.g., « *do you understand ?* », Graesser & al., 1995). Person & Graesser (2003) montrent que, malgré l'usage répandu, les marqueurs de vérification de compréhension ne permettent pas de vérifier le degré réel de compréhension des apprenants. Ils observent en fait que les apprenants qui répondent ne pas comprendre obtiennent de meilleurs scores (un résultat analogue est reporté par Chi, Bassok, Lewis, Reinmann & Glaser, 1989) et ils attribuent ce résultat à la capacité limitée des apprenants de juger eux-mêmes de leur propre compréhension (cf. Glenberg, Wilkinson & Epstein, 1982).

iv) Tuteur : Les marqueurs de validation de compréhension

Ces marqueurs, qui prennent surtout la forme d'acquiescements, sont utilisés par le tuteur pour donner son accord à ce que l'apprenant dit ou veut faire (Chi, 1997). Dans la littérature, la classification des acquiescements n'est pas unanime. Par exemple, Brandle (1998), qui étudie les dialogues tutoriels dans l'apprentissage de la chimie, propose une catégorisation de l'acquiescement selon le critère de la *polarité* : l'acquiescement à *polarité positive* (e.g. Apprenant = « *CO augmente* » (réponse correcte) Tuteur = « *C'est bien* »), l'acquiescement à *polarité neutre* (e.g., A : « *CO augmente* » (réponse incomplète), T : « *quoi d'autre ?* ») et à *polarité négative* (e.g., A : « *CO augmente* (réponse négative), T : « *c'est faux* »). Graesser & al. (1995) observent que, comme un nombre élevé d'acquiescements négatifs pourrait inciter les apprenants à se décourager, les tuteurs produisent plutôt des acquiescements nuancés dans le cas d'une fausse réponse.

1.2.3.2 Les actes de parole dédiés à la co-construction du contenu du dialogue

Nous décrivons d'abord les actes de parole de l'apprenant (i-ii) et ensuite les actes de parole généralement utilisés par le tuteur (iii-iv).

i) Apprenant : Les actes de parole consacrés à demander de l'aide au tuteur

* *La question*. Le tuteur peut comprendre les faiblesses et les besoins formatifs du novice en interprétant le contenu de ses questions. D'après Graesser & Person (1994), l'apprenant pose des questions car il n'a aucune représentation (ou bien elle n'est pas adéquate) de l'état du problème (i.e., quoi faire ?), des modes d'action (i.e., comment le faire ?) et de la nature des entités mises en œuvre (i.e., une méthode ou une théorie ?). Pour Graesser & al. (1995), ces types de questions représentent 29% de la quantité globale des questions de l'apprenant.

ii) Apprenant : Les actes de parole destinés à « apprendre »

* *La question*. Pour l'apprenant, la question n'est pas uniquement un moyen pour demander de l'aide, mais aussi pour faire valider ses connaissances par le tuteur (e.g. « *it is true that I have in mind is correct* ? ») ou pour faire vérifier le travail accompli (Graesser & Person, 1994).

iii) Tuteur : Les actes de parole destinés à connaître les besoins de l'apprenant

**La question*. Le tuteur peut connaître les objectifs et les difficultés de l'apprenant par l'intermédiaire de questions. Après que l'apprenant lui a répondu, le tuteur interprète sa réponse pour émettre des hypothèses sur ses besoins formatifs, en d'autres termes produire un *diagnostic* (Salembier, 1993). Par exemple, dans leur étude sur la nature de la question produite par le tuteur dans le domaine d'apprentissage de la physique, Porayska-Pomsta, Mellish & Pain (2000) observent que le tuteur utilise la « question test » (e.g. « *this button shifts colors, isn't it ?* ») et surtout la « question fermée » (e.g. « *do you know what PH is* ? ») pour savoir ce que l'apprenant connaît, ce qu'il fait ou ce qu'il a l'intention de faire.

iv) Tuteur : Les actes de parole destinés à « faire apprendre »

Les recherches sur le dialogue tutoriel montrent que le tuteur utilise des actes de parole différents (Chi, Siler, Jeong, Yamauchi & Hausmann, 2001). Dans cette section, nous évoquons les actes les plus pertinents pour notre étude.

* *L'explication*. Sur le plan de la linguistique, d'après Cahour (1997) l'explication implique un référent-cible (i.e. ce qu'il faut expliquer), une relation explicative (i.e. la relation causale et modale) et le contenu de l'explication. Bien que l'explication soit souvent

utilisée par le tuteur, son efficacité pour l'apprentissage n'est pas certaine. Par exemple, Chi & al. (2001), comparent le score d'un test de vingt-deux étudiants impliqués dans l'apprentissage des notions de physiologie humaine. Une moitié interagit avec un tuteur qui fournit des explications, l'autre moitié est assistée par un tuteur qui ne peut pas donner d'explication mais qui doit utiliser des phrases-indices, que les expérimentateurs ont précédemment choisies après une analyse préalable de corpus. Le score d'apprentissage des novices est équivalent dans les deux conditions. Chi & al. (2001) suggèrent qu'en condition « sans explications », le novice pose un nombre de questions et de tentatives d'auto-explications[16] plus important qu'en condition « avec explications ». En d'autres termes, en condition « sans explication » l'apprenant produit surtout des actes de parole consacrés à la co-construction du contenu du dialogue (« *…that make a substantive contribution with respect to the content…* », Chi & al., 2001 : 482). En revanche, en condition « avec explications », l'apprenant produit un nombre de marqueurs consacrés au processus (régulateurs et répliques courtes) plus important qu'en condition « sans explications », montrant ainsi que sa contribution se focalise sur le processus d'intercompréhension.

* *L'allusion.* L'allusion est une manière de réveiller une idée dans l'esprit de l'interlocuteur sans l'exprimer de manière directe. Pour Chi & al. (2001), l'allusion favorise l'apport constructif au dialogue. D'après Horaek & Fiedler (2001), le tuteur privilégie l'allusion : *a)* lorsqu'il veut suggérer à l'apprenant comment cerner le problème ; *b)* après que l'apprenant a donné une réponse fausse (plutôt que le corriger ou lui donner une évaluation négative de sa réponse).

* *La question*. Le tuteur ne pose pas de questions seulement pour connaître les besoins de l'apprenant mais aussi pour favoriser son apprentissage. Person & Graesser (2003) observent que : *a*) plus la question du tuteur est profonde et articulée, plus l'apprenant produit une réponse détaillée et cohérente, montrant ainsi son niveau de

[16] L'auto-explication est une explication que l'apprenant propose sans que le tuteur lui demande de manière directe de préciser les relations causales ou modales entre des phénomènes (Chi & al.,1989).

compréhension[17] ; *b*) lorsque le tuteur pose une question indirecte à la suite d'une question de l'apprenant, l'apprenant construit la réponse au problème par lui-même (Di Paolo, Graesser, Hacker & White, 2002 ; Hume, Michael, Rovick & Evens, 1993 ; Porayska & al., 2000).

* *La correction*. Le tuteur utilise la correction pour signaler à l'apprenant l'erreur commise (Di Paolo & al., 2002) et pour lui fournir des indications sur la procédure à suivre pour obtenir la solution correcte (Hume & al., 1993). Pour cela, la correction ne permet pas à l'apprenant de prendre pleinement conscience de son erreur, car il se limite à appliquer les instructions sans modifier les fausses conceptions responsables de son erreur (VanLehn, Siler, Murray & Bagget, 2003).

* *La motivation*. D'après Lepper, Woolverton, Mumme, & Gurtner (1993), les composants émotionnels et motivationnels occupent une place fondamentale dans l'interaction tutorielle. Le tuteur est censé donner à l'apprenant de la confiance en ses moyens, provoquer en lui de la curiosité et de l'intérêt pour l'apprentissage. La motivation se concrétise ainsi par des formes linguistiques d'encouragement et de soutien au travail de l'apprenant.

Synthèse

La communication humaine est multimodale, car elle s'appuie sur des indices différents, perçus et produits en même temps par les interlocuteurs. Les indices non-verbaux interviennent surtout dans la régulation du processus de compréhension réciproque (alternance et temps de parole) et peuvent aider aussi la construction du contenu. On pourrait donc supposer que lorsque les locuteurs ne peuvent plus disposer des indices non-verbaux (i.e., condition de communication audio-seul), la qualité de l'interaction se dégrade. En revanche, comme nous le verrons au chapitre 2 et 3, en situation de communication médiatisée les

[17] Chi & al. (2001) observent que plus la question du tuteur est profonde, plus la réponse de l'apprenant l'est. En effet, ils trouvent une corrélation positive entre le gain d'apprentissage des apprenants et le niveau de profondeur de leurs réponses. Les réponses des sujets sont classées selon deux niveaux : un *niveau de surface*, où l'apprenant répond à une question sur la définition d'une notion

caractéristiques du dispositif de communication ne permettent que la perception simultanée d'un nombre limité d'indices non-verbaux : cela limite la propriété multimodale de la communication et structure différemment l'interaction dialogique entre les partenaires distants.

Les indices verbaux participent au processus d'intercompréhension et sont responsables de la co-construction du contenu du dialogue. Nous avons décrit les indices verbaux utilisés par le tuteur et par l'apprenant dans le dialogue tutoriel. Pour comprendre les besoins de l'apprenant, le tuteur a besoin d'informations à partir desquelles il peut inférer le type d'aide à fournir à l'apprenant. Si on considère l'asymétrie des compétences dans la situation tutorielle, le type de contribution de l'apprenant reflète le contrôle que le tuteur exerce sur l'interaction. Dans ce sens, certains indices produits par le tuteur ne permettent à l'apprenant que de contribuer au processus (e.g. explication, question fermée, correction), alors que d'autres lui donnent la possibilité d'enrichir le contenu (e.g. allusion, question ouverte). Cependant, nous ne pouvons pas généraliser cette considération, car les études citées concernent surtout l'apprentissage de connaissances déclaratives (i.e., notions de physiologie humaine, physique) et non l'apprentissage de connaissances procédurales.

Conclusion

L'activité de communication implique une action *sur* autrui ainsi qu'une interaction *avec* autrui. Les modèles ostensifs-inférentiels et coopératif suggèrent que cette activité consiste à interpréter et produire des indices de communication dans une logique de coopération, c'est-à-dire que les locuteurs co-construisent leur interaction, en choisissant, modifiant et reformulant leurs énoncés jusqu'à atteindre un niveau de compréhension satisfaisant. Cette co-construction est influencée par les représentations que chaque locuteur se fait de l'enjeu de la situation, des compétences et de l'identité de son partenaire, et elle est construite sur les indices de communication disponibles. Par exemple, en situation de communication audio à distance, les locuteurs ne disposent que des indices verbaux et para-

et un *niveau profond*, où l'apprenant répond à une question exigeant des connexions de type logico-causale.

verbaux, alors qu'en situation de face à face, les indices non-verbaux sont disponibles (cf. Clark & Brennan, 1991).

Pour mieux comprendre la fonction de ces indices, nous nous sommes appuyés sur une situation de communication précise, le dialogue tutoriel. Notre domaine de recherche étant la communication médiatisée synchrone, nous devons nous interroger sur les fonctions des indices verbaux et non-verbaux en situation distale : *Comment co-construisent-ils le processus et le contenu de leur interaction ? Faut-il se voir pour se comprendre ?*

Lorsque les partenaires sont délocalisés, ils communiquent par l'intermédiaire de dispositifs dédiés à la transmission du signal audio et vidéo. Chaque dispositif véhicule des indices communicationnels, à savoir les indices verbaux et non-verbaux : certains dispositifs ne transmettent que les indices verbaux (e.g., le téléphone, la communication radio), alors que d'autres véhiculent aussi les indices non-verbaux (e.g. audio-vidéo conférence, vidéo-téléphone).

Dans les chapitres suivants, nous voulons comprendre le rôle des indices non-verbaux sur les activités de communication à distance. Notamment, dans le chapitre 2 nous abordons les recherches concernant *l'image-vidéo du partenaire distant*, qui véhicule les indices non-verbaux kinésiques ; le chapitre 3 est dédié à *l'image-vidéo de l'espace de travail du partenaire distant*, véhiculant les indices non-verbaux ostensifs-inférentiels. A la suite de ces deux chapitres, nous avançons des hypothèses de recherche relatives à l'impact de l'image-vidéo sur l'interaction tutorielle à distance.

Chapitre 2 - Voir le partenaire distant : un avantage ou une entrave pour l'activité de communication ?

Peut-on prévoir si les indices relatifs au partenaire distant influencent l'activité de communication ? A l'état actuel, le domaine de recherche de la communication médiatisée ne propose que deux perspectives théoriques principales : la théorie de la présence sociale et de la richesse des média.

La *théorie de la présence sociale* (Short, Williams & Christie, 1976) postule que les sujets se construisent des impressions différentes de la « présence sociale[18] » du partenaire distant selon la modalité de communication utilisée. La *présence sociale* n'est pas un attribut défini *a priori*, mais elle résulte d'une impression subjective associée à un moyen de communication. Ainsi, chaque moyen de communication se distingue des autres par la quantité d'indices verbaux et non-verbaux transmis : plus un moyen de communication favorise la transmission

[18] Le concept de « présence sociale » ne doit pas être confondu avec celui de « coprésence » ou avec celui de « télé-présence ». Le premier, issu des recherches menées en psychologie sociale par Short & al. (1976), concerne les technologies de la communication et l'appropriation subjective du medium de communication ; le deuxième, issu de la sociologie de Goffman (1963), focalise sur les règles de comportement humain lors d'une interaction physiquement localisée dans un même environnement (Durlach & Slater, 2000 ; Slater, Sadagic & Schroeder, 2000) ; enfin, le troisième, utilisé en cybernétique par Minsky (1980), définit les environnements où les utilisateurs manipulent à distance des objets physiques par l'intermédiaire de robots.

de ces indices, plus le niveau de « présence sociale » perçu par les sujets est élevé et plus ils communiquent efficacement. Le corollaire de la théorie est le suivant : les moyens de communication qui ne véhiculent pas les indices non-verbaux (e.g. l'écrit, le téléphone) ne favorisent pas la communication de manière aussi efficace que la condition « face à face ».

La *théorie de la richesse des media* affirme que les moyens de communication qui véhiculent le plus d'indices (verbaux et non-verbaux) facilitent l'interprétation d'une information et en réduisent l'ambiguïté (Daft & Lengel, 1986 ; Trevino, Lengel & Daft, 1990). Pour Daft & Lengel (1986) trois paramètres principaux influent sur l'interprétation d'une information :

- *la vitesse de retour de l'information*[19] : plus le feedback est court, plus le temps d'interprétation d'une information est court ;
- *le niveau d'interactivité*, résultant du nombre de modalités sensorielles simultanées (i.e. le texte, l'audio, l'audio-vidéo) et de la qualité de leur transmission ;
- *l'expressivité*, comme la clarté de l'intonation, la mimique de l'expression faciale, etc.

En synthèse, d'après ces deux positions théoriques, il resulte que plus une technologie de communication se rapproche de la situation « face à face », plus la communication est efficace. Or, en situation de communication médiatisée, le taux de transmission du signal audio et vidéo est un paramètre qui détermine la perception des indices verbaux et non-verbaux, et donc influe sur l'interprétation des informations. Par exemple, Munhall, Gribble, Sacco & Ward (1996) suggèrent que 15-17 frames par seconde (fps)[20] sont suffisants pour percevoir les mouvements faciaux en situation de visioconférence (cf. Kies, 1997 ; Mullin, Smallwood, Watson & Wilson, 2001 ; Watson, 1999 pour une revue des questions). Vitkovitch & Barber (1994) suggèrent que 16.6 fps est le seuil minimal de transmission du signal vidéo pour pouvoir lire les lèvres du partenaire distant. Des études conduites pour évaluer les effets de transmission du signal vidéo sur les activités de communication à distance indiquent qu'au-dessous de 12 fps :

[19] On peut noter que ce critère est proche du critère de *séquentialité* de Clark & Brennan (1991).

[20] L'unité de mesure « frames par seconde » indique le nombre d'images affichées en une seconde. Plus le nombre est élevé, plus l'animation est fluide, 25 images par seconde étant la fréquence minimale nécessaire pour percevoir une impression de fluidité dans une vidéo.

- les partenaires distants ne construisent pas un dialogue aussi interactif qu'en condition de « face à face », mais ils jouent plutôt des monologues alternés, caractérisés par des tours de parole peu nombreux, très longs et peu d'interruptions (Masoodian, Apperley & Frederickson, 1995; O'Conaill & al., 1993;O'Malley, Langton, Anderson, Doherty-Sneddon & Bruce,1996) ;

- la compréhension réciproque est moins immédiate, car les marqueurs verbaux consacrés à la vérification de la compréhension du partenaire distant sont plus nombreux qu'au dessus de 12 fps (cf. Matarazzo & Sellen, 2000 ; Sanford & al., 2004).

Malgré l'apport de ces recherches, la technologie à bande large assure désormais une qualité très élevée de transmission du signal audio et vidéo. On pourrait donc supposer que les moyens de communication véhiculant à la fois les indices verbaux et non-verbaux (kinésiques et ostensifs-inférentiels) offrent les meilleures conditions pour rendre efficace toute activité de communication. Cependant, de nombreux résultats expérimentaux contredisent cette hypothèse: comme nous le montrerons, parfois des moyens de communication « pauvres » (véhiculant moins d'indices non-verbaux) facilitent les interactions communicatives plutôt que des moyens de communication « riches » (véhiculant plus d'indices verbaux et non-verbaux). Ainsi, nous retenons que d'autres questions de recherche sont à envisager, concernant notamment les caractéristiques spécifiques des indices non-verbaux :

- *Est-ce que l'image du partenaire distant favorise ou entrave l'interaction entre les partenaires* ?

- *Est-ce que l'image complète du partenaire distant en est responsable ? Ou bien peut-on considérer des indices spécifiques de cette image-vidéo ?*

- *Par exemple, les regards échangés entre les partenaires influencent-ils leur interaction ?*

Ces questions ne sont pas approfondies par la littérature du domaine de la communication médiatisée, où les questions de recherche se cantonnent à évaluer globalement les apports et les entraves de l'image-vidéo (cf. Foulon-Molenda, 2002), suivant une tradition issue de premières recherches expérimentales (Chapanis, 1975 ; Ochsman & Chapanis, 1974). Ces recherches pionnières se proposaient d'évaluer l'impact de l'image-vidéo du partenaire distant sur la performance de réalisation de la tâche plutôt que sur les interactions communicatives, montrant que l'image-vidéo du partenaire distant n'induit aucune

amélioration sur la réalisation des tâches collaboratives, à l'exception des tâches de négociation (Reid, 1977 ; Short & al., 1976; Williams, 1977).

Afin d'éclaircir les questions de recherche que nous venons d'évoquer, nous avons sélectionné des études expérimentales portant sur les activités collaboratives[21] à distance. Notre sélection concerne le domaine des systèmes informatiques de support au travail coopératif, ainsi que de la psychologie ergonomique et de la psychologie sociale appliquée aux technologies de la communication. Les tâches expérimentales utilisées sont assez variées, allant de la planification et de la négociation (tâches plutôt déclaratives) à la réparation d'objets (plutôt procédurale). Concernant les catégories théoriques utilisées par les auteurs pour interpréter les résultats, la plupart des auteurs évoquent principalement les suivantes :

a) l'*interactivité,* mesurée de manière directement proportionnelle au nombre de mots, de tours de parole et d'interruptions produit par les locuteurs ;

b) l'*efficacité,* évaluée selon le rapport entre le score de réalisation de la tâche et la quantité de mots produits[22].

Comme nous le montrons, selon les auteurs, le volume des dialogues est calculé soit sur la base de seuls mots produits, soit des mots et des interruptions ou aussi des tours de parole.

Notre revue de la littérature est ainsi structurée : après une première section consacrée aux recherches expérimentales qui étudient l'apport de l'image-vidéo du partenaire distant (2.1), nous présentons des études expérimentales qui s'interrogent sur les effets de l'image-vidéo relative au visage seul et des effets de l'image-vidéo relative au corps du partenaire distant (2.2). Enfin, la dernière section propose une synthèse des études focalisées sur l'influence du regard mutuel et de la direction du regard du partenaire distant (2.3).

[21] Pour notre objet d'étude, les termes « collaboratif » et « coopératif » seront utilisés de manière interchangeable dans la suite du manuscrit. Ils désignent des types d'activités humaines où deux ou plus individus communiquent, partagent des savoirs et des savoir-faire, et opèrent sur leur environnement afin de répondre à des attentes externes (e.g. la consigne expérimentale) et internes (e.g. atteindre un but inhérent à l'environnement cognitif personnel).

[22] Cependant, Fussell, Kraut & Siegel (2000 : 22) définissement l'efficacité dialogique comme « … *the ease and brevity of referring expressions and other utterances.* ».

2.1 Faut-il se voir pour bien coopérer ?

L'étude de terrain réalisée par Isaacs & Tang (1994) a pour objectif de comprendre les mécanismes de régulation des interactions entre les membres distants, appartenant à un même groupe de travail. Leur hypothèse est que, dans les activités humaines à fort engagement personnel (e.g. conflit, négociation, jugements exogène et endogène, etc.), l'image des partenaires distants influencerait la structure et le contenu des interactions dialogiques du groupe. Pour le vérifier, des groupes de cinq ingénieurs participent à des réunions de travail, relatives à des projets en cours de conception. Les membres de chaque groupe participent aux réunions selon trois conditions de communication :

c1. « audio seul » : les membres du groupe de travail communiquent uniquement par audioconférence, grâce à un signal audio transmis en continu ;

c2. « audio & vidéo du partenaire distant » : les membres du groupe communiquent par un dispositif audio-vidéo doté d'un taux élevé de transmission du signal vidéo et d'un signal audio transmis en continu (*full-duplex*) ;

c3. « face à face » : les membres du groupe sont réunis autour d'une table en face à face (condition contrôle).

Les enregistrements audio sont transcrits et codés, les enregistrements vidéo sont analysés afin de calculer le nombre de regards et les types de gestes.

Les effets sur les interactions communicatives. Sur le plan du processus de communication, les tours de parole sont plus courts et plus nombreux en c2 (condition « audio & vidéo du partenaire distant ») qu'en c1 (« audio seul »). Sur le plan du contenu, les actes de parole consacrés à vérifier et à confirmer la compréhension du partenaire sont plus nombreux en condition « audio seul » qu'en condition « face à face » et « audio & vidéo du partenaire distant ». D'après l'analyse des enregistrements audio-vidéo, les membres du groupe montrent leur désaccord surtout en produisant des indices non-verbaux (i.e., les mouvements de la tête, les mouvements faciaux) plutôt que par des indices verbaux.

En synthèse, sur le plan de la structure des interactions communicatives, les auteurs considèrent que lorsque les indices non-verbaux du partenaire distant sont visibles :

- les participants peuvent facilement contrôler l'intercompréhension ;

- la gestion de l'alternance de la parole est facilitée (tours de parole moins longs) et le dialogue est plus interactif ;
- la gestion du conflit entre les différents points de vue est facilitée, permettant aux interlocuteurs d'afficher leur accord ou désaccord aux autres membres du groupe.

Il semblerait donc que l'image-vidéo du partenaire distant soit nécessaire pour favoriser les activités humaines coopératives du type « réunions de projet ».

Cependant, ces résultats doivent être évalués à la lumière des limites de la méthode d'analyse des données : en effet, le codage des dialogues est réalisé par un nombre non spécifié de codeurs et la valeur de l'accord inter-juge n'est pas précisée.

Doherty-Sneddon, Anderson, O'Malley, Langton, Garrod & Bruce/1 (1997)[23] supposent que si le partenaire distant est visible et, de plus, connu personnellement par les autres participants, alors le dialogue sera plus interactif. Pour vérifier cette hypothèse, Doherty-Sneddon & al./1 (1997) utilisent la tâche expérimentale consistant à reproduire un parcours sur une carte (*map task*, cf. Boyle, Anderson & Newlands, 1994). La tâche impose deux rôles : un instructeur et un sujet-guidé. Le premier possède une carte géographique, sur laquelle un chemin a été tracé. Son but est de fournir des indications précises au sujet-guidé, qui doit reproduire le même chemin sur un autre exemplaire de la même carte. Les deux cartes diffèrent seulement pour quelques détails (ex : la position d'un arbre, etc.) et ces différences obligent l'instructeur et le sujet-guidé à établir une entente commune. Pour l'expérimentation, soixante-quatre binômes (tous étudiants) sont recrutés et répartis de manière aléatoire selon deux conditions :

c1. « audio seul» : les deux sujets ne peuvent pas se voir et ne peuvent communiquer que via un signal audio transmis en continu ;

c2. « audio & vidéo du partenaire distant » : le sujet-instructeur et le sujet-guidé peuvent se voir par un dispositif de visioconférence, avec un signal vidéo de

[23] Doherty-Sneddon & al. (1997) réalisent deux expérimentations : dans cette section nous résumons la première expérimentation. Comme pour d'autres études présentées, nous indiquons avec l'expression /1, /2, etc. la position de l'étude citée à l'intérieur de la référence bibliographique.

qualité analogique, et communiquent grâce à un signal audio transmis en continu.

Chaque sujet répète la tâche quatre fois : deux fois en tant qu'instructeur et deux fois en tant qu'exécutant. Pour chaque rôle, le sujet accomplit la tâche une fois avec un partenaire connu, une fois avec un partenaire inconnu. Les mesures récoltées concernent les dialogues (transcrits, codés et analysés selon la méthode de Kowtko, Isard & Doherty-Sneddon (1991)), la durée des regards échangés entre les partenaires pendant la session et la réalisation de la tâche (dont le score est calculé selon la différence géométrique entre le parcours dessiné par le sujet-guidé et le parcours imprimé sur la carte de l'instructeur : plus la différence est petite, moins le score est important).

Les effets sur les interactions communicatives. En ce qui concerne la structure des dialogues, il en résulte que les tours de parole, les mots, les interruptions et les marqueurs phatiques d'attention sont plus nombreux en c1 qu'en c2. Sur le plan du contenu, les actes de parole destinés à vérifier la compréhension du partenaire sont plus nombreux en condition «audio seul » qu'en condition « audio & vidéo du partenaire distant ».

Les effets sur la réalisation de la tâche. Le score de réalisation de la tâche est équivalent dans les deux conditions expérimentales, donc aucune différence n'est constatée.

Les auteurs suggèrent que, à parité de performance de la tâche, la condition « audio & vidéo du partenaire distant » est la plus efficace, car elle contient moins de mots (et de tours de parole). Pour appuyer leur argumentation, ils analysent les regards échangés entre les partenaires dans la phase la plus difficile de la tâche, c'est-à-dire lorsque le sujet-instructeur et le sujet-guidé doivent faire face aux différences entre leur carte respective. Il en résulte que la fréquence des regards échangés entre les deux sujets augmente de manière significative. La raison plausible de cette augmentation est la suivante : la tâche de co-construction du parcours met en jeu une relation asymétrique sur le plan des connaissances, car le sujet-instructeur connaît un plus grand nombre d'informations que le sujet-guidé. Ainsi, le sujet-instructeur fournit plus d'informations que le sujet-guidé, et il utilise le regard afin de vérifier que le sujet-guidé comprenne ses indications. Cet argument semble cohérent avec le fait que les actes de parole pour contrôler la compréhension du partenaire sont plus nombreux en condition « audio seul » qu'en condition « audio & vidéo du partenaire distant ».

En résumé, d'après Doherty-Sneddon & al. 1/(1997), la condition « audio & vidéo du partenaire distant » est la plus efficace car le regard permet de vérifier la compréhension d'autrui (du locuteur vers l'allocutaire) et de manifester sa propre compréhension (de l'allocutaire vers le locuteur).

Anderson, Newlands, Mullin, Fleming, Doherty-Sneddon & Van Der Velden/1 (1996) supposent que l'image du partenaire distant favoriserait des activités coopératives complexes, où les individus doivent rechercher de l'information, planifier, prendre des décisions et négocier avec une autre personne. Ils construisent ainsi une tâche expérimentale impliquant deux rôles : un agent de voyage et un client. Le client doit organiser un voyage en avion, et il doit optimiser son propre parcours, visitant le plus de villes possible. Dans cet objectif, il doit respecter certaines contraintes : le temps limité à 15 minutes, un budget limité et les correspondances des vols des compagnies aériennes. L'agent de voyage doit le renseigner sur les horaires et les prix des vols. Les sujets « clients » (trente étudiants) et les sujets « agents de voyage » (trois chercheurs associés au projet) sont répartis de manière aléatoire dans les trois conditions expérimentales suivantes :

c1. « audio seul » : l'agent et le client ne peuvent communiquer que grâce à un signal audio transmis en continu ;

c2. « audio & vidéo du partenaire distant » : l'agent et le client se voient (visage et buste) grâce à une connexion vidéo câblée et ils communiquent par voie orale grâce à un lien audio transmis en continu ;

c3. « face à face » : l'agent et le client sont réunis autour d'une table, chacun peut voir le visage et le buste de l'autre.

L'agent et le client ont à leur disposition une carte des aéroports des USA : en format papier en condition face à face, en format électronique dans les deux autres conditions. Les données expérimentales concernent les dialogues entre l'agent et le client (analysés par leur structure et par leur contenu), le score de réalisation de la tâche (selon le nombre de villes indiquées sur le plan du voyage proposé par le client) et les impressions subjectives des participants (les sujets doivent évaluer sur une échelle de 0 à 5 la facilité de communication et la facilité d'apporter des changements à leur itinéraire).

Les effets sur les interactions communicatives. En ce qui concerne la structure des dialogues, il en résulte que les mots et les tours de parole en c2 (condition « audio & vidéo du partenaire distant ») sont du même ordre de grandeur qu'en c1 (« audio seul »). En revanche, les dialogues sont plus courts (en mots et en tours de parole) en condition « face à face » qu'en c1. Les interruptions sont moins nombreuses (environ 30%) en condition « audio & vidéo du partenaire distant » qu'en condition « audio seul ». Plus précisément, les agents de voyage parlent moins en condition « face à face » qu'en condition « audio seul ». En ce qui concerne le contenu, les solutions proposées par le client en « audio & vidéo du partenaire distant » et « audio seul » sont moins nombreuses que les solutions proposées en condition « face à face ».

Les effets sur la réalisation de la tâche. Aucune différence significative n'est constatée parmi les trois conditions expérimentales.

Les effets sur les impressions subjectives. D'après les résultats, les sujets-clients préfèrent communiquer avec leur agent de voyage en condition « face à face » plutôt que par une liaison audio seule ou audio-vidéo (ces dernières conditions ne diffèrent pas de manière significative).

Sur le plan de la structure du dialogue, lorsque l'image du partenaire distant est disponible, les tours de parole sont plus courts et le nombre d'interruptions moins nombreux que dans la condition « audio seul ». D'après les auteurs, grâce à l'image du partenaire distant, le niveau d'interactivité du dialogue est plus important qu'en condition « audio seul » et aussi important qu'en condition « face à face » (les tours de parole et les interruptions sont du même ordre de grandeur). Sur le plan du contenu, les solutions proposées par les sujets sont plus nombreuses en « face à face » que dans les autres conditions, où le nombre des solutions est équivalent. Concernant la réalisation de la tâche, le nombre de villes visitées est équivalent quelle que soit la condition de communication. En d'autres termes, d'une part la condition « face à face » favorise une productivité de solutions plus élevée que les conditions à distance (peut-être parce qu'en « face à face » les sujets disposent d'un plan au format papier *vs*. au format électronique dans les autres conditions), d'autre part les solutions proposées sont de qualité équivalente entre les différentes conditions. Sur le plan des impressions subjectives, les sujets préfèrent interagir en condition « face à face » que dans les deux conditions à distance.

Une interprétation possible pourrait s'appuyer sur le fait que les sujets ne sont entraînés ni à l'utilisation du dispositif audio-vidéo ni à la manipulation de la carte en format électronique.

Daly-Jones & al. (1998) étudient l'impact de l'image du partenaire distant sur une activité de prise de décision collaborative. Leur hypothèse de recherche est que l'image du partenaire distant facilite le processus de négociation, car les expressions du visage et les mimiques sont visibles et permettent à chacun de mieux observer le niveau d'accord et d'attention de l'autre. Les auteurs s'attendent donc à ce que l'image du partenaire distant induise des dialogues plus fluides (tours de parole courts, moins d'interruptions et de questions). Pour vérifier ces hypothèses, ils construisent deux expérimentations.

Dans la première expérimentation, Daly-Jones & al./1 (1998), huit binômes (tous étudiants, dont 25% de sexe féminin) doivent analyser dix dossiers administratifs d'étudiants universitaires et en sélectionner trois, dans un temps limité à trente minutes. D'après la procédure expérimentale, les sujets savent qu'il s'agit d'une sélection fictive et que leur réponse n'a pas de conséquences réelles. Le plan expérimental (plan-intra à mesures répétées) prévoit que chaque sujet accomplisse la tâche selon les deux conditions expérimentales (effet d'ordre contrôlé) :

c1. « audio seul » : les deux partenaires ne peuvent communiquer que grâce à un signal audio transmis en continu ;

c2. « audio & vidéo du partenaire » : les deux partenaires se voient (visage et buste) grâce à une connexion vidéo câblée et ils communiquent par voie orale grâce à un lien audio transmis en continu.

Chaque sujet dispose d'un écran pour observer le partenaire distant (sur lequel est placée une caméra permettant la connexion vidéo) et d'un autre écran pour la consultation en commun et synchronisée des dossiers en format électronique, grâce à un logiciel de partage d'écran (Aspects®). Les deux écrans sont orientés selon un angle de 90°. Les données expérimentales sont :

a) les dialogues, analysés selon une approche conversationnelle (nombre de tours de parole, de mots, d'interruptions) et pragmatique (limité à l'acte de parole de type « question », e.g. « *tu es d'accord avec moi ?* », « *tu penses d'accepter ce formulaire* ? ») ;

b) le temps de réalisation de la tâche (le temps employé à trouver un accord sur les trois dossiers) ;

c) le sentiment relatif au degré de présence sociale (les sujets doivent répondre à des questions en échelle de 0 à 5 portant sur leur perception de l'état d'attention du partenaire distant).

Les effets sur les interactions communicatives. Le nombre et la longueur des tours de parole, les interruptions et les acquiescements ne varient pas de manière significative dans les deux conditions expérimentales. Par ailleurs, les questions sont moins nombreuses en c1 (« audio seul ») qu'en c2.

Les effets sur la réalisation de la tâche. Le temps nécessaire pour trouver un accord est équivalent dans les deux conditions expérimentales.

Les effets sur les impressions subjectives. Les sujets évaluent la perception de l'état attentionnel du partenaire distant meilleure en condition « audio & vidéo » plutôt qu'en condition « audio seul ».

D'après les résultats, l'image du partenaire distant n'influence pas le volume des échanges mais plutôt leur contenu (moins de questions). Ce résultat peut être interprété comme l'indice d'une amélioration du processus de compréhension réciproque. Cependant cette interprétation ne serait pas correcte : en fait, lors du codage, la catégorie « question » comprend à la fois les questions posées pour obtenir la confirmation d'une information (demander pour confirmer) et les questions dont la finalité est de connaître une information nouvelle (demander pour savoir). Par ailleurs, en condition « audio & vidéo », les sujets ont l'impression que leur partenaire est plus attentif. Bien que dans l'activité de négociation collaborative les indices non-verbaux soient importants pour la coordination du processus de communication entre plusieurs participants (Goodwin, 1981), les résultats de l'expérimentation de Daly-Jones & al./1 (1998) montrent que la condition « audio seul », véhiculant les seuls indices verbaux, est suffisante pour permettre la compréhension réciproque entre les partenaires distants.

Dans la deuxième expérimentation, Daly-Jones & al./2 (1998) étudient l'impact de l'image du partenaire distant dans des groupes de travail. Ils proposent ainsi la même tâche (sélection de trois dossiers pour l'attribution de bourses d'études), les mêmes conditions

expérimentales (« audio seul » *vs*. « audio & vidéo du partenaire distant ») et les mêmes données expérimentales (les dialogues, le temps pour trouver un accord et les impressions subjectives des participants). Le seul changement concerne l'organisation des participants. En effet, trente-deux sujets (dont 60 % de sexe masculin) sont organisés en huit quatuors (chacun composé de sujets du même sexe), qui répètent la tâche dans les deux conditions expérimentales (effet d'ordre contrôlé).

Les effets sur les interactions communicatives. Les tours de parole sont plus nombreux en condition « audio & vidéo du partenaire » qu'en condition « audio seul », alors qu'ils sont plus longs en condition « audio seul » qu'en condition « audio & vidéo ». Les interruptions et les acquiescements sont plus nombreux en condition « audio & vidéo» qu'en condition « audio seul ». Le nombre des questions est équivalent dans les deux conditions expérimentales.

Les effets sur la réalisation de la tâche. Le temps employé pour trouver un accord sur les trois dossiers est équivalent dans les deux conditions expérimentales.

Les effets sur les impressions subjectives. Les sujets estiment plus positivement la présence de leur partenaire distant et ils le jugent plus attentif en condition « audio & vidéo » qu'en condition « audio seul ».

D'après les auteurs, la condition « audio & vidéo du partenaire distant » semble induire un niveau d'interactivité conversationnelle équivalent à la condition de communication « face à face » (tours de parole courts et nombreux, acquiescements nombreux). En ce qui concerne les impressions subjectives des participants, comme pour la première expérimentation, l'image-vidéo du partenaire distant augmente l'intensité du sentiment de présence sociale et de l'état attentionnel du partenaire.

Synthèse

Nous avons présenté les études portant sur l'effet de l'image-vidéo du partenaire distant sur des activités de communication différentes : la conception collaborative (Isaacs & Tang, 1994), la reproduction d'une configuration spatiale (Doherty-Sneddon & al./1**,** 1997), la planification d'un voyage (Anderson & al./1**,** 1996) et la prise de décision collaborative (Daly-Jones & al./1**,** 1998).

Quel est l'impact de l'image-vidéo du partenaire distant sur les interactions communicatives ?

Globalement, les résultats de ces études indiquent que l'image-vidéo du partenaire distant favorise la compréhension réciproque et la gestion du dialogue (Clark, 1996). Plus spécifiquement, des contradictions apparaissent sur le plan du processus de communication, concernant notamment (Tableau 2-1) :

i) le nombre des tours de parole : par exemple, Daly-Jones & al./2 (1998) et Isaacs & Tang (1994) constatent que les tours de parole sont plus nombreux lorsque les partenaires distants sont visibles plutôt que quand ils ne le sont pas, alors que Doherty-Sneddon & al./1 (1997) montrent le contraire ;

ii) le nombre des interruptions : pour Anderson & al./1, (1996), Doherty-Sneddon & al./1 (1997) et Isaacs & al. (1994), les interruptions sont moins nombreuses en condition « audio & vidéo du partenaire » qu'en condition « audio seul », alors que les résultats de Daly-Jones & al./2 (1998) indiquent le contraire.

Sur le plan du contenu, les partenaires produisent un nombre inférieur de questions pour vérifier la compréhension d'autrui en condition « audio & vidéo » (cf. Doherty-Sneddon & al./1, 1997 ; Isaacs & Tang, 1994) : il semble donc que les regards échangés entre les partenaires se substituent aux indices verbaux consacrés à la vérification. Donc, l'image-vidéo du partenaire distant semble améliorer à la fois le processus et la compréhension réciproque. Cette interprétation doit cependant être nuancée à cause de profondes différences méthodologiques de ces études : le type de tâche expérimentale, les méthodes de codage (codage réalisé par un seul juge *vs*. codage à plusieurs) et le nombre des sujets distants (quintet & quatuor *vs*. binôme).

L'image-vidéo du partenaire distant améliore-t-elle la performance de la tâche ?

Les résultats des études présentées suggèrent deux considérations :

a) globalement, même si l'image-vidéo du partenaire distant est disponible les sujets n'améliorent pas la réalisation de la tâche, confirmant ainsi de nombreuses recherches (Ochsman & Chapanis, 1974 ; Williams, 1977) ;

b) même pour les tâches à fort enjeu personnel (e.g. négociation dans la prise de décision, cf. Daly-Jones & al. ,1998), l'image-vidéo du partenaire n'induit aucun effet.

L'image du partenaire distant influence-t-elle les impressions des participants ?

Seulement deux études se penchent sur les impressions subjectives (Tableau 2-1) et en plus selon des dimensions d'investigation différentes : d'après une étude, les sujets estiment pouvoir communiquer aussi bien sans qu'avec l'image-vidéo du partenaire distant (Anderson & al./1**,** 1996), alors que d'après l'autre les sujets apprécient davantage l'interaction avec leur partenaire en condition « audio & vidéo » qu'en « audio seul » (Daly-Jones & al., 1998).

Pour répondre à notre question de recherche initiale, « *Est-ce qu'il est préférable de voir le partenaire distant pour mieux coopérer ?* », ces résultats nous suggèrent que l'image-vidéo du partenaire facilite la coordination entre les locuteurs, leur permet de connaître l'état d'attention d'autrui, mais n'améliore pas la réalisation de la tâche.

De nombreuses questions restent encore ouvertes. Par exemple, l'image du partenaire se compose de différents éléments visuels, comme le corps et sa gestualité, le visage et ses expressions, le regard et sa direction. Chacun de ses éléments joue un rôle lorsque les partenaires interagissent en « face à face » (Argyle & Graham, 1977 ; Hall, 1971). Est-ce que ces éléments sont également pertinents lorsque les sujets communiquent à distance ? Si oui, comment ? Nous abordons ces questions dans la section suivante.

	Tâche	**Conditions**	**Nombre de sujets**	**Interactions communicatives**		**Impressions subjectives**
				Effets sur le processus	**Effets sur le contenu**	
Isaac & Tang (1994)	Conception collaborative	c1) audio seul c2) audio & vidéo du partenaire distant c3) face à face	3 quintéts	Les tours de parole sont plus courts et plus nombreux, les interruptions sont moins nombreuses en c2 que c1	Les actes de parole destinés à la vérification et à la confirmation de la compréhension sont moins nombreux en c2 qu'en c1	---
Doherty-Sneddon & al./1 (1997)	Reproduction d'une configuration spatiale (*map task*)	c1) audio seul c2) audio & vidéo du partenaire distant	64 binômes	Les tours de parole, les mots, les interruptions et les marqueurs phatiques d'attention sont moins nombreux en c2	Les actes de parole destinés à la vérification et à la confirmation de compréhension sont moins nombreux en c2	---
Anderson & al./1 (1996)	Planification d'un itinéraire	c1) audio seul c2) audio & vidéo du partenaire distant c3) face à face	30 binômes	Les interruptions sont moins nombreuses en c2 qu'en c1	---	Les sujets préfèrent communiquer en c3 qu'en c1 ou en c2
Daly-Jones & al1/. (1998)	Prise de décision collaborative	c1) audio seul c2) audio & vidéo du partenaire distant	8 binômes	Aucune différence n'est constatée.	Les questions sont moins nombreuses en c2	Les sujets préfèrent communiquer avec leur partenaire et évaluent mieux son état d'attention en c2 qu'en c1
Daly-Jones & al./2 (1998)	Prise de décision collaborative	c1) audio seul c2) audio & vidéo du partenaire distant	8 quatuors	Les interruptions et les tours de parole sont plus nombreux en c2 ; les tours de parole sont plus longs en c1	Le nombre des questions est équivalent en c1 et c2 ; les acquiescements sont plus nombreux en c2	

Tableau 2-1 : Les effets de l'image-vidéo du partenaire distant sur les interactions communicatives et les impressions subjectives.

2.2 Peut-on ne voir que le visage du partenaire distant pour bien coopérer ?

Lors de l'interaction avec autrui, les indices non-verbaux relatifs au corps du partenaire (i.e., les gestes, la posture du corps, les mimiques, etc.) sont des ressources d'information importantes pour la gestion du dialogue et pour l'intercompréhension entre les locuteurs (Kendon, 1967 ; McNeill, 1992). Par exemple, l'indice non-verbal de type kinésique (§ ch.1) comme le paume de la main du locuteur, ouverte sur l'axe horizontal et adressée à son allocutaire en fin de son tour de parole, invite l'allocutaire à prendre la parole à son tour. Également, le doigt du locuteur indiquant un lieu ou un objet incite l'allocutaire à regarder dans une direction précise, relative à une même référence extralinguistique. Ces indices non-verbaux sont visibles en condition de communication « face à face », alors qu'en situation « audio & vidéo médiatisé » les caractéristiques du dispositif de communication peuvent en limiter la visibilité (Bruce, 1996). Ainsi, l'image-vidéo du partenaire distant peut montrer :

a) *le visage seul* : la caméra est dirigée vers le visage seul, limitant la vue aux expressions du visage, aux mouvements faciaux et oculaires du locuteur ;
b) *la visage et buste* : la prise de la caméra est limitée au visage et au buste, révélant seulement le comportement facial, gestuel et postural du locuteur.

Dans cette section, nous présentons des études expérimentales relatives à ces deux cas de visibilité de l'image-vidéo du partenaire distant.

Dans l'étude de O'Malley & al./2 (1996), douze binômes sont recrutés pour accomplir la tâche consistant à reproduire un parcours sur une carte géographique (*map task*, Boyle & al., 1994). Les sujets distants travaillent en même temps sur une version de la carte géographique en format électronique, grâce à une application informatique permettant l'édition graphique partagée. Chaque binôme accomplit la tâche dans deux conditions expérimentales (effet d'ordre contrôlé):

c1. « audio seul », où le binôme peut communiquer oralement par un signal audio transmis en continu ;
c2. *soit* en : a) condition « audio & vidéo du visage du partenaire », où chaque partenaire communique par un signal audio transmis en continu et ne voit que le visage de l'autre, *soit* en b) condition « audio & video

du visage et buste du partenaire », où la partie supérieure du corps est visible (visage et buste).

Pour chaque condition, chaque sujet du binôme effectue la tâche une fois en tant que sujet-instructeur, une fois en tant que sujet-guidé (le rôle est un facteur contrôlé aléatoirement). Dans les deux conditions où le partenaire distant est visible, les sujets disposent de deux écrans : un écran est dédié à la visualisation de la carte géographique en version électronique, l'autre écran à l'observation du partenaire distant. Une vidéo-caméra et un microphone ouvert sont placés entre les deux écrans, afin d'assurer la connexion audio-vidéo. Les données récoltées sont de deux types : les dialogues (transcrits et codés selon l'approche conversationnelle) et le score de réalisation de la tâche (calculé par différence géométrique entre le parcours dessiné par le sujet-guidé et le parcours original sur la carte du sujet-instructeur).

Les effets sur les interactions communicatives. Les tours de parole et les mots sont significativement plus nombreux lorsque les deux partenaires se voient que lorsqu'ils ne se voient pas. Quand ils se voient, les dialogues sont plus longs en condition « audio & vidéo du visage » qu'en condition « audio & vidéo du visage-buste ». Le nombre des interruptions est équivalent dans les différentes conditions expérimentales.

Les effets sur la réalisation de la tâche. Aucune différence significative n'est constatée pour le score et le temps de réalisation de la tâche.

Sur le plan théorique, ces résultats confirment que lorsqu'on voit le partenaire distant les tours de parole et les mots produits sont plus nombreux (cf. O'Conaill & al., 1993 ; Matarazzo & Sellen, 2000). Une interprétation possible est que les indices visuels relatifs au comportement expressif du visage du partenaire distant peuvent distraire les locuteurs lors de l'interaction.

Sur le plan applicatif, il apparaît que si on veut réduire la longueur des dialogues et améliorer ainsi l'efficacité de l'interaction, il vaut mieux alors montrer le buste et le visage de chaque locuteur plutôt que le visage seul.

Anderson & al./2 (1996) veulent vérifier les résultats de l'expérimentation de O'Malley & al./2, (1996). Ils supposent que lorsqu'un dispositif de communication permet de voir les mouvements faciaux et gestuels du partenaire distant, la réalisation des activités collaboratives est favorisée. La tâche est la suivante : un agent de voyage doit aider un client à organiser un voyage en avion, afin de visiter le nombre plus élevé de villes (cf., Anderson &

al./1**,** 1996). Pour réaliser cette tâche, les sujets disposent d'un temps fixé (15 minutes), d'un budget limité et ils doivent respecter les correspondances des vols des compagnies aériennes. L'agent de voyage doit renseigner le client sur les horaires et les prix des vols. Trente-six sujets volontaires et trois « agents de voyage » (trois chercheurs associés au projet) sont recrutés et repartis de manière aléatoire dans les trois conditions expérimentales (plan expérimental intersujet) :

c1. « audio seul » : l'agent et le client ne peuvent communiquer que par voie orale grâce à un signal audio transmis en continu ;

c2. « audio & vidéo du visage » : chaque membre du binôme peut communiquer oralement avec son partenaire et il ne peut voir que le visage de son partenaire;

c3. « audio & vidéo du visage-buste » : chaque membre du binôme peut communiquer oralement avec son partenaire et voir son visage et son buste.

Un canal audio transmis en continu assure la communication orale entre le sujet-client et l'agent de voyage dans chaque condition. Les données récoltées concernent les dialogues (transcrits et analysés selon leur structure et leur contenu), la qualité de réalisation de la tâche (évaluée selon le nombre des villes insérées dans le voyage) et les impressions subjectives des participants sur la facilité de communication avec l'agent de voyage et sur la facilité de modifier l'itinéraire (questions en échelle de 0 à 5).

Les effets sur les interactions communicatives. D'après l'analyse de la structure des dialogues, aucune différence significative n'est constatée entre les trois conditions expérimentales pour le nombre des tours de parole, des mots et des interruptions. L'analyse du contenu révèle que le nombre des solutions de voyage proposées est équivalent dans les trois conditions.

Les effets sur la réalisation de la tâche. Le score ne diffère pas de manière significative entre les différentes conditions.

Les effets sur les impressions subjectives. Les sujets sont invités à juger la facilité de communication avec l'agent de voyage et la facilité d'apporter des changements au plan de voyage pendant la session expérimentale. Aucune différence significative n'est constatée entre les trois conditions expérimentales.

D'après les résultats, il en résulte que la tâche de planification d'itinéraires n'est influencée ni par la visibilité du partenaire distant, ni par les types d'indices non-verbaux (visage seul ou visage-buste). Plus précisément, ces résultats suggèrent que :

a) une activité complexe comme celle de la planification d'itinéraires, intégrant la négociation et la prise de décision, peut être effectuée de manière efficace sans voir le partenaire distant (*vs*. la théorie de la présence sociale) ;
b) sur le plan applicatif, lorsqu'il n'est pas possible d'assurer une liaison vidéo, une liaison audio en continu peut se révéler suffisante pour accomplir la tâche.

Dans la même direction de recherche, Grayson & Coventry (1998) comparent les effets des indices non-verbaux relatifs aux seules expressions faciales *vs*. les indices relatifs aux expressions faciales et gestuelles. La tâche expérimentale est la suivante : un client d'une banque doit choisir comment investir de l'argent avec l'aide d'un conseiller financier, qui doit à son tour convaincre le client d'accepter l'offre la plus avantageuse pour la banque. Pour l'expérimentation, trente étudiants sont recrutés pour jouer les rôles du client (15 hommes et 15 femmes, étudiants universitaires) et un étudiant pour jouer le rôle du conseiller financier (ce dernier reçoit des consignes précises sur l'offre à proposer au client). Les sujets sont répartis de manière aléatoire dans deux conditions expérimentales :

c1. « audio & vidéo du visage », où le client et le conseiller ne voient que le visage de leur partenaire ;
c2. « audio & vidéo du visage-buste », où les sujets voient le visage et le buste de leur partenaire.

Dans chaque condition, le client et le conseiller financier sont placés dans deux pièces différentes. Ils communiquent grâce à une liaison audio transmise en continu et ils se voient grâce à une caméra placée à côté de l'écran (angle non précisé). Les mesures relevées concernent les dialogues (transcrits et analysés selon la structure conversationnelle).

Les effets sur les interactions communicatives. Les mots, les tours de parole et les interruptions sont significativement plus nombreux en condition « audio & vidéo du visage » qu'en condition « audio & vidéo du visage-buste ». En revanche, les marqueurs phatiques d'attention sont équivalents dans les deux conditions.

Suivant la littérature, les auteurs interprètent ces résultats selon le critère d'interactivité : les mots, les interruptions et les tours de parole étant plus nombreux, la condition « audio & vidéo du visage » permet aux locuteurs de construire un processus de communication plus interactif que la condition « audio & vidéo du visage-buste ». Conséquemment, s'appuyant sur les travaux de Schober & Clark (1989), ils considèrent que plus l'interactivité est importante, plus la compréhension des locuteurs est profonde. Nous

retenons qu'il aurait été préférable d'appuyer cette argumentation par un codage des actes de parole (notamment les actes dédiés à la confirmation et à la vérification du partenaire) et par un questionnaire post-test pour vérifier le niveau de compréhension des offres présentées par le « chargé de la clientèle ».

Synthèse

Dans cette section, nous avons présenté trois études expérimentales qui comparent l'impact de l'image-vidéo relative au visage du partenaire distant et de l'image-vidéo relative au visage-buste dans des activités collaboratives différentes (Tableau 2-2) : la reproduction d'une configuration spatiale (O'Malley & al./2, 1996), la planification d'un itinéraire de voyage (Anderson & al./2, 1996) et le choix d'un investissement financier (Grayson & Coventry, 1998). Les activités collaboratives proposées sont plutôt de type procédural, ce qui réduit la portée de ces résultats.

Est-ce que l'image-vidéo du visage seul et l'image-vidéo du visage-buste influencent les interactions communicatives ? Et si oui, quelle est la plus efficace ?

Globalement, les dialogues contiennent plus de mots, de tours de parole et d' interruptions lorsque les partenaires distants ne voient l'image-vidéo que du visage seul plutôt que du visage & buste (cf. Grayson & Coventry, 1998 ; O'Malley & al./2, 1996). Ce résultat, infirmant sur le plan théorique la théorie de la présence sociale, doit cependant être relativisé à cause d'une différence méthodologique entre les études. Notamment, le codage des tours de parole n'est pas effectué selon les mêmes règles : à la différence de O'Malley & al./2 (1996), pour Grayson & Coventry (1998) chaque marqueur phatique d'attention est un tour de parole.

Quels sont les effets de l'image-vidéo du visage et du visage-buste sur la réalisation de la tâche ?

D'après les résultats, l'image-vidéo relative au visage seul ou au visage-buste du partenaire n'entrave ni n'améliore la qualité de réalisation de la tâche (Anderson & al./2, 1996 ; O'Malley & al./2, 1996). De plus, ces résultats confirment les études présentées dans la section précédente (§ 2.1), c'est-à-dire que l'image du partenaire distant n'influence pas la performance de réalisation de la tâche. Considérant le résultat relatif aux interactions communicatives, à égalité de score de réalisation, il en résulte que les dialogues sont moins longs lorsque les locuteurs peuvent voir le visage et le buste de leur partenaire distant plutôt que le visage seul. Donc, sur le plan applicatif, si on veut privilégier l'efficacité d'une

situation de communication à distance, il est préférable de montrer le buste et le visage du partenaire distant plutôt que le visage seul.

Quels sont les effets sur les impressions subjectives ?

Des trois études présentées, seuls Anderson & al./2 (1996) analysent les impressions personnelles des utilisateurs. D'après leurs résultats, les participants apprécient de manière équivalente la qualité de l'interaction quelle que soit la condition de communication.

En résumé, indépendamment du type d'activité coopérative, si on veut favoriser l'interactivité du processus de communication, il est préférable que chaque locuteur ne voie que le visage du partenaire distant. En revanche, si on veut privilégier l'efficacité, alors il est préférable que le visage et le buste soient visibles aux partenaires distants.

	Tâche	**Conditions**	**Sujets**	**Interactions communicatives**	**Réalisation de la tâche**	**Impressions subjectives**
O'Malley & al./2 (1996)	Reproduction d'une configuration spatiale (*map task*)	c1) audio seul c2 a) audio & vidéo du visage *ou* b) audio & video du visage-buste	12 binômes	Les mots et les tours de parole sont plus nombreux en c2a qu'en c2b et qu'en c1.	Aucune différence significative	---
Anderson & al2/. (1996)	Planification d'un itinéraire	c1) audio seul c2) audio & vidéo du visage c3) audio & video du visage-buste	36 binômes	Aucune différence significative n'est constatée	Aucune différence significative	Aucune différence significative
Grayson & Coventry (1998)	Investissement d'une somme d'argent	c1) audio & vidéo du visage c2) audio & video du visage-buste	30 binômes	Les mots, les interruptions et les tours de parole plus nombreux en « audio & vidéo du visage ».	---	---

Tableau 2-2 : Les effets de l'image-vidéo relative au visage *vs.* au visage & buste.

2.3 Faut-il se regarder dans les yeux pour se comprendre ?

Le regard mutuel et la connaissance de la direction du regard (ou visée)[24] du partenaire sont deux indices non-verbaux fondamentaux pour l'intercompréhension (Argyle & al., 1968).

Le regard mutuel (ou réciproque) fournit un retour d'information continu et immédiat entre nous et notre partenaire : notre interlocuteur sait qu'on le regarde, nous-mêmes nous savons qu'il nous regarde et notre interlocuteur sait que nous savons qu'il nous regarde. Cela permet la régulation de la prise de parole entre les interlocuteurs tout au long de leur interaction (Goodwin, 1981 ; Kendon, 1967), car les individus se regardent dans les yeux jusqu'à 50% du temps de dialogue en situation de face à face. Cette valeur varie en fonction du type d'activité collaborative (Vertegaal, Slagter, Van der Veer & Nijholt, 2001) et de la nature des indices non-verbaux à disposition des partenaires distants. Par exemple, Watts, Monk & Daly-Jones (1996) montrent que le temps dédié aux regards réciproques baisse d'environ de la moitié si les partenaires peuvent observer aussi des éléments appartenant à l'environnement de travail de leur partenaire (§ ch.3).

L'autre indice non-verbal, la connaissance de la visée d'autrui, permet à des interlocuteurs de construire facilement leurs référentiels communs pendant la réalisation d'une activité collaborative (Clark, 1996). En fait, nous pouvons savoir à peu près la direction vers laquelle notre interlocuteur oriente son regard (visée partielle) ou bien connaître le référent spécifique observé par notre partenaire (visée absolue) (Monk & Gale, 2002).

Malgré leur importance pour le processus de construction des référentiels communs, ces deux indices ne sont pas encore suffisamment supportés par les dispositifs actuels de visioconférence (Figure 2-1). En effet, souvent la direction de la vue de la caméra ne coïncide pas exactement avec la ligne du regard du partenaire distant, créant une sorte de « regard latéral », la parallaxe, soit verticale soit horizontale[25].

[24] Cette distinction est analogue à la proposition de Stefik, Foster, Bobreow, Kahn, Lanning & Suchman (1987) pour les collecticiels entre la « congruence de vue stricte » (What You See is What I See) et la « congruence de vue élargie » (What You See is Almost What I See).

[25] Sur le problème de la parallaxe visuelle, Ostberg, Lindstrom & Renhall (1989) précisent que "*... a camera positioned at the periphery of the display screen introduces a significant parallax - seeing into the (virtual) eyes of your interlocutor does not mean that he/she will be seeing into your (virtual) eyes. Direct eye contact is lost, and the resulting eye-gaze may actually be a source of misinformation*".

En ce qui concerne la connaissance de la visée d'autrui, la plupart des dispositifs permettent de ne voir que la partie supérieure du corps (visage et buste) du partenaire distant, ainsi le champ visuel n'est pas assez large pour montrer ce que le partenaire distant observe.

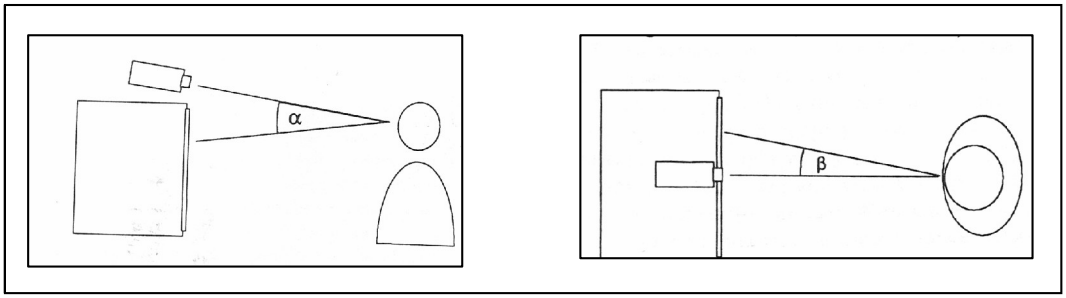


Figure 2-1 : Le regard latéral vertical (à gauche) et le regard latéral horizontal (à droite) (adapté de Mühlbach, Bocker & Prussog, 1995).

Pour pallier ces difficultés, plusieurs solutions ingénieuses ont été proposées. Par exemple, le système « Video-Tunnel » avec le miroir démi-argenté (Buxton & Moran, 1990 ; Smith, O'Shea, O'Malley, Scanlon & Taylor, 1991), le système à vision stéréoscopique (Mühlbach, Bocker & Prussog, 1995) ou encore le système permettant la connaissance partielle du regard d'autrui (Velichkovsky, 1995). Il s'agit en tout cas des dispositifs complexes, coûteux, non encore fiables et qui ne satisfont pas les utilisateurs (Mühlbach & al., 1995), mais qui néanmoins peuvent améliorer la communication implicite entre les interlocuteurs et augmenter leur engagement dans la réalisation de la tâche (Smith & al., 1991 ; Joiner, Scanlon, O'Shea & Blake, 2002).

Dans cette section, nous présentons des études expérimentales relatives à l'impact du regard réciproque (§ 2.3.1) et de la connaissance de la visée d'autrui (§ 2.3.2) sur les activités collaboratives à distance.

2.3.1 Le regard réciproque

L'hypothèse de recherche de Sellen (1995) est que le regard mutuel favorise la régulation de l'interaction conversationnelle (Rutter & Stephenson, 1977). Quarante-huit sujets sont recrutés, organisés en douze quatuors afin de débattre sur un thème d'intérêt public : ils doivent argumenter leur point de vue et convaincre les autres participants de l'accepter. Chaque groupe doit accomplir cette tâche selon trois conditions expérimentales :

c1) « face à face » : les membres du groupe sont réunis autour d'une table ;

c2) « audio & vidéo du visage-buste sans regard mutuel » : les participants sont placés dans des pièces différentes, où chacun dispose d'un seul écran pour voir

les autres sujets en même temps (leur buste et leur visage) et d'une connexion audio ouverte en continu pour leur parler ;

c3) « audio & vidéo du visage-buste avec regard mutuel » : chaque participant dispose d'une liaison audio pour parler avec un seul interlocuteur à la fois et d'un dispositif vidéo pour le voir et établir un regard réciproque avec lui (une caméra est placée sur chaque écran, cas de la parallaxe verticale).

Chaque session expérimentale dure environ 15 minutes, le temps de discussion étant limité à 5 minutes par session. L'auteur s'attend à ce que dans les deux conditions de communication permettant le regard mutuel (en « face à face » et à distance), les dialogues soient plus interactifs, c'est-à-dire que les tours de parole soient plus nombreux et plus courts, et que les sujets soient davantage satisfaits de la qualité de la condition de communication. Les dialogues sont transcrits et analysés et les sujets sont interviewés après chaque session.

Les effets sur les interactions communicatives. En ce qui concerne la longueur et le nombre de tours de parole, aucune différence significative n'est constatée entre les trois conditions expérimentales. En revanche, les superpositions en début de phrase sont plus nombreuses en « face à face » que dans les deux conditions à distance.

Les effets sur les impressions subjectives. Les participants préfèrent communiquer en condition « face à face » ou « avec regard mutuel » plutôt qu'en condition « sans regard mutuel ».

Contrairement aux attentes, la condition « avec regard mutuel » ne favorise pas l'interactivité du dialogue par rapport à la condition « sans regard mutuel ». Sellen (1995) justifie ce résultat par les limites techniques du dispositif. D'après l'auteur, dans la situation «avec regard mutuel » l'angle de vue de la caméra ne permettrait pas aux interlocuteurs de bénéficier au mieux du regard réciproque. Cependant, si on admet cette explication, on aurait dû s'attendre à ce que les superpositions en début de phrase soient plus nombreuses en condition « avec regard mutuel » plutôt qu'en « face à face ». En effet, comme la condition « face à face » permet aux locuteurs de se regarder de manière réciproque, le nombre des superpositions en début de phrase est généralement réduit. Or, les résultats contredisent cette explication. Il semble donc que le regard mutuel n'améliore ni ne diminue le niveau d'interactivité entre les locuteurs en situation de communication à distance à plusieurs.

Selon la même ligne de recherche, O'Malley & al./1 (1996) supposent que le regard mutuel est nécessaire pour assurer un haut niveau d'interactivité entre les locuteurs distants.

Pour vérifier leur hypothèse, plusieurs binômes[26] sont recrutés pour reproduire un parcours sur une carte géographique (*map task*) et ils sont répartis de manière aléatoire dans les trois conditions expérimentales :

c1. « audio seul » : le binôme communique par voie orale ;

c2. « audio & vidéo du visage-buste sans regard mutuel » : les membres du binôme peuvent se voir l'un l'autre (leur buste et leur visage) et communiquer par voie orale, grâce à une liaison audio ouverte en continu ;

c3. « audio & vidéo du visage-buste avec regard mutuel » : les participants communiquent par voie orale (liaison ouverte en continu), observent la partie supérieure du corps de leur partenaire distant et établissent un regard réciproque, grâce à une version modifiée du système Video-Tunnel de Smith & al. (1991).

Les auteurs s'attendent à ce que les dialogues soient plus fluides et à ce que le score de réalisation de la tâche soit plus élevé dans la condition expérimentale « avec regard mutuel ». Les données recueillies se composent des transcriptions des dialogues, du temps employé et du score de réalisation de la tâche (obtenu par différence géométrique entre le parcours reproduit par le sujet guidé et le parcours tracé sur la carte originale du sujet-instructeur).

Les effets sur les interactions communicatives. Le nombre de mots, de tours de parole et d'interruptions est significativement plus important en condition « avec regard mutuel » que dans les deux autres conditions.

Les effets sur la réalisation de la tâche. Le temps et le score de réalisation de la tâche sont équivalents dans les trois conditions expérimentales.

Les auteurs suggèrent deux interprétations :

a) le critère de l'efficacité communicationnelle : à égalité de score de performance, la condition « avec regard mutuel » est la moins efficace des trois, à cause du nombre plus important de mots et de tours de parole ;

b) le niveau de familiarité et de confiance des sujets avec le dispositif (cf. Anderson & al., 1996). D'après les auteurs, les sujets seraient plus prudents en condition de

[26] Comme pour O'Conaill & al. (1993), le nombre des sujets participants à l'expérimentation n'est pas précisé.

communication à distance et le seraient moins en « face à face ». En d'autres termes, ils adopteraient un comportement « à faible risque communicationnel » en situation de communication à distance : cette conduite se manifeste par des énoncés où les références sont explicitées (i.e. peu d'anaphore et de déictiques[27]). En revanche, en condition « face à face », ils assumeraient un comportement à « haut risque communicationnel », qui se manifeste par un usage élevé d'implicites.

Cela expliquerait comment le regard mutuel augmente l'interactivité du dialogue, sans pour autant améliorer la performance de réalisation de la tâche. Néanmoins, afin de rendre plausible cette interprétation, les auteurs auraient dû analyser les niveaux de construction du référentiel commun utilisé par les sujets. Par exemple, ils auraient pu envisager une analyse du niveau de langage implicite (usage des déictiques et des anaphores) et explicite (les descriptions et les occurrences du nom du référent).

Doherty-Sneddon & al./2 (1997) supposent que le regard réciproque entre les locuteurs distants améliore leur processus de communication lors d'une activité collaborative consistant à reproduire un parcours sur une carte (*map task*). Trente-six binômes sont recrutés et repartis dans les trois conditions expérimentales :

c1. « audio seul » : le sujet-instructeur et le sujet-guidé ne communiquent que par voie orale ;

c2. « audio & vidéo du visage-buste sans regard mutuel » : chacun peut voir le buste supérieur et le visage du partenaire distant, mais sans établir le contact des yeux ;

c3. « audio & vidéo du visage-buste avec regard mutuel » : les partenaires peuvent se parler, se voir et établir un contact des yeux grâce au dispositif « Videotunnel » (Smith & al., 1991).

Les auteurs s'attendent à ce que les dialogues soient plus interactifs (i.e. avec tours de parole plus nombreux, plus courts et avec moins d'interruptions) et plus fluides (i.e. contenant un nombre inférieur d'acquiescements) lorsque les partenaires distants peuvent se voir et *a fortiori* lorsqu'ils peuvent disposer du regard mutuel. Chaque sujet doit réaliser la tâche deux

[27] Un déictique est « une expression dont le référent ne peut être déterminé que par rapport à la situation de communication » (du glossaire DALI, groupe d'étude sur le dialogue).

fois : une fois en tant que sujet-instructeur et une fois en tant que sujet-guidé (effet d'ordre contrôlé). Contrairement à la procédure expérimentale de Boyle et al. (1994), les participants doivent accomplir la tâche par ordinateur. Pour faire cela, chaque sujet dispose de deux écrans juxtaposés : l'un, relié à un ordinateur, visualisant la carte en version électronique, et l'autre destiné à afficher l'image du partenaire distant. Pour chaque condition, les signaux vidéo et audio (ce dernier ouvert en continu) sont disponibles, respectivement, grâce à une caméra placée entre les deux écrans (ce qui n'empêche pas la parallaxe horizontale) et à un microphone. Les données recueillies comprennent les transcriptions des dialogues, le temps et le score de réalisation de la tâche (obtenu par différence géométrique entre le parcours reproduit par le sujet-guidé et le parcours original tracé sur la version électronique du sujet-instructeur).

Les effets sur les interactions communicatives. Sur le plan du processus de communication, les hypothèses ne sont pas validées : les tours de parole sont plus longs et les interruptions plus nombreuses en condition « audio & vidéo du visage-buste avec regard mutuel » qu'en condition « audio seul ». Les mots et les tours de parole ne diffèrent pas de manière significative dans les conditions « audio & vidéo du visage-buste sans regard mutuel» et « audio seul ». En revanche, leurs hypothèses sont confirmées en ce qui concerne les acquiescements : leur nombre est plus important lorsque les partenaires distants ne se voient pas plutôt que lorsqu'ils se voient, mais aucune différence significative n'est constatée avec la condition « audio & vidéo du visage-buste avec regard mutuel ». En ce qui concerne le contenu des interactions, l'analyse en jeux conversationnels montre que les actes de parole destinés à vérifier la compréhension du partenaire sont plus nombreux en condition « audio seul » que dans les deux conditions « audio & vidéo du visage-buste ». Enfin, les auteurs décomptent les regards échangés entre les interlocuteurs : il s'avère que leur nombre est plus élevé (plus de 56%) en condition « avec regard mutuel » qu'en condition « sans regard mutuel ».

Les effets sur la réalisation de la tâche. Le temps et le score de réalisation de la tâche ne diffèrent pas significativement entre les trois conditions expérimentales.

Ces résultats montrent que le regard réciproque ne favorise ni l'interactivité ni la fluidité des dialogues entre les individus distants. Les partenaires se parlent et s'interrompent l'un l'autre davantage lorsqu'ils peuvent établir un regard mutuel que lorsqu'ils ne le peuvent pas. Considérant aussi le nombre important de regards échangés en condition « avec regard mutuel », les auteurs interprètent ces résultats comme un comportement excessif (en ang.,

overgazing) dû à l'« effet de nouveauté » du dispositif de communication. D'autre part, les sujets ont gardé la parole plus longtemps en condition « avec regard mutuel » qu'en condition « sans regard mutuel ». Il faut cependant tenir compte du fait que la condition « avec regard mutuel » est caractérisée par le phénomène de la parallaxe horizontale (cas b), la caméra étant placée entre les deux écrans.

Dans la section suivante nous abordons l'impact de la direction du regard d'autrui en situation de communication audio-vidéo médiatisée.

2.3.2 Connaître la direction du regard du partenaire distant

Monk & Gale (2002) comparent l'impact du regard mutuel et de la connaissance de la visée d'autrui sur une activité collaborative. Leur hypothèse est que la connaissance de la visée d'autrui favorise la construction des référentiels communs entre les locuteurs, facilitant l'identification de la référence de l'énoncé émis par le locuteur. En d'autres termes, ils supposent que si chacun connaît la direction du regard d'autrui, alors les actes de parole dédiés à la vérification et à la validation de la compréhension seront moins nombreux. La tâche expérimentale est la suivante : un sujet-instructeur doit décrire la position d'un couple de points A et B (dessinés sur une photo parmi d'autres) à un sujet-guidé, et ce dernier a la tâche d'identifier ces deux points sur une photo similaire mise à sa disposition. D'après la consigne expérimentale, la première réponse du sujet-guidé est considérée comme définitive. Pour chaque session expérimentale, cinq photos au contenu différent sont présentées (i.e. circuit électrique, plante d'une maison, image d'une molécule chimique, etc.). Chaque sujet utilise un écran d'ordinateur pour observer la photo (dispositif de type Clearboard 2 ; Ishii, Kobayashi & Grudin, 1993) et un écran pour regarder le partenaire distant.

Les vingt-quatre binômes sont répartis de manière aléatoire dans les trois conditions expérimentales :

c1. « audio seul » : les sujets communiquent uniquement par voie orale ;

c2. « audio & vidéo du visage-buste avec regard mutuel » : les sujets peuvent se regarder réciproquement dans les yeux, grâce au dispositif «VideoTunnel » (Smith & al., 1991) ;

c3. « audio & vidéo du visage-buste avec regard mutuel & direction de son regard» : les sujets peuvent se regarder réciproquement dans les yeux et percevoir la direction du regard du partenaire distant.

Dans les trois conditions expérimentales, une liaison audio (ouverte en continu) assure la communication orale entre les participants. Les dialogues sont transcrits et analysés selon la méthode des jeux conversationnels (Kotwko & al., 1991).

Les effets sur les interactions communicatives. Sur le plan du processus de communication, les mots et les tours de parole sont significativement moins nombreux (55%) en condition « avec regard mutuel & direction de son regard » qu'en condition « avec regard mutuel » et « audio seul ». En revanche, aucune différence significative n'est constatée entre ces deux dernières conditions expérimentales. Les interruptions sont du même ordre de grandeur dans les trois conditions expérimentales. Sur le plan du contenu, les actes de parole dédiés à la validation de la compréhension d'autrui et dédiés à la vérification de sa propre compréhension sont moins nombreux en condition « avec regard mutuel & direction de son regard » qu'en condition « avec regard mutuel » et « audio seul ».

Les effets sur la réalisation de la tâche. Les résultats montrent un taux global de réussite très élevé, sans aucune différence significative entre les trois conditions expérimentales.

D'après ces résultats, la condition qui assure le regard mutuel et la connaissance de la visée d'autrui est plus efficace que la condition « regard mutuel seul ». En effet, à score égal, les dialogues sont moins longs et contiennent moins d'actes de parole destinés à vérifier la compréhension du partenaire distant. Il faut noter que le score de réalisation de la tâche est très élevé (effet plafond) : les auteurs considèrent que la contrainte de la consigne expérimentale (une seule réponse valide) a influencé les sujets à prendre tout leur temps avant de répondre et ils ont ainsi fourni presque toujours la réponse correcte. Les auteurs suggèrent que pour une activité collaborative fortement structurée et exigeant une qualité élevée de réalisation, un dispositif de communication audio-vidéo doit permettre aux locuteurs distants de connaître la direction du regard d'autrui.

Vertegaal, van der Veer & Vons (2000) conduisent une étude expérimentale afin d'évaluer l'impact de la connaissance de la direction du regard d'autrui et du regard mutuel sur les interactions à distance. Selon la procédure expérimentale, des groupes de trois personnes doivent résoudre une série de casse-tête linguistiques dans un temps limité à 15 minutes. Chacun d'eux reçoit un morceau de la phrase et ils doivent coopérer pour trouver la combinaison correcte. Dans chaque trinôme, deux sujets sont compères (une femme et un homme) recrutés et entraînés pour jouer un script comportemental prédéfini par les

expérimentateurs, alors que le troisième sujet est un volontaire, ne connaissant pas le rôle des autres. Globalement cinquante-six sujets-volontaires son recrutés (tous étudiants, dont 60% hommes, âge moyen de 22 ans). Chaque sujet (Figure 2-2) est placé dans une pièce différente, où il dispose d'un écran d'ordinateur dédié à la visualisation de la phrase du casse-tête et de deux moniteurs affichant l'image des deux partenaires distants. Une caméra est placée sur l'écran du milieu (cas de la parallaxe verticale et horizontale).

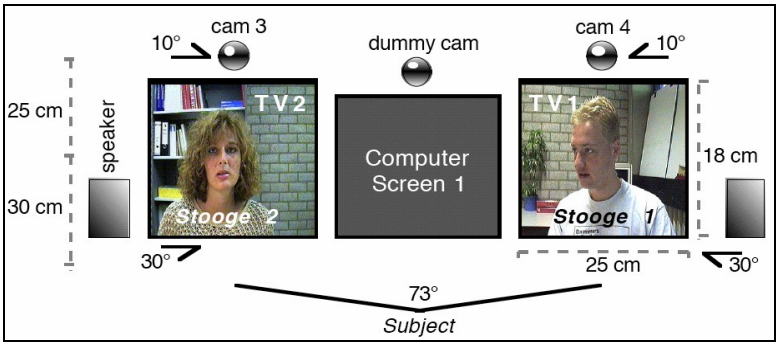


Figure 2-2 : L'écran du sujet-volontaire dans Vertegaal & al. (2000).

En revanche, à l'insu du sujet-volontaire, les deux compères se trouvent dans la même pièce, chacun disposant d'un écran d'ordinateur et d'un moniteur pour voir le sujet-volontaire.

Selon le plan expérimental à échantillons indépendants, les cinquante-six trinômes sont répartis de manière aléatoire dans les trois conditions expérimentales, où les deux compères se comportent suivant des consignes précises :

c1. « audio & vidéo du visage-buste avec regard mutuel » : chaque compère oriente son propre regard uniquement face à la caméra, son buste et son visage sont visibles au sujet-volontaire sur son écran ;

c2. « audio & vidéo du visage-buste avec regard mutuel & sa visée » : chaque compère est face à la camera et peut orienter son propre regard dans toutes les directions, permettant ainsi au sujet-volontaire de percevoir sa visée ainsi que les autres indices globaux ;

c3. « audio & arrêt-image du visage-buste & sa visée » : à tout instant, chaque compère sélectionne un arrêt-image de lui-même, qui s'affiche en temps réel à l'écran du sujet-volontaire. Le compère peut choisir parmi trois arrêt-images différents : le regard dirigé vers la caméra ou le regard dirigé vers l'autre compère ou bien le regard dirigé vers l'écran de son propre ordinateur.

Ainsi le sujet-volontaire ne peut ni voir les mimiques faciales ni établir un regard mutuel avec le compère.

Dans chaque condition, les sujets disposent d'une liaison audio ouverte en continu. Il faut remarquer que ces trois conditions sont asymétriques : le compère voit toujours le sujet-volontaire grâce à la caméra placée entre les deux écrans. Les données récoltées se composent des dialogues transcrits, du score de réalisation de la tâche calculé selon le nombre des solutions correctes et d'un questionnaire d'évaluation globale[28].

Les effets sur les interactions communicatives. Les tours de parole sont plus nombreux (environ 25%) en c3 (condition « audio & arrêt-image du visage-buste & sa visée ») qu'en c1 (« audio & vidéo du visage-buste avec regard mutuel ») et c2 (« audio & vidéo du visage-buste avec regard mutuel & sa visée »). Les formes déictiques destinées à identifier l'interlocuteur (i.e., « … *you* …») sont plus nombreuses en condition c2 qu'en condition c1. Pour compléter l'analyse, les auteurs décomptent les regards, dont le nombre est plus important en condition c3 qu'en condition c2. Les auteurs constatent aussi une corrélation positive entre le nombre de tours de parole et le nombre de regards échangés entre le sujet-volontaire et les deux acteurs.

Les effets sur la réalisation de la tâche. Le nombre de solutions correctes ne diffère pas de manière significative parmi les trois conditions.

Les effets sur les impressions subjectives. Les sujets affirment communiquer plus facilement avec leur interlocuteur en condition c3 que dans les autres conditions.

Il faut remarquer que les formes linguistiques destinées à identifier l'interlocuteur sont plus nombreuses en condition « audio & vidéo du visage-buste avec regard mutuel & sa visée». Le sujet ne nomme pas la personne, mais il la regarde en disant « you » : cela montre qu'un indice non-verbal (e.g. le regard, un geste déictique, etc.) réduit l'ambiguïté des expressions référentielles langagières. Si l'on interprète les mêmes résultats selon le critère d'efficacité, la condition « audio & vidéo du visage-buste avec regard mutuel » est plus efficace, car à égalité de score de réalisation de la tâche, le nombre de tours de parole est inférieur que dans les autres conditions. Cependant, les auteurs choisissent le critère d'interactivité du dialogue : ils suggèrent qu'un dispositif dédié à la communication à distance

[28] Les auteurs ne précisent ni les dimensions ni les items du questionnaire.

doit montrer la direction du regard du partenaire distant afin de rendre le processus de communication le plus interactif (plus de tours de parole).

Synthèse

Nous avons présenté cinq études expérimentales relatives à l'impact du regard mutuel et de la visée d'autrui sur des activités de communication à distance : l'argumentation et la conversation autour d'un thème d'intérêt générale (Sellen, 1995), la reproduction d'une configuration spatiale (Doherty-Sneddon & al./2, 1997 ; O'Malley & al./1, 1996), l'identification d'un couple de points sur une image complexe (Monk & Gale, 2002) et la résolution de casse-tête linguistiques (Vertegaal & al., 2000). A l'exception de l'étude réalisée par O'Malley & al./1 (1996), les dispositifs expérimentaux sont affectés par le problème de la parallaxe visuelle. Nous remarquons que les activités présentées dans cette section sont pour moitié de type procédural et pour moitié de type déclaratif.

Est-ce que les interactions sont plus interactives si on se voit dans les yeux ?

Nous discutons les résultats selon deux dimensions : le processus de communication et le contenu des interactions.

* *Processus de communication*. Il est bien difficile d'extraire des tendances générales, car les résultats sont sujets à des interprétations contradictoires, même pour une même activité de communication. Par exemple pour la reproduction d'une configuration spatiale (Doherty-Sneddon & al./2**,** 1997 ; O'Malley & al./1, 1996), il en résulte qu'avec le regard mutuel *soit* les tours de parole sont plus longs (Doherty-Sneddon & al./2, 1997) *soit* ils sont plus nombreux (O'Malley & al./1, 1996 ; Monk & Gale, 2000). Dans le premier cas, le regard mutuel n'assure pas l'interactivité du processus de communication, alors qu'il l'assure dans le second cas. En revanche, si on considère le critère d'efficacité, dans un cas (cf. O'Malley & al./1, 1996) le regard mutuel n'améliore pas l'efficacité de la condition de communication, alors qu'il l'améliore dans un autre (cf. Monk & Gale, 2000) et même pour des tâches où le langage n'est pas utilisé pour décrire des relations spatiales (Vertegaal & al., 2000).

* *Contenu des interactions*. Les actes de parole destinés à vérifier la compréhension du partenaire distant sont moins nombreux lorsque le dispositif vidéo permet *soit* uniquement le regard mutuel (Doherty-Sneddon & al./2, 1997) *soit* le regard mutuel et la visée du partenaire distant (Monk & Gale, 2002). Par ailleurs, lorsque les partenaires peuvent établir un regard mutuel, les regards échangés entre les locuteurs sont plus nombreux que lorsqu'il n'est pas disponible (Doherty-Sneddon & al./2, 1997 ; Vertegaal & al., 2000). En termes d'efficacité, il

vaut donc mieux se voir afin de réduire les actes de parole finalisés à valider et à vérifier la compréhension. Cependant, la condition « audio & vidéo avec regard mutuel » n'apporte pas un avantage important par rapport à la condition « audio & vidéo sans regard mutuel ».

Est-ce que la réalisation de la tâche est facilitée par le regard mutuel ou par la connaissance de la visée du partenaire ?

D'après les résultats, la performance dans la tâche n'est affectée ni par le regard réciproque, ni par la connaissance de l'orientation du regard du partenaire distant. Bien que le regard réciproque augmente le nombre d'interruptions (Doherty-Sneddon & al./2, 1997 ; O'Malley & al./1, 1996), la performance de réalisation de la tâche ne change pas de manière significative, résultat qui est en contre tendance par rapport aux résultats de Sanford & al. (2004) et O'Malley & al./1 (1996), où il y a une corrélation entre le nombre d'interruptions et la performance de réalisation de la tâche.

Quels sont les effets sur les impressions subjectives des participants distants ?

Bien que peu d'études s'interrogent sur le vécu des participants (Tableau 2-3), il en résulte que les partenaires distants préfèrent se regarder dans les yeux (Sellen, 1995) et connaître la direction du regard d'autrui (Vertegaal & al., 2000) plutôt que se voir par l'intermédiaire d'une simple image-vidéo.

En résumé, lorsque « le regard mutuel » et « le regard mutuel & sa visée » sont disponibles, l'interactivité du processus de communication augmente, la compréhension réciproque est favorisée, alors que le score de réalisation de la tâche ne change pas. Du point de vue de l'efficacité, les conditions permettant le « regard mutuel » et la « direction du regard » sont donc équivalentes.

	Tâche	**Conditions**	**Sujets**	**Interactions communicatives**		**Impressions subjectives**
Sellen (1995)	Argumentation et conversation autour d'un thème	c1) face à face c2) audio & vidéo visage-buste sans regard mutuel c3) audio & vidéo visage-buste avec regard mutuel	12 quatuors	Les superpositions sont plus nombreuses en c1 qu'en c2 ou c3	---	Les sujets préfèrent communiquer avec leur propre interlocuteur en c3
O'Malley & al./1 (1996)	Reproduction d'une configuration spatiale (*map task*)	c1) audio seul c2) audio & vidéo visage-buste sans regard mutuel c3) audio & vidéo visage-buste avec regard mutuel	(Non précisé)	Les mots, les tours de parole et les interruptions sont plus nombreux en c3	---	---
Doherty-Sneddon & al./2 (1997)	Reproduction d'une configuration spatiale (*map task*)	c1) audio seul c2) audio & vidéo visage-buste sans regard mutuel c3) audio & vidéo visage-buste avec regard mutuel	36 binômes	Les tours de parole sont plus longs, les mots et les interruptions plus nombreux en c3 ; les acquiescements sont plus nombreux en c1	Les actes de parole destinés à vérifier la compréhension du partenaire sont plus nombreux en c1	---
Monk & Gale (2002)	Repérage d'un couple de points sur une image complexe	c1) audio seul c2) audio & vidéo visage-buste avec regard c3) audio & vidéo visage-buste avec regard mutuel & direction du regard	24 binômes	Les mots et les tours de parole sont plus nombreux en c3.	Les actes de parole destinés à vérifier et à valider la compréhension du partenaire sont plus nombreux en c2	---
Vertegaal & al. (2000)	Résolution de casse-tête linguistiques	c1) audio & vidéo visage-buste avec regard c2) audio & vidéo visage-buste avec regard mutuel & sa visée c3) audio & arrêt-image visage-buste & direction du regard	56 trinômes	Les tours de parole sont plus nombreux en c3. Les formes déictiques sont plus nombreuses en c2.	---	Les sujets préfèrent communiquer avec leur propre interlocuteur en c3

Tableau 2-3 : Les effets du regard mutuel et de la connaissance de la direction du regard du partenaire distant.

Conclusion

Dans ce chapitre, nous avons résumé des études concernant des activités de communication à distance, dans l'objectif de comprendre l'impact de l'image-vidéo du partenaire distant sur trois indicateurs principaux : les interactions communicatives, la réalisation de la tâche et les impressions subjectives des sujets. Les recherches présentées dans la section 2.1 suggèrent que l'image du partenaire distant assure une qualité de communication satisfaisante, favorise la compréhension réciproque sans pour autant induire des améliorations à la réalisation de la tâche (surtout pour les tâches de type déclaratif). Nous nous sommes ensuite interrogés sur les effets spécifiques des indices non-verbaux véhiculés par l'image-vidéo du partenaire distant (Tableau 2-4) :

* *visage seul ou visage-buste* (§ 2.2). Afin que le processus de communication soit plus interactif, il est préférable que le visage seul du partenaire distant soit visible. En revanche, si on veut privilégier l'efficacité, alors il vaut mieux montrer le visage et le buste plutôt que le visage seul.

Facteurs expérimentaux / **Variables dépendantes**	**Audio & vidéo du partenaire distant**	**Visage-buste vs. Visage seul**	**Regard mutuel et visée du partenaire**
Processus de communication	Réduction des interruptions	Le visage seul augmente l'interactivité	Réduction des interruptions
Compréhension réciproque	Favorise la construction des référentiels communs	---	Favorisent la construction des référentiels
Réalisation de la tâche	Inutile	Le visage-buste améliore l'efficacité	Inutiles
Impressions subjectives	Construit la présence sociale	Equivalents	Construisent la présence sociale

Tableau 2-4 - Synthèse des effets des facteurs expérimentaux relatifs à l'image-vidéo du partenaire distant.

* *le regard mutuel et la direction du regard d'autrui* (§ 2.3). Globalement, les résultats des études suggèrent que le regard mutuel assure l'interactivité, alors que la connaissance de la direction du regard assure plutôt l'efficacité. Cependant, ces tendances concernent surtout les tâches de type procédural. En ce qui concerne la performance, ni le regard mutuel ni la visée d'autrui ne permettent d'améliorer la

réalisation de la tâche. En revanche, les sujets estiment apprécier globalement ces dispositifs pour interagir avec leur partenaire distant.

Nous discutons la portée de ces résultats selon trois aspects majeurs.

Théorique.

D'une part, les résultats de Sellen (1995) et Vertegaal & al. (2000) confirment la théorie de la présence sociale : les participants apprécient davantage la condition de communication audio-vidéo, dont les caractéristiques se rapprochent le plus de la condition « face à face ». D'autre part, ces résultats l'infirment, car le score de réalisation de la tâche n'est pas influencé par la disponibilité de l'image-vidéo, quelle que soit la taille du groupe[29]. Cela suggère que les deux positions théoriques ne mettent pas à disposition des outils pour prévoir correctement les effets des indices non-verbaux sur les activités de communication à distance. A ce propos, Whittaker (2002) considère qu'il serait nécessaire de se doter d'un ensemble de descripteurs très fins pour construire une taxonomie adéquate des tâches de communication.

Méthodologique.

Les différences entre les études sont importantes, notamment pour :

- les dispositifs expérimentaux : par exemple, la disposition des caméras montrant l'image-vidéo n'est pas la même, avec la parallaxe parfois verticale parfois horizontale, ce qui peut induire de la gêne sociale chez les participants (cf. Huang, Olson & Olson, 2002) ;
- les variables indépendantes parfois confondues, comme la qualité de transmission de l'image-vidéo et le décalage entre le signal audio et vidéo ;
- les variables dépendantes quantitatives ne sont pas identiques : par exemple, le volume du dialogue (nécessaire pour évaluer le niveau d*'interactivité* d'une condition de communication) est parfois calculé sur la base du nombre des interruptions[30] (Rutter &

[29] A ce propos, Foulon-Molenda (2000) suggère que l'image du partenaire distant améliore la performance de réalisation lorsque le groupe est composé d'au moins trois individus. Cependant, cette interprétation s'appuie sur les résultats de l'étude d'Olson, Olson & Meader (1995), où le score de la performance des trinômes est plus élevé en condition « avec image-vidéo » qu'en condition « sans image », mais aucune différence statistique n'est constatée.

[30] Cependant, l'interprétation du nombre d'interruptions est dépendante du contrat de communication : par exemple, dans une conversation générique, un nombre élevé d'interruptions est un

Stephenson, 1977 ; Sellen, 1995) parfois sur la base du nombre des mots et des tours de parole (Doherty-Sneddon & al., 1997 ; O'Conaill & al., 1993 ; O'Malley & al., 1996). De plus, certains ne tiennent pas compte des interjections (Sellen, 1995), alors que d'autres oui (O'Conaill & al., 1993) ;

- les variables dépendantes qualitatives souvent non homogènes et non exhaustives (e.g. les catégories du codage des dialogues, les items des questionnaires) ;
- les caractéristiques des sujets, comme leur sexe (parfois non équilibré, plutôt des hommes ou plutôt des femmes)[31] et leur nombre (binômes *vs*. trinômes *vs.* etc.).

Ces différences méthodologiques limitent la portée de ces résultats, à la fois sur le plan de l'explication théorique (Olson & Olson, 1997) et de leur application à la conception de dispositifs destinés à la communication audio-vidéo.

Applicatif.

Comme nous l'avons indiqué précédemment, globalement l'image-vidéo du partenaire distant favorise le processus de communication entre les partenaires distants. Plus spécifiquement :

- si on veut améliorer l'efficacité de l'interaction, alors il sera préférable que le dispositif de communication offre au partenaire distant l'image-vidéo du visage-buste d'autrui ;
- si on veut favoriser le processus de communication (réduction des interruptions) et la compréhension réciproque (construction des référentiels communs), alors le dispositif de communication devra permettre aux partenaires distants d'échanger des regards mutuels et de connaître la direction du regard d'autrui.

indicateur d'un haut niveau d'interactivité (Sellen, 1995), alors que dans un dialogue finalisé à la réalisation d'une tâche c'est un indicateur des difficultés du processus de compréhension réciproque (O'Conaill & al., 1993).

[31] En situation dyadique, les femmes regardent leur interlocuteur plus souvent que les hommes (Argyle & al., 1968 ; Duncan & Fiske, 1977).

Problématique de recherche et hypothèses de travail

Les résultats de ces études ouvrent de nombreuses questions. Nous appuyant sur cette revue de la littérature, nous présentons des hypothèses de travail concernant l'impact de l'image-vidéo du partenaire distant.

Est-ce que l'image du tuteur et de l'apprenant modifie le volume des échanges ?

Cette question de recherche est fréquemment abordée dans la littérature, car le volume des dialogues est une mesure utilisée pour évaluer l'interactivité et l'efficacité des conditions de communication. Les différences évoquées précédemment relatives aux variables quantitatives se manifestent sur le plan de l'interprétation de l'impact de l'image-vidéo. En effet, lorsque des partenaires distants se voient :

a) *soit* le volume des dialogues est aussi important que lorsque les partenaires ne se voient pas (Daly-Jones & al./1, 1998 ; Masoodian & al., 1995 ; Sellen, 1995) ;
b) *soit* les dialogues sont plus longs (contiennent plus de mots et de tours de parole) que lorsque les partenaires ne se voient pas (Daly-Jones & al./2**,** 1998 ; Doherty-Sneddon & al./2, 1997 ; Isaac & Tang, 1994 ; Monk & Gale, 2002 ; O'Malley & al., 1996).

Deux explications sont proposées dans la littérature :

a) *sans l'image-vidéo du partenaire distant, le volume des dialogues est plus important* car les locuteurs doivent produire davantage d'énoncés pour contrôler le processus de communication et s'assurer de leur intercompréhension (Clark & Brennan, 1991) ;
b) *avec l'image-vidéo du partenaire, le volume des dialogues est plus important car le visage d'autrui agirait comme un facteur attrayant*, encourageant les locuteurs à exploiter[32] davantage cette modalité de communication. Monk & Gale (2002) suggèrent que les sujets impliqués dans les études, non habitués à la communication médiatisée, utiliseraient le dispositif de communication audio-vidéo presque de manière ludique et ils se parleraient davantage afin de tester les limites et les apports du dispositif.

[32] « … *the benefits of having access to visual signals may be counterbalanced by users becoming distracted by their partner's face and so using significantly more speech to achieve a comparable level of task success*" (Anderson, O'Malley, Doherty-Sneddon, Langton, Newlands, Mullin, Fleming & Van der Velden (1997 : 142).

Nous retenons que lors d'une interaction de type tuteur-apprenant, en vertu de ses caractéristiques (i.e., asymétrie de connaissances, relation d'assistance), l'image-vidéo d'autrui serait un facteur attrayant, et nous supposons donc une augmentation du volume des dialogues. A ce propos, nous retenons qu'un dialogue interactif n'est pas seulement riche en mots. En effet, on pourrait penser à deux personnes qui parlent longtemps, produisant un nombre important de mots mais qui n'alternent que rarement le tour de parole (e.g., un interview). Pour cela, nous pensons que les mesures quantitatives traditionnelles doivent être accompagnées d'un indicateur relatif à la longueur des tours de parole (rapport entre la quantité de mots et la quantité de tours de parole). De plus, dans le volume du dialogue, nous retenons qu'il est opportun de considérer aussi une mesure relative à la structure des interactions dialogiques, quelle que soit la grille d'analyse utilisée[33]. En effet, on pourrait penser à une situation dans laquelle des mots sont prononcés mais sans constituer une structure dialogique.

Donc, nous pensons que :

→ l'image-vidéo du partenaire distant induirait un volume du dialogue plus important (en nombre de mots, de tours de parole et des énoncés) et des tours de parole plus courts.

Concernant le volume des échanges, la littérature sur les activités d'assistance expert-novice a aussi étudié la fréquence d'aide spontanée que le tuteur apporte à l'apprenant (cf. Kraut, Fussell & Siegel, 2003). Cet indicateur se révèle pertinent dans le dialogue tutoriel, où l'apprenant ne maîtrise pas bien les connaissances nécessaires pour réaliser des exercices et, généralement en face à face, il sollicite souvent le tuteur pour lui demander de l'aide. En situation de dialogue tutoriel en « face à face », Fox (1993) observe que les expressions faciales de l'apprenant (montrant son implication et ses phases de difficultés) incitent le tuteur à intervenir auprès de lui sans que l'apprenant lui demande explicitement de l'aide.

→ Nous supposons que lorsque le tuteur peut observer le visage de l'apprenant, il peut anticiper ses requêtes d'aide et l'aider ainsi de manière spontanée.

[33] Comme nous le présentons de manière détaillée au chapitre 4, nous avons choisi de découper les protocoles verbaux issus de deux expérimentations en énoncés, chaque énoncé étant composé d'au moins un acte de parole et d'au moins un argument.

Est-ce que l'image-vidéo du partenaire distant influence le dialogue tutoriel ?

Comme nous l'avons indiqué au chapitre 1, les indices verbaux et non-verbaux influent sur la compréhension réciproque. L'interlocuteur peut ainsi manifester son incompréhension ou bien son accord et son désaccord. Sur le plan linguistique, le locuteur *vérifie* généralement la compréhension de l'allocutaire par des questions directes (e.g. « as-tu compris ? ») ou par des fins de phrases interrogatives et ce dernier lui *manifeste* son accord ou désaccord au moyen de marques de validation (e.g. acquiescement, réplique courte, etc.) (Clark & Brennan, 1991). Sur le plan visuel, le co-locuteur peut aussi manifester sa compréhension au locuteur par un hochement de tête, une mimique ou un geste, par un ensemble d'indices non-verbaux qui sont produits à la place ou en même temps qu'un indice verbal. En effet, l'image du partenaire distant permet de réduire le nombre de questions destinées à vérifier la compréhension d'autrui (cf. Doherty-Sneddon & al./1, 1997 ; Isaacs & al., 1994), car le regard entre les partenaires se substitue aux indices verbaux favorisant la compréhension réciproque.

→ On peut donc supposer que lorsque les indices non-verbaux relatifs à l'image du partenaire (tuteur et apprenant) ne sont pas à la disposition des locuteurs, les indices verbaux destinés à vérifier et à valider la compréhension sont plus nombreux que lorsque les partenaires distants se voient.

Lorsque le tuteur propose au novice des explications :

a) *soit* il lui délivre des actes de parole dédiés à la correction : le novice ne peut alors que répondre que par des acquiescements et des répliques courtes, et ne peut pas montrer son véritable niveau de compréhension ;

b) *soit* le tuteur utilise des énoncés en forme d'allusion et de question ouverte : il offre alors au novice la possibilité de contribuer au dialogue de manière constructive (e.g. par des réponses longues, des auto-explications).

→ Nous supposons que l'image du partenaire distant, réduisant le coût de contrôle de compréhension mutuelle, permet au tuteur et à l'apprenant de focaliser davantage sur les contenus d'apprentissage et donc de produire plutôt des actes de parole intrinsèques au dialogue tutoriel.

Est-ce que la qualité de réalisation de la tâche est différente lorsque le tuteur et l'apprenant se voient ?

D'après les études présentées, le visage et le buste du partenaire distant facilitent la réalisation de la tâche plutôt que le visage seul, mais n'apportent aucun avantage significatif par rapport à la condition « sans image-vidéo ». Cependant, ces études, mesurant surtout le temps de réalisation de la tâche ou attribuant un score de réalisation, ne nous donnent aucune indication générale de l'effet de l'image-vidéo sur des performances d'apprentissage. En revanche, des recherches conduites en situation de communication « face à face » (§ ch.1) montrent que, quand un enseignant et un apprenant peuvent échanger des regards, plus la fréquence des regards est élevée plus la performance d'apprentissage (en termes de rappel et résolution de problèmes) est meilleure (Fry & Smith, 1975 ; Otteson & Otteson, 1980).

→ Nous pensons que lorsque le tuteur et l'apprenant peuvent se voir, le score de réalisation de la tâche des apprenants augmente.

Est-ce que le tuteur et l'apprenant préfèrent se voir lorsqu'ils coopèrent à distance ?

Seulement deux études suggèrent que les individus préfèrent voir leur partenaire distant plutôt que ne pas le voir (Daly-Jones & al., 1998 ; Masoodian & al., 1995) et, par ailleurs, les items des questionnaires ne sont pas relatifs à une même dimension. D'après Clark & Brennan (1991), les propriétés physiques d'un environnement de communication permettent aux co-locuteurs de connaître réciproquement leur niveau d'attention et de compréhension, d'où l'importance de s'interroger sur cette dimension. Nous appuyant sur la théorie de la richesse sociale, nous supposons que

→ l'image-vidéo du partenaire distant peut induire une évaluation plus positive de leurs contributions à l'interaction.

En résumé, **lorsque le tuteur et l'apprenant peuvent se voir, nous supposons que :**

- le volume de leurs échanges verbaux sera plus important (H1) ;

- le tuteur interviendra davantage auprès de l'apprenant de manière spontanée (H2) ;

- le nombre d'indices verbaux consacrés au processus de compréhension réciproque diminuera (H3) ;

- le nombre d'indices verbaux inhérents au dialogue tutorial augmentera, le dialogue tutoriel sera ainsi focalisé sur les contenus à apprendre (H4);

- le niveau de réalisation de la tâche sera plus élevé (H5) ;

- l'évaluation de leurs interactions communicatives sera plus positive (H6)

Chapitre 3 - Partager un espace commun avec le partenaire distant : un avantage ou une entrave pour communiquer ?

Après avoir présenté des études dédiées aux effets de l'image-vidéo du partenaire distant sur les activités de communication médiatisée, dans ce chapitre nous abordons le rôle de l'image-vidéo relative à l'environnement de travail des partenaires distants. Lorsque les partenaires sont co-présents, ils disposent d'un même espace physique : chacun peut indiquer et manipuler des objets, et chacun peut voir ce que l'autre est en train de faire. En d'autres termes, ces objets et ces actions interviennent dans la construction de l'interaction, permettant la coordination et la négociation des tours de parole (Brassac & Gregori, 2001) et aussi la focalisation de l'attention des participants.

En situation de communication médiatisée, l'image-vidéo relative aux objets et aux actions doit permettre à chaque locuteur de savoir ce que son partenaire distant fait, regarde et désigne, de façon à coordonner leurs actions et à construire des représentations partagées et synchronisées (Navarro, 2001). Pour ce faire, les technologies offrent de partager :

a) un *espace visuel commun*, où des partenaires distants peuvent observer en même temps les mêmes entités, réduisant ainsi les ambiguïtés dans la communication, facilitant la coordination de leurs actions et le maintien de l'attention du groupe sur le déroulement de l'activité (Nardi & al., 1997).

b) un *espace de travail commun*, où des partenaires distants peuvent à la fois observer les actions de chacun et agir en même temps sur ces mêmes entités. « Agir » implique que les participants peuvent modifier les propriétés de ces ressources, selon les droits d'accès imposés par le dispositif et choisis par le groupe de travail (Ishii & al., 1993 ; Minneman & Bly, 1991; Tang, 1991; Whittaker, Brennan & Clark, 1991).

Comme pour l'image-vidéo du partenaire distant, peu de recherches se sont intéressées au partage des ressources lors de l'interaction tuteur-apprenant à distance. Nous avons donc eu recours à des études expérimentales les plus pertinents par rapport à notre objet d'étude. Dans un premier temps, nous abordons les études principales relatives à l'espace visuel commun (§ 3.1), puis nous résumons des études relatives à l'espace de travail commun (§ 3.2). Quelques études que nous présentons comparent des conditions de communication « face à face » *vs.* médiatisée : notre but étant de comprendre les mécanismes sociocognitifs impliqués en situation de communication médiatisée, nous ne présentons que les résultats pertinents pour notre perspective de recherche.

3.1 Le partage d'un espace visuel commun.

Karsenty (1999) s'interroge sur l'impact de l'espace visuel commun sur l'activité d'assistance à distance et il suppose qu'un espace visuel commun favorise le processus de compréhension réciproque entre des personnes distantes. Dans la tâche expérimentale, un utilisateur-novice d'un logiciel d'écriture doit mettre en forme un texte dans un temps limité (trente minutes), assisté par un utilisateur-expert du logiciel. La mise en forme comporte la réalisation de dix micro-tâches environ, comprenant, par exemple l'usage de la commande pour mettre en italique des phrases du texte, en gras, etc. Globalement, quinze utilisateurs-novices (étudiants, âge moyen de 23 ans) et deux utilisateurs-experts sont recrutés et repartis de manière aléatoire dans trois conditions expérimentales :

c1) « sans espace visuel commun & audio en continu » : le novice et l'expert sont délocalisés, communiquent uniquement grâce à une liaison audio ouverte en continu et chacun dispose d'un ordinateur contenant le logiciel d'écriture ;

c2) « avec espace visuel commun & audio en continu » : le binôme communique par une liaison audio ouverte en continu et le tuteur peut observer en temps réel

sur son écran l'activité de l'utilisateur-novice, grâce à un logiciel permettant le partage d'écran (Timbuktu®) ;

c3) « avec espace visuel commun en face à face » : le novice et l'expert sont assis à côté l'un de l'autre, avec le premier qui travaille sur un ordinateur et le deuxième qui répond à ses questions, sans intervenir directement sur la réalisation de la tâche.

Les dialogues sont enregistrés et retranscrits, mais seules les séquences où le sujet-novice demande de l'aide[34] à l'expert sont codées au moyen d'une grille structurée selon deux catégories :

a) les « niveaux d'information contextuelle » :

- le « niveau implicite », où les tours de parole ne contiennent aucune référence à l'espace de travail (e.g. « *comment cadrer le titre du document ?* ») ;
- le « niveau déictique », où les tours de parole renvoient aux éléments du contexte (e.g. « *est-ce que je déplace mon curseur ici* ? ») ;
- le « niveau explicite », où les objets du contexte sont nommés explicitement (e.g. « *ah la fenêtre a disparu*») ;

b) les « éléments du contexte » qui comprennent :

- l'état de l'interface, relatif à l'interface du logiciel d'édition (i.e. les boutons, le curseur) ;
- l'état d'avancement, relatif à l'état de la tâche en cours (i.e. les lignes, le format d'un paragraphe) ;
- les actions effectuées par le novice (e.g. écrire, effacer, etc.) ;
- les objectifs du novice (e.g. implicites ou explicités par ses mots ou par ses actions).

Voici les principaux résultats.

[34] Karsenty (1999) définit une « interaction d'aide » comme une séquence dialogique où le sujet-novice demande de l'aide à l'expert, ce qui est pour Kraut & al. (2003) une aide réactive. Cette séquence peut aboutir soit à une autre requête d'explication de la part du novice soit à la mise en oeuvre de la solution proposée par le tuteur.

Les effets sur les interactions communicatives. Sur le plan du processus, les tours de parole sont significativement plus nombreux en condition « sans espace visuel commun & audio » que dans les deux autres conditions. Sur le plan du contenu :

- le niveau déictique est significativement plus fréquent en conditions « avec espace visuel commun » (c2 et c3) qu'en condition « sans espace visuel commun » ;
- le niveau explicite est significativement moins fréquent en « face à face » qu'en condition « sans espace visuel commun & audio » ;
- le niveau implicite est utilisé de manière équivalente dans les trois conditions expérimentales.

Les questions de l'expert (sur l'état de l'interface, sur l'état d'avancement et sur les actions du novice) sont plus nombreuses en condition « sans espace visuel commun » qu'« avec espace visuel commun ».

Les effets sur la réalisation de la tâche. Le score de performance est calculé selon le temps nécessaire à accomplir chaque sous-tâche. Aucune différence significative n'est constatée entre les différentes conditions expérimentales.

Ces résultats montrent que :

a) pour le *processus* de communication, lorsque les partenaires distants n'ont pas accès à un espace visuel commun, leurs échanges sont plus longs ;

b) pour le *contenu* :

i) l'espace visuel commun réduit le nombre de tours de parole de « niveau explicite » et augmente le nombre de tours de parole de « niveau déictique », facilitant la construction des référentiels communs ;

ii) l'espace visuel commun réduit le nombre des questions relatives aux étapes déjà accomplies.

Kraut & al. (2003)[35] conduisent deux expérimentations sur l'activité d'assistance expert-novice dédiée à la réparation d'objets. L'objectif de leur recherche est d'analyser comment des personnes coopérant à distance utilisent un espace visuel commun pour construire leurs référentiels communs. Dans la première expérimentation, Kraut & al./1 (2003) avancent les hypothèses suivantes :

i. si la transmission du signal audio est de qualité faible (en alterné), alors le processus de communication est moins interactif et la réalisation de la tâche plus difficile ;
ii. avec un espace visuel commun, la compréhension réciproque et la construction des référentiels communs entre les partenaires est facilitée.

Pour les vérifier, ils construisent une situation expérimentale dans laquelle un expert doit aider un novice à effectuer un certain nombre de réparations sur une bicyclette. Selon la procédure expérimentale, les compétences en réparation du novice sont vérifiées à l'avance par un pré-test (test de Eckstrom pour équilibrer les capacités visuelles et spatiales des sujets). Selon la consigne, l'expert doit répondre au mieux à toutes les questions du novice (aide réactive) et l'assister spontanément lorsqu'il considère que le novice opère de manière incorrecte (aide spontanée ou proactive). Les 32 binômes recrutés pour l'expérimentation (deux experts et soixante-quatre novices, dont 69% de sexe féminin, sans aucune expérience en réparation de bicyclette) sont placés dans des pièces différentes et ils sont répartis de manière aléatoire dans les trois conditions expérimentales (plan inter) :

c1) « sans espace visuel commun & audio en continu » : l'expert communique avec le novice grâce à une connexion audio ouverte en continu (full-duplex) ;
c2) « avec espace visuel commun & audio en continu » : l'expert et le novice communiquent grâce à une liaison audio transmise en continu (full duplex);

[35] Ces deux expérimentations sont aussi décrites par Kraut, Miller & Siegel (1996) et par Fussell, Kraut & Siegel (2000).

c3) « avec espace visuel commun & audio alterné » : l'expert et le novice communiquent par voie orale grâce à une connexion audio en alterné (half duplex) ; l'expert dispose d'un écran pour observer les actions du novice (grâce à une caméra placée sur le casque du novice) et pour consulter le manuel en ligne dédié à la réparation du vélo; le novice dispose d'un écran sur lequel la partie supérieure du corps de l'expert apparaît (visage et buste) (Figure 3-1).



Figure 3-1 : Dispositif expérimental (Kraut & al., 2003)

Les données récoltées sont de deux types : les interactions dialogiques et la réalisation de la tâche. Les dialogues sont transcrits et codés au moyen d'une grille structurée en trois catégories : les questions (relatives à la tâche à réaliser), les descriptions (le novice décrit son état d'avancement à l'expert, qui lui décrit les actions à effectuer) et les acquiescements (de validation de l'expert vers le novice et de compréhension du novice vers l'expert). Le score de performance de la tâche est calculé selon le nombre des réparations effectuées et le temps employé.

Les effets sur les interactions communicatives. En ce qui concerne l'impact de la qualité de transmission du signal audio (continu *vs.* alterné) sur les interactions communicatives, aucune différence significative n'est constatée entre les différentes conditions expérimentales. En revanche, l'espace visuel commun influe sur le volume et sur la coordination du dialogue. Les mots et les aides spontanées de l'expert sont significativement moins nombreux en condition « sans espace visuel commun » que dans les deux autres conditions. Le novice décrit son état d'avancement et ses propres objectifs plus souvent en condition « sans espace visuel commun » plutôt qu'« avec espace visuel commun ». Les acquiescements verbaux de l'expert sont significativement plus nombreux en condition « sans espace visuel commun » qu'en condition « avec espace visuel commun ».

La réalisation de la tâche. Le nombre et la qualité de réparations effectuées, ainsi que le temps employé, ne diffèrent pas dans les trois conditions expérimentales.

D'après les résultats, d'une part, la qualité de la transmission du signal audio (continu *vs*. alterné) n'influe ni sur la réalisation de la tâche ni sur les interactions communicatives ; d'autre part, lorsque les participants accèdent à un espace visuel commun, le novice décrit moins souvent son état d'avancement et l'expert apporte plus fréquemment son aide de manière spontanée. La portée de ces résultats doit cependant être évaluée par rapport à des limites méthodologiques :

i. le plan expérimental aurait pu être complété par une autre condition (« sans espace visuel & audio alterné ») qui aurait apporté des indications complémentaires sur la liaison audio en basse qualité ;
ii. concernant le dispositif expérimental, la micro-caméra installée sur le casque du novice fournit à l'expert une image de qualité très faible, souvent dégradée lorsque le novice bouge la tête rapidement ;
iii. sur le plan méthodologique, la grille de codage utilisée pour analyser les dialogues ne semble pas assez bien structurée, comme en témoigne la valeur faible de l'indice inter-juge (kappa de Cohen k =.40). En effet, en accord avec la littérature relative au codage des corpus dialogiques (Carletta, Isard, Isard, Kowtko, Doherty-Sneddon & Anderson, 1996), la valeur de l'indice inter-juge (comprise entre 0 et 1) nous donne une appréciation sur la validité générale du codage. Dans l'expérimentation de Kraut & al./1 (2003), cette valeur est inférieure à la valeur moyenne ($0.41 < K < 0.6$).

Dans la deuxième expérimentation, Kraut & al./2 (2003) comparent de nouvelles modalités de communication sur la même tâche expérimentale, c'est-à-dire la réparation de vélo. Les auteurs intègrent les résultats de la première expérimentation (la qualité de transmission du signal audio n'influence ni la réalisation de la tâche ni les interactions communicatives) et paramètrent le signal audio ouvert en continu dans toutes les conditions expérimentales. Ils modifient aussi le plan expérimental (plan intra-sujets) et la grille de codage des dialogues. Ainsi, sous l'assistance d'un expert, vingt-cinq sujets (novices en réparation de bicyclette et étudiants universitaires dont 68% de sexe masculin) doivent réparer une partie spécifique du

vélo, assistés par un expert distant (plan intra-sujet, chaque sujet accomplit la tâche dans chaque condition expérimentale) :

c1) « sans espace visuel commun & audio continu » : l'expert et le novice communiquent par voie orale grâce à une connexion audio ouverte en continu ;

c2) « avec espace visuel commun & audio continu » : par rapport à la condition c2, l'expert peut observer sur son écran les actions du novice grâce à une micro-caméra placée sur le casque du novice ;

c3) « avec espace visuel commun en face à face » : le novice travaille à côté de l'expert.

Les données récoltées concernent des indicateurs relatifs à la réalisation de la tâche (la qualité des réparations effectuées et le temps employé pour les réaliser) et au corpus issu des interactions dialogiques. La grille utilisée pour le codage des dialogues est structurée selon le formalisme de type <prédicat, argument>, où :

- le <prédicat> comprend les questions, les réponses, les acquiescements verbaux et les assertions ;
- l'<argument> comprend les déictiques, les étapes procédurales, l'état de la tâche et l'état attentionnel du locuteur (e.g. « *I don't understand what you're talking about* »).

Deux juges réalisent le codage, dont la valeur de l'indice de fiabilité est très élevée (k =.90).

Les effets sur les interactions communicatives. En ce qui concerne le type d'aide apporté par l'expert (spontanée *vs.* réactive), les aides spontanées sont significativement plus nombreuses dans les deux conditions dotées d'un espace visuel commun. Globalement, les dialogues sont significativement moins longs en condition de « face à face » que dans les deux conditions à distance (qui ne diffèrent pas de manière significative entre elles).

D'après le codage effectué, il en résulte que :

- pour le prédicat, les acquiescements verbaux de l'expert sont significativement plus nombreux en condition à distance qu'en « face à face » ; aucune différence significative n'est constatée entre les trois conditions expérimentales pour les assertions et les réponses ;
- pour l'argument, l'expert suggère au novice les étapes procédurales à suivre plus souvent en condition de « face à face » qu'à distance (différence significative); en revanche les énoncés relatifs à l'état attentionnel du novice sont significativement plus

nombreux en condition « espace visuel commun avec audio en continu » que dans les deux autres conditions. En ce qui concerne les expressions déictiques, elles sont significativement plus nombreuses en condition « avec espace visuel commun en face à face » qu'en condition « espace visuel commun » et qu'en condition « sans espace visuel commun ».

Les effets sur la réalisation de la tâche. Le temps de réalisation est significativement plus court (25%) en condition « face à face » que dans les deux conditions à distance, qui ne diffèrent pas significativement entre elles. En ce qui concerne la qualité de résolution de la tâche, aucune différence significative n'est constatée entre les trois conditions expérimentales.

Parmi les résultats à retenir, nous soulignons que lorsque l'expert et le novice accèdent à un espace visuel commun : a) la qualité de réalisation de la tâche n'est pas améliorée ; b) les aides spontanées et les expressions déictiques sont plus nombreuses.

On pourrait donc supposer que grâce à l'espace visuel commun : i) l'expert détecte mieux les difficultés du novice et intervient sans attendre une demande d'aide de sa part ; ii) la construction des référentiels communs est facilitée.

L'expérimentation de Gergle, Kraut & Fussell (2004)[36] a pour objectif de comprendre les caractéristiques spécifiques qu'un espace visuel commun doit posséder pour favoriser les activités de communication à distance. Les auteurs supposent que, pour des tâches à haute complexité visuelle, un espace visuel commun améliore l'entente entre les partenaires distants et aussi la performance de réalisation. Pour vérifier cette hypothèse, ils construisent une situation expérimentale où un sujet-novice doit recomposer des puzzles grâce à l'aide d'un sujet-expert distant. Le puzzle est en version électronique et les deux partenaires sont placés dans deux pièces différentes, chacune équipée d'un ordinateur et d'un écran. L'écran du sujet-novice est séparé en deux zones (Figure 3-2): l'une est consacrée à afficher les pièces restantes du puzzle, l'autre est consacrée au puzzle construit au fur et à mesure que l'apprenant place les pièces. L'écran du sujet-expert se compose d'une partie consacrée à visualiser la solution du

puzzle et d'une partie consacrée à montrer le puzzle construit par l'apprenant (zone d'espace visuel commun).

Les variables indépendantes de l'expérimentation sont les suivantes :

a) le type d'espace visuel commun (facteur intra-sujet à trois modalités) : sans espace visuel commun *vs*. avec espace visuel commun décalé temporellement de 3 secondes (entre le mouvement d'une pièce sur l'écran du novice et l'affichage de la pièce sur l'écran de l'expert) *vs*. avec espace visuel commun en temps réel ;

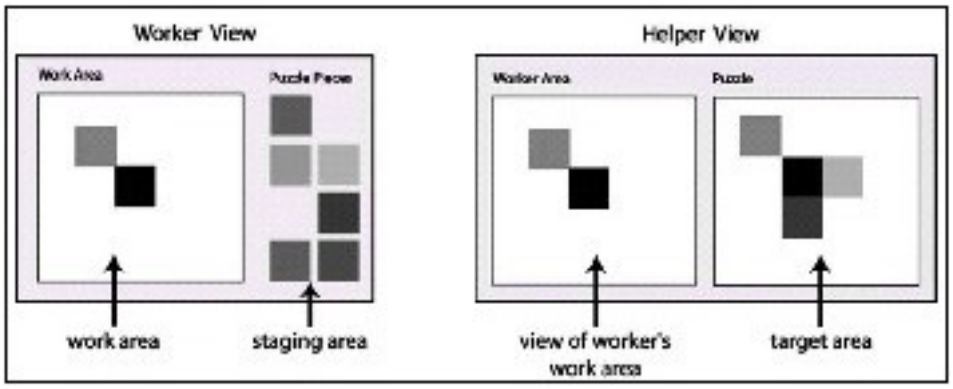


Figure 3-2 : L'espace visuel commun utilisé par Gergle & al. (2004).

b) la difficulté du puzzle (facteur inter-sujet) à deux modalités : couleur changeante (la couleur des pièces change toutes les 5 secondes) *vs*. couleur invariante (la couleur ne change pas tout au long de la tâche) ;

c) la nature des pièces du puzzle (facteur intra-sujet) à deux modalités : pièces juxtaposées *vs*. pièces emboîtées ;

Les sujets recrutés sont douze binômes (étudiants), repartis de manière aléatoire en deux moitiés : l'une doit recomposer les puzzles dont les pièces changent de couleur, l'autre doit construire les puzzles dont les pièces ne changent pas de couleur. Les binômes de chaque moitié doivent construire quatre puzzles dans chacune des six conditions expérimentales (n=3 niveaux d'espace visuel commun X n=2 types de pièces du puzzle). Les données retenues par

[36] La même expérimentation est décrite par Kraut, Gergle & Fussell (2002).

les auteurs sont les dialogues (enregistrés, transcrits et codés, avec accord inter-juges k = .90) et le temps de réalisation de la tâche.

Les effets sur les interactions communicatives. Sur le plan du processus, les mots sont significativement moins nombreux lorsque l'espace visuel commun est disponible en temps réel. Sur le plan du contenu :

- sans espace visuel commun, les acquiescements sont significativement plus nombreux (plus chez le novice que chez l'expert) que dans les autres conditions ;
- concernant les déictiques, les auteurs distinguent les pronoms déictiques (e.g. *ici*, *là*, etc.) et les expressions déictiques spatiales (e.g. en bas, à droite, etc.) : les premiers sont significativement plus nombreux lorsque l'espace visuel commun est disponible en temps réel, alors que les secondes sont significativement plus nombreuses dans chaque condition « avec espace visuel commun » et lorsque la nature du puzzle comporte des pièces à juxtaposer.

Les effets sur la réalisation de la tâche. Le temps de réalisation des puzzles est significativement plus court lorsque l'expert dispose de l'espace visuel commun en temps réel. Par ailleurs, lorsque les couleurs du puzzle ne changent pas et l'espace visuel commun est en temps réel, le temps de réalisation de la tâche est significativement plus court que lorsque les couleurs du puzzle ne changent pas et l'espace visuel commun n'est pas disponible ou retardé. Quand la couleur change, le temps de réalisation est significativement plus court quand l'espace visuel commun est disponible en temps réel plutôt que lorsqu'il ne l'est pas.

Ces résultats indiquent que l'espace visuel commun en temps réel améliore l'efficacité des dialogues, car, à parité de score de performance, les mots sont moins nombreux. En effet, les experts et les novices utilisent des pronoms déictiques plutôt que des expressions déictiques pour indiquer la position des objets sur l'écran. De plus, les experts s'assurent que les instructions données au novice sont exécutées et comprises simplement en observant leurs actions. Ainsi, les résultats soutiennent l'hypothèse d'un modèle coopératif de communication (Schober & Clark, 1989) : en fait, le novice modifie sa stratégie de dialogue en fonction de l'information à disposition de l'expert, notamment en augmentant le nombre d'acquiescements verbaux pour manifester sa compréhension lorsque l'expert ne peut pas l'observer.

Comme nous l'avons montré précédemment (cf. Kraut & al., 2003), les dispositifs mobiles permettant un espace visuel commun peuvent parfois entraver la coopération à distance. La cause principale serait le manque d'adéquation entre la direction de la vue de la caméra qui capture l'espace du partenaire distant et la visée de la personne qui regarde la caméra. Pour pallier ces limites, Fussell, Setlock & Kraut (2003) utilisent d'autres dispositifs permettant un espace visuel commun, notamment une caméra-vidéo frontale mouvant avec eye-tracker et une caméra fixe orientée sur le partenaire distant et sur son espace de travail. La première offre une vue détaillée et rapprochée sur les objets du champ visuel du partenaire distant, alors que la deuxième offre une vue plus globale de son espace de travail. Les auteurs avancent l'hypothèse que ces deux dispositifs permettraient à chaque participant distant de percevoir plus facilement l'état d'avancement du travail de l'autre, améliorant ainsi leur entente commune et donc la qualité de réalisation de la tâche. Pour vérifier leur hypothèse, ils construisent une situation expérimentale dans laquelle 38 sujets doivent assembler des briques de Lego™ pour construire un robot, assistés par un expert distant. Chaque novice doit construire cinq parties différentes d'un même robot dans chacune de ces trois conditions expérimentales (plan intra-sujet) :

c1) « sans espace visuel commun & audio continu» : l'expert et le novice ne communiquent que par voie orale ;

c2) « avec espace visuel commun fixe & audio continu» : une caméra fixe est placée dans la pièce sur le côté latéral postérieur droit, permettant à l'expert d'observer sur son propre écran la partie supérieure du corps du novice et les briques placées sur le plan du travail ;

c3) « avec espace visuel commun mouvant & audio continu » : grâce à une micro-caméra dotée d'un dispositif eye-tracker et placée sur le front du novice, l'expert peut observer sur son propre écran les objets du champs visuel du novice ;

c4) « avec espace visuel commun mouvant & fixe & audio continu» (Figure 3-3): l'expert peut observer dans une fenêtre de son propre écran les objets du champs visuel du novice (comme en c2) et dans une autre fenêtre la partie supérieure du corps du novice et son plan de travail (comme en c3) ;

c5) « avec espace visuel commun en face à face » : l'expert et le novice se trouvent dans la même pièce.

Pour construire chaque partie du robot il faut environ 10 minutes, pour un total de 50 minutes environ de passation expérimentale pour chaque sujet-novice. Dans les conditions c1, c2 et c4, le novice peut voir sur un écran le visage et le buste de l'expert distant, grâce à une caméra placée au-dessus de l'écran de l'expert. Dans toutes les conditions, l'expert dispose sur son écran d'une aide en ligne contenant les instructions pour la construction du robot.

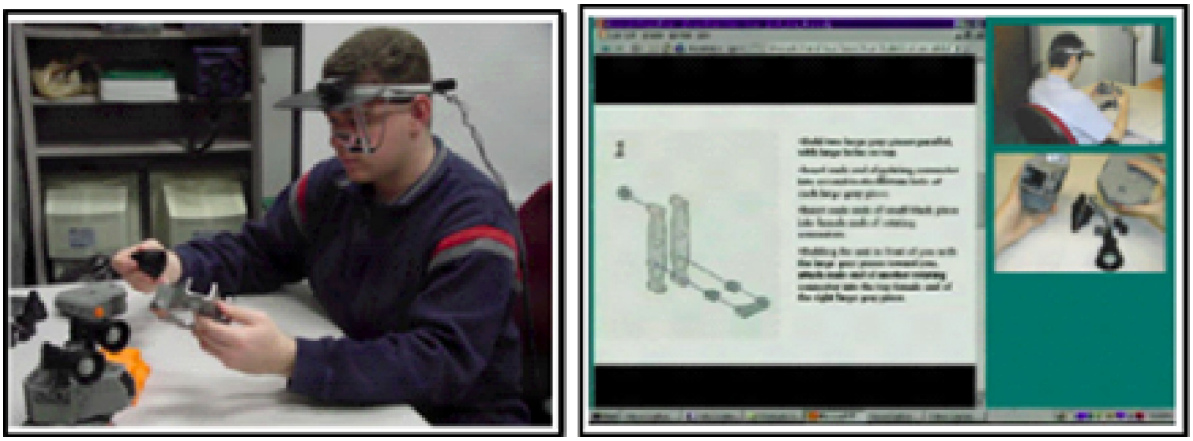


Figure 3-3 : A gauche : le novice en train de construire le robot. A droite : l'écran de l'expert avec l'espace visuel commun fixe (en haut à droite) et mouvant (en bas à droite) (Fussell & al., 2003)

Les variables dépendantes quantitatives sont la quantité de mots produits lors de l'interaction et le temps de réalisation de la tâche. Un questionnaire est soumis à la fin de chaque session expérimentale au sujet-novice et au sujet-expert. Le sujet-expert est invité à évaluer (sur une échelle allant de 1 à 5) le succès de la collaboration (e.g. « *my partner and I worked well together on these* »), l'apport de l'espace visuel commun (e.g. « *It was important to me to be able to see what my partner was doing* ») et l'aide apportée au sujet-novice.

Les effets sur les interactions communicatives. Au niveau du processus de communication, l'expert et le novice produisent significativement moins de mots en condition « face à face » que dans les autres conditions. De plus, les mots produits en condition « espace visuel commun mouvant & fixe » et en condition « espace visuel commun fixe » sont significativement moins nombreux que les mots produits en condition « espace visuel commun mouvant ».

Les effets sur la réalisation de la tâche. Le temps de réalisation de la tâche est significativement plus court en condition « face à face » que dans les autres conditions. Concernant les autres conditions à distance, les sujets construisent plus rapidement le robot (différence significative) avec un « espace visuel commun fixe » plutôt qu'en condition « sans espace visuel commun ». Aucune différence significative n'est constatée entre les autres conditions.

Les effets sur les impressions subjectives. Les experts et les novices apprécient davantage leur performance (différence significative) en condition « face à face » plutôt que dans les autres conditions. En condition de communication à distance, les participants (experts et novices confondus) préfèrent collaborer de manière significative en condition « avec espace visuel commun fixe » et « avec espace visuel commun mouvant & fixe » plutôt qu'en condition « avec espace visuel commun mouvant » et « sans espace visuel commun ». L'expert estime que son aide est significativement plus pertinente en condition « face à face » plutôt que dans les autres conditions ; il préfère la condition « avec espace visuel commun fixe » ou « avec espace visuel commun mouvant & fixe » que la condition « avec espace visuel commun mouvant » (différence significative).

Globalement, les résultats montrent que lorsque les partenaires distants disposent d'un « espace visuel commun », ils réalisent la tâche plus rapidement que lorsqu'ils n'en disposent pas. Contrairement aux hypothèses des auteurs, l' « espace visuel mouvant » n'améliore pas les performances de réalisation de la tâche. En ce qui concerne l'efficacité de la condition expérimentale, il en résulte que pour un score de performance équivalent, l' « espace visuel commun mouvant » est moins efficace que l' « espace visuel commun fixe » (avec ce dernier, les sujets produisent un nombre moins important de mots). D'après l'ensemble des résultats relatifs aux impressions subjectives des participants, il en résulte que la vidéo-caméra fixe est préférée à la vidéo-camera frontale. Pour expliquer ces résultats, les auteurs suggèrent que le mouvement continu de la vidéo-caméra frontale induit une charge cognitive importante chez l'expert et en conséquence une perte d'efficacité. En particulier, les auteurs soulignent la difficulté de construction d'un référentiel commun.

Synthèse

Dans cette section, nous avons présenté quatre études expérimentales relatives à l'impact de l'espace visuel commun sur l'activité d'assistance expert-novice dans des domaines différents : la mise en forme d'un document (Karsenty, 1999), la réparation de vélo (Kraut & al., 2003), la composition de puzzles (Gergle & al., 2004) et l'assemblage de pièces (Fussell & al., 2003). Les résultats de ces études sont à évaluer par rapport à trois facteurs :

a) la tâche expérimentale : les sujets sont impliqués surtout dans des tâches procédurales ;

b) le champ de vision offert par l'espace de travail est réduit à cause des caractéristiques du dispositif (e.g., la résolution et les dimensions de l'image-vidéo, le point de la focale, la parallaxe, etc.) et chaque sujet ne dispose que d'informations sélectionnées et non exhaustives du contexte d'interaction[37] de son partenaire distant ;

c) le codage des dialogues n'est pas uniforme (e.g. valeur de l'accord inter-juges, nombre de juges, catégories d'analyse non exhaustives).

Est-ce que l'« espace visuel commun » influence les interactions communicatives ?

Quant au processus de communication, les résultats sont assez homogènes (Tableau 3-1) : on constate qu'avec l'espace visuel commun, les dialogues sont plus courts et les interventions spontanées de l'expert auprès du novice sont plus nombreuses (cf. Karsenty, 1999 ; Kraut & al./2, 2003). Quant au contenu, lorsque l'espace visuel commun est disponible, les expressions déictiques sont plus nombreuses (cf. Gergle & al., 2004 ; Karsenty, 1999 ; Kraut & al./2, 2003). Cela suggère donc que la construction des référentiels communs est favorisée. Cependant, l'espace visuel commun doit répondre à des spécifications minimales : l'image doit être transmise sans aucun délai temporel (cf. Gergle & al., 2004) et la caméra doit

[37] Nous ne nous attardons pas ici sur le débat autour de la notion de « contexte » dans la conception des systèmes dédiés au travail coopératif. Nous rappelons seulement que, suivant le paradigme proposé par Borges, Brézillon, Pino & Pomerol (2005), le contexte d'interaction est l'ensemble des informations relatives aux actions que les participants sont en train de réaliser.

être fixe (et orientée vers autrui et ses objets de travail) plutôt que mouvante (cf. Fussell & al., 2003).

Est-ce que l'espace visuel commun améliore la réalisation de la tâche ?

Globalement les résultats sont contrastés. En effet,

- *soit* l'espace visuel commun n'a aucun effet ni sur la qualité (cf. Kraut & al./2, 2003) ni sur le temps de réalisation de la tâche (cf. Karsenty, 1999 ; cf. Kraut & al./1, 2003),
- *soit* il réduit le temps de réalisation (Fussell & al., 2003 ; Gergle & al., 2004 ; Kraut & al./2, 2003).

Les sujets préfèrent-ils coopérer via un espace visuel commun ?

Parmi les études présentées, seulement Fussell & al. (2003) s'interrogent sur l'effet de l'espace visuel commun sur les impressions subjectives des participants (Tableau 3-1). Nous nous limitons à rappeler que tous les participants préfèrent coopérer au moyen d'un espace visuel fixe plutôt que mouvant. De plus, l'expert estime que son aide est plus pertinente et adéquate aux exigences du novice en condition « avec espace visuel commun fixe » et « avec espace visuel commun mouvant & fixe » plutôt qu'en condition « avec espace visuel commun mouvant ».

Si on interprète l'ensemble de ces résultats selon le critère de l'efficacité dialogique (ratio « performance de réalisation de la tâche / volume des échanges »), alors ces dialogues sont sensiblement plus efficaces lorsque l'espace visuel commun est disponible selon une configuration fixe et en temps réel.

Les technologies informatiques mettent aussi à disposition des logiciels permettant à des utilisateurs distants de manipuler des objets communs en temps réel. Un certain nombre de recherches, dont nous présentons une synthèse dans la section suivante, s'interroge sur les apports de ces logiciels.

	Tâche	Conditions expérimentales	Sujets	Interactions communicatives		Réalisation de la tâche
Karsenty (1999)	Edition d'un document	c1) sans espace visuel commun & audio continu c2) avec espace visuel commun & audio continu c3) avec espace visuel commun & face à face	15 binômes	Les tours de parole sont plus nombreux en c3 qu'en c1 ou c2.	Les déictiques sont moins nombreux en c2 et c3 ; les questions de l'expert sur l'état d'avancement du novice sont plus nombreuses en c3.	Aucune différence significative.
Kraut & al./1 (2003)	Réparation de bicyclette	c1) sans espace visuel commun & audio continu c2) avec espace visuel commun & audio continu c3) avec espace visuel commun & audio alterné	32 binômes	Les aides spontanées de l'expert sont moins nombreuses en c1 qu'en c2 ou en c3 ; les mots et les tours sont plus nombreux en c1 qu'en c2 et en c3	Les descriptions du novice sur son état d'avancement et sur ses propres objectifs sont plus nombreuses en c1 ; les acquiescements de l'expert sont plus nombreux en c1.	Aucune différence significative.
Kraut & al./2 (2003)	Réparation de bicyclette	c1) sans espace visuel commun & audio continu c2) avec espace visuel commun & audio continu c3) avec espace visuel commun & face à face	25 binômes	Les dialogues sont moins longs en c3 ; les aides spontanées sont moins nombreuses en c1 qu'en c3 et c2.	Les acquiescements de l'expert sont moins nombreux en c3 ; L'expert décrit plus d'étapes procédurales en c3 ; Les expressions déictiques sont plus nombreuses en c3	Le temps de réalisation est plus court (25%) en c3 ; La qualité de réalisation ne diffère pas entre les conditions.
Gergle & al. (2004)	Composition de puzzles	Espace visuel (sans *vs*. décalé de 3 sec. *vs*. en temps réel) X Difficulté du puzzle (couleur changeante *vs*. invariante)	12 binômes	Les mots sont moins nombreux lorsque l'espace visuel commun est en temps réel.	Les acquiescements sont plus nombreux sans l'e.v.c. ; les pronoms déictiques sont plus nombreux lorsque l'e.v.c. est disponible en temps réel	Le temps de réalisation est plus court avec l'e.v.c. en temps réel, même si la couleur change.
Fussell & al. (2003)	Construction d'un robot	c1) sans espace visuel commun & audio continu c2) avec espace visuel commun fixe & audio continu c3) avec esp. visuel comm. mouvant & audio cont. c4) avec esp. vis. com. mouvant & fixe & audio con. c5) avec espace visuel commun & face à face	38 binômes	Les mots sont moins nombreux en c5 que dans les autres conditions et moins nombreux en c2 qu'en c3.	---	Le temps de réalisation est plus court en c5 plutôt qu'en c2 et qu'en c1.

Tableau 3-1 : Les effets de l'espace visuel commun sur les interactions communicatives et sur la réalisation de la tâche.

3.2 Le partage d'un espace de travail commun.

Un *espace de travail commun* permet aux partenaires distants non seulement de visualiser mais aussi d'agir sur les mêmes ressources pour réaliser des activités coopératives. Est-ce que la possibilité d'« agir » constitue un avantage supplémentaire à l'espace visuel commun ?

Whittaker, Geelhoed & Robinson (1993) conduisent trois expérimentations afin de comprendre comment un espace de travail commun tel qu'un tableau électronique peut être un support pour des activités collaboratives à distance.

Pour la première expérimentation, Whittaker & al./1 (1993) recrutent huit binômes (experts dans la rédaction de notes de travail et chacun d'eux appartenant à des équipes de travail différentes) pour rédiger un document électronique. Le document doit contenir les objectifs du groupe de travail auquel ils appartiennent, et doit être rédigé dans un temps limite d'environ sept minutes. Les sujets doivent accomplir cette tâche dans deux conditions (effet d'ordre contrôlé) :

c1) « sans espace de travail commun & audio en continu » : seulement un membre du binôme peut écrire sur un tableau électronique, et tous les deux discutent sur le contenu et l'ordre des objectifs grâce à une liaison audio ouverte en continu ;

c2) « avec espace de travail commun & audio en continu » : les deux sujets peuvent écrire sur le même document électronique, tout en communiquant oralement grâce à la liaison audio ouverte en continu.

Les dialogues enregistrés sont analysés sur le plan de la structure (tours de parole) et du contenu (le nombre d'idées évoquées). La réalisation de la tâche est mesurée selon le temps et la qualité de réalisation du document électronique. A la fin de chaque session expérimentale, les participants sont invités à remplir un questionnaire relatif à leur sentiment de satisfaction globale. Le questionnaire comprend les items suivants (la valeur de l'échelle d'évaluation n'est pas précisée par les auteurs): 1) « *how satisfied were you with what you as a team produced* ?, 2) « *how happy were you with the cooperation between you and your partner* ? », 3 « *How well-suited was the system to the task* ? ».

Les effets sur les interactions communicatives. Les tours de parole sont significativement plus nombreux en condition « sans espace de travail commun » qu'en condition « avec espace de travail commun ».

Les effets sur la réalisation de la tâche. Ni le temps ni la qualité de réalisation ne sont significativement différents dans les deux conditions expérimentales.

Les effets sur les impressions subjectives. Les sujets sont plus satisfaits de la qualité de leur performance en condition « espace de travail commun & audio en continu » que dans les autres conditions (différence significative).

Dans la deuxième expérimentation, Whittaker & al./2 (1993) construisent une tâche de rédaction collaborative plus complexe. Il s'agit de lire un long document de travail (500 mots), détecter les parties du texte qui ne sont pas homogènes et les modifier de manière à rendre la lecture cohérente pour des enfants de 9 ans. Huit binômes sont recrutés (tous appartenant au staff administratif du laboratoire Hewlett Packard de Bristol, UK) et ils ont à disposition quinze minutes pour accomplir la tâche. Les auteurs s'attendent à ce que l'espace de travail commun favorise la réalisation de la tâche, étant donnée la longueur du texte. Les sujets doivent accomplir la tâche dans les mêmes conditions que la première expérimentation (« sans espace de travail commun & audio en continu » et « avec espace de travail commun & audio en continu»). Les données recueillies concernent les dialogues (nombre de mots), la réalisation de la tâche (nombre de paragraphes édités) et les impressions subjectives des participants (les mêmes items que l'expérimentation précédente).

Les effets sur les interactions communicatives. Le nombre de tours de parole ne diffère pas de manière significative dans les deux conditions expérimentales.

Les effets sur la réalisation de la tâche. Le nombre de paragraphes édités ne diffère pas de manière significative dans les deux conditions.

Les effets sur les impressions subjectives. Les sujets sont plus satisfaits de leur travail lorsqu'ils accomplissent la tâche en condition « avec espace de travail commun » (différence significative).

Dans la troisième expérimentation de Whittaker & al. (1993), huit binômes doivent choisir comment disposer les meubles d'un appartement. Chaque membre du binôme dispose

d'un ordinateur où des meubles différents sont proposés. Les deux ordinateurs composent un espace de travail commun et le binôme doit optimiser le nombre des meubles (chaises, tables, sofas) selon l'espace disponible. Les deux sujets disposent d'une carte en deux dimensions montrant les dimensions de l'appartement et des meubles. Le plan et les conditions expérimentales sont les mêmes que dans les deux expérimentations précédentes. Au regard des caractéristiques visuelles requises par la tâche, les auteurs s'attendent à ce que l'espace de travail commun favorise les interactions communicatives (moins de tours de parole) ainsi que la performance de la tâche. A la fin de chaque session expérimentale, les sujets sont interrogés sur leur sentiment de satisfaction (mêmes items que les expérimentations précédentes).

Les effets sur les interactions communicatives. Les tours de parole sont significativement plus nombreux en condition « sans espace de travail commun » qu'en condition « avec espace de travail commun ».

Les effets sur la réalisation de la tâche. Le score de réalisation de la tâche (calculé selon le nombre des éléments placés dans chaque pièce) ne diffère pas de manière significative dans les deux conditions expérimentales à la fin du temps accordé. Néanmoins, après les sept premières minutes, les sujets obtiennent un score plus élevé en condition « avec espace de travail commun » plutôt qu'en condition « sans espace de travail commun ». On note aussi que 63% des sujets en condition « avec espace de travail commun » atteignent un score très élevé (environ 90% du score total) contre seulement 25% des sujets de la condition « sans espace de travail commun ».

Les effets sur les impressions subjectives. Concernant le sentiment de satisfaction globale des participants, aucune différence significative n'est constatée entre les deux conditions.

Les résultats de ces trois expérimentations sont contrastés. En effet, dans la première (tâche de rédaction collaborative) et la troisième expérimentation (tâche d'optimisation de l'espace), avec l'espace visuel commun les dialogues sont plus courts et la réalisation de la tâche est équivalente. En revanche, dans la deuxième expérimentation (tâche de rédaction collaborative), ni le volume des dialogues ni le score de réalisation ne sont influencés par la disponibilité de l'espace visuel commun. Si on considère le critère d'efficacité, on peut suggérer que l'espace visuel commun augmente l'efficacité du dialogue. Cependant, si on

considère simplement la ralisation de la tâche, aucune amélioration de la performance de réalisation n'est constatée dans les trois expérimentations.

D'autres études ont évalué l'impact de l'espace de travail commun sur les tâches d'apprentissage. Par exemple, Gnisci, Papa & Spedaletti (1999) ont réalisé une expérimentation pour évaluer l'apport d'une plateforme multimédia pour l'enseignement synchrone à distance. Leur hypothèse de recherche s'inspire de la théorie de la richesse des média (Daft & Lengel, 1986). Dans la situation expérimentale, un tuteur anime un cours sur les notions fondamentales pour utiliser un ordinateur. Le public est constitué de trois binômes, placés dans des pièces distantes. L'enseignant anime le cours, structuré en cinq unités didactiques, et à la fin de chaque unité, les trois binômes doivent résoudre des exercices pratiques (dont la nature n'est pas précisée par les auteurs), sous le contrôle pédagogique de l'enseignant. A la fin de la session de travail, les participants remplissent un questionnaire. Globalement, dix-huit binômes (dont 60% de sexe féminin) et un tuteur sont recrutés, préalablement entraînés à l'usage des fonctionnalités des dispositifs de communication et repartis de manière aléatoire selon deux conditions expérimentales :

c1. « sans espace de travail commun & vidéo du partenaire avec audio continu » : le tuteur et les apprenants communiquent grâce à un canal audio (transmis en continu) et ils peuvent se voir l'un l'autre en temps réel grâce à une liaison vidéo (les auteurs ne précisent pas l'étendue de l'image du partenaire distant) ;

c2. « avec espace de travail commun & espace visuel commun & vidéo du partenaire avec audio continu » : le tuteur et les trois binômes communiquent grâce à une liaison audio-vidéo et ils disposent d'un espace de travail commun, grâce auquel ils peuvent dessiner des schémas et écrire, et d'un espace visuel commun, composé d'une une caméra-vidéo qui permet de montrer à tous les participants leurs documents de travail.

Les mesures récoltées consistent en un post-test pour mesurer le score d'apprentissage des sujets et d'un questionnaire (en échelle Likert) pour mesurer leurs impressions sur les fonctionnalités du dispositif et sur le contexte socio-relationnel.

Les effets sur la réalisation de la tâche. Les sujets améliorent leur score dans les deux conditions et le gain est significativement plus important en condition « avec espace de travail commun & vidéo du partenaire avec audio continu ».

Les effets sur les impressions subjectives. Les impressions des sujets relatives à leur satisfaction, au sentiment de présence du partenaire distant, au climat d'apprentissage et à la facilité d'usage des fonctionnalités de communication ne diffèrent pas de manière significative dans les deux conditions expérimentales.

Contrairement aux hypothèses, les résultats montrent qu'un dispositif de communication à distance constitué d'un espace visuel commun et d'un espace de travail commun n'améliore ni l'efficacité de l'apprentissage ni l'évaluation des sujets portant sur le climat socio-relationnel. Plus précisément, le score d'apprentissage est moins élevé dans la condition « avec espace visuel commun & espace de travail ».

Ces résultats ne sont ni expliqués, ni justifiés par les auteurs par rapport à leurs hypothèses de recherche. Nous pouvons avancer une explication en termes de surcharge cognitive provoquée par la richesse des modalités de présentation de l'information (Sweller & Chandler, 1994 ; Kalyuga, Chandler & Sweller, 1999) : les apprenants auraient un score faible au post-test car pendant le cours ils ont dû partager leur attention visuelle entre les documents affichés dans l'espace visuel commun, les objets à modifier dans l'espace de travail et la vidéo de l'enseignant.

Synthèse

Dans cette section, nous avons présenté quatre études expérimentales relatives à l'impact de l'espace de travail sur la rédaction de textes, la résolution de problèmes de type spatial (Whittaker & al., 1993) et l'apprentissage (Gnisci & al., 1999).

Quel est l'impact de l'espace de travail commun sur les interactions communicatives ?

Les mesures relatives aux interactions communicatives ne concernent que la structure des dialogues, qui sont plus courts (et donc plus efficaces à égalité de performance, cf. Whittaker & al./1-3, 1993) grâce à l'espace de travail commun (Tableau 3-2).

	Tâche	**Conditions**	**Sujets**	**Interactions communicatives**	**Réalisation de la tâche**	**Impressions subjectives**
Whittaker & al./1 (1993)	Edition d'un document	c1) sans espace de travail commun & audio continu c2) espace de travail commun & audio continu	8 binômes	Les tours de parole sont moins nombreux en c2	Aucune différence significative.	Les sujets sont plus satisfaits de la qualité de leur travail en c2
Whittaker & al./2 (1993)	Edition d'un document par rapport à un lecteur cible	c1) sans espace de travail commun & audio continu c2) espace de travail commun & audio continu	8 binômes	Aucune différence significative	Aucune différence significative.	Les sujets sont plus satisfaits de la qualité de leur travail en c2
Whittaker & al./3 (1993)	Optimisation d'un espace bi-dimensionnel	c1) sans espace de travail commun & audio continu c2) espace de travail commun & audio continu	8 binômes	Les tours de parole sont moins nombreux en c2	Le score augmente plus rapidement en c2, mais aucune différence finale n'est constatée.	Aucune différence significative
Gnisci & al. (1999)	Résolution de problèmes dans une situation d'apprentissage	c1) sans espace de travail commun & vidéo & audio continu c2) espace de travail commun & espace visuel commun & vidéo & audio continu	18 binômes avec 1 tuteur	---	Le score est plus élevé en c1	Aucune différence significative

Tableau 3-2 : Les effets de l'espace de travail commun sur les activités de communication médiatisées.

Quelle est influence de l'espace de travail commun sur la réalisation de la tâche ?

Les résultats de ces études sont assez contrastés (Tableau 3-2). Deux différences principales émergent :

a) pour des tâches de résolution de problèmes visuels (comme l'optimisation d'un espace), l'espace de travail commun favorise de meilleures performances dans un délai temporel très court (cf. Whittaker & al./3, 1993);

b) pour des tâches d'apprentissage, lorsque les participants disposent de « l'espace de travail commun & l'espace visuel commun & l'image-vidéo du partenaire distant » les performances sont de moindre qualité (cf. Gnisci & al., 1999). Cela suggère que les

participants, devant plusieurs sources d'informations visuelles co-temporelles, sont dans une situation de surcharge cognitive dans le traitement de l'information.

Quel est l'impact de l'espace de travail commun sur les impressions subjectives ?

Les résultats sont contrastés (cf. Tableau 3-2). Les sujets préfèrent travailler avec un espace de travail commun plutôt que sans lorsqu'ils sont impliqués dans une activité plutôt déclarative (écriture collaborative) (cf. Whittaker & al./1-2,1993), autrement aucune préférence spécifique n'est manifestée (cf. Gnisci & al., 1999 ; Whittaker &al./3).

Conclusion

L'objectif de ce chapitre était de comprendre l'impact de l'espace visuel commun et de l'espace de travail commun (Tableau 3-3). En ce qui concerne le premier, les résultats dans l'ensemble montrent que le processus d'interlocution est facilité (dialogues plus courts, plus de déictiques pour la construction des référentiels communs), le temps de réalisation de la tâche est réduit et les participants apprécient de coopérer avec un espace visuel commun fixe.

En revanche, les résultats des études concernant l'espace de travail commun indiquent que d'une part, les dialogues sont plus courts, et que, d'autre part, la réalisation de la tâche n'est améliorée et les participants ne sont plus satisfaits.

Facteurs expérimentaux **Variables dépendantes**	**Espace visuel commun**	**Espace de travail commun**
Processus de communication	Réduction du volume des dialogues	Réduction du volume des dialogues
Compréhension réciproque	Favorise la construction des référentiels communs	---
Réalisation de la tâche	Réduit le temps de réalisation ; aucun effet sur la qualité	Aucune amélioration du temps et du score de réalisation
Impressions subjectives	Evaluation positive des participants à la coopération	Appréciation mitigée des participants

Tableau 3-3 : Synthèse des effets de l'espace visuel commun et de l'espace de travail commun.

Sur le plan *méthodologique*, ces études présentent des limites similaires aux études du chapitre précédent, notamment :

a) les variables dépendantes, non homogènes et non exhaustives sur le plan qualitatif ;

b) les caractéristiques de l'échantillon expérimental (e.g., nombre non équilibré, cf. Gnisci & al., 1999).

Sur le plan *théorique*, globalement ces résultats soutiennent la théorie de la richesse des média. En effet, la théorie affirme que plus les indices (verbaux et non-verbaux) mis à disposition sont nombreux, plus l'incertitude d'interprétation d'une information se réduit. Les résultats montrent en fait que l'espace visuel commun favorise la compréhension réciproque, notamment pour la construction des référentiels.

Sur le plan *applicatif*, l'aspect plus intéressant est qu'un espace visuel mouvant n'apporte pas le benéfice équivalent à celui d'un espace visuel fixe (cf. Fussell & al., 2003 ; Kraut & al., 2003).

Problématique et hypothèses de recherche

Nous avons présenté ces études afin d'éclaircir l'influence de l'espace visuel et de l'espace de travail commun sur les activités de communication à distance. Malgré les résultats de ces recherches, des questions demeurent nombreuses et une sélection s'impose. Par rapport à notre objet d'étude, à savoir l'interaction médiatisée synchrone tuteur-apprenant, nous retenons comme opportun de considérer surtout les questions concernant l'espace visuel commun. En effet, comme nous l'avons évoqué au premier chapitre, dans le « contrat de communication tutorielle » le tuteur fait en sorte que l'apprenant devienne autonome. En d'autres termes, le tuteur se devrait de structurer le dialogue avec l'apprenant afin que ce dernier puisse comprendre et apprendre et le tuteur ne devrait ni agir à la place de l'apprenant, ni lui fournir une assistance directe (i.e., résoudre le problème à sa place). Cette perspective oriente notre recherche sur les effets de l'espace visuel commun : le tuteur ne peut qu'observer ce que l'apprenant fait (sans agir à sa place) et il ne peut que co-construire une interaction avec lui.

***Est-ce que l'espace visuel commun influe sur le volume des dialogues* ?**

Bien que les recherches présentées concernent des tâches plutôt procédurales que déclarative, les résultats montrent que le volume des dialogues est moins important lorsque les partenaires distants disposent d'un espace visuel commun plutôt que lorsqu'ils n'en disposent pas (Fussell & al., 2003 ; Whittaker & al., 1993). Autrement dit, à performance égale, l'espace

visuel commun améliore l'efficacité du dialogue. Les interprétations suggèrent que le volume des dialogues diminue car les participants utilisent surtout des indices non-verbaux pour désigner les objets appartenant à leur espace de travail (cf. Gergle & al., 2004 ; Karsenty, 1999 ; Kraut & al./2, 2003).

→ Il peut donc être supposé que le volume des dialogues soit moins important lorsque l'espace visuel commun est disponible.

Par ailleurs, les études de Karsenty (1999) et Kraut & al. (2003) portant sur l'activité d'assistance expert-novice, montrent que, grâce à l'espace visuel commun, l'expert aide de manière plus spontanée le novice.

→ On peut donc supposer que, lorsque l'espace visuel commun est disponible, le tuteur puisse aider davantage l'apprenant de manière spontanée.

Est-ce que l'espace visuel commun influence la compréhension réciproque ?

D'après notre revue de la littérature, deux réponses sont envisageables :

a) D'une part, il apparaît que l'espace visuel commun permet aux partenaires distants de comprendre plus rapidement l'état de travail de leur partenaire distant (moins de questions sur l'avancement de l'activité, cf. Karsenty, 1999 ; Kraut & al., 2003) et d'identifier plus rapidement les objets de leur environnement de travail (plus d'expressions déictiques et moins de descriptions explicites, cf. Gergle & al., 2004 ; Karsenty, 1999 ; Kraut & al., 2003).
b) D'autre part, les acquiescements (employés pour valider la compréhension ou pour donner son propre accord) et les énoncés dédiés à la manifestation de sa propre compréhension sont plus nombreux lorsque l'espace visuel commun n'est pas disponible (cf. Gergle & al., 2004 ; Kraut & al./2, 2003).

Une interprétation proposée par la littérature suggère que l'espace visuel commun favorise la construction d'un « *contexte cognitif partagé* » entre les participants (Karsenty, 1999). En d'autres termes, chaque participant possèderait une connaissance actualisée des actions en cours des autres partenaires, ce qui lui permettrait de synchroniser sa propre contribution et son action avec le travail des autres participants (Falzon, 1994 ; Schmidt, 1994). Nous supposons que :

→ L'espace visuel commun favoriserait la compréhension réciproque entre le tuteur et l'apprenant pendant l'interaction.

Par ailleurs, lorsque les représentations cognitives de la tâche sont « synchronisées » (Navarro, 2001), le tuteur et l'apprenant sont en mesure de co-construire leur interaction surtout grâce aux énoncés intrinsèques au dialogue tutoriel (Graesser & Person, 1994).

→ On peut donc s'attendre à ce que le tuteur et les apprenants émettent un nombre plus important d'énoncés intrinsèques au dialogue tutoriel.

***Est-ce que l'espace visuel commun influe sur la qualité de réalisation de la tâche commune* ?**

Les recherches présentées se consacrent surtout à mesurer les effets de l'espace visuel commun sur le temps de réalisation (Fussell & al., 2003 ; Gergle & al., 2004), qui diminue lorsque les partenaires en disposent. D'après les études sur le dialogue tutoriel (§ ch.1), il s'avère que plus les participants produisent des énoncés intrinsèques au dialogue tutoriel (e.g., question et explication de l'apprenant ; allusion du tuteur, etc), plus la qualité de l'apprentissage est élevée. Supposant que l'espace visuel commun pourrait favoriser la production d'énoncés intrinsèques au dialogue tutoriel,

→ On peut supposer que l'espace visuel commun pourrait faciliter la qualité de l'apprentissage.

Est-ce que l'espace visuel commun influence les impressions subjectives des participants ?

D'après les études présentées, il apparaît que les participants distants apprécient davantage la qualité des interactions lorsqu'ils disposent d'un espace visuel commun (cf. Whittaker & al., 1993), ainsi que la pertinence de leur contribution (cf. Fussell & al., 2003).

→ On peut donc s'attendre à ce que l'apprenant et le tuteur évaluent plus positivement la qualité de leur contribution.

Nous allons opérationnaliser ces hypothèses de recherches dans la deuxième étude expérimentale dédiée à l'impact de l'espace visuel commun sur l'interaction tutorielle (ch. 5).

PARTIE EXPERIMENTALE

Chapitre 4 - Etude I : les effets de l'image-vidéo du partenaire sur l'interaction tuteur-apprenant[38]

L'objectif de cette première expérimentation est d'étudier les effets des indices non-verbaux de type kinésique, véhiculés par l'image-vidéo du partenaire distant, sur l'interaction tutorielle à distance. En accord avec argumentations exposées à la fin du chapitre 2, dans cette première étude expérimentale nous voulons vérifier que si l'image-vidéo du partenaire distant est disponible, alors :

- le volume des échanges verbaux est plus important ;
- le tuteur intervient davantage auprès de l'apprenant de manière spontanée ;
- le nombre d'indices verbaux consacrés au processus de compréhension diminue;
- le nombre d'indices verbaux consacrés au dialogue tutoriel augmente ;
- le niveau de réalisation de la tâche des apprenants est plus élevé ;
- les interactions communicatives sont évaluées plus positivement par les locuteurs.

[38] Une partie des résultats présentés dans ce chapitre a fait l'objet d'une communication de recherche (Tajariol, Dubois & Adam, 2002) et d'une publication (Tajariol, Adam & Dubois, 2003).

4.1 Méthodologie

4.1.1 La population

Nous avons recruté deux populations de participants. Concernant les apprenants, nous avons recruté 48 étudiants en Psychologie (répartis de manière équivalente en Licence et en Maîtrise), d'âge moyen 24.1 ans (SD = 6.2), avec 75% de femmes, tous francophones et novices en programmation et en notation HTML. Leur participation étant volontaire, ils ont bénéficié d'un crédit pour leur examen de Psychologie Sociale. Afin de créer des groupes homogènes, chaque apprenant a répondu à un pré-test (test à choix multiple, Tableau 4-1).

Quel est le formalisme correct pour obtenir un saut de ligne entre « il pleut » et « à Grenoble » ? a. <B> il pleut à Grenoble </B> b. <P> il pleut </P> à Grenoble c. </P> il pleut à Grenoble <P> d. je ne sais pas
Quel est le formalisme correct pour centrer une phrase dans une page de texte? a. <P> … … </P> b. <CENTER> … … </CENTER> c. </CENTER> … … <CENTER> d. je ne sais pas
Quel est le formalisme correct pour créer un lien hypertextuel interne orienté vers une partie d'un même document? a. <A HREF="#départ"> … </A> b. <A NAME ="#départ"> … </A> c. <A HREF=#départ> d. je ne sais pas
Quel est le formalisme correct pour obtenir ***Parc de la Chartreuse*** (en italique souligné gras)? a. <I><B><A HREF="#départ"> Parc de la Chartreuse </B></I> b. <I><U> <B> Parc de la Chartreuse </B> </U> </I> c. </I></U> </B> Parc de la Chartreuse <B> <U> <I> d. je ne sais pas

Tableau 4-1 : Items du pré-test pour les apprenants

Pour chaque réponse, le score est le suivant :

0 point pour la réponse « je ne sais pas »;

1 point pour la réponse qui présente une erreur sémantique importante (i.e. un type de balise qui ne sert pas à mettre en forme de manière correcte) ;

2 points pour la réponse avec une erreur de syntaxe (i.e. un slash de plus) ;

3 points pour la réponse correcte.
Le score obtenu est équivalent dans les conditions expérimentales (t (46) = .09 ; n.s.), ainsi que leur âge (t (46) = 1.318 ; n.s.).

Pour la population des tuteurs, 12 enseignants (six hommes et six femmes, d'âge moyen 30.9 ans (SD = 7.7)) ont été recrutés (2 maîtres de conférences, 4 doctorants-moniteurs de 3ième année, 2 doctorants-moniteurs de 2ième année et 4 moniteurs de 1ière année). La moitié d'entre eux (âge M = 35.5, S.D. = 8) a une expérience confirmée dans l'enseignement supérieur (supérieure à 2 ans), alors que l'autre moitié a peu d'expérience de l'enseignement (âge M = 25.5, SD = 2.2). Leur formation de base est l'informatique ; tous maîtrisent plusieurs langages de programmation ainsi que les méthodes de réalisation des pages HTML. Aucun d'eux n'a jamais participé à de sessions de tutorat synchrone médiatisé. Pour la participation à l'expérimentation, de la durée d'une heure, chaque tuteur perçoit une rémunération de 25€.

4.1.2 Le matériel

4.1.2.1 Le cours

La tâche demandée au tuteur consiste à assister deux apprenants en situation de travail pratique. L'objectif des apprenants consiste à apprendre des notions de base de la notation HTML et à les utiliser pour réaliser une page HTML. Le tuteur se doit de les assister, en répondant à leurs questions, et d'assurer leur réussite. Chaque apprenant dispose d'un support du cours de quatre pages (annexe I), contenant les définitions des balises HTML, leur notation formelle, des exemples de mise en forme et des exercices à réaliser.

4.1.2.2 Le dispositif technique

Pendant la phase expérimentale, les deux apprenants et le tuteur sont installés dans trois salles différentes. Chaque salle dispose d'un ordinateur connecté au réseau Internet et d'un dispositif audio-vidéo dédié à la communication. Ce dernier est composé d'une caméra, avec microphone annexe, d'un téléviseur de taille 14 pouces qui affiche l'image (le buste et le visage) de la personne distante et permet à chacun, apprenant et tuteur, de voir et d'entendre son partenaire.

Les deux apprenants ne communiquent pas entre eux, le seul flux de communication possible est entre le tuteur et l'apprenant. Le tuteur décide de parler à l'un d'eux grâce à un

dispositif spécifique[39]. Chaque apprenant peut solliciter l'intervention de l'enseignant par l'intermédiaire d'un logiciel d'écriture textuelle synchrone (MSN Messenger). Afin de favoriser la co-direction du regard entre la caméra (qui capture l'image du locuteur) et le moniteur (qui l'affiche), la camera est placée au-dessus du moniteur[40]. Chaque partenaire est enregistré grâce à un magnétoscope relié à la caméra. Pour la réalisation des pages en HTML, les apprenants disposent d'un même éditeur (Edit Plus) installé sur chaque ordinateur. Pour la visualisation de ces pages, chaque ordinateur est équipé du navigateur Internet Explorer.

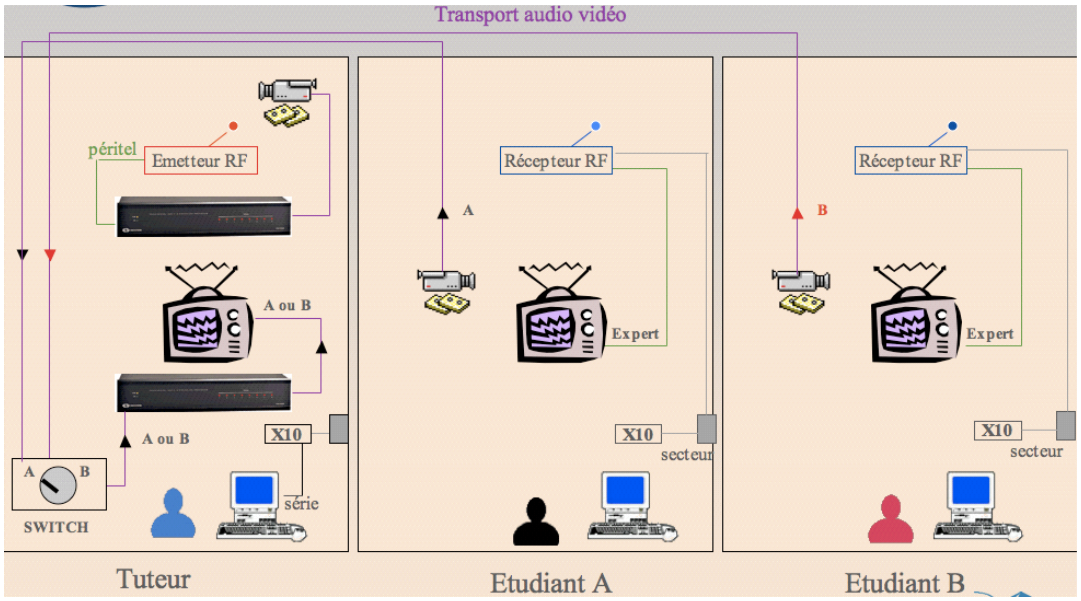


Figure 4-1 : Schéma technique de la situation expérimentale

[39] Nous remercions l'équipe MULTICOM, notamment Brigitte Meillon et Martine Harsche, pour avoir mis à disposition leurs locaux et pour avoir développé le dispositif de commutation en technologie x10. La norme X10 utilise des transmetteurs spécifiques qui émettent un signal codé à travers le réseau électrique d'un bâtiment. Les récepteurs X10 branchés entre les appareils et les prises de courant détectent ce signal et agissent en fonction du message réçupéré.

[40] Le dispositif technique à notre disposition n'empêche pas de reproduire la parallaxe visuelle verticale (§ ch. 2).

4.1.3 La procédure

La procédure expérimentale est organisée en trois phases, celles-ci se répartissant sur une durée totale d'une heure d'expérimentation. Pour chaque session expérimentale, un tuteur suit deux apprenants à la fois.

Phase introductive (15 minutes). Le tuteur et les deux apprenants sont réunis dans un bureau avec l'expérimentateur, présentés les uns aux autres et ils sont invités à lire la consigne expérimentale (cf. Annexes A). Puis les apprenants sont invités à remplir un test à choix multiple (§ Tableau 4.1) afin de mesurer leur niveau de connaissance de la notation HTML et pour les répartir ainsi de manière équivalente dans les deux groupes expérimentaux. Une fois le test rempli, à l'aide d'une check-list l'expérimentateur leur explique les commandes des logiciels à utiliser pendant la session expérimentale, de manière à réduire la variabilité de la compétence parmi les apprenants. Les questions relatives à la connaissance de la notation HTML sont rapidement corrigées et les apprenants sont affectés à un des deux groupes expérimentaux, selon leur niveau de connaissance en HTML afin de rendre les deux groupes équilibrés par rapport à leur sexe, leur âge et leurs connaissances de la notation HTML. Ils sont ensuite informés des moyens de communication disponibles pendant la session et ils sont invités à rejoindre l'une des deux salles.

Phase expérimentale (35 minutes). Dans chaque salle, les apprenants trouvent un ordinateur, le moniteur et la caméra, ainsi que le support papier du cours html, contenant les définitions et les exemples des balises (annexe I). D'après la consigne expérimentale, ils doivent mettre en format HTML une page de texte non formatée. Les apprenants ont pour consigne de solliciter l'aide du tuteur si nécessaire, et celui-ci a la tâche d'assister et de suivre les deux étudiants pendant la session. A la fin du temps accordé, le travail s'arrête, quel que soit le niveau de réalisation atteint par chaque apprenant.

Phase de post-test (10 minutes). Le cours en format papier est retiré et chaque apprenant est invité à répondre aux mêmes questions du pré-test (Tableau 4.1) et à un questionnaire destiné à recueillir ses impressions subjectives (cf. 4.7.1.2). Le tuteur également répond à un questionnaire relatif aux impressions ressenties pendant la session expérimentale.

4.1.4 Les conditions expérimentales

Notre but étant de comprendre l'impact de l'image-vidéo du partenaire distant sur les interactions dialogiques, nous avons construit la situation expérimentale suivante :

c1) « *audio seul* » *:* le tuteur peut dialoguer avec l'apprenant sans le voir ;

c2) « *audio & image-vidéo du partenaire distant* » *:* le tuteur et l'apprenant peuvent communiquer oralement et chacun d'eux peut voir la partie supérieure du corps d'autrui (buste & visage).

Nous rappelons que le tuteur ne peut parler qu'avec un seul apprenant à la fois. Le plan expérimental est de type intra-sujets : chaque tuteur accomplit la tâche d'assistance dans chacune de deux conditions. L'ordre de passation des sujets-enseignants a été contrôlé selon une procédure aléatoire. Nous avons veillé à que le niveau d'étude des apprenants ($3^{ème}$ et $4^{ème}$ année de Psychologie) et leur niveau de connaissance de la notation HTML soient répartis de manière équivalente dans les deux conditions expérimentales. Le plan expérimental résultant est le suivant : $\underline{\mathbf{T}}_{\mathbf{12}}$ ***** $\mathbf{C}_{\mathbf{2}}$**,** où T est le facteur « tuteur », C le facteur « condition de communication ». Ce choix méthodologique se justifie en raison du nombre limité de tuteurs à notre disposition, sachant qu'un plan de type inter-sujets (i.e. un tuteur réalise la tâche dans une seule condition expérimentale) aurait offert de conditions strictement comparables (sans aucun biais dû à la répétition de la tâche, cf. McCarthy & Monk, 1994; Fussell & al. 2000).

4.1.5 Justifications des choix méthodologiques

a) Choix de l'activité de communication. Quelques études présentées aux chapitres précédents possèdent des analogies avec l'activité tutorielle, car un sujet-expert est invité à assister un sujet-novice pour accomplir la tâche (cf. Fussell & al., 2003 ; Karsenty, 1999 ; Kraut & al., 2003). Cependant, dans ces recherches on demande à des sujets de jouer à l'expert plutôt qu'au tuteur : c'est-à-dire que le sujet-expert aide le sujet-novice à résoudre son problème. En revanche, nous avons demandé à de véritables experts, ayant tous une expérience d'enseignement, d'animer, de suivre et d'expliquer à l'apprenant le comment et le pourquoi des notions et de leur application pour réaliser une page html.

b) Choix des conditions expérimentales. La plupart des recherches expérimentales présentées comparent et évaluent les conditions de communication à distance avec la

condition en « face à face ». La raison de ce choix est que la condition en « face à face » est considérée comme la condition de communication naturelle pour toute activité humaine (Bavelas, Hutchinson, Kenwood & Matheson, 1997). En revanche, en accord avec Whittaker (2002), nous pensons qu'il est préférable d'étudier et de comparer des conditions de communication à distance afin d'évaluer les effets réels induits par la distance.

c) *Choix des variables dépendantes*. En accord avec McCarthy & Monk (1994) et Monk, McCarthy, Watts & Daly-Jones (1996), nous retenons que la seule mesure de la réalisation de la tâche ne permet ni d'appréhender les processus impliqués ni de décrire en détail les effets des modalités de communication sur l'activité humaine. Ainsi, nous pensons qu'il est nécessaire de mesurer aussi les effets des modalités de communication sur le processus et le contenu des interactions communicatives, ainsi que s'interroger sur les impressions subjectives des participants pendant l'interaction à distance (Monk & Watts, 2000). En effet, une expérimentation implique généralement des sujets non habitués à l'usage des technologies de la communication et pour une courte période de temps : comme les sujets tendent à focaliser toutes leurs ressources sur la tâche demandée, il est fort probable qu'ils arrivent tous à accomplir la tâche. Pour cela, il est nécessaire de choisir d'autres indicateurs susceptibles de mesurer les effets des variables indépendantes.

En résumé, ces considérations nous amènent à suivre une démarche expérimentale dans laquelle :

a) seulement des conditions de communication à distance sont étudiées et comparées ;
b) trois types de variables sont considérées : les interactions communicatives, la réalisation de la tâche et les impressions subjectives des sujets participants.

4.1.6 Méthodologie d'analyse du corpus

Les enregistrements de chaque session expérimentale ont été retranscrits *in extenso*. La transcription a été découpée selon la structure suivante, que nous avons construite nous inspirant à la fois du modèle genevois (Moeschler, 1985) et interactionnel (Kerbrat-Orecchioni, 1998) : séquence → tours de parole → énoncés → <actes de parole, argument>. Généralement, une séquence est constituée d'au moins un tour de parole. Chaque tour de parole est composé

d'au moins un énoncé, qui contient au moins un acte de parole. Nous allons maintenant décrire ces unités de codage, avec des exemples extraits de notre corpus.

Une **séquence** débute avec la prise de parole de l'apprenant ou du tuteur, et se termine avec un accord réciproque de deux interlocuteurs. La séquence débute avec une « ouverture » (Sacks & al., 1974) (généralement de ce type : A « *salut j'ai une question à te poser* », T « *oui j'écoute* ») et se termine avec une « fermeture » (généralement : A « *merci* », T « *de rien* »). Nous inspirant de la littérature relative au dialogue tutoriel, nous distinguons les *séquences proactives*, où le tuteur intervient sans que l'apprenant ne l'interpelle et les *séquences réactives*, où en revanche le tuteur intervient suite à une requête de l'apprenant.

Un **tour de parole** débute avec la prise de la parole d'un locuteur et se termine lorsqu'il arrête de parler. A l'intérieur de chaque tour de parole, nous avons identifié chaque énoncé (E) produit par le locuteur (e.g., un tour de parole contenant deux énoncés : T « *tu prends l'icone en haut à droite* [E1] *et tu la déplaces sur la grande icone au milieu de l'écran* [E2] »).

Pour le découpage des **énoncés**, chaque énoncé est interprété après avoir lu l'énoncé qui le suit dans l'enchaînement du dialogue. A l'intérieur, nous avons identifié les **marqueurs dédiés au processus** de compréhension et les **actes de parole dédiés à la co-construction du contenu**. Nous avons aussi repéré l'**argument** qui correspond à une spécification de l'acte de parole, son référent. Pour la catégorie de l'argument, nous avons considéré ces référents :

- *l'objectif* (obj), ce que l'apprenant déclare vouloir faire (e.g. : A « *euh j'aimerais en fait mettre euh parc naturel de la chartreuse et présentation en gras* ») ;
- *la procédure* (pro), c'est-à-dire la description des moyens et des règles nécessaires pour réaliser une action ;
- *l'état de l'environnement de l'apprenant* (eta), c'est-à-dire l'ensemble des éléments qui apparaissent sur son écran et sur son poste informatique ;
- *l'action accomplie par l'apprenant* (apa);
- *la tâche prescrite* (tp), l'ensemble des tâches que l'apprenant doit réaliser.

Pour coder les indices verbaux du tuteur et de l'apprenant, nous nous sommes inspirés à la fois de la littérature relative au dialogue tutoriel (§ ch.1) et du principe de l'intention communicative de Grice (1975) (e.g., Quelle est l'intention du locuteur ? Quelle est la fonction de son indice verbal ?).

Dans les sections suivantes, nous détaillons la grille de codage concernant les indices verbaux de l'apprenant et du tuteur.

4.1.6.1 Les indices verbaux utilisés par l'apprenant.

Pour la contribution de l'apprenant, nous avons identifié les marqueurs qu'ils utilisent pour le processus de compréhension réciproque et les actes de parole consacrés à la co-construction du contenu.

* *Les marqueurs de l'apprenant dédiés au processus de compréhension réciproque*.

a) Marqueurs d'acceptation de l'énoncé du tuteur. L'apprenant montre avoir compris (ou d'être d'accord avec) ce que le tuteur vient de dire. Il s'agit généralement d'acquiescements, dont nous avons identifié deux types :

- a-contextuel (e.g.: A « *oui, d'accord* »), où l'apprenant accepte simplement ce que le tuteur dit ;
- contextuel, où l'apprenant répète textuellement le(s) mot(s) du tuteur (e.g. : T « *tu cliques sur le menu à droite* », A « *ah le menu à droite* »).

La forme contextuelle offre une information plus précise au tuteur. En effet, grâce aux marqueurs d'acceptation contextuels, le tuteur sait quelle information a été validée par l'apprenant ; en revanche, les marqueurs d'acceptation a-contextuels donnent au tuteur une indication plutôt globale de ce que l'apprenant a accepté. Dans notre corpus, le nombre des marqueurs contextuels est bien faible par rapport aux marqueurs a-contextuels, donc nous ne les avons pas inclus parmi nos variables dépendantes[41].

b) Marqueurs de vérification de l'énoncé du tuteur. Lorsque l'apprenant n'est pas sûr de comprendre ce que le locuteur vient de dire, il utilise les marqueurs de vérification. En accord avec les quatre niveaux de compréhension réciproque proposés par le modèle de Clark & Wilkes-Gibbs (1986), nous avons repéré ces marqueurs de vérification :

[41] La somme de marqueurs d'acceptation est de 14, dont 10 unités en condition « audio seul » (M = 0.8, S.D. = 1), et 4 unités en condition « audio & vidéo » (M = 0.3, S.D. = 0.5).

- a-contextuels, lorsque l'apprenant n'a pas entendu correctement l'énoncé prononcé par le tuteur (niveau 1 du modèle ; e.g. : A « … *vous pouvez répéter* ?», « *Comment* ? », « *Quoi ?* ») ;
- contextuels, lorsque l'apprenant a entendu correctement l'énoncé du tuteur mais il veut vérifier sa compréhension et il demande au tuteur de confirmer (peut-être parce qu'il ne connaît pas la signification du mot, niveau 2 du modèle de Clark & Wilkes-Gibbs, 1986) (e.g., T « *tu ouvres la fenêtre et tu cliques sur le bouton ouvrir* », A « *bouton ouvrir ?* »).

Par les marqueurs de vérification contextuels, l'apprenant demande au tuteur de confirmer une information ponctuelle, alors que les marqueurs de vérification a-contextuels ont une portée plutôt globale. En raison du nombre faible d'effectifs, les marqueurs de vérification contextuels ne sont pas retenus pour notre étude[42].

* *Les actes de parole de l'apprenant dédiés à la co-construction du contenu*

c) Informer le tuteur sur son état d'avancement. L'apprenant renseigne le tuteur sur les étapes et les actions accomplies, sur l'état de son avancement (A « *cette partie je dois encore la terminer* ») et sur ses erreurs (e.g.: A « *ah l'erreur vient de là je cliquais pas au bon endroit* »).

d) Demander de l'aide au tuteur. L'apprenant demande de l'aide au tuteur sur :

- le contenu du cours (e.g.: A « *la partie HEAD doit être dans mon texte* ? «) ;
- la tâche prescrite (e.g. : A :« *je peux passer à l'exercice suivant* ? ») ;
- une procédure ou une commande (e.g. : A : « *je ne trouve pas la commande source du menu affichage* ») ;
- une étape réalisée afin d'avoir une évaluation ou une simple confirmation de la part du tuteur (e.g. : A « *ce que j'ai fait jusqu'à maintenant est correct selon vous* ? »).

[42] Le nombre de marqueurs destinés à vérifier l'énoncé du tuteur sont au total 17, dont 13 en condition « audio seul » (M = 1.1, S.D. = 1.7) et 4 unités en condition « audio et vidéo » (M = 0.3, S.D. = 0.5).

	Catégorie pragmatique	Forme verbale
Processus	Marqueurs d'acceptation de l'énoncé du tuteur	Acquiescements a-contextuels
	Marqueurs de vérification de l'énoncé du tuteur	Acquiescements a-contextuels
Contenu	Actes de parole pour demander de l'aide au tuteur	Requête : procédure, objectif, action réalisée
	Actes de parole pour informer le tuteur sur son état d'avancement	Description de : état d'avancement, actions accomplies, objectifs, procédures, tâche prescrite

Tableau 4-2 : Tableau récapitulatif des indices verbaux utilisés par l'apprenant.

4.1.6.2 Les indices verbaux utilisés par le tuteur.

Nous présentons les marqueurs du tuteur dédiés au processus de compréhension réciproque et les actes de parole inhérents au contenu tutoriel. Etant donné les caractéristiques du dialogue tutoriel (asymétrie des connaissances, objectifs complémentaires), le tuteur dirige le déroulement du dialogue. Cela comporte des différences profondes par rapport aux indices verbaux de l'apprenant.

* *Les marqueurs du tuteur dédiés au processus de compréhension réciproque*.

a) Marqueurs d'acceptation de l'énoncé de l'apprenant. Le tuteur donne son accord à ce que l'apprenant vient de dire. Son acceptation se présente sous forme d'acquiescements, que nous avons codés en :

- a-contextuels, constitués de formes neutres (e.g. : T « *oui d'accord* ») ;
- contextuels, où le tuteur répète textuellement ce que l'apprenant a dit (e.g. : A « *et je clique sur la petite flèche* », T « *oui sur la petite flèche* »)

En raison du nombre faible d'effectifs, les marqueurs d'acceptation contextuels ne sont pas retenus pour notre étude[43].

[43] Le nombre de marqueurs d'acceptation contextuels sont au total 13, dont 5 en condition « audio seul » et 8 en condition « audio et vidéo ».

b) Marqueurs de vérification de l'énoncé de l'apprenant. Le tuteur vérifie qu'il a bien compris ce que l'apprenant vient de lui dire. Nous avons identifié seulement quelques marqueurs de type a-contextuel (e.g. : T « *pardon* ? », « *Qu'est-ce que vous avez dit ?* »), mais leur nombre reste fort faible pour les intégrer dans l'analyse[44].

c) Marques de vérification de la compréhension de l'apprenant. Lors du déroulement de l'interaction verbale, le tuteur vérifie le degré de compréhension de l'apprenant. La vérification porte sur chaque information nouvelle et se manifeste généralement par des questions fermées (i.e. : T : « *c'est clair maintenant ?* » ou bien : « *ok ?* », « *c'est bon ?* ») et, rarement, par des queues de phrase interrogatives.

** Les actes de parole du tuteur dédiés à la co-construction du contenu du dialogue tutoriel*

c) S'informer sur l'état d'avancement de l'apprenant. Pour assister l'apprenant, le tuteur doit acquérir des informations nécessaires pour se représenter l'état d'avancement de l'apprenant. Dans ce but, il peut obtenir des informations concernant :

- les étapes accomplies par l'apprenant (e.g. : T « *as-tu ouvert le fichier sans le faire glisser sur l'application ?* ») ;
- les actions réalisées (e.g. : T « *Tu as essayé de le placer sur l'icone rouge* ? ») ;
- l'état de son environnement de travail (e.g. : T « *Tu as une flèche sur ta droite tu la vois ?* ») ;
- son objectif (e.g. : T « *Tu veux mettre en gras ou en italique ?* »)

d) Aider l'apprenant. Après avoir obtenu les informations nécessaires, généralement le tuteur propose son aide à l'apprenant. Nous appuyant sur la littérature relative au dialogue tutoriel (Chi, 1997 ; Graesser & Person, 1994), nous avons codé l'aide du tuteur en :

[44] Le nombre de marqueurs de vérification de l'énoncé de l'apprenant sont 7, dont 5 en condition « audio seul » et 2 en condition « audio et vidéo ».

d1) *Aide informative*. Le tuteur apporte de nouvelles informations à l'apprenant, mais sans lui expliquer les relations causales et modales de ces informations. Le tuteur fournit à l'apprenant des informations sur :

- une commande logicielle ou un élément du dispositif (e.g. : A « *euh c'est-à-dire que je mets P et P et après BR?* », T « *exactement* ») ;
- un terme spécifique (e.g. : T « *le bureau est l'espace où tu as tes objets* ») ;
- une procédure (e.g. : T « *tu fermes la fenêtre et tu ouvres l'autre* ») ;
- la tâche prescrite par la consigne (e.g. : T « *on vous demande de faire l'exercice-1 d'abord*»).

d2) *Aide explicative*. Le tuteur propose des explications à l'apprenant, de sorte que l'apprenant puisse créer des connexions logiques entre les nouvelles informations et ses connaissances antérieures. D'après notre corpus, le tuteur produit des explications concernant :

- les contenus du cours (e.g. : T « *le document HTML se compose d'une structure logique et d'une partie de texte et les balises font partie de la structure logique* ») ;
- les procédures (e.g. : T « *tu ne peux pas mettre là ta balise car tu n'as pas fermé l'autre avant et la structure est imbriquée* »).

d3) *Aide corrective*. La correction est un acte de parole généralement délivré à la suite d'une erreur de l'apprenant et qui ne comporte pas d'explications sur les raisons de l'erreur. D'après le corpus, le tuteur apporte des corrections :

- à une procédure (e.g. : A « *pour ouvrir il faut que je clique une fois c'est ça ?* », T « *non tu dois cliquer deux fois* ») ;
- aux contenus du cours (e.g. : A « *ce que j'écris en dehors des balises apparaît à l'écran* » T « *non c'est bien le contraire les balises font que le texte apparaît à l'écran* ») ;
- à un objectif (e.g.: A« *mais je peux mettre un coefficient devant* ?» T « *non absolument pas* »).

d4) *Aide évaluative*. L'intention du tuteur est d'évaluer l'état d'avancement de l'apprenant (i.e., l'étape accomplie, la procédure mise en oeuvre, l'objectif, etc.).

La valence du jugement peut être positive (e.g. : T « *là tu as bien travaillé sur cette partie là* ») ou bien négative (e.g. : T « *ça ne va pas du tout* »).

d5) *Aide allusive*. La fonction de l'allusion est d'orienter l'apprenant vers une information qui peut favoriser sa compréhension sans pour autant lui fournir la réponse correcte. Lors du codage de notre corpus, nous n'avons trouvé qu'une seule allusion (A « *quelle est la balise pour le gras ?* », T « *penses aux initiales de la commande mais en anglais* », A « *hum je ne connais rien en anglais* », T « *ok c'est B* »).

Etant le nombre d'effectifs faible (n=1), nous n'avons pas inclus cette catégorie dans l'étude expérimentale.

e) Encourager l'apprenant. Le tuteur offre un soutien moral à l'apprenant et il l'encourage à terminer les exercices (e.g. : A « *c'est moi qui me suis trompée* », T « *oui c'est pas grave allez* »).

	Catégorie pragmatique	**Forme verbale**
Processus	Marques d'acceptation de l'énoncé de l'apprenant	Acquiescements a-contextuels
	Marques de vérification de l'énoncé de l'apprenant	Acquiescements a-contextuels
	Marques de vérification de la compréhension de l'apprenant	Questions fermées
Contenu	Actes de parole pour connaître l'état d'avancement de l'apprenant	Questions sur : les étapes accomplies, les actions réalisées, l'état de son environnement de travail, ses objectifs
	Actes de parole pour aider l'apprenant	- Aide Informative : commande, procédure, tâche prescrite - Aide Explicative : objectif, procédure, contenu cours, tâche prescrite - Aide Corrective : procédure, contenu cours, objectif - Aide Evaluative : procédure, objectif, action accomplie - Aide Allusive
	Actes de parole pour encourager l'apprenant	Encouragement : tâche prescrite

Tableau 4-3 : Tableau récapitulatif des indices verbaux utilisés par le tuteur.

4.1.6.3 Les justifications relatives à la méthode suivie pour le codage des dialogues.

Pour construire cette grille, nous nous sommes inspirés de la littérature relative à l'analyse des protocoles verbaux en situation dynamique (Amalberti & Hoc, 1998) et des études relatives au dialogue tutoriel (Chi, 1997 ; Graesser, Person & Magliano, 1995 ; Pilkington, 1992 ; Porayska-Pomsta & al., 2000 ; Van Lehn & al., 2003). Concernant la méthode de construction, nous avons essayé de définir des catégories mutuellement exclusives.

Cependant, comme toute méthode de codage, la distance sémantique entre une forme et l'autre peut ne pas être égale (Chi & al., 2001). Par exemple, les marqueurs consacrés à la validation des énoncés du tuteur du type « *acquiescement a-contextuel »* et « *confirmation* » n'ont pas toujours obtenu une valeur d'accord inter-juges acceptable. Afin de rendre compte de la validité de notre grille de codage, nous avons voulu calculer un indice de fiabilité. A cette fin, nous avons recensé[45] trois tests de fiabilité dans la littérature (Carletta & al., 1996).

Nous nous sommes limités au calcul du test de reproductibilité, étant donné la dimension du corpus (environ 350 pages) et du coût en temps du codage (un ratio de 3 heures de codage pour 1 heure d'enregistrement réel). Le test de reproductibilité est une mesure inter-codeurs, calculant la variance entre les codeurs sur une même catégorie. Le corpus a été analysé par trois codeurs (deux étudiants en maîtrise de psychologie et moi-même) et une partie des transcriptions a fait l'objet d'une analyse commune (110 pages sur les 550). Le test de reproductibilité se base sur le calcul de la valeur du coefficient *K* (Cohen, 1960 ; Siegel & Castellan, 1988). Nous avons considéré trois intervalles pour interpréter la valeur *K* du test de reproductibilité (Carletta & al., 1996 ; Krippendorf, 1980) : $K < 0$ nulle ; $0 < K < 0.2$ très faible ; $0.2 < K < 0.4$ faible ; $0.41 < K < 0.6$ moyenne ; $0.61 < K < 0.8$ importante ; $0.81 < K < 1$ presque parfaite (voir encadré n.2 pour le calcul de *K*). Nous avons obtenu une valeur de reproductibilité moyenne pour l'ensemble des catégories relatives aux indices verbaux produits

[45] Les autres tests de validité (Carletta & al., 1996) sont :

a) *le test de stabilité*, qui est une mesure de la variance du codage intra-codeur (un même codeur tout au long du processus de codage) ;
b) *le test de précision*, qui est une mesure de variance de codage à partir d'un standard pré-établi *a priori*.

par l'apprenant (K = .58), une valeur importante pour les catégories des indices verbaux du tuteur (K =.67) et élevée pour les catégories relatives à l'argument (K = .82).

Encadré n.2 : calcul de la valeur *K*.

Nous avons limité le calcul de l'accord inter-juges aux actes de parole et aux marqueurs les plus importants : pour le tuteur, la question, l'acquiescement, la correction ; pour l'apprenant, la question et l'information. La valeur de K est calculée comme suit : K = (Po – Pc) / (1 – Pc), où Po représente les proportions réelles d'accord entre les juges et Pc représente la proportion qu'on peut attribuer au hasard. La valeur de Pc est ainsi calculée : Pc = (Pc accord positif + Pc accord négatif) / Total, où :
Pc accord positif = (les catégories positives du codeur 1 / Total catégories codées) * (les catégories positives du codeur 2 / Total catégories codées) * Total ;
Pc accord négatif = (les catégories négatives du codeur 1 / Total catégories codées) * (les catégories négatives du codeur 2 / Total catégories codées) * Total.
La valeur de Po est déterminée par : Po = Somme des valeurs d'accord sur la diagonal / Total.
Voici l'exemple relatif au codage des acquiescements a-contextuels (Aac) et contextuels (Ac):

		Codeur 2 Aac	Codeur 2 Ac	Codeur 1 Total
Codeur 1	Aac	131	34	165
Codeur 1	Ac	30	287	317
Codeur 2	Total	161	321	482

Pc accord positif = (165 / 482) * (161 / 482) * 482 = 55
Pc accord négatif = (317 / 482) * (321 / 482) * 482 = 211
Pc = (55 + 211) / 482 = 0.55
Po = (131+287) / 482 = 0.87
K = (0.87 – 0.55) / (1 – 0.55) = 0.71

4.1.7 Opérationnalisation des variables dépendantes

Les hypothèses générales que nous voulons vérifier concernent les mesures relatives aux interactions communicatives (H1-H4), au score d'apprentissage (H5) et aux appréciations des participants sur la qualité des interactions (H6).

4.1.7.1 Les indicateurs des interactions communicatives

Le volume[46] *des dialogues.* Cela comprend l'ensemble de mots, de tours de parole et d'énoncés produits par le tuteur et l'apprenant pendant chaque session expérimentale.

Les interventions du tuteur. Nous avons catégorisé les interventions du tuteur en :

a) *spontanées*, où le tuteur intervient sans que l'apprenant ne l'ait sollicité ;

b) réactives, où le tuteur intervient à la suite d'une requête que l'apprenant à exprimée au moyen du dispositif de communication textuelle (Messenger).

Les indices verbaux. D'après notre grille de codage, nous retenons les catégories suivantes :

i) les indices verbaux destinés au **processus de compréhension réciproque** :

- pour l'apprenant, les marqueurs destinés à accepter et à vérifier l'énoncé du tuteur ;

- pour le tuteur, les marqueurs consacrés à accepter et à vérifier l'énoncé de l'apprenant, ainsi que les marqueurs pour vérifier le degré de compréhension de l'apprenant ;

ii) les indices verbaux orientés à la **co-construction du contenu** :

- pour l'apprenant, les actes de parole consacrés à demander de l'aide au tuteur et pour le renseigner sur son état d'avancement ;

- pour le tuteur, les actes de parole pour se renseigner sur l'état d'avancement de l'apprenant, les actes de parole consacrés à l'aider et à l'encourager.

4.1.7.2 Les indicateurs concernant les impressions subjectives des participants

Nous avons également voulu prendre en considération l'aspect subjectif de l'interaction entre les partenaires. Ainsi, à la fin de chaque session expérimentale, le tuteur et les deux apprenants sont invités, au moyen d'un questionnaire, à évaluer leurs contributions. Nous inspirant de la notion de « conscience interpersonnelle » (McCarthy & Monk, 1994 ; Watts & al., 1996), qui suggère que les environnements de communication induisent les participants à

[46] Ce volume correspond, en moyenne, à 15 minutes de dialogue effectif sur la durée de 35 minutes de chaque session.

ressentir le niveau de présence sociale du partenaire distant et le niveau d'attention réciproque, nous avons ajouté les dimensions de la clarté et de la pertinence des interactions communicatives. En plus, nous avons décliné ces dimensions selon deux directions : a) l'évaluation que chaque participant porte sur la contribution de son partenaire (impression centrifuge) ; b) la contribution que chaque sujet porte sur sa propre contribution (impression centripète).

a) Les impressions sur la contribution d'autrui (impressions centrifuges).

La contribution d'autrui est mesurée à l'aide de quatre questions sur une échelle allant de 1 (réponse à valence négative) à 7 (réponse à valence positive). Pour le tuteur, les questions concernent la clarté des questions et des explications formulées par l'apprenant, son niveau d'attention et la pertinence de ses réponses (Tableau 4-4).

Items	**Valeurs**
Jugez-vous que l'apprenant a posé des questions	1. confuses – 7. Claires
Estimez-vous qu'il a su s'expliquer de manière	1. confuse – 7. Claire
Estimez-vous qu'il a été attentif à vos indications	1. pas du tout – 7. très attentif
Estimez-vous qu'il a répondu de manière	1. pas du tout – 7. très pertinente

Tableau 4-4 : Les items relatifs à l'appréciation du tuteur sur la qualité de la contribution de l'apprenant

De manière complémentaire, l'apprenant est invité à évaluer le degré de clarté des questions et des explications proposées par le tuteur, son niveau d'attention manifesté pendant l'interaction et la pertinence des suggestions offertes (Tableau 4-5).

Items	**Valeurs**
Jugez-vous que votre tuteur a posé des questions	1. confuses – 7. claires
Estimez-vous que ses explications étaient	1. confuses – 7. claires
Estimez-vous qu'il a été attentif pendant le dialogue	1. pas du tout – 7. très attentif
Estimez-vous que ses suggestions ont été	1. pas du tout – 7. très pertinentes

Tableau 4-5 : Les items relatifs à l'appréciation de l'apprenant sur la qualité de la contribution du tuteur

b) Les impressions sur sa propre contribution (impressions centripètes).

Nous avons mesuré les impressions de chaque sujet sur sa propre contribution grâce à quatre questions en échelle de Likert en 7 points, avec 1 comme réponse négative et 7 comme réponse positive. Le tuteur est invité à évaluer la clarté de ses propres questions et explications, le niveau de son degré d'attention et la pertinence de ses explications (Tableau 4-6).

Items	Valeurs
Estimez-vous avoir posé des questions de manière	1. confuse – 7. claire
Estimez-vous avoir fourni des explications	1. confuses – 7. claires
Pendant son discours vous avez été attentif	1. pas du tout – 7. très attentif
Estimez-vous avoir fourni des explications	1. pas du tout – 7. très pertinentes

Tableau 4-6 : Les items relatifs à l'appréciation du tuteur sur la qualité de sa propre contribution.

De même, l'apprenant est invité à évaluer la clarté de ses propres questions et des explications qu'il a proposées au tuteur, l'intensité de son propre niveau d'attention et la pertinence de ses explications (Tableau 4-7).

Items	Valeurs
Estimez-vous avoir posé des questions de manière	1. confuse – 7. claire
Estimez-vous avoir su exposer vos difficultés de manière	1. confuse – 7. claire
Pendant ses explications, vous avez été attentif	1. pas du tout – 7. très attentif
Estimez-vous avoir posé des questions	1. pas du tout – 7. très pertinentes

Tableau 4-7 : Les items relatifs à l'appréciation de l'apprenant sur la qualité de sa propre contribution.

4.1.7.3 Les indicateurs du score de performance

Nous avons considéré deux mesures pour évaluer la performance de réalisation de la tâche. La première mesure nous permet d'avoir une estimation de ce que l'apprenant a retenu des notions théoriques présentées sur la notation html, alors qu'avec la deuxième nous avons une évaluation du travail qu'il a réalisé grâce aux règles apprises sur la notation html.

a) Le pré-test & le post-test.

Nous avons mesuré le gain d'apprentissage des apprenants : les mêmes questions du pré-test ont été soumises au sujet-apprenant à la fin de chaque séance expérimentale (post-test).

b) Les traces du travail des apprenants.

A la fin de chaque session, le fichier de travail créé par chaque apprenant est enregistré et conservé. Chaque fichier est noté suivant la grille suivante (score maximum possible est de 186, étant le nombre de balises n = 62) :

0 point pour une balise manquante ;

1 point pour une balise dont la syntaxe n'est pas correcte mais appliquée aux éléments pertinents du texte (e.g., une balise non-fermée) ;

2 points pour une balise correcte mais appliquée à une partie non demandée du texte (e.g., une balise qui met en noir toute une phrase et non un seul mot) ;

3 points pour une balise correcte et appliquée aux éléments pertinents du texte.

4.2 Hypothèses opérationnelles

Nous détaillons nos hypothèses opérationnelles, relatives aux interactions langagières (H1-H4), au score de réalisation de la tâche (H5) et aux impressions subjectives des participants (H6).

(H1) Le volume des échanges

Nous supposons que lorsque les locuteurs distants se voient, le dialogue est plus interactif, c'est-à-dire que le volume de leurs échanges est plus important et les tours de parole sont plus courts que lorsqu'ils ne se voient pas. Ainsi, nous nous attendons à ce que :

H1a : les mots, les tours de parole et les énoncés soient plus nombreux en condition «audio & image-vidéo » plutôt qu'en condition « audio seul ».

H1b : les tours de parole soient plus courts en condition « audio & image-vidéo » plutôt qu'en condition « audio seul ».

(H2) Les interventions spontanées

Nous supposons qu'en condition audio-vidéo, le tuteur peut observer le visage de l'apprenant, et donc il peut déceler son niveau de concentration, sa motivation et son intérêt. Ainsi, le tuteur peut suivre l'activité de l'apprenant et intervenir auprès de lui sans attendre qu'il l'interpelle et il lui demande de l'aide. Nous supposons donc que le tuteur interviendra davantage de manière spontanée lorsque le visage de l'apprenant est visible :

H2 : le nombre d'interventions spontanées du tuteur est plus important en condition « audio & image-vidéo » qu'en condition « audio seul ».

(H3) Le processus de compréhension réciproque

La littérature suggère que les indices non-verbaux relatifs au partenaire distant favorisent le processus d'intercompréhension. Ainsi nous supposons que *sans* l'image-vidéo:

- l'apprenant, pour accepter l'énoncé du tuteur, utiliserait un nombre plus élevé de marqueurs d'acceptation et, pour vérifier d'avoir compris l'énoncé du tuteur, il utiliserait un nombre plus élevé de marqueurs de vérification ;
- le tuteur, pour accepter l'énoncé de l'apprenant, produirait un nombre plus important de marqueurs d'acceptation et, pour vérifier le degré de compréhension de l'apprenant, il produirait un nombre plus important de marqueurs de vérification ;

Plus précisément, nous nous attendons à ce que :

- L'apprenant produise plus de marqueurs consacrés à accepter (H3a) et à vérifier (H3b) l'énoncé du tuteur en condition « audio seul » qu'en condition « audio & image-vidéo ».

- Le tuteur produise plus de marqueurs destinés à accepter (H3c) et à vérifier (H3d) l'énoncé de l'apprenant en condition « audio seul » qu'en condition « audio & image-vidéo ».

(H4) Le contenu du dialogue tutoriel

Comme les indices non-verbaux favorisent le processus de compréhension mutuelle, nous supposons qu'en condition « audio & image-vidéo » le dialogue tutoriel sera plutôt orienté aux éléments concernant la tâche (objectif, état d'avancement) et à sa réalisation (demande d'aide de l'apprenant, réponses et encouragement du tuteur). En d'autres termes, ils utiliseraient davantage des actes de parole intrinsèques au dialogue tutoriel, focalisant ainsi leur

attention sur les contenus spécifiques de la tâche à réaliser, des commandes du logiciel à utiliser, etc. Ainsi, nous nous attendons à ce que :

- L'apprenant produise plus d'actes de parole consacrés à demander de l'aide au tuteur (H4a) et à informer le tuteur sur son état d'avancement (H4b) en condition « audio & image-vidéo » qu'en « audio seul ».

- Les actes de parole du tuteur destinés à connaître l'état d'avancement de l'apprenant (H4c), à aider (H4d) et à encourager l'apprenant (H4e) soient plus nombreux en condition « audio & image-vidéo » qu'en condition « audio seul ».

(H5) Le score de performance des apprenants

D'après la littérature relative au dialogue tutorial en « face à face », les regards échangés entre un enseignant et l'apprenants facilitent la performance de réalisation de la tâche des apprenants (§ ch.1). Nous supposons donc que lorsque le tuteur et l'apprenant peuvent se voir l'un l'autre, le score obtenu par les apprenants est plus élevé. Nous nous attendons à ce que :

H5 : le score de performance de l'apprenant réalisé en condition « audio & image-vidéo » soit plus élevé du score en condition « audio seul ».

(H6) L'évaluation subjective des participants

Nous nous intéressons à l'influence de l'image-vidéo du partenaire distant sur les impressions des participants distants, notamment leurs impressions sur la qualité de leur contribution et sur le degré d'attention de leur partenaire. Nous appuyant sur la théorie de la présence sociale, nous supposons que l'image du partenaire distant induise une évaluation plus positive de la qualité de l'interaction. Nous opérationnalisons cette hypothèse selon deux directions : l'appréciation que chaque participant donne sur la contribution de son partenaire (*impression centrifuge*) et l'appréciation que chaque participant donne sur sa propre contribution (*impression centripète*).

L'évaluation de la contribution d'autrui (impression centrifuge)

D'après la littérature, les indices non-verbaux aident les locuteurs à rendre moins ambigus et plus clairs leurs échanges verbaux. Nous supposons donc qu'en condition « avec

image-vidéo » le tuteur et l'apprenant puissent apprécier davantage la qualité de leurs contributions :

H6a : le tuteur juge de manière plus positive la contribution de l'apprenant en condition « audio & image-vidéo » qu'en condition « audio seul ».

H6b : l'apprenant juge de manière plus positive la contribution du tuteur en condition « audio & image-vidéo » qu'en condition « audio seul ».

L'évaluation de sa propre contribution (impression centripète)

L'image d'autrui favorise le processus de compréhension réciproque. Nous supposons donc que les locuteurs évaluent plus positivement leur interaction lorsqu'ils peuvent se voir que lorsqu'ils ne le peuvent pas :

H6c : le tuteur juge sa propre contribution meilleure en condition « audio & image-vidéo» qu'en condition « audio seul ».

H6d : l'apprenant juge sa propre contribution meilleure en condition « audio & image-vidéo» qu'en condition « audio seul ».

4.3 Les résultats

Nous présentons les résultats suivant l'ordre de présentation de nos hypothèses. Pour chaque hypothèse, nous avons vérifié leurs conditions d'application et nous avons indiqué les cas de non-respect, afin de justifier nos choix pour les tests de statistique. Pour chaque test statistique, suivant les conventions APA, nous avons calculé la valeur de la taille de l'effet[47] afin de préciser l'amplitude de la différence entre les deux conditions expérimentales.

4.3.1 Les effets de l'image-vidéo sur le volume des dialogues (H1)

Nous devons vérifier que (H1a) le volume des dialogues est plus important en condition « avec image-vidéo » qu'en condition « audio seul ». D'après les résultats, aucune différence

[47] Pour le calcul de la taille de l'effet dans la comparaison d'échantillons indépendants (test *t* de Student), nous avons utilisé la formule d = (M1 - M2) / [(SD1+SD2) / 2]. Pour les analyses univariées,

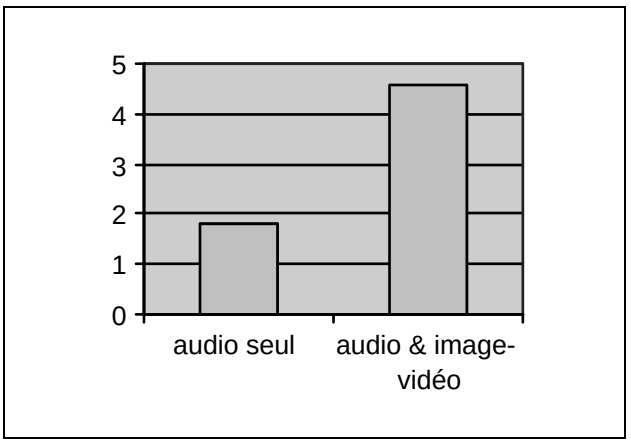
significative n'est constatée entre les deux conditions expérimentales, ni pour le nombre de mots (M = 1784.7 (586.8) *vs.* 1513.3 (444.4), t(22)= 1.277; n.s., *d* = 0.52), ni pour le nombre d'énoncés (M = 326.6 (123) *vs.* 285.4 (114.9), t(22) = 0.847; n.s., *d* = 0.34), ni pour le nombre de tours de parole (M = 210.1 (67) *vs.* 202.7 (105.2) ; t(22) = 0.2; n.s., *d* = 0.08). L'hypothèse H1a, affirmant que le volume des dialogues est plus important en condition « avec image-vidéo » qu'en « audio seul », ne peut pas être retenue.

Nous devons vérifier que (H1b) les tours de parole sont plus courts en condition « avec image-vidéo » qu'en condition « audio seul ». Aucune différence significative du ratio « nombre de mots / nombre de tours de parole » n'est constatée entre les deux conditions expérimentales (M = 8.9 (3.7) *vs.* 9.3 (4.2), t(22) =0.246 ; n.s., *d* = 0.1). L'hypothèse H1b n'est pas donc confirmée.

4.3.2 Les effets de l'image-vidéo sur les interventions du tuteur (H2).

Nous devons vérifier que les interventions spontanées du tuteur sont plus nombreuses en condition « audio & avec image-vidéo » qu'en condition « audio seul ». D'après les résultats (Graphique 4-1), le nombre d'interventions spontanées est significativement plus élevé en condition « audio & image-vidéo » qu'en condition « audio seul » (M = 4.6 (2) *vs.* 1.8 (.9), t(22)= 4.27 ; p <.001, d = 1.8). L'hypothèse H2 est donc validée : en condition « image-vidéo », le tuteur intervient plus souvent de manière spontanée.

nous avons calculé la mesure η^2 = SC effet / (SC effet + SC erreur) (où SC est la somme des carrés) (Cohen, 1988).



Graphique 4-1 : Moyennes du nombre d'interventions spontanées du tuteur

4.3.3 Les effets de l'image-vidéo sur le processus de compréhension mutuelle (H3)

Nous supposons que l'image-vidéo réduit la production de marqueurs linguistiques nécessaires au processus de compréhension réciproque.

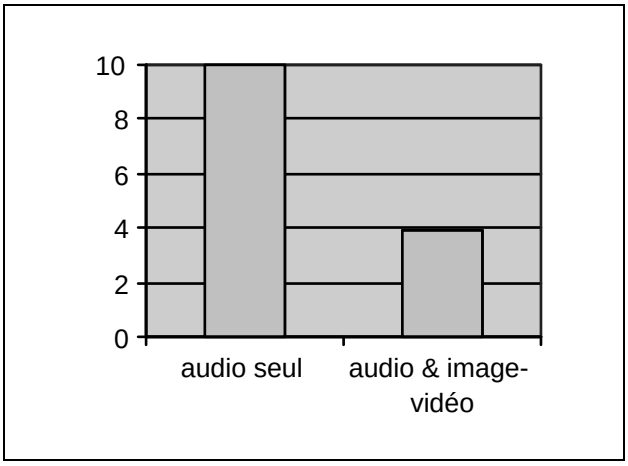
(H3a) Cette hypothèse affirme que l'apprenant utilise un nombre plus important de marqueurs verbaux pour accepter l'énoncé du tuteur en condition « audio seul » qu'en condition « audio & image-vidéo ». D'après les résultats, le nombre d'acquiescements est équivalent dans les deux conditions expérimentales (M = 37.2 (17.2) *vs*. 37.9 (20.5), $t(22) = .09$; n.s., $d = 0.08$). L'hypothèse H3a est donc rejetée.

(H3b) Nous devons vérifier que l'apprenant produit plus de marqueurs pour vérifier l'énoncé du tuteur en condition « audio seul » qu'en condition « audio & image-vidéo ». Les résultats indiquent que le nombre de marqueurs de vérification est équivalent dans les deux conditions (M = 15.9 (7.4) *vs*. 16.2 (8.8), $t(22) = .09$; n.s., $d = 0.08$). L'hypothèse H3b n'est pas validée.

(H3c) Nous avons formulé l'hypothèse que les marqueurs verbaux du tuteur dédiés à l'acceptation de l'énoncé de l'apprenant sont plus nombreux en condition « audio seul » qu'en condition « audio & image-vidéo ». Les résultats indiquent que les acquiescements ne diffèrent

pas significativement dans les deux conditions expérimentales (M = 21.3 (23.5) *vs*. 12.9 (8.6), t(22) = 1.16 ; n.s., *d* = 0.47). L'hypothèse H3c est donc invalidée.

(H3d) Cette hypothèse affirme que, pour vérifier la compréhension de l'apprenant, le tuteur produit un nombre plus important de marqueurs en condition « audio seul » qu'en condition « audio & image-vidéo ». D'après l'analyse des résultats (Graphique 4-2), cette hypothèse est confirmée. En effet, le nombre de marqueurs produit par tuteur pour vérifie la compréhension de l'apprenant est plus important en condition « audio seul » qu'en condition « audio & image-vidéo » (M = 10 (6.4) *vs*. 3.9 (3.6), *t*(22) = 2.85 ; *p* <. 01, *d* = 1.17). L'hypothèse H3d est donc validée.



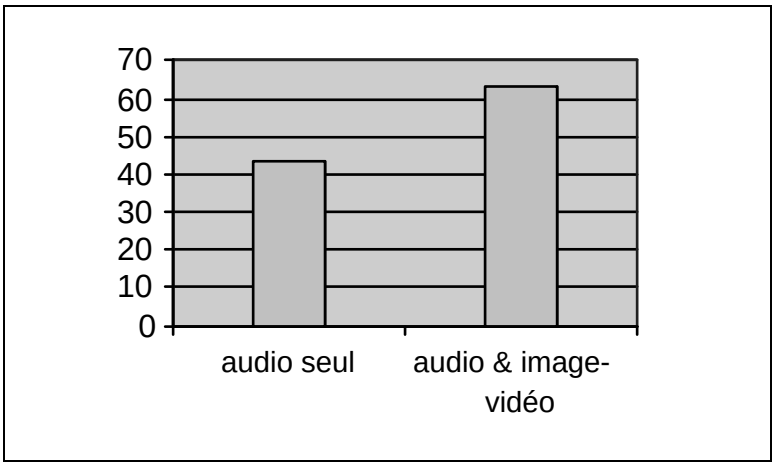
Graphique 4-2 : Moyennes du nombre de marqueurs verbaux du tuteur consacrés à vérifier la compréhension de l'apprenant.

4.3.4 Les effets de l'image-vidéo sur le contenu du dialogue tutoriel (H4)

Nous vérifions les hypothèses concernant les effets de l'image-vidéo du partenaire distant sur les actes de parole produits par l'apprenant (H4a, b) et par le tuteur (H4c, d, e).

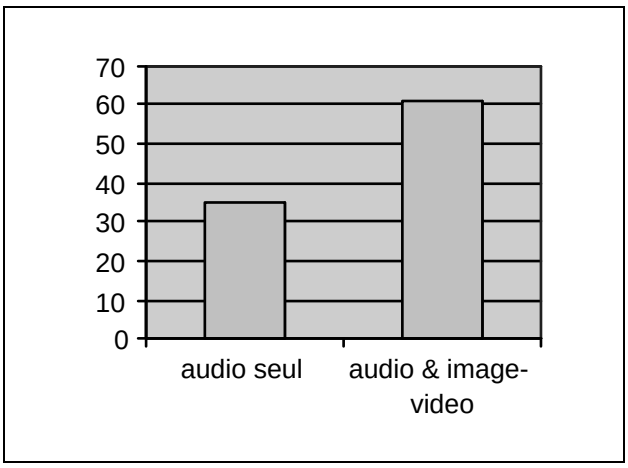
(H4a) Nous devons vérifier que les actes de parole de l'apprenant consacrés à questionner le tuteur sont plus nombreux en condition « audio & image-vidéo » qu'en condition « audio seul ». Aucune différence significative n'est constatée entre les deux conditions expérimentales (M = 26.2 (19.9) *vs*. 20.5 (12.5), t(22) = .846 ; n.s., *d* = 0.34). L'hypothèse H4a n'est pas validée.

(H4b) Nous devons vérifier que les actes de parole de l'apprenant consacrés à informer le tuteur de l'état d'avancement sont plus nombreux en condition « audio & image-vidéo » plutôt qu'en condition « audio seul ». L'analyse révèle (Graphique 4-3) que les actes de parole de l'apprenant consacrés à informer le tuteur sur son état d'avancement sont plus nombreux en condition « audio & image-vidéo » qu'en condition « audio seul » (M = 63.6 (28.4) *vs*. 43.8 (14.4), t(22) = 2.15 ; $p <$. 05, d = 0.88). L'hypothèse H4b est validée.



Graphique 4-3 : Moyennes du nombre d'actes de parole de l'apprenant destinés à informer le tuteur sur l'état d'avancement.

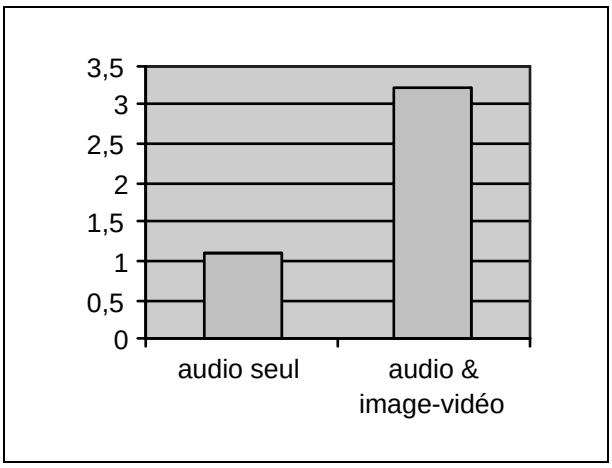
(H4c) Nous devons également vérifier que les actes de parole du tuteur consacrés à connaître l'état d'avancement de l'apprenant sont plus nombreux en condition « audio & image-vidéo » qu'en condition « audio seul ». D'après les résultats (Graphique 4-4), le nombre d'actes de parole que le tuteur consacre à la connaissance de l'état d'avancement de l'apprenant sont plus nombreux en condition « audio & image-vidéo » qu'en condition « audio seul » (M=60.7 (33.2) *vs*. 35.3 (8.1), t(22) = 4.27 ; $p <$. 02, d = 1.05). L'hypothèse H4c est validée.



Graphique 4-4 : Moyennes du nombre d'actes de parole du tuteur consacrés à connaître l'avancement de l'apprenant.

(H4d) Nous avons émis l'hypothèse que les actes de parole du tuteur dédiés à l'aide à l'apprenant sont plus nombreux en condition « audio & image-vidéo » qu'en condition «audio seul ». Les résultats indiquent que le nombre des actes de parole de type aide informative (M = 90.2 (43.5) *vs*. 77.9 (38.5), t(22) = .736 ; n.s., *d* =0.37), aide explicative (M = 9.1 (7.6) *vs*. 11.8 (7.1), t(22) = .915 ; n.s., *d* = 0.37), aide corrective (M = 4.2 (5.7) *vs*. 2.7 (3.5), t(22) = .777 ; n.s., *d* = 0.32) et évaluative (M = 1.5 (1.9) *vs*. 1.8 (1.5), t(22) = .449 ; n.s., *d* = 0.17) est équivalent dans les deux conditions. L'hypothèse H4d est donc rejetée.

(H4e) Nous avons émis l'hypothèse que les actes de parole du tuteur consacrés à encourager l'apprenant sont plus nombreux en condition « audio & image-vidéo » qu'en condition « audio seul ». L'analyse révèle que le tuteur produit un nombre plus important d'actes de parole destinés à encourager l'apprenant lorsqu'il peut voir son visage (M = 3.2 (2.3) *vs.* 1.1 (1.2), t(16.3) = 2.9 ; $p < .01$, *d* = 1.14) que lorsqu'il ne le peut pas (Graphique 4-5). L'hypothèse H4e, validée, affirme que lorsque l'image-vidéo du partenaire distant est disponible, le tuteur encourage davantage l'apprenant.

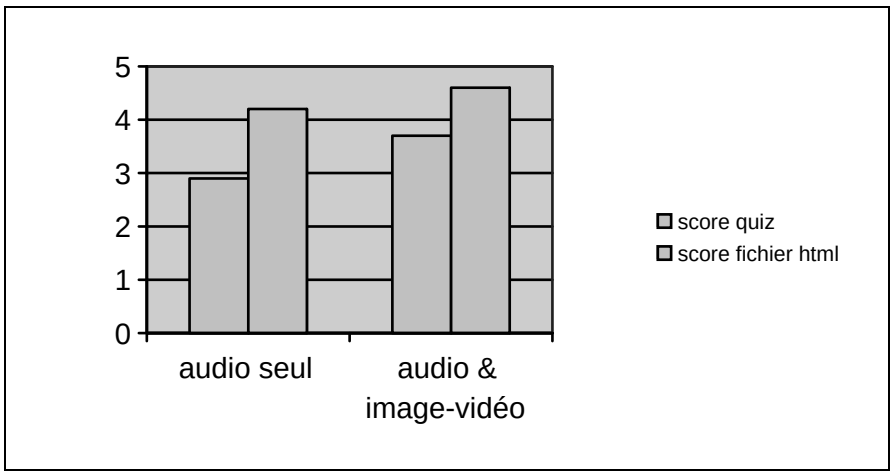


Graphique 4-5 : Moyennes du nombre d'actes de parole du tuteur consacrés à encourager l'apprenant.

4.3.5 Les effets de l'image-vidéo sur le score de performance (H5)

Nous devons vérifier que le score de l'apprenant en condition « audio & image-vidéo » est plus élevé que le score réalisé en condition « audio seul ». Les deux scores[48] montrent deux tendances différentes (Graphique 4-6). Le score de performance au test est significativement plus élevé lorsque le tuteur et les apprenants peuvent se voir que lorsqu'ils ne le peuvent pas (M = 2.9 (1.2) *vs*. 3.7 (1), t(46) = 2.53 ; p <.01, d = 0.44). En revanche, aucune différence signicative entre les deux conditions n'est constatée pour le score relatif à la qualité du fichier html produit par les apprenants (M = 4.2 (1.4) *vs*. 4.6 (1.5), t(46) = .83 ; n.s., d =1.1). Il en résulte ainsi que la condition « audio & image-vidéo » améliore l'apprentissage des notions relatives au cours html, alors qu'elle ne facilite pas leur mise en œuvre : l'hypothèse H5 n'est pas entièrement validée.

[48] Afin de pouvoir mieux interpréter les résultats, nous avons transformé la valeur des deux scores : le score des réponses au questionnaire est ramené à un note sur 10 points, le score obtenu dans la réalisation de la page html est ramené à une note sur 20 points.



Graphique 4-6 : Moyennes des scores de performance des apprenants.

4.3.6 Les effets de l'image-vidéo sur l'évaluation subjective des contributions (H6).

Nous avons formulé l'hypothèse que l'image du partenaire distant améliore l'évaluation que les participants font de leur interaction communicative. Nous vérifions cette hypothèse, en présentant les résultats relatifs d'abord aux impressions centrifuges (évaluation sur la contribution d'autrui) puis aux impressions centripètes (évaluation sur sa propre contribution).

** Les effets de l'image-vidéo sur les impressions centrifuges*.

Afin de décider le traitement statistique le plus adéquat, nous avons réalisé des analyses préliminaires afin de vérifier la consistance interne des items composant chaque dimension.

Pour les réponses des tuteurs, la valeur du coefficient d'homogénéité (alpha de Cronbach) des réponses est satisfaisante (α=.82), nous permettant de regrouper les quatre items en un seul indice (score composite relatif aux impressions centrifuges). La matrice des corrélations inter-items (Tableau 4-8) montre que les items sont corrélés, à l'exception de items relatifs à la clarté des questions et de l'état d'attention de l'apprenant.

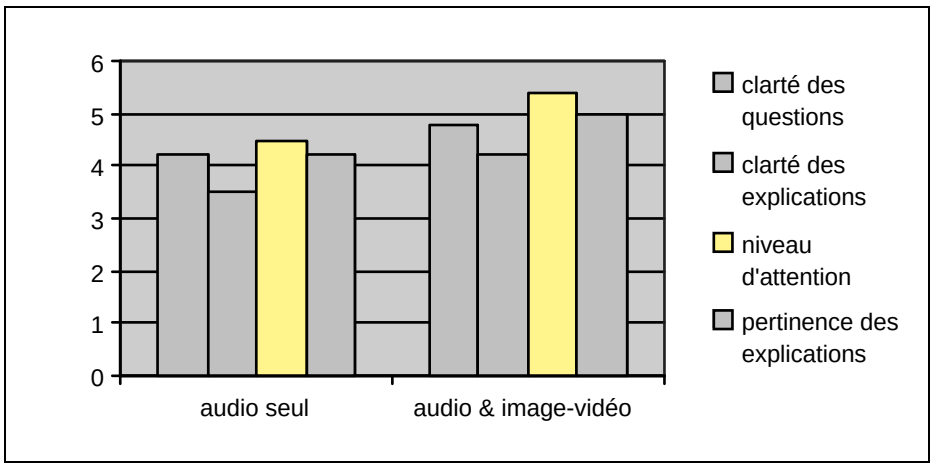
Pour les réponses des apprenants, la valeur du coefficient d'homogénéité est très faible (α=.45) et chaque item sera traité individuellement.

	L'apprenant a su s'expliquer de manière (1.confuse-7.claire)	L'apprenant a été attentif à vos indications (1. pas du tout – 7. très attentif)	L'apprenant a répondu de manière (1.pas du tout - 7.très pertinente)
L'apprenant a posé des questions 1. confuses- 7.claires	.50**	.21	.43*
L'apprenant a su s'expliquer de manière 1.confuse-7.claire	-	.83**	.67**
L'apprenant a été attentif à vos indications 1.pas du tout - 7.très attentif		-	.59**

Tableau 4-8 : Matrice des corrélations entre les items relatifs au jugement du tuteur sur l'apprenant.
N = 24, seuils de significativité des corrélations : ** p <.01, * p <.05.

(H6a) Nous devons vérifier que le tuteur apprécie davantage la contribution de l'apprenant en condition « audio & image-vidéo » qu'en condition « audio seul ». L'analyse des réponses au test ne révèle aucune différence significative entre les deux conditions (M = 17.7 (4.7) *vs*. 19.2 (4), t(22) = .79 ; n.s., *d* = 0.34). L'hypothèse H6a, affirmant que le tuteur évalue plus positivement la contribution de l'apprenant lorsqu'ils se voient, n'est pas confirmée.

(H6b) Nous vérifions que l'apprenant évalue la contribution du tuteur de manière plus positive en condition « audio & image-vidéo » qu'en condition « audio seul ». D'après les résultats (Graphique 4-7), les apprenants jugent que les questions (M = 4.8 (1.4) *vs*. 4.2 (0.87), t(46) = 2.13 ; *p* <.04, *d* = 0.51) et les explications (M = 4.2 (0.9) *vs*. 3.5 (1.3), t(46) = 2.31 ; *p* <.02, *d* = 0.63) du tuteur sont plus claires en condition « avec image-vidéo » qu'en condition « sans image-vidéo ». De plus, les apprenants estiment que le tuteur est plus attentif (M= 5.4 (1.4) *vs*. 4.5 (1), t(46) = 2.51 ; *p* <.02, *d* = 0.74) et ses explications plus pertinentes (M= 5 (1.5) *vs*. 4.2 (1), t(46) = 2.04 ; *p* <.001, *d* = 0.63). L'hypothèse H6b est donc validée : l'apprenant juge plus positivement la contribution du tuteur lorsqu'il le voit.



Graphique 4-7 : Moyennes des réponses de l'apprenant sur la contribution du tuteur.

** Les effets de l'image-vidéo sur l'évaluation des impressions centripètes.*

D'après l'analyse préliminaire des réponses du tuteur, tous les items résultent corrélés (Tableau 4-9). La valeur du coefficient d'homogénéité est satisfaisante (α=.83) et ce résultat nous permet de traiter les données en un seul vecteur d'analyse (score composite).

	Estimez-vous avoir fourni des explications 1.confuses – 7.claires	Pendant le discours de l'apprenant vous étiez 1.pas du tout - 7.très attentif	Estimez-vous avoir fourni des explications 1.pas du tout - 7. très pertinentes
Estimez-vous avoir posé des questions 1.confuses-7.claires	.55**	.63**	.56**
Estimez-vous avoir fourni des explications 1.confuses - 7.claires	-	.51*	.77**
Pendant le discours de l'apprenant vous étiez 1.pas du tout - 7.très attentif		-	.48*

Tableau 4-9 : Matrice des corrélations inter-items des impressions centripètes des tuteurs. N = 24, seuils de significativité des corrélations : ** *p* <.01, * *p* <.05

Pour les apprenants, la valeur est faible (α=.49), ce qui induit un traitement individuel pour chaque item.

(H6c) Nous devons vérifier que le tuteur juge que sa propre contribution dialogique est plus positive en condition « audio & image-vidéo » qu'en condition « audio seul ». Les résultats ne montrent aucune différence significative entre les deux conditions expérimentales (M=16.7 (5.3) *vs*. 18.2 (4.1) ; t(22) =.81, n.s., *d* = 0.31). L'hypothèse H6c est donc invalidée.

(H6d) Nous voulons vérifier si l'apprenant évalue plus positivement sa propre contribution en condition « audio & image-vidéo » plutôt qu'en condition « audio seul ». Nous avons effectué des comparaisons de moyennes pour chaque item et aucune différence significative entre les conditions n'est constatée. L'hypothèse H6d, affirmant que l'apprenant évalue plus positivement sa propre contribution lorsque les partenaires distants se voient, est invalidée.

4.4 Discussions des résultats.

L'objectif de cette première étude était de vérifier les effets de l'image-vidéo du partenaire distant sur certains indicateurs relatifs au dialogue tutoriel. Nous avons donc comparé les effets de deux conditions expérimentales (audio seul *vs*. audio & image-vidéo du partenaire) sur les mesures relatives au volume et au contenu des dialogues, sur le score de performance des apprenants et sur les impressions subjectives des participants.

Est-ce que l'image-vidéo du partenaire distant influe sur le volume des échanges verbaux ?

La réponse est négative. En effet, ni le volume des échanges n'augmente, ni la longueur des tours de parole ne se réduit lorsque les participants distants se voient (H1). Ce résultat, qui par ailleurs confirme certaines recherches (cf. Anderson & al., 1996 ; Daly-Jones & al./1,1998; Masoodian & al., 1995), est surprenant si l'on considère que la gestion des tours de parole est généralement moins fluide lorsqu'on ne voit pas le visage de notre interlocuteur (cf. Whittaker & O'Conaill, 1997). Le type d'activité collaborative semblerait en être un facteur déterminant : en effet, dans une tâche mobilisant la compétence linguistique des sujets pour décrire des relations spatiales, le volume des dialogues est moins important lorsque les sujets se voient que lorsqu'ils ne se voient pas (cf. Doherty-Sneddon & al., 1997). Il semblerait donc que, selon le type d'activité collaborative, les interlocuteurs régulent le niveau d'interactivité, mais indépendamment du niveau de présence sociale induit par les moyens de communication (cf. les impressions subjectives centripètes et centrifuges).

Est-ce que l'image-vidéo du partenaire distant influence les interventions spontanées du tuteur ?

Les résultats montrent que les interventions spontanées du tuteur sont plus nombreuses lorsque les partenaires distants se voient que lorsqu'ils ne se voient pas (H2). Les expressions du visage et les autres indices non-verbaux de l'apprenant induisent une attitude proactive du tuteur, qui intervient davantage sans attendre une requête explicite de l'apprenant. Ce résultat, tout en appuyant la théorie de la présence sociale (Short & al., 1976), confirme ce que Fox (1993) a observé en situation de tutorat « face à face », à savoir que le tuteur intervient avant que l'apprenant n'explicite verbalement sa requête d'aide.

Est-ce que l'image-vidéo facilite le processus de compréhension réciproque ?

La facilitation du processus de compréhension réciproque se manifesterait par une diminution de marqueurs verbaux consacrés à accepter ce que le partenaire distant énonce. D'après les résultats, le tuteur (H3c) et l'apprenant (H3a) produisent un nombre de marqueurs verbaux d'acceptation équivalent dans les deux conditions de communication. Qu'ils se voient ou pas, ils manifestent leur compréhension réciproque au moyen d'un nombre équivalent de marqueurs verbaux. Ce comportement langagier, qu'on retrouve dans d'autres études (cf. Daly-Jones & al., 1998 ; Kraut & al., 2003), pourrait s'expliquer par le fait que les sujets sélectionnés pour l'expérimentation ne sont pas des usagers confirmés des modalités de communication de type « audio-vidéo conférence », et ils renforceraient l'information offerte par leurs expressions faciales par des marqueurs verbaux explicites.

Concernant les marqueurs produits par les tuteurs, il est à noter que les valeurs des écarts types des acquiescements sont assez importantes (S.D. = 23.5 *vs*. 8.6), ce qui suggère une importante différence inter-individuelle. Sur le plan de l'analyse des données, des sessions d'autoconfrontation[49] auraient permis d'expliciter les raisons de cette différence. Cependant, la

[49] Le but de la technique de l'auto-confrontation est de faire expliciter aux sujets les raisonnements et les représentations qu'ils avaient pendant la session expérimentale. La technique consiste à présenter aux sujets l'enregistrement audio-vidéo de la session expérimentale et de s'arrêter sur les passages que l'expérimentateur considère intéressants ou lorsque le sujet propose spontanément de les commenter. Malgré ses limites (cf. Cahour, 2003), cette technique permet au cherch-eur/euse de confirmer ce qu'il/elle a inféré à partir de l'observation du comportement non-verbal des sujets.

disponibilité limitée des sujets impliqués dans l'expérimentation a empêché ces analyses complémentaires.

Nous avons supposé que l'image-vidéo du partenaire distant facilite la compréhension de l'énoncé produit par le partenaire distant, en réduisant le nombre de marqueurs verbaux consacrés à la vérification. D'après les résultats (H3b), l'apprenant vérifie l'énoncé du tuteur par un nombre équivalent de marqueurs verbaux, que l'image-vidéo du tuteur soit disponible ou non. Il se peut que, à cause des défauts de la configuration technique du dispositif de communication (i.e., la parallaxe), l'apprenant préfère se rassurer sur le contenu des énoncés du tuteur, étant donné aussi qu'il doit l'intégrer progressivement dans sa représentation cognitive de la tâche.

En revanche, pour vérifier si l'apprenant a compris (H3d), le tuteur utilise un nombre inférieur de marqueurs verbaux lorsqu'il peut voir l'apprenant. Ce résultat rejoint d'autres études expérimentales (cf. Doherty-Sneddon & al./1, 1997 ; Tang & Isaacs, 1993) et confirmerait la théorie de la présence sociale. Lorsque l'image de l'apprenant distant est disponible, le tuteur peut observer ses expressions faciales et induire ainsi son état de compréhension, sans devoir toujours vérifier la compréhension de l'apprenant par des marqueurs verbaux.

Est-ce que l'image-vidéo influe sur le contenu du dialogue tutoriel ?

Les résultats montrent que l'image-vidéo du partenaire distant influe sur le contenu des échanges tuteur-apprenant. Tout d'abord, le tuteur interroge l'apprenant plus souvent sur son état d'avancement lorsqu'il peut l'observer que lorsqu'il ne peut pas (H4c), obligeant ainsi l'apprenant à lui répondre davantage (H4b). On peut supposer que les expressions faciales des apprenants évoquent des difficultés, induisant ainsi le tuteur à contrôler davantage leur état d'avancement et s'assurer de leur progression. Il s'agirait peut-être d'une sorte d'*effet accrocheur* (dont les raisons auraient dû être vérifiées et explicitées par une analyse en auto-confrontation) pour lequel les tuteurs encouragent davantage l'apprenant lorsqu'ils peuvent l'observer (H4e). En d'autres termes, le niveau plus élevé de présence sociale inciterait le tuteur à un suivi pédagogique plus constant et à renforcer son rôle tutoriel dans l'interaction (cf. Fox, 1993).

En revanche, les questions de l'apprenant et les réponses du tuteur sont du même ordre de grandeur dans les deux conditions expérimentales (H4a & H4d). Compte tenu que les énoncés produits par le tuteur sont significativement plus nombreux que ceux produits par l'apprenant, ce résultat confirme les observations conduites par Fox (1993) sur le tutorat « face à face ». En effet, comme pour le dialogue tutoriel en « face à face », l'interaction médiatisée tuteur-apprenant demeure asymétrique, c'est-à-dire que le tuteur gère de manière plutôt individuelle le rythme du dialogue et le choix des contenus.

Est-ce que l'image-vidéo du partenaire distant améliore la performance de réalisation de la tâche ?

Les résultats relatifs à la performance de réalisation de la tâche (H5) montrent deux tendances divergentes. D'une part, le score aux réponses du post-test est significativement plus élévé lorsque le tuteur et les apprenants se voient ; d'autre part, le score de réalisation du fichier html est du même ordre de grandeur quelle que soit la modalité de communication. Le test concernait les contenus théoriques (format papier) : il se peut que les explications du tuteur sur les balises aient été mieux mémorisées en condition « audio & image-vidéo » plutôt qu'en condition « audio seul ». Cela semblerait être cohérent avec ce que Fry & Smith (1975) observent en situation d'enseignement présentiel, c'est-à-dire que les échanges des regards entre l'enseignant et les apprenants facilitent la mémorisation du contenu à apprendre. Il faut cependant nuancer cette interprétation par le fait que, dans notre procédure expérimentale, nous demandons aux sujets-apprenants de lire le cours sur papier et ensuite de demander de l'aide au tuteur.

Le score de réalisation du fichier html est du même ordre de grandeur dans les deux conditions expérimentales, ce qui confirme la littérature sur les tâches de type procédural (cf. Anderson & al., 1996 ; Daly-Jones & al., 1998 ; Masoodian & al., 1995), à savoir que voir autrui n'améliore pas la réalisation de la tâche.

Enfin, si on considère l'efficacité de la condition de communication (calcul du ratio entre le volume des échanges verbaux et le score de réalisation de la tâche), les résultats suggèrent que, pour une tâche d'assistance tutorielle, la modalité de communication « audio & image-vidéo » est aussi efficace que la modalité « audio seul ».

Est-ce que l'image-vidéo du partenaire distant influe sur l'évaluation des interactions ?

Les résultats concernant les impressions centrifuges confirment l'asymétrie des rôles du tuteur et de l'apprenant. D'une part, quelle que soit la modalité de communication, les appréciations du tuteur sur la contribution de l'apprenant ne diffèrent pas (H6a). D'autre part, lorsqu'ils peuvent se voir, l'apprenant estime que la contribution du tuteur est plus pertinente, plus claire et que le tuteur est globalement plus attentif (H6b). Il semblerait donc qu'entre des sujets distants possédant des niveaux de compétences différents, le sujet moins compétent évalue de manière plus positive la contribution du sujet plus compétent. Une explication possible de ce dernier résultat est que l'image du tuteur pendant l'interaction lui permettrait de rendre moins ambiguë la perception de nouveaux mots dont la compréhension n'est pas immédiate (cf. McGurk & MacDonald, 1976). En revanche, quel que soit leur niveau de compétence, les sujets évaluent leur propre contribution positivement de manière équivalente dans les deux conditions expérimentales (H6c & H6d).

En résumé, du point de vue théorique ces résultats ne soutiennent que de manière mitigée les théories de la présence et de la richesse des média. En effet, l'image-vidéo du partenaire distant ne favorise que la compréhension réciproque entre le tuteur et l'apprenant et n'améliore que légèrement le score d'apprentissage des apprenants. De plus, seuls les apprenants estiment que l'image-vidéo rend le contenu de l'interaction plus claire et plus pertinente. Au regard des applications possibles de ces résultats dans la conception de plates-formes dédiées au tutorat à distance, nous suggérons que si on veut privilégier la compréhension réciproque et l'apprentissage, l'image-vidéo apporte peu d'avantages mais significatifs.

Est-il avantageux pour le dialogue tutoriel d'utiliser des indices non-verbaux kinésiques-inférentiels ? Notre deuxième étude expérimentale veut répondre à la question de recherche suivante : *Quels sont les effets d'un espace visuel commun, permettant au tuteur d'observer ce que l'apprenant fait mais sans agir à sa place, sur l'interaction tutorielle distante ?*

Chapitre 5 - Etude II : les indices non-verbaux relatifs à l'espace visuel commun.

Cette deuxième expérimentation est consacrée à comprendre l'impact de l'espace visuel commun, véhiculant les indices non-verbaux de type ostensif-inférentiel, sur l'interaction médiatisée tuteur-apprenant. Plus précisément, nous voulons vérifier que lorsque l'espace visuel commun est disponible :

- le volume global des échanges verbaux diminue ;
- le tuteur intervient davantage auprès de l'apprenant de manière spontanée ;
- le nombre d'indices verbaux consacrés au processus de compréhension diminue ;
- le nombre d'indices verbaux consacrés à la co-construction du contenu augmente, le dialogue tutoriel se focalisant surtout sur les contenus à apprendre;
- le niveau de réalisation de la tâche des apprenants est plus élevé ;
- l'évaluation par les participants de la qualité des interactions communicatives est plus positive.

5.1 Méthodologie

La tâche expérimentale proposée aux participants est la même de la première expérimentation, ainsi que le matériel relatif au cours html et la procédure expérimentale.

5.1.1 La population

Nous avons recruté 72 étudiants en psychologie (de nouveaux sujets = 48, et 24 appartiennent à l'étude n.1). Leur âge moyen est de 23.8 ans (S.D.=5.1), avec 70% de femmes, tous francophones et novices en programmation et en notation html. Afin de construire des groupes expérimentaux homogènes, les sujets ont répondu à un pré-test (les mêmes questions que l'étude 1).

Les tuteurs recrutés sont au nombre de 12 (six hommes et six femmes, d'âge moyen 30.9 ans (S.D. = 7.7)) ont été recrutés (2 maîtres de conférences, 4 doctorants-moniteurs de 3ème année, 2 doctorants-moniteurs de 2ème année et 4 doctorants-moniteurs de 1ère année). La moitié d'eux (âge M = 35.5, S.D. = 8) a une expérience confirmée dans l'enseignement supérieur (supérieure à 2 ans), alors que l'autre moitié a peu d'expérience de l'enseignement (âge M = 25.5, S.D. = 2.2). Leur formation de base est l'informatique ; tous maîtrisent plusieurs langages de programmation ainsi que les techniques de conception de pages en HTML. Pour sa participation à l'expérimentation, chaque tuteur perçoit une rémunération de 25€.

5.1.2 Le dispositif technique

Le dispositif technique de la première expérimentation est enrichi d'une nouvelle fonctionnalité. En fait, le tuteur dispose d'un logiciel d'observation à distance, Vnc Viewer©, lui permettant de visualiser le travail de chaque apprenant dans une fenêtre d'écran différente. Chaque action produite par le tuteur et les apprenants est enregistrée à l'aide du logiciel Snag-It© sous un format vidéo compressé.

5.1.3 Les conditions expérimentales

De la première étude, nous avons conservé la condition « audio & image-vidéo du partenaire » afin de pouvoir comparer l'apport des indices non-verbaux kinésiques (véhiculés par l'image-vidéo du partenaire distant) et l'apport des indices non-verbaux ostensifs-inférentiels (véhiculés par l'espace visuel commun). Dans chaque condition, le tuteur peut parler à un seul apprenant à la fois, grâce à un dispositif de sélection.

Nous avons ainsi construit les conditions expérimentales suivantes :

c1) « *audio & image-vidéo du partenaire distant* » : le tuteur et l'apprenant peuvent communiquer oralement et se voir ;

c2) « *audio & espace visuel commun* » : le tuteur et l'apprenant ne peuvent que communiquer oralement et le tuteur peut observer, à tout instant sur son propre écran, ce que chaque apprenant fait dans son espace de travail ;

c3) « *audio & image-vidéo & espace visuel commun* » : le tuteur et l'apprenant peuvent se voir, communiquer oralement et le tuteur peut observer l'espace de travail des deux apprenants (Figure 5-1).

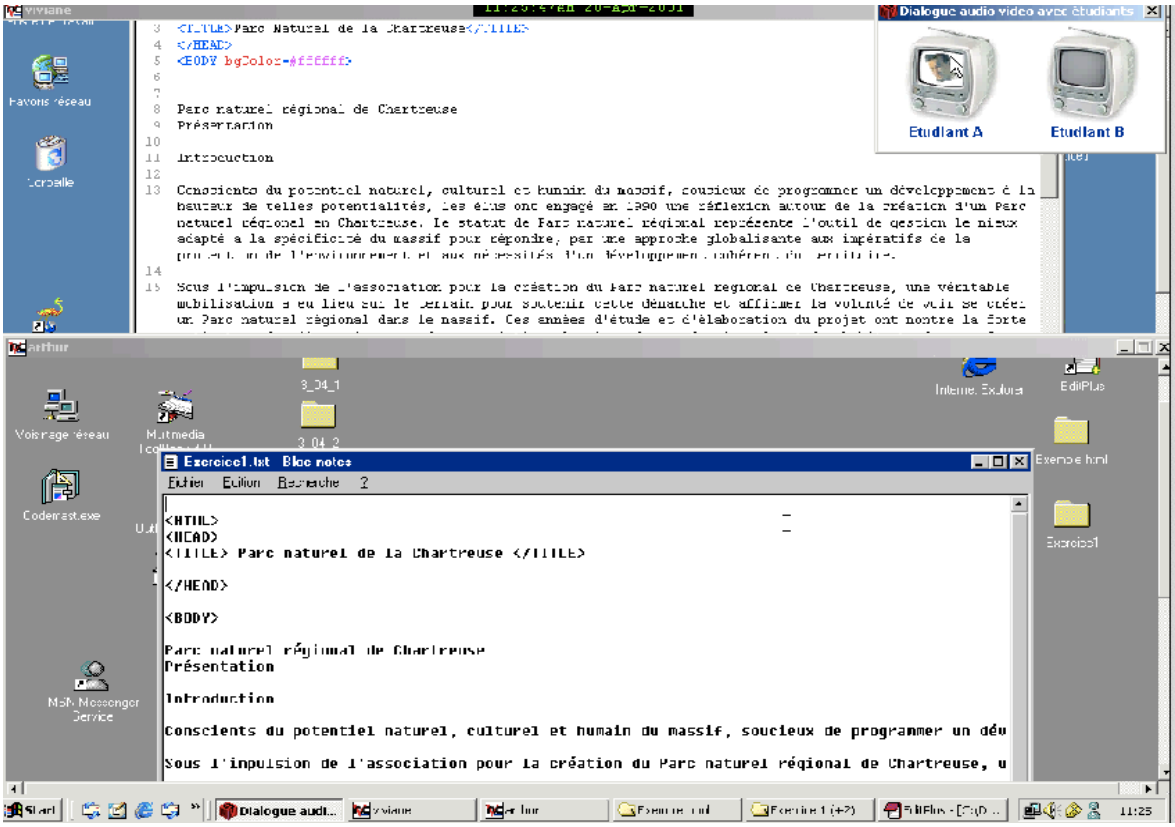


Figure 5-1 : Image relative à l'écran de l'ordinateur du tuteur : la partie supérieure est réservée à l'apprenant A, la partie inférieure à l'apprenant B.

Nous avons mis en place un plan expérimental de type intra-sujets : chaque tuteur accomplit la tâche d'assistance dans chacune de trois conditions. L'ordre de passation des sujets-enseignants a été contrôlé selon une procédure aléatoire. Les apprenants ont été repartis dans les trois conditions expérimentales : 24 apprenants pour chaque condition, répartis de manière équivalente par rapport à leur âge ($F(2, 69) = 1.73$; n.s.), à leurs réponses au pré-test

($F(2, 69) = 0.16$; n.s.) et à leur niveau d'étude (3ème et 4ème année de psychologie). Le plan expérimental résultant pour les tuteurs est le suivant $\underline{\mathbf{T}_{12}}$ * $\mathbf{C}_3$ **,** où douze tuteurs accomplissent la tâche dans chaque condition expérimentale.

5.1.4 La grille de codage

Nous avons utilisé la structure de découpage et la grille de codage de l'étude précédente, structurée selon les niveaux suivants : séquence → tours de parole → énoncés → <actes de parole, argument>. Concernant la catégorie *argument*, qui est une spécification de l'acte de parole, nous avons aussi codé les référents implicites, c'est-à-dire les expressions déictiques (e.g. : A « *je mets l'icone **en bas*** ? ») et les pronoms déictiques (e.g. : A « *je le mets **ici*** ? »). Ce choix s'impose, étant donné nos hypothèses sur la fonction de l'espace visuel commun dans l'interaction verbale. Pour les marqueurs verbaux et les actes de parole de l'apprenant et du tuteur, nous avons utilisé la même grille que l'étude précédente.

5.1.5 Les variables dépendantes

Nous avons retenu les mêmes variables que la première expérimentation, c'est-à-dire le *score de performance des apprenants* et les *impressions subjectives*. Pour les indicateurs concernant les *interactions communicatives*, nous avons ajouté les expressions et les pronoms déictiques.

5.2 Hypothèses opérationnelles

Nous présentons maintenant les hypothèses opérationnelles, concernant les interactions verbales (H1-H4), le score de réalisation de la tâche (H5) et les impressions subjectives (H6).

(H1) <u>Le volume des échanges</u>.

Les études sur les activités collaboratives montrent que lorsque les partenaires distants partagent un espace visuel commun (evc), le volume de leurs échanges diminue (Gergle & al., 2004 ; Karsenty, 1999 ; Kraut & al., 2003). Nous appuyant sur ces études, nous nous attendons à ce que :

H1a : les mots, les tours de parole et les énoncés soient moins nombreux lorsque le tuteur et l'apprenant disposent d'un espace visuel commun que lorsqu'ils n'en disposent pas.

H1b : les tours de parole soient plus courts lorsque l'espace visuel commun est disponible plutôt que lorsqu'il ne l'est pas.

(H2) Les interventions spontanées

Nous supposons que lorsque le tuteur peut observer ce que l'apprenant fait, il peut intervenir davantage de manière spontanée sans attendre d'être sollicité par l'apprenant. Nous nous attendons à ce que :

H2 : le nombre d'interventions spontanées du tuteur soit plus élevé lorsque le tuteur et l'apprenant disposent d'un espace visuel commun plutôt que lorsqu'ils n'en disposent pas.

(H3) Le processus de compréhension mutuelle

Nous supposons que l'espace visuel commun (evc) favorise les processus de compréhension mutuelle. D'après la littérature, grâce à un evc, les participants distants peuvent contrôler les actions que chacun d'eux effectue. Cela permet une meilleure compréhension réciproque car les participants peuvent trouver une adéquation entre les mots énoncés et les actions réalisées par les partenaires (Karsenty, 1999). Nous supposons que lorsque le tuteur dispose de l'evc, il peut contrôler si l'apprenant a compris ses conseils et ses explications, simplement en observant sur son écran comment l'apprenant les applique et sans nécessairement vérifier la compréhension de l'apprenant. Par conséquent, l'apprenant produira moins de marqueurs verbaux consacrées à montrer son niveau de compréhension au tuteur et le tuteur utilisera moins de marqueurs verbaux consacrés à valider la contribution de l'apprenant. Nous nous attendons à ce que :

L'apprenant produise un nombre inférieur de marqueurs verbaux destinés à accepter (H3a) et à vérifier (H3b) l'énoncé du tuteur lorsque l'espace visuel commun est disponible que lorsqu'il ne l'est pas

Le tuteur produise un nombre inférieur de marqueurs verbaux destinés à accepter (H3c) et à vérifier (H3d) ce que l'apprenant vient de dire lorsque l'espace visuel commun est disponible que lorsqu'il ne l'est pas

Concernant les expressions déictiques, la littérature suggère que les participants distants réduisent le temps de construction des référentiels communs. En effet, pour identifier et désigner un objet, les partenaires distants utilisent moins d'expressions verbales explicites (comme le nom de l'objet ou l'indication de l'endroit où l'objet se trouve) et plus de pronoms (e.g. « celui-ci ») et d'expressions déictiques (e.g. « en bas, ici, là ») (Gergle & al., 2004 ; Karsenty, 1999 ; Kraut & al., 2003). Ainsi, nous nous attendons à ce que :

H3e : le nombre d'expressions déictiques soit plus important lorsque l'espace visuel commun est disponible plutôt que lorsqu'il ne l'est pas.

(H4) Le contenu du dialogue tutoriel

Nous supposons que lorsque le tuteur peut observer l'activité de l'apprenant, il sera moins contraint de le questionner pour avoir des informations sur l'état d'avancement de son travail. Par conséquent, l'apprenant produira un nombre inférieur d'actes de parole pour le renseigner sur son propre état d'avancement. Par ailleurs, nous supposons que le tuteur et l'apprenant peuvent co-construire leur interaction verbale grâce à des actes de parole intrinsèques au dialogue tutoriel (i.e. le cours, les balises, etc.). Ainsi, nous supposons que l'apprenant posera un nombre plus important de questions et que le tuteur produira plus d'actes de parole pour l'aider. De plus, comme le tuteur peut observer l'activité en temps réel, il pourra mieux comprendre les difficultés de l'apprenant, de manière à l'aider plus efficacement et à l'encourager davantage. Nous nous attendons à ce que :

L'apprenant produise un nombre inférieur d'actes de parole consacrés à demander de l'aide au tuteur (H4a) et à l'informer sur l'état d'avancement de sa tâche (H4b) lorsque l'espace visuel commun est disponible que lorsqu'il ne l'est pas.

Le tuteur produise un nombre supérieur d'actes de parole destinés à connaître l'état d'avancement de l'apprenant (H4c), à l'aider (H4d) et à l'encourager (H4e) lorsque l'espace visuel commun est disponible que lorsqu'il ne l'est pas.

(H5) Le score de performance des apprenants

Comme l'evc permettrait au tuteur et à l'apprenant de construire plus facilement leurs référentiels communs et de co-construire leur dialogue principalement sur le contenu d'apprentissage de la session tutorielle, nous supposons que l'apprenant pourrait en bénéficier au plan de la réalisation de la tâche. Ainsi, nous supposons que :

H5 : le score de performance de l'apprenant sera plus élevé lorsque l'espace visuel commun est disponible que lorsqu'il ne l'est pas.

(H6) Les impressions subjectives des participants

Nous supposons que l'espace visuel commun, permettant au tuteur et à l'apprenant de construire plus facilement leur compréhension mutuelle, les induirait à évaluer plus positivement la qualité de leur interaction communicative. Nous opérationnalisons cette hypothèse selon deux directions : l'évaluation de la contribution du partenaire distant (impression centrifuge) et de sa propre contribution (impression centripète).

** L'évaluation de la contribution d'autrui (impression centrifuge)*

Nous déclinons cette hypothèse en deux hypothèses opérationnelles :

H6a : le tuteur évalue de manière plus positive la contribution de l'apprenant lorsque l'espace visuel commun est disponible.

H6b : l'apprenant évalue de manière plus positive la contribution du tuteur lorsque l'espace visuel commun est disponible.

** L'évaluation de sa propre contribution (impression centripète)*

Nous formulons deux hypothèses relatives au jugement que le tuteur et l'apprenant portent chacun sur leur propre contribution :

H6c : le tuteur juge sa propre contribution de manière plus positive lorsque l'espace visuel commun est disponible.

H6d : l'apprenant juge sa propre contribution de manière plus positive lorsque l'espace visuel commun est disponible.

5.3 Les résultats

Nous présentons les résultats suivant l'ordre de présentation de nos hypothèses opérationnelles. Pour chaque hypothèse, nous avons vérifié leurs conditions d'application et nous avons indiqué les cas où ces conditions ne sont pas respectées. Pour la vérification des hypothèses, nous avons utilisé des comparaisons *a priori* (contrastes linéaires) et nous avons reporté la valeur de la taille de l'effet.

5.3.1 Les effets de l'espace visuel commun sur le volume des échanges verbaux.

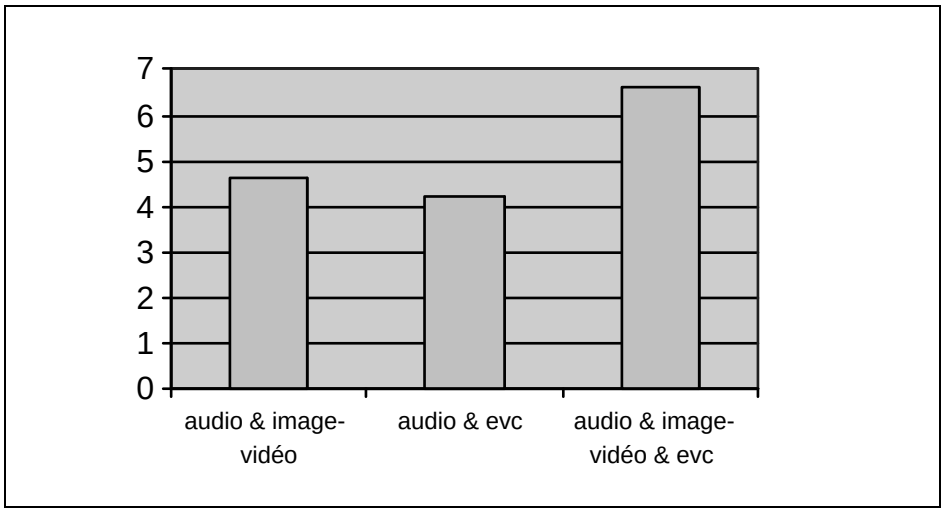
(H1a) Nous avons supposé que les mots, les tours de parole et les énoncés sont moins nombreux lorsque le tuteur et l'apprenant disposent d'un evc plutôt que lorsqu'ils n'en disposent pas. L'analyse multivariée conduite sur les trois variables dépendantes ne montre aucun effet significatif ($F(6, 64) = 1.606$, n.s., $\eta^2 = 0.131$). L'hypothèse H1a est donc invalidée.

(H1b). Nous avons avancé l'hypothèse que les tours de parole sont plus courts lorsque l'espace visuel commun est diponible. L'analyse univariée conduite sur le ratio « nombre de mots / nombre de tours de parole » n'indique aucun effet significatif de la condition expérimentale ($F(2, 33) = 0.970$, n.s., $\eta^2 = 0.076$). L'hypothèse H1b n'est donc pas confirmée.

5.3.2 Les effets de l'espace visuel commun sur les interventions spontanées du tuteur

(H2) Nous devons vérifier que les interventions spontanées du tuteur sont plus nombreuses lorsque l'espace visuel commun est disponible plutôt que lorsqu'il ne l'est pas. L'analyse univariée indique un effet global de la condition sur le nombre d'interventions spontanées du tuteur ($F(2, 33) = 4.671$; $p < .02$, $\eta^2 = 0.22$). Les comparaisons *a priori* indiquent que le nombre d'interventions spontanées du tuteur est significativement plus important en condition « audio & image-vidéo & evc » qu'en condition « audio & image-vidéo » (M= 6.6 (1.9) *vs.* 4.6 (2), $t(33)=2.366$; $p < .02$, $d = 1.02$). En revanche, aucune différence n'est constatée entre la condition « audio & image-vidéo » et « audio & evc » (M= 4.6 (2) *vs.* 4.2 (2.25), $t(33)=.47$; n.s., $d = 0.19$). Il en résulte que le tuteur intervient davantage de manière spontanée

lorsque l'image-vidéo et l'espace visuel commun sont disponibles en même temps : l'hypothèse H2 n'est donc pas entièrement validée.



Graphique 5-1 : Moyennes du nombre d'interventions spontanées du tuteur.

5.3.3 Les effets de l'espace visuel commun sur le processus de compréhension mutuelle

Nous vérifions les hypothèses concernant les effets de l'espace visuel commun (evc) sur le processus de compréhension mutuelle.

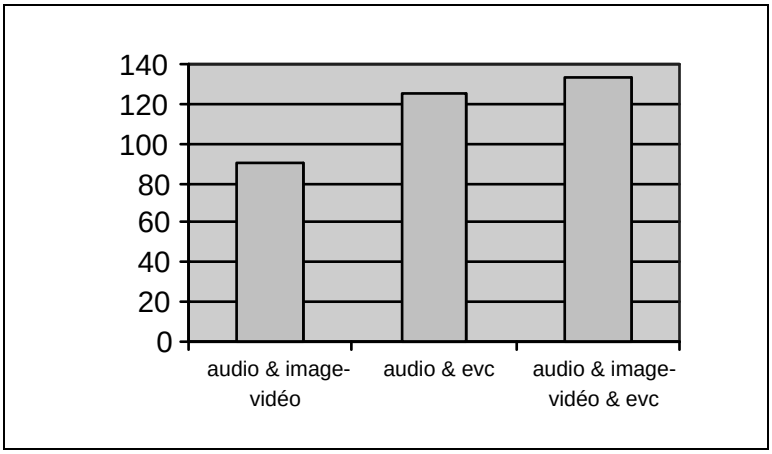
(H3a) Nous devons vérifier que, pour accepter ce que le tuteur vient de dire, l'apprenant produit un nombre inférieur de marqueurs verbaux lorsque le tuteur et l'apprenant disposent d'un evc que lorsqu'ils n'en disposent pas. L'analyse univariée conduite sur les marqueurs verbaux de l'apprenant ne révèle aucune différence significative ($F(2, 33) = .263$; n.s., $\eta^2 = 0.05$). L'hypothèse H3a est invalidée, car l'apprenant produit un nombre équivalent de marqueurs verbaux dans les trois conditions de communication.

(H3b) Nous supposons que, pour vérifier d'avoir compris ce que le tuteur vient de dire, l'apprenant produit moins de marqueurs verbaux lorsque l'espace visuel commun est disponible. L'analyse univariée conduite sur les marqueurs verbaux de l'apprenant ne révèle aucune différence significative ($F(2, 33) = .637$; n.s., $\eta^2 = 0.02$). L'hypothèse H3b est donc invalidée.

(H3c) Nous devons vérifier que le nombre de marquers du tuteur destinés à accepter l'énoncé de l'apprenant est plus important lorsque l'evc est disponible que lorsqu'il ne l'est pas. L'analyse univariée effectuée sur les acquiescements ne révèle aucun effet significatif ($F(2, 33) = 0.40$; n.s., $\eta^2 = 0.02$). L'hypothèse H3c n'est pas donc validée.

(H3d) Cette hypothèse affirme que le tuteur consacre moins de marqueurs verbaux à vérifier la compréhension de l'apprenant lorsque l'espace visuel commun est disponible que lorsque il ne l'est pas. L'analyse univariée ne montre aucune différence significative entre les trois conditions ($F(2, 33) = .911$; n.s., $\eta^2 = 0.05$). Il en résulte ainsi que, quelle que soit la condition de communication, le tuteur vérifie la compréhension de l'apprenant par un nombre équivalent de marqueurs verbaux : l'hypothèse H3d est donc invalidée.

(H3e) Nous devons vérifier que le nombre d'expressions déictiques est plus important lorsque le tuteur et l'apprenant disposent d'un espace visuel commun plutôt que lorsqu'ils n'en disposent pas. L'analyse de la variance révèle un effet significatif de la condition expérimentale sur la variable considérée ($F(2, 33) = 5.453$, $p < .01$, $\eta^2 = 0.99$). Les contrastes linéaires précisent que les expressions déictiques sont significativement plus nombreuses en condition « audio & image-vidéo & evc » qu'en condition « audio & image-vidéo » (M=134.3 (46.2) *vs.* 90.9 (29.3); t(18.6)=2.75, $p < .01$, $d = 1.12$) et plus nombreuses en condition « audio & evc » qu'en condition « audio & image-vidéo » (M=125.2 (21.6) *vs.* 90.9 (29.3) ; t(20.2)=3.27, $p < .005$, $d = 1.33$). L'hypothèse H3e, affirmant que les expressions déictiques produites sont plus nombreuses lorsque l'espace visuel commun est disponible, est donc validée.



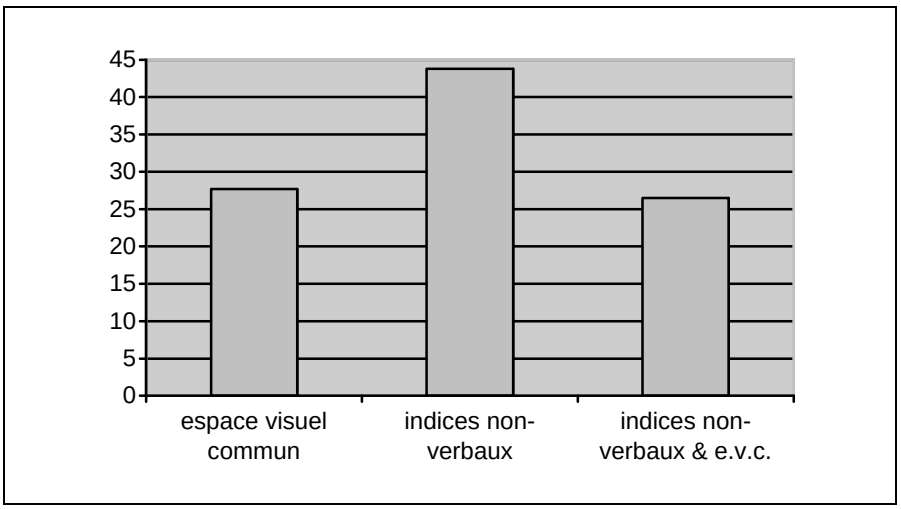
Graphique 5-2 : Moyennes du nombre des expressions déictiques.

5.3.4 Les effets de l'espace visuel commun sur la co-construction du contenu du dialogue tutoriel.

Nous devons valider les hypothèses des effets de l'espace visuel commun sur les actes de parole produits par l'apprenant (H4a, b) et produits par le tuteur (H4c, d, e).

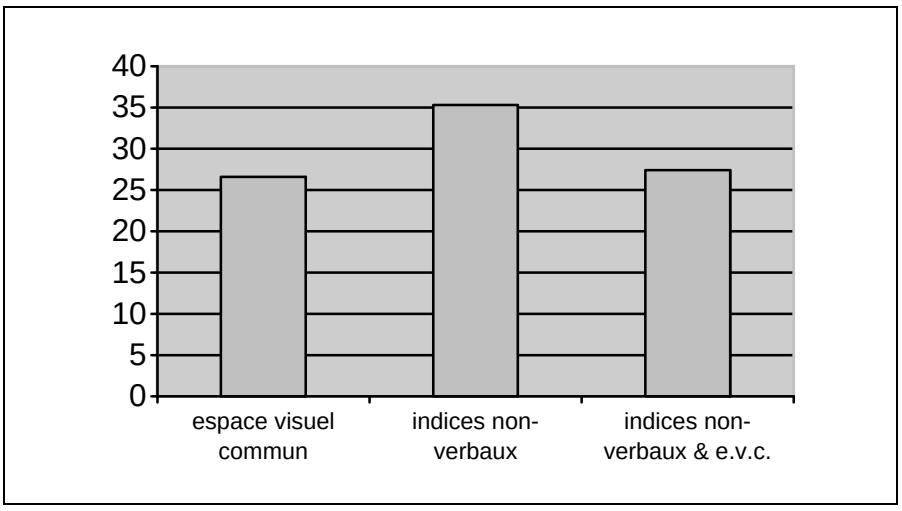
(H4a) Nous devons vérifier que l'apprenant pose au tuteur un nombre de questions plus important lorsque l'espace visuel commun est disponible que lorsqu'il ne l'est pas. Aucune différence n'est constatée entre les trois conditions expérimentales ($F(2, 33) = 0.348$; n.s., $\eta^2 = 0.02$). L'hypothèse H4a est donc invalidée.

(H4b) Nous devons vérifier que les actes de parole de l'apprenant destinés à informer le tuteur sur son état d'avancement sont moins nombreux lorsque le tuteur et l'apprenant disposent d'un espace visuel commun plutôt que lorsqu'ils n'en disposent pas. L'analyse univariée montre que le nombre d'actes de parole destinés à informer le tuteur sur son propre état d'avancement est différent selon les trois conditions expérimentales ($F(2, 33) = 4.529$; $p < .02$, $\eta^2 = 0.21$). En particulier, les actes de parole produits par l'apprenant sont moins nombreux lorsque le tuteur et l'apprenant peuvent se voir et disposer de l'espace visuel commun plutôt que lorsqu'ils ne peuvent que se voir (M= 26.5 (14.1) *vs*. 43.8 (14.4), $t(22) = 2.5$; $p < .01$, $d = 1.21$). De plus, lorsque le seul espace visuel commun est disponible, les actes de parole de l'apprenant sont moins nombreux que lorsque le tuteur et l'apprenant peuvent se voir l'un l'autre (M= 27.7 (18.4) *vs*. 43.8 (14.4), $t(22) = 2.6$; $p < .005$, $d = 0.97$). L'hypothèse H4b est vérifiée : lorsque l'espace visuel commun est disponible l'apprenant réduit le nombre d'actes de parole destinés à informer le tuteur.



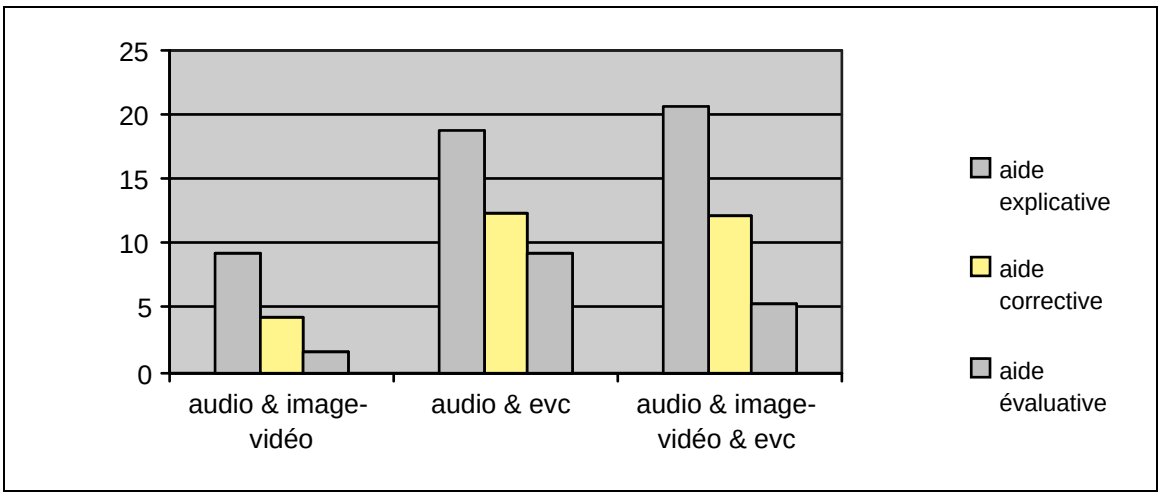
Graphique 5-3 : Moyennes du nombre des actes de parole de l'apprenant consacrés à informer le tuteur.

(H4c) Nous supposons que le tuteur produit moins de questions consacrées à se renseigner sur l'état d'avancement de l'apprenant lorsqu'il dispose d'un evc que lorsqu'il n'en dispose pas. Les résultats de l'analyse univariée indiquent que le tuteur produit un nombre différent de questions selon la condition expérimentale ($F(2, 33) = 4.959$; $p < .01$, $\eta^2 = 0.23$), bien que la taille de l'effet soit faible. Les contrastes linéaires précisent que le tuteur, afin de se renseigner sur l'état d'avancement de l'apprenant, produit moins d'actes de parole lorsqu'il voit l'apprenant et son environnement de travail plutôt que lorsqu'il ne voit que l'apprenant (M= 27.4 (7.6) *vs.* 35.3 (8.1), $t(22)=2.8$; $p < .01$, $d = 1$). De plus, le tuteur pose moins de questions lorsqu'il peut observer seulement l'environnement de travail plutôt que l'image seule de l'apprenant (M= 26.6 (6.8) *vs.* 35.3 (8.1), $t(22)=2.6$; $p < .005$, $d = 1.16$). L'hypothèse H4c est vérifiée : l'espace visuel commun permet au tuteur de réduire le nombre d'actes de parole destinés à connaître l'état d'avancement de l'apprenant.



Graphique 5-4 : Moyennes du nombre d'actes de parole du tuteur consacrés à connaître l'avancement de l'apprenant.

(H4d) Nous avons avancé l'hypothèse que le tuteur produit plus d'actes de parole consacrés à aider l'apprenant lorsque l'espace visuel commun est disponible que lorsqu'il ne l'est pas. Nous avons d'abord vérifié la corrélation entre les actes de parole consacrés à aider l'apprenant et l'analyse ne révèle qu'une corrélation positive entre les actes de parole informatifs et explicatifs (r = 0.43, p < 0.01). L'analyse multivariée conduite sur les quatre types d'actes de parole montre un effet simple du facteur « condition » (F(8, 62) = 3.772 ; p<.001). D'après l'analyse univariée conduite sur chacun des actes de parole, on ne constate qu'un effet tendanciel pour les actes de parole informatifs (F(2, 33) = 2.8; p = 0.7, η^2= 0.14). En revanche, on constate un effet significatif de la condition expérimentale sur les aides de type explicatif (F(2, 33) = 5.68; p <.01, η^2= 0.26), correctif (F(2, 33) = 5.7 ; p <.01, η^2= 0.26) et évaluatif (F(2, 33) = 10.2; p <.0001, η^2= 0.38). L'analyse des contrastes montre que les actes de parole explicatifs sont significativement plus nombreux en condition « audio & image-vidéo & espace visuel commun » qu'en condition « audio & image-vidéo » (M = 20.6 (11.2) *vs*. 9.1 (7.6), t(22) = 3.1 ; p <.003, d= 1.2) et en condition « audio & espace visuel commun » qu'en condition « audio & image-vidéo » (M= 18.7 (7.7) *vs*. 9.1 (7.6), t(22) = 2.6 ; p <.01, d= 1.25).



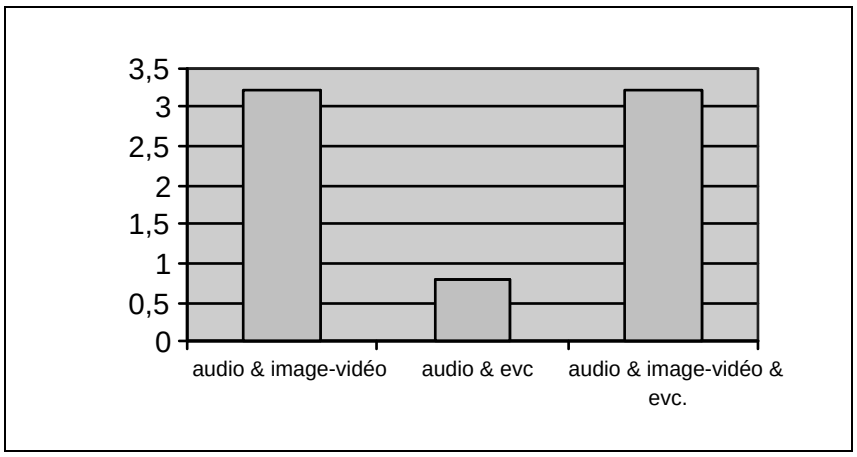
Graphique 5-5 : Moyennes des actes de parole du tuteur consacrés à aider l'apprenant.

Pour les aides de type correctif[50], leur nombre est plus important en condition « audio & image-vidéo & espace visuel commun » qu'en condition « audio & image-vidéo » (M = 12 (3.7) *vs*. 4.2 (5.7), t(19) = 3.9 ; p <.0005, d =1.62) et en condition « audio & espace visuel commun » qu'en condition « audio & image-vidéo » (M = 12.2 (9.1) *vs*. 4.2 (5.7), t(18.5) = 2.5 ; p <.01, d = 1.05). En ce qui concerne les actes de parole évaluatifs[51], ils sont significativement plus nombreux en condition « audio & espace visuel commun » qu'en condition « audio & image-vidéo » (M =9.2 (3.5) *vs*. 1.5 (1.9), t(17) = 6.8 ; p <.0001, d = 2.73) et plus nombreux en condition « audio & image-vidéo & espace visuel commun » qu'en condition « audio & image-vidéo » (M = 5.3 (6.1) *vs*. 1.5 (1.9), t(33) = 2.2 ; p <.03, d = 0.84). L'hypothèse H4d est partiellement validée, car aucun effet général n'a été constaté pour les actes de parole de type informatif.

(H4e) Enfin, nous avons à vérifier que le tuteur encourage l'apprenant plus souvent lorsque l'espace visuel commun est disponible que lorsqu'il ne l'est pas. L'analyse univariée

[50] Pour les actes de parole de type correctif, la distribution des valeurs ne respecte pas l'homogénéité des variances. D'après Field (2000) et Howell (1998), les effectifs étant égaux, le non-respect de l'homogénéité des variances n'entraîne pas de difficultés d'interprétation des valeurs obtenues.

révèle que le nombre d'actes de parole consacrés à encourager l'apprenant (Graphique 5-8) diffère significativement entre les trois conditions expérimentales ($F(2, 33) = 4.813$; $p < .01$, $\eta^2 = 0.22$). Contrairement à nos hypothèses, les comparaisons *a priori* montrent que le tuteur encourage l'apprenant moins souvent en condition « audio & espace visuel commun » qu'en condition « audio & image-vidéo » (M = 0.8 (1.1) *vs.* 3.2 (2.3), t(33)=2.69 ; $p < .01$, $d = 1.33$) et qu'en condition « audio & image-vidéo & espace visuel commun » (M = 0.8 (1.1) *vs.* 3.2 (2.8), t(33)= 2.69 ; $p < .01$, $d = 1.12$). Bien que le tuteur puisse observer l'activité de l'apprenant, il ne l'encourage pas davantage : l'hypothèse H4e n'est pas donc validée.



Graphique 5-6 : Moyennes des actes de parole du tuteur destinés à encourager l'apprenant.

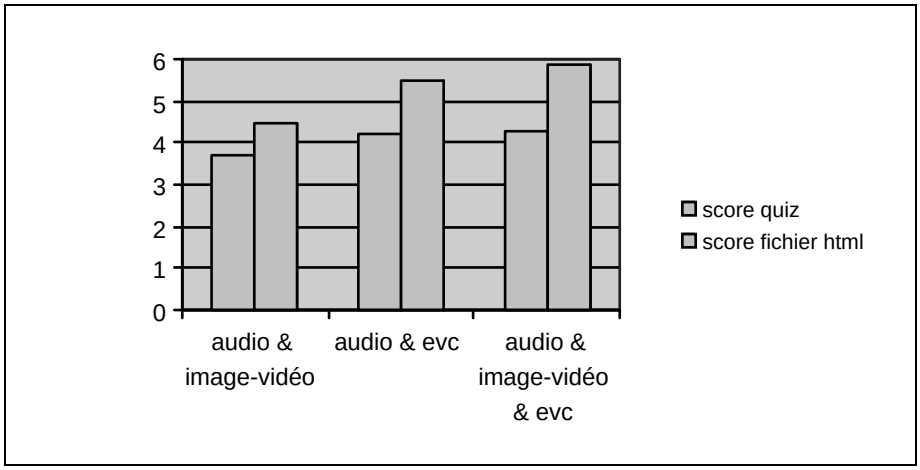
5.3.5 Les effets de l'espace visuel commun sur la performance de réalisation de la tâche

(H5) Nous avançons l'hypothèse que le score de l'apprenant est plus élevé lorsque le tuteur et l'apprenant disposent d'un espace visuel commun que lorsqu'ils n'en disposent pas. Comme dans l'étude précédente, nous avons transformé les scores réels des apprenants en une note sur 10 pour les réponses au test post-tâche (score max test = 12) et une note sur 20 pour la réalisation du fichier html (score max fichier = 186). Une analyse univariée révèle un effet

[51] La distribution des actes de parole destinés à l'évaluation ne respecte pas l'homogénéité des variances. Comme pour le cas précédent, nous considérons que, les effectifs étant égaux, le non-respect

global de la condition expérimentale sur le score obtenu au test ($F(2, 69) = 3.119$; $p < .05$, $\eta^2=0.08$) et sur le score obtenu pour la réalisation du fichier html ($F(2, 69) = 5.776$; $p < .005$, $\eta^2= 0.14$).

Concernant les réponses au test, les contrastes linéaires précisent que le score obtenu est significativement plus élevé en condition « audio & image-vidéo & espace visuel commun » qu'en condition « audio & image-vidéo » (M= 4.3 (.7) *vs*. 3.7 (1), t(69)=2.38 ; $p<.02$, $d = 0.7$). En revanche, on ne constate qu'un effet tendanciel entre la condition « audio & espace visuel commun » et la condition « audio & image-vidéo » (M= 4.2 (1) *vs*. 3.7 (1), t(69)=1.85 ; $p=.07$, $d = 0.5$).



Graphique 5-7 : Moyennes des scores de performance des apprenants.

En ce qui concerne la qualité de réalisation du fichier html, il en résulte que le score des apprenants est plus élevé en condition « audio & image-vidéo & espace visuel commun » qu'en condition « audio & image-vidéo » (M= 5.9 (1.6) *vs*. 4.5 (1.5), t(69)=3.27 ; $p<.002$, $d = 0.88$) et plus élevé en condition « audio & espace visuel commun » qu'en condition « audio & image-vidéo » (M= 5.5 (1.2) *vs*. 4.5 (1.5), t(69)=2.4 ; $p < .02$, $d = 0.73$).

Il en résulte que la qualité de réalisation du fichier html est plus importante lorsque l'espace visuel commun est disponible, alors que le score au test post-tâche ne s'améliore que

---

de l'homogénéité des variances n'entraîne pas de difficulté d'interprétation.

lorsque l'espace visuel commun est couplé avec l'image-vidéo du partenaire distant. L'hypothèse H5 est donc partiellement validée.

5.3.6 Les effets de l'espace visuel commun sur l'évaluation des contributions des participants

Nous supposons que lorsque l'espace visuel commun est disponible, les participants évaluent plus positivement la qualité de leurs contributions. Nous présentons d'abord les résultats relatifs aux impressions centrifuges, puis ceux relatifs aux impressions centripètes.

** Les effets sur l'évaluation des impressions centrifuges.*

Pour les réponses des tuteurs, la valeur du coefficient d'homogénéité (alpha de Cronbach) est satisfaisante (α=.77), nous permettant de regrouper les quatre items en un seul indice (score composite relatif aux impressions centrifuges des tuteurs). La matrice des corrélations inter-items (Tableau 5-1) montre que les items sont corrélés, à l'exception de items relatifs à la clarté des questions et de l'état d'attention de l'apprenant.

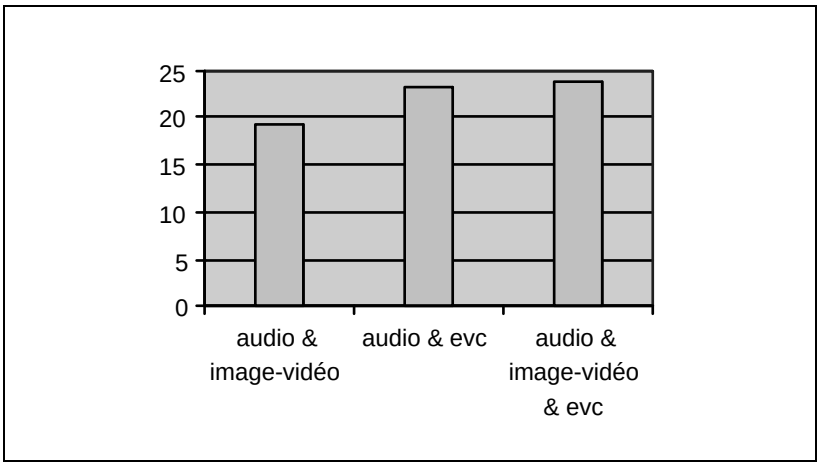
	L'apprenant a su s'expliquer de manière 1.confuse-7.claire	L'apprenant a été attentif à vos indications 1.pas du tout - 7.très attentif	L'apprenant a suivi vos indications de manière 1.inefficace – 7.efficace
L'apprenant a posé des questions 1. confuses- 7.claires	.43**	.21	.35*
L'apprenant a su s'expliquer de manière 1.confuse-7.claire	-	.83**	.53**
L'apprenant a été attentif à vos indications 1.pas du tout - 7.très attentif		-	.52**

Tableau 5-1: Matrice des corrélations inter-items des impressions centrifuges des tuteurs N = 36, seuils de significativité des corrélations : ** p <.01, * p <.05

Pour les réponses des apprenants, la valeur du coefficient est très faible (α=.30) et, comme pour l'étude précédente, chaque item sera traité individuellement.

(H6a) Nous vérifions que le tuteur évalue de manière plus positive la contribution de l'apprenant lorsque l'espace visuel commun est disponible. L'analyse montre un effet global de la condition sur l'ensemble des items ($F(8, 62) = 2.629$; $p < .02$, $\eta^2 = 0.25$) et un effet de la

condition expérimentale sur le score composite mesurant l'impression du tuteur à l'égard de la contribution de l'apprenant ($F(2, 33) = 8.6$; $p < .001$, $\eta^2 = 0.34$). Les comparaisons *a priori* montrent que le tuteur apprécie la contribution de l'apprenant davantage en condition « audio & image-vidéo & evc » qu'en condition « audio & image-vidéo » (M=23.7 (2.4) *vs*. 19.2 (4), t(17.84)=3.32 ; $p < .005$, $d = 1.36$) et davantage en condition « audio & evc » qu'en condition « audio & image-vidéo » (M=23.2 (1.8) *vs*. 19.2 (4), t(15.42)=3.12 ; $p < .008$, $d = 1.29$).

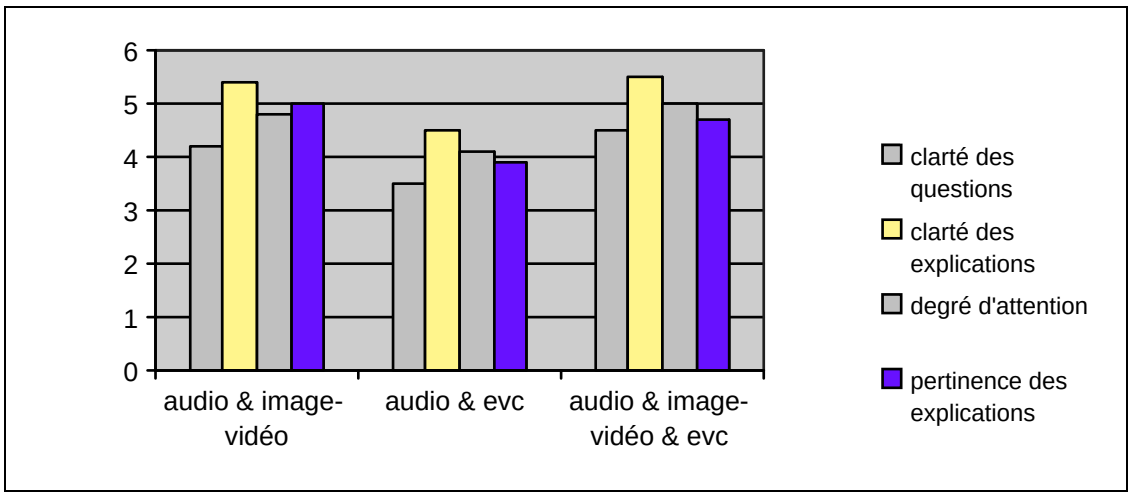


Graphique 5-8 : Moyennes des valeurs relatives aux impressions centrifuges exprimées par le tuteur.

L'hypothèse H6a est donc validée : lorsque le tuteur peut observer l'activité de l'apprenant sur son écran, il juge que la contribution de l'apprenant est globalement meilleure que lorsqu'il ne peut pas l'observer. Plus précisément, les questions et les explications de l'apprenant sont jugées plus claires, ses réponses plus pertinentes et son niveau d'attention plus élevé.

(H6b) Concernant les impressions centrifuges des apprenants, nous supposons qu'ils apprécient davantage la contribution du tuteur lorsque l'espace visuel commun est disponible. L'analyse indique un effet global sur l'ensemble des quatre items ($F(8, 134) = 3.38$; $p < .001$, $\eta^2 = 0.17$). Les analyses univariées conduites sur chaque item de manière individuelle montrent un effet significatif de la condition expérimentale sur les évaluations que l'apprenant fait à propos de la clarté des questions ($F(2, 69) = 3.854$; $p < .03$, $\eta^2 = 0.10$) et des explications du

tuteur ($F(2, 69) = 4.424$; $p < .02$, $\eta^2 = 0.11$), de son niveau d'attention ($F(2, 71) = 3.867$; $p < .03$, $\eta^2 = 0.10$) et de la pertinence de ses explications ($F(2, 71) = 4.22$; $p < .02$, $\eta^2 = 0.11$).



Graphique 5-9 : Moyennes des impressions centripètes de l'apprenant.

Les comparaisons *a priori* indiquent que l'apprenant juge le tuteur plus attentif (M = 4.8 (1.1) *vs.* 4.1 (1.1), t(69) =2.116 ; $p < 0.4$, $d = 0.63$), que ses questions (M = 4.2 (0.9) *vs.* 3.5 (1.4), t(39.79) =2.049 ; $p < .05$, $d = 0.59$) et ses explications (M = 5.4 (1.5) *vs.* 4.5 (1.3), t(69) =2.297 ; $p < .03$, $d = 0.64$) sont moins claires et moins pertinentes (M = 5 (1.5) *vs.* 3.9 (1.2), t(69) =2.74 ; $p < .009$, $d = 0.81$) en condition « audio & image-vidéo » qu'en condition « audio & espace visuel commun ».

Aucune différence significative n'est constatée entre les conditions « audio & image-vidéo & espace visuel commun » et « audio & image-vidéo ». L'hypothèse H6b, affirmant que l'apprenant évalue plus positivement la contribution d'autrui lorsque l'espace visuel commun est disponible, n'est donc pas validée.

** Les effets sur l'évaluation des impressions centripètes.*

En ce qui concerne les impressions centripètes des tuteurs, nous avons calculé l'analyse des corrélations inter-items (Tableau 5-2) et le coefficient d'homogénéité de leurs réponses, dont la valeur satisfaisante ($\alpha = .78$) permet l'agrégation des quatre items en un seul vecteur.

	Estimez-vous avoir fourni des explications 1.confuses - 7.claires	Pendant le discours de l'apprenant vous étiez 1.pas du tout - 7.très attentif	Estimez-vous avoir fourni des explications 1.non pertinentes – 7. Pertinentes
Estimez-vous avoir posé des questions 1.confuses-7.claires	.42*	.61**	.37*
Estimez-vous avoir fourni des explications 1.confuses - 7.claires	-	.53**	.52**
Pendant le discours de l'apprenant vous étiez 1.pas du tout - 7.très attentif		-	.40*

Tableau 5-2 : Matrice des corrélations inter-items des impressions centripètes des tuteurs. N = 36, seuils de significativité des corrélations : ** p <.01, * p <.05.

En revanche, la valeur du coefficient des réponses des apprenants est assez faible (α=.50) et les items seront traités de manière individuelle.

(H6c) Nous devons vérifier si le tuteur évalue sa propre contribution de manière plus positive lorsqu'il dispose de l'espace visuel commun. Nous avons effectué d'abord une analyse sur les quatre items mesurant l'impression du tuteur sur sa propre contribution. L'analyse ne montre aucun effet significatif de la condition sur l'ensemble des variables considérées (F(8 , 62) = 1.347 ; n.s., η^2= 0.15) et l'analyse univariée conduite sur le score composite ne montre qu'un effet tendanciel (F(2 , 33) = 2.979 ; p =.06, η^2= 0.15). Il apparaît donc que l'espace visuel commun n'influe pas sur les impressions centripètes du tuteur : l'hypothèse H6c est donc invalidée.

(H6d) Concernant les impressions centripètes de l'apprenant, d'après l'analyse des quatre items, il en résulte qu'aucun effet global n'est constaté (F(8, 134) = 0.376 ; n.s., η^2= 0.22). L'hypothèse H6d, selon laquelle les impressions centripètes de l'apprenant seraient plus positives lorsque l'espace visuel commun est disponible, est donc invalidée.

5.4 Discussion des résultats.

L'objectif de cette deuxième étude était de comparer les effets de l'espace visuel commun sur l'interaction tutorielle, notamment sur trois indicateurs précis : le volume et le contenu des dialogues, les impressions subjectives des participants et le score de performance des apprenants.

Quel est l'impact de l'espace visuel commun sur le volume des échanges verbaux ?

Contrairement à nos hypothèses, l'espace visuel commun n'induit aucune réduction significative du volume des dialogues (H1a) ni de la longueur des tours de parole (H1b). Ces résultats, qui sont en contre tendance par rapport à d'autres études expérimentales (cf. Fussell & al., 2003 ; Gergle & al., 2004 ; Karsenty, 1999 ; Kraut & al., 2003), suggèrent que le niveau d'interactivité du dialogue tutoriel est indépendant du type d'indices non-verbaux dont les partenaires disposent.

Est-ce que l'espace visuel commun influence le nombre d'interventions spontanées du tuteur ?

Concernant les interventions spontanées du tuteur (H2), l'influence de l'espace visuel commun est partielle. En effet, le tuteur intervient davantage lorsqu'il dispose à la fois de l'image-vidéo de l'apprenant et de l'espace visuel commun. Lorsque le tuteur ne dispose que du seul espace visuel commun, le nombre de ses interventions spontanées est équivalent à quand il ne peut que voir l'image-vidéo de l'apprenant. Ces résultats suggèrent donc que les seuls indices ostensifs-inférentiels (véhiculés par l'espace visuel commun) n'induisent pas le tuteur à intervenir sans attendre l'appel d'aide de l'apprenant. Ce constat est conforté par des analyses complémentaires, qui montrent que le nombre d'interventions spontanées du tuteur est plus important en condition « audio & image-vidéo & evc » qu'en condition « audio & evc » (M=6.6 (1.9) *vs.* 4.2 (2.2), t(33)=2.859 ; p <.007, d = 1.16). En d'autres termes, le tuteur intervient davantage de manière spontanée quand il dispose à la fois des indices ostensifs-inférentiels et kinésiques. Sur le plan théorique, cette complémentarité des indices non-verbaux semble donc appuyer la théorie de la richesse des média (Daft & Lengel, 1986), alors que d'autres recherches, analogues à la notre, mettent en valeur surtout l'apport de seuls indices non-verbaux ostensifs-inférentiels (cf. Karsenty, 1999 ; Kraut & al., 2003).

Est-ce que l'espace visuel commun favorise la compréhension mutuelle ?

Les résultats suggèrent que l'espace visuel commun n'améliore pas la compréhension mutuelle par rapport à l'image-vidéo du partenaire distant. En effet, le nombre des marqueurs verbaux de l'apprenant destinés à accepter l'énoncé produit par le tuteur (H3a) et à vérifier d'en avoir compris le contenu (H3b) est équivalent quel que soit le type d'indice non-verbal

disponible au tuteur et à l'apprenant. De même, le tuteur accepte le contenu de l'énoncé de l'apprenant (H3c) et vérifie si l'apprenant a compris (H3d) par un nombre équivalent de marqueurs verbaux, quel que soit le type d'indice non-verbal disponible, confirmant ainsi les résultats de l'étude 1. De manière analogue, VanLehn & al., (2003) observent qu'en situation de dialogue tutoriel en face à face, le tuteur s'assure par l'intermédiaire d'actes de vérification que l'apprenant a compris ce que le tuteur lui a dit, bien que les indices kinésiques soient bien visibles en face à face.

Malgré ces considérations, des études que nous avons citées où des sujets distants doivent décrire des relations spatiales entre les objets (Gergle & al., 2004 ; Kraut & al., 2003), montrent que le nombre de marqueurs dédiés à la validation de la compréhension du partenaire distant est moins important lorsque les participants distants disposent d'un espace visuel commun. Il se peut donc que les caractéristiques spécifiques de la tâche collaborative (plutôt que les propriétés des indices non-verbaux) induisent des processus de compréhension mutuelle différents.

En ce qui concerne les expressions déictiques (H3e), leur nombre est plus important lorsque le tuteur et l'apprenant disposent de l'espace visuel commun, qu'il soit couplé ou non avec l'image-vidéo du partenaire distant. Lorsque l'espace visuel commun est disponible, l'apprenant sait que le tuteur peut observer son activité à tout instant : cette « conscience mutuelle » de la situation les induit à utiliser une stratégie de description de type déictique. Ce résultat confirme donc qu'un espace visuel commun facilite la construction des représentations partagées entre les participants distants (cf. Karsenty, 1999 ; Kraut & al., 2003) et suggère que les indices non-verbaux ostensifs-inférentiels améliorent le processus de compréhension réciproque dans les activités de communication (Clark & Brennan, 1991).

Quelle influence de l'espace visuel commun sur les actes de parole dédiés au contenu tutoriel ?

Lorsque l'espace visuel commun est disponible, le tuteur pose un nombre inférieur de questions pour se renseigner sur l'état d'avancement de l'apprenant (H4c) et ce dernier réduit en conséquence le nombre d'actes de parole consacrés à le décrire (H4b). Ces résultats, confirmant d'autres études expérimentales (cf. Karsenty, 1999 ; Kraut & al., 2003), suggèrent que le tuteur utilise l'espace commun pour superviser les actions effectuées par l'apprenant.

Comme pour l'étude 1, les actes de parole de l'apprenant consacrés à interroger le tuteur sur les contenus d'apprentissage sont équivalents quelle que soit la modalité de communication (H4a).

Il en résulte aussi que grâce à l'espace visuel commun, le tuteur aide davantage l'apprenant, au moyen d'actes de parole de type explicatif, correctif et évaluatif (H4d). Ce résultat montre ainsi que l'espace visuel commun est indispensable dans une situation de dialogue tutoriel où ce sont principalement des compétences procédurales qui sont mobilisées.

En revanche, les actes de parole pour encourager l'apprenant sont plus nombreux lorsque le tuteur peut voir l'apprenant, sans ou avec l'espace visuel commun (H4e). Ce résultat indique que l'espace visuel commun oriente le dialogue tuteur-apprenant sur un plan « fonctionnel »[52] (plus d'actes de parole concernant la tâche) plutôt que relationnel (moins d'actes de parole d'encouragement) et suggère que l'aspect relationnel du dialogue tutoriel est accentué lorsque les indices kinésiques sont disponibles.

Est-ce que l'espace visuel commun améliore la performance de réalisation de la tâche ?

Les résultats montrent que globalement l'espace visuel commun permet aux apprenants d'améliorer leur performance (H5). Plus précisément, les apprenants obtiennent un score plus élevé, au test et à la mise en page html, lorsque l'espace visuel commun est disponible seul ou avec l'image-vidéo du partenaire. Cependant, la portée de ces résultats est à considérer à la lumière de la valeur de la taille de l'effet des tests statistiques, qui est légèrement en dessous du seuil « fort » (0.80). Néanmoins, la littérature n'est pas unanime sur l'apport réel d'un espace visuel commun à la réalisation de la tâche : certaines études indiquent la réduction de temps de réalisation de la tâche plutôt que la qualité de réalisation (cf. Fussell & al., 2003 ; Gergle & al., 2004 ; Kraut & al./2, 2003) alors que d'autres ne montrent aucun bénéfice (cf. Karsenty, 1999 ; Kraut & al./1, 2003). Vu ces résultats, du point de vue du critère de l'efficacité du dialogue

[52] Dans le domaine de l'ergonomie cognitive du dialogue, Savoyant et Leplat (1983) définissent les communications fonctionnelles comme des «*...communications regardant directement le contenu du travail réalisé, excluant ainsi celles qui sont prioritairement centrées sur les relations humaines* » (1983 :248).

(rapport entre le volume des dialogues et le score de performance), les indices ostensifs-inférentiels sont plus efficaces que les indices kinésiques.

Est-ce que l'espace visuel commun influe sur les impressions centrifuges et centripètes des participants ?

Nous avons vérifié si, disposant de l'espace visuel commun, le tuteur et l'apprenant estiment plus positivement la contribution d'autrui (impressions centrifuges) et leur propre contribution personnelle (impressions centripètes) pendant l'interaction. Les résultats ne montrent aucun effet significatif sur les impressions centripètes, ni du tuteur ni de l'apprenant (H6c & H6d).

En revanche, concernant les impressions centrifuges, le tuteur estime que l'apprenant est plus attentif, que ses questions et ses explications sont plus pertinentes, ses réponses plus claires lorsque le tuteur dispose d'un espace visuel commun que lorsqu'il n'en dispose pas (H6a). Sur le plan théorique, ce résultat soutient l'idée que lorsque l'on dispose de plusieurs indices non-verbaux (Daft & Lengel, 1984), l'ambiguïté interprétative de l'information se réduit car la pertinence de l'information est élevée (cf. Sperber & Wilson, 1986). En ce qui concerne l'apprenant, il apprécie plus positivement la contribution du tuteur lorsqu'ils peuvent se voir l'un l'autre que lorsque l'espace visuel commun seul est disponible (H6b). Sous un autre point de vue, ce résultat suggère que le sujet possédant un niveau inférieur de compétences évalue plus positivement celui qui possède le niveau de compétences plus élevé lorsqu'ils peuvent se voir l'un l'autre.

En résumé, l'espace visuel commun facilite la compréhension réciproque des référents communs de l'activité, la co-construction des contenus du dialogue tutoriel, améliore la performance de réalisation de la tâche et induit le sujet moins compétent à évaluer de manière positive la contribution du sujet plus compétent. Par rapport à l'image-vidéo du partenaire distant, les apports de l'espace visuel commun semblent plus nombreux, surtout lorsque l'image-vidéo du partenaire distant est disponible en même temps. Cet aspect sera l'objet de nos conclusions finales.

Conclusions

L'objectif de cette recherche était de comprendre l'impact des indices non-verbaux sur une activité de communication à distance spécifique, notamment le dialogue tuteur-apprenant. Après avoir défini l'activité de communication comme une activité à la fois cognitive et sociale, nous avons décrit plus précisément les composants langagièrs du dialogue tutoriel, notamment du processus de compréhension réciproque et de co-construction du contenu tutoriel (ch.1). Nous avons ensuite dressé un état de l'art des recherches portant sur les effets des modalités de communication audio et audio-vidéo sur les activités de communication synchrone médiatisée (ch.2 et ch.3).

Nous appuyant sur les résultats de ces recherches, nous avons construit deux hypothèses générales : a) les indices non-verbaux kinésiques facilitent l'interaction dialogique et le sentiment de proximité sociale entre les partenaires distants ; b) les indices non-verbaux ostensifs-inférentiels facilitent la compréhension réciproque et la réalisation de l'activité commune. Pour les vérifier, nous avons mis en place deux études expérimentales : dans la première, nous nous sommes intéressés aux indices non-verbaux kinésiques (ch.4) et dans la deuxième étude nous avons abordé les indices non-verbaux ostensifs-inférentiels (ch.5). Nous discutons maintenant les implications des résultats obtenus, soulignant les limites méthodologiques de notre démarche.

L'apport des indices non-verbaux

D'après la première étude, il en résulte que les indices non-verbaux kinésiques :

- facilitent la compréhension réciproque (réduction du nombre d'énoncés du tuteur pour valider et vérifier la compréhension de l'apprenant, H3) ;
- favorisent la mémorisation de règles procédurales (le score des réponses au test post-tâche est plus élevé, H5) ;
- induisent les participants à se focaliser davantage sur les étapes accomplies et les étapes à réaliser concernant la tâche à résoudre (i.e. plus d'énoncés du tuteur et de l'apprenant concernant l'état d'avancement de l'apprenant, H4b et H4c) ;
- renforcent l'aspect relationnel de l'interaction (nombre plus important d'énoncés d'encouragement du tuteur, H4e) ;
- favorisent le suivi de l'activité du partenaire distant (nombre plus important d'interventions spontanées du tuteur, H4a) ;
- induisent un sentiment plus positif sur la qualité de la contribution du sujet possédant des compétences inférieures (les apprenants évaluent plus positivement la contribution du tuteur, H6).

Ces résultats confirment partiellement notre hypothèse de travail et soutiennent la théorie de la *présence sociale* (Short & al., 1976), affirmant que chaque moyen de communication induit un sentiment différent de la présence sociale du partenaire distant et que plus les caractéristiques de la modalité de communication sont proches de la condition « face à face », plus le sentiment de présence sociale est renforcé.

Dans la deuxième étude, nous avons comparé les apports des indices kinésiques et des indices ostensifs-inférentiels. Les résultats principaux indiquent qu'un espace visuel commun :

- favorise la construction des référentiels communs (nombre important d'expressions déictiques, H3e) ;
- permet au tuteur d'observer l'activité de l'apprenant (moins de questions concernant son état d'avancement, H4c) ;
- permet de focaliser le contenu de l'interaction dialogique sur la tâche (plus d'énoncés du tuteur pour informer l'apprenant, H4d) plutôt que sur la relation (le tuteur encourage moins l'apprenant, H4e).
- améliore l'efficacité de la communication (meilleure performance de réalisation de la tâche à parité de volume des dialogues, H5).

Il apparaît que l'espace visuel commun permet à des partenaires distants de produire des inférences à partir d'une même « représentation externe » (Zhang & Norman, 1994), facilitant la synchronisation de leurs « représentations mentales partagées » (Pavard & Karsenty, 1997).

Par ailleurs, ces résultats indiquent que lorsque les indices non-verbaux kinésiques et ostensifs-inférentiels co-occurrent :

- le partenaire distant interpelle plus fréquemment son interlocuteur (on constate un nombre plus important d'interventions spontanées du tuteur, H2) ;
- le partenaire plus compétent (le tuteur) juge que la contribution du partenaire moins compétent (l'apprenant) est plus attentive, plus claire et plus pertinente (H6).

Ces résultats suggèrent que lorsque les technologies de communication véhiculent les deux types d'indices non-verbaux, le tuteur assure un suivi pédagogique plus attentif et apprécie davantage la contribution de l'apprenant. Ces résultats soutiennent donc la *théorie de la richesse des média* (Daft & Lengel, 1986), qui affirme que plus le nombre d'indices verbaux et non-verbaux véhiculés est élevé, moins l'interprétation des énoncés du partenaire distant est incertaine et plus la compréhension mutuelle est facile.

Les limites de notre démarche

La portée de ces résultats est cependant limitée par la démarche expérimentale que nous avons choisie. Sur le plan global, comme nous l'avons rappelé précédemment (§ ch.4), les expérimentations *in vitro* sollicitent des sujets parfois faiblement motivés et impliqués dans la réalisation d'une tâche pendant une durée de temps souvent très limitée (Holland, Hutchins & Kirsh, 2000 ; Monk, Nardi, Gilbert, Mantei & McCarthy, 1993). Dans notre cas spécifique, les sujets-apprenants n'ont disposé que de quelques minutes pour apprendre à manipuler le dispositif de communication audio-vidéo. Cette contrainte temporelle peut avoir inhibé leurs intentions de s'adresser au tuteur (ce qui a été évoqué librement par cinq apprenants à la fin de la session).

De plus, dans chaque session expérimentale le nombre des étudiants est limité à deux, ce qui réduit la généralisation de ces résultats par rapport à une situation de « classe virtuelle », composée d'une dizaine d'apprenants distants au maximum (Laurillard, 1993). Enfin, sur le

plan matériel, des contraintes techniques ne nous ont pas permis de mettre les sujets en condition de se regarder mutuellement dans les yeux (cf. problème de la parallaxe).

Les implications pour la conception des dispositifs consacrés à assister la communication tutorielle

Les résultats de la deuxième étude indiquent que, grâce aux indices non-verbaux ostensifs-inférentiels, le tuteur assure un suivi pédagogique plus attentif. Il semblerait donc qu'un espace visuel commun soit la fonctionnalité indispensable dans une plate-forme logicielle consacrée à la formation interactive à distance. D'après les autres résultats, il apparaît qu'il est préférable de mettre à disposition des apprenants distants l'image du formateur distant plutôt que sa seule voix. Si on devait choisir entre l'un ou l'autre type d'indice non-verbal, quel type d'indice suffirait pour assurer que le dialogue tutoriel soit efficace et interactif ?

D'après l'ensemble des résultats, il apparaît que les indices non-verbaux ostensifs-inférentiels sont à privilégier par rapport aux indices kinésiques : donc, il vaut mieux priviliegier un espace visuel commun plutôt qu'un espace de communication interpersonnelle vidéo. Néanmoins, il ne suffit pas de mettre à disposition un espace visuel générique. En effet, les expérimentations de Fussell & al. (2003) et Kraut & al., (1996, 2003) montrent que les dispositifs mobiles permettant un espace visuel commun (e.g., une caméra-vidéo frontale) peuvent parfois entraver la coopération à distance, notamment car : *a*) les partenaires distants ne focalisent pas leur attention sur un même élément du champ visuel ; *b*) la direction du regard du sujet portant la caméra-vidéo frontale n'est pas mise en correspondance avec la direction du regard du sujet distant qui observe la scène sur son propre écran. Pour le point *a*), Fussell, Seatlock & Parker (2003) ont proposé un système de *tracking visuel portatif* : lorsque le sujet fixe un point de son champ visuel, un symbole s'affiche sur l'écran de son partenaire distant. Pour atténuer les difficultés du point *b*), Berlage & Sohlenkamp (1999) proposent qu'un dispositif se dote :

- du point de focalisation, indiquant le focus visuel du locuteur ;
- du point de deixis, indiquant le point ou l'objet qui est pointé par le locuteur ;
- du point actuel, indiquant le niveau d'attention de l'utilisateur ;
- du point de travail, indiquant l'endroit où le locuteur veut changer quelque chose.

En effet, d'après nos observations et l'analyse des dialogues, même si l'espace visuel commun est disponible, parfois les interlocuteurs doivent préciser quelles parties ou quels éléments de l'écran sont les référents communs.

Pour cela, un espace visuel commun devrait :

- favoriser l'usage d'expressions déictiques dans les descriptions spatiales, surtout pour des environnements de travail composés de nombreux objets, dont le nom n'est pas connu par tous les participants (e.g. situation tutorielle) ou dont le nom est de prononciation difficile (cf. Bly, 1988 ; Whittaker & al., 1991) ;
- faciliter la définition d'une sorte d'« agenda » des choses à réaliser et réalisées de manière progressive (trace de l'activité) ;
- apporter une information directe : par exemple, lorsque l'on supervise le travail du partenaire distant, on ne soit pas obligé de lui poser des questions sur ce qu'il est en train de faire. Les indices ostensifs-inférentiels sont ainsi utilisés à la place des indices verbaux (Nardi & al., 1997).

Des suggestions pour la poursuite de ces recherches

Nous envisageons des perspectives de recherches à différents niveaux.

a) Validation générale des résultats. Tout d'abord, l'augmentation du nombre d'apprenants est la *conditio sine qua non* pour vérifier la portée de ces résultats. Par exemple, d'après l'étude n.1, il résulte que les indices non-verbaux kinésiques n'affectent ni le niveau d'interactivité ni le taux d'efficacité (on obtient un même ratio entre le volume des dialogues et le niveau de réalisation de la tâche).

b) Forme de visualisation de l'information relative à l'activité de l'apprenant. Cette perspective ouvre une question de recherche fondamentale : comment observer en même temps l'activité d'une *classe* virtuelle de douze étudiants distants ? Quelle forme de visualisation est la plus adéquate pour rendre efficace et interactif le dialogue tutoriel ? Au niveau général, nous pouvons envisager deux types de visualisation : l'un plutôt *analogique*, l'autre *symbolique*. Par visualisation *analogique*, nous indiquons le cas où il y a une correspondance en temps réel entre ce que le tuteur observe et ce que l'apprenant fait, comme dans les deux études présentées. Cependant, cette forme de visualisation qui permet au tuteur de voir chaque déplacement de la souris et chaque balise écrite par l'apprenant (cf. ch5), semble plus adaptée à

des situations pédagogiques où le nombre d'étudiants est réduit à un binôme ou à un trinôme (cf. Baudin, Owerzarski, Cames, Villemur, Owerzaski, Diaz & Schmidt, 1998), (Figure 6-1).

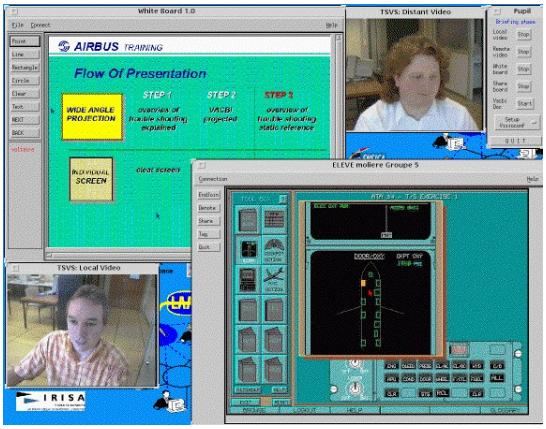


Figure 6-1 : La forme de visualisation analogique utilisée dans le projet Topase

Par visualisation *symbolique*, nous désignons le cas où toute action de l'apprenant sur le système est interprétée par une application logicielle qui montre au tuteur un nombre défini de descripteurs relatifs à l'activité de l'apprenant, ou *traces*[53]. Cette solution est proposée par le projet FORMID de l'équipe ARCADE (Clips-Imag), permettant au tuteur le suivi synchrone de l'activité d'un groupe d'apprenants distants (Figure 6-2).

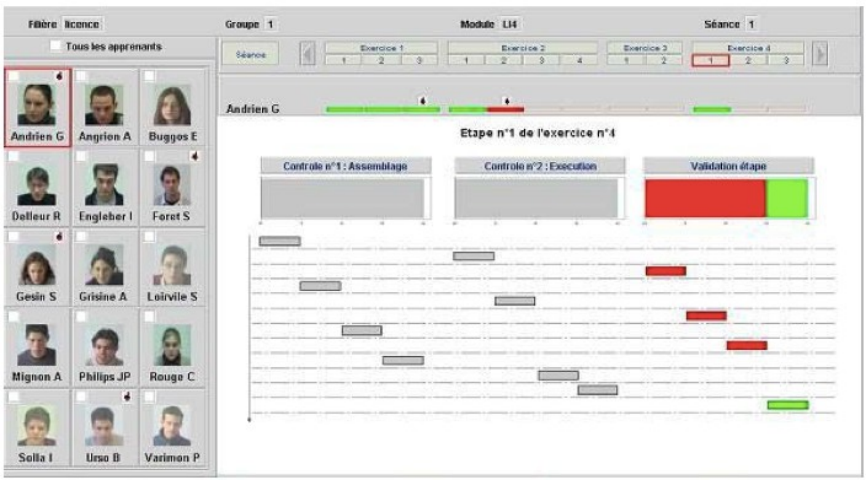


Figure 6-2 : La forme de visualisation symbolique de FORMID

[53] Settouti (2006) définit les traces comme une « *séquence temporelle d'observés* ». Il s'agit de « *données issues d'observation directe ou indirecte permettant la régulation, le contrôle, l'analyse et la compréhension de l'activité d'apprentissage* ».

c) *Charge cognitive*. Quelle que soit la forme de visualisation, il est impératif de s'interroger sur quelle forme de visualisation implique une charge cognitive moins importante. En effet, dans une situation de visualisation analogique (cf. *mediaspace*, Mackay, 1999), on peut facilement imaginer que l'attention visuelle du tuteur est vite surchargée. Cela laisse supposer que pour faciliter la communication synchrone tuteur-apprenant en situation de *classe virtuelle*, il serait préférable de proposer des formes de visualisation symbolique plutôt qu'analogique (cf. Dyck & Gutwin, 2002 ; Nunnari & Simone, 2002). Cependant, les interfaces à représentation symbolique induisent aussi des problèmes de surcharge de l'attention visuelle (Hoc, 1996).

d) Interprétation des expressions faciales. Une autre perspective de recherche concerne l'effet du visage et des expressions faciales de l'apprenant. Les résultats de l'étude 1 montrent que le tuteur intervient davantage de manière spontanée auprès de l'apprenant lorsqu'il peut en voir le visage. Cela implique que le tuteur interprète les expressions faciales produites par l'apprenant et décide d'apporter de l'aide à l'apprenant avant qu'il la demande : quelle est la nature de ces expressions faciales ? Est-ce que le tuteur intervient uniquement lorsqu'une certaine expression faciale se produit ? Considérons maintenant une *classe* virtuelle composée de plusieurs apprenants : afin d'aider le tuteur à interpréter les expressions faciales, pourrait-on automatiser la boucle « interprétation des expressions faciales / déclenchement de l'intervention de l'apprenant » ? Dans cette direction, des recherches sont en cours pour comprendre comment les expressions faciales produites par un sujet sont jugées et catégorisées par un observateur distant (Tcherkassof, Bollon, Dubois, Pansu, Paignon & Adam, 2004).

e) Les outils théoriques. Les théories disponibles ne permettent pas de prévoir finement l'influence des modalités de communication sur les activités médiatisées. D'après Whittaker (2002), ces théories ne considèrent pas que chaque indice non-verbal incite à un comportement communicatif spécifique. Cette position demanderait donc d'identifier et de définir des descripteurs spécifiques pour décrire les activités de communication et construire ainsi une taxonomie des tâches de communication en rapport aux indices non-verbaux.

f) Les outils méthodologiques pour l'analyse du corpus. Sur le plan de l'analyse du corpus dialogique, d'autres analyses complémentaires sont à envisager. Par exemple, l'analyse des *hésitations* pourrait fournir un autre indicateur du coût des technologies de communication sur la co-construction du terrain commun d'entente (cf. Brennan & Ohaeri, 1999). Une autre dimension à explorer est celle des *déictiques* : on pourrait les catégoriser selon des tâches différentes et déterminer si certaines formes déictiques sont corrélées à un type de tâche spécifique (e.g., plutôt procédurale *vs.* plutôt déclarative) ou bien à d'autres indices non-verbaux (i.e., est-ce qu'on produit une forme déictique lorsqu'on utilise la souris pour déplacer des objets ?)

Bibliographie générale

Abric, J.C. (1999). *Psychologie de la communication : théories et méthodes*. Paris : Colin.

Amado, G., & Guittet, A. (2003). *Dynamique des communications dans les groupes*. Paris : Colin.

Amalberti, R., & Hoc, J.-M. (1998). Analyse des activités cognitives en situation dynamique : pour quels buts? Comment? *Le Travail Humain, 61, 3*, 209-234.

Anderson, A.H., Newlands, A., Mullin, J., Fleming, A.M., Doherty-Sneddon, G., & Van Der Velden, J. (1996). Impact of video-mediated communication on simulated service encounters. *Interacting with computers, 8,* 193-206.

Anderson, A. H., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., Fleming, A. M., & Van der Velden, J. (1997). The impact of VMC on collaborative problem solving: An analysis of task performance, communicative process, and user satisfaction. In K. E. Finn, A. J. Sellen, & S. B. Wilbur (Eds.), *Video-mediated Communication* (pp. 133-155). Hillsdale, NJ: Lawrence Erlbaum Associates.

Argyle, M., & Graham, J. (1977). The Central Europe Experiment : looking at persons and looking at things. *Journal of Environmental Psychology and Nonverbal Behaviour, 1*, 6-16.

Argyle, M., Lalljee, M., & Cook, M. (1968). The effects of visibility on interaction in the dyad. *Human Relations, 28*, 289-304.

Austin, J. L. (1962). *How to do things with words.* Oxford : Oxford University Press.

Bange, P. (1992). *Analyse conversationnelle et théorie de l'action*. Paris : Didier-Hatier.

Baudin V., Owerzarski S., Cames J.L., Villemur T., Owerzaski P., Diaz M. & Schmidt J.F. (1998). *Conception d'un environnement de télé-formation synchrone : Projet TOPASE*. Rapport de Recherche n. 98245. Toulouse : LAAS – CNRS.

Bavelas, J., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin 21*, 394-405.

Bavelas, J., Hutchinson, S., Kenwood, C., & Matheson, D. H. (1997). Using face-to-face dialogue as a standard for other communication systems. *Canadian Journal of Communication, 22*, 5-24.

Beall, A.C., Bailenson, J.N., Loomis, J., Blascovich, J., & Rex, C. (2003). Non-Zero-Sum Mutual Gaze in Collaborative Virtual Environments. In *Proceedings of Human-Computer Interaction International* (pp.1108-1112). Crete : Erlbaum Associates.

Beattie, G. (1981). The regulation of speaker turns in face-to-face conversation: some implications for conversation in sound-only communication channels. *Semiotica, 34*, 55-70.

Berlage, T., & Sohlenkamp, M. (1999). Visualizing Common Artefacts to Support Awareness. *Computer-Mediated Cooperation, Computer Supported Cooperative Work, (8),* 3, 207-238.

Bly, S. (1988). A use of drawing surfaces in collaborative settings. In *Proceedings of Conference on Computer Supported Cooperative Work* (pp.250-256). New York: ACM Press.

Boyle, E. A., Anderson, A. H., & Newlands, A. (1994). The effects of eye contact on dialogue and performance in a co-operative problem solving task. *Language and Speech, 37,* 1-20.

Borges, M., Brézillon, P., M., Pino, J., & Pomerol, J.-Ch. (2005). Bringing context to CSCW. In W. Shen, T. Li, Z. Lin, J.-P. Barthès, W. Zeng, S. Li & C. Yang (Eds.), *Computer Supported Cooperative Work in Design* (pp. 45-54). Berlin : Springer-Verlag.

Brassac, C. (2000). La conception située et distribuée : un point de vue de psychologue des processus cognitifs collaboratifs. In *Actes de la Septième Ecole d'été de l'Association pour la Recherche Cognitive*, vol. 3, pp.1-56.

Brassac, C., & Grégori, N. (2001). Situated and Distributed Design of a Computer Teaching Device. *Journal of Design Sciences and Technology, 8* (2), 11-31.

Brennan, S. E. (1990). Conversation as direct manipulation: an iconoclastic view. In B.K. Laurel (Ed.), *The Art of Human-Computer Interface Design* (pp.393-404. Reading, MA:Addison-Wesley.

Brennan, S.E. (1998). The grounding problem in conversations with and through computers. In Fussell, S.R., & Kreuz, R.J. (Eds.) *Social and cognitive psychological approaches to interpersonal communication* (pp.201-224). Lawrence Erlbaum : Hillsdale.

Brennan, S.E., & Ohaeri, J.O. (1999). Why do electronic conversations seem less polite? The costs and benefits of hedging. In *Proceedings of International Joint Conference on work activities, coordination and collaboration – WACC* (pp. 227-235). San Francisco, CA : ACM.

Bromberg, M. (2004). Contrat de communication et co-construction de sens. In M. Bromberg & A. Trognon (Eds.), *Psychologie sociale et communication*, (pp. 95-109). Paris:Dunod.

Bruce, V. (1996). The role of the face in communication: implications for video-phone design. *Interacting with computers, 8*, 166-176.

Bruner, J. (1983). *Le développement de l'enfant. Savoir faire, Savoir dire*. Paris : Presses Universitaires de France.

Buxton, W., & Moran, T. P. (1990). EuroPARC's Integrated Interactive Intermedia Facility (IIIF): Early Experiences. In *Proceedings of IFIP WG8.4 Conference on Multi-user Interfaces and Applications*. Heraklion, Crete.

Cahour, B., & Falzon, P. (1992). Assistance à l'opérateur et modélisation de sa competence. *Intellectica, 2* (2), 159-186.

Cahour, B. (1997). Difficultés et réglages de l'intercompréhension en fonction du contexte visuel. In P. Falzon & L. Karsenty (Eds.), *Dialogue et cooperation* (pp.137-150). Rapport d'étude du PRC Sciences cognitives, Paris : CNAM.

Cahour, B. (2003). Accès aux affects en situation d'interaction professionnelle. In *Actes du colloque de Psychologie Ergonomique Epique'03,* (pp.10-25).Paris:Université Paris V.

Carletta, J., Isard, A., Isard, S., Kowtko, J.C., Doherty-Sneddon, G., & Anderson, A.H. (1996). The reliability of a dialogue structure coding scheme. *Computational Linguistics, 23*(1), 13-31.

Carstensen, P. H., & Schmidt, K. (1999). *Computer Supported Cooperative Work: New Challenges to Systems Design*. Working Paper 43, Lyngby: Technical University of Denmark, Center for Tele-Information.

Cavanaugh, J. (2001). The effectiveness of interactive distance education technologies in K-12 learning: A meta-analysis. *International Journal of Educational Telecommunications 7*(1), 73-88.

Chabrol, C., & Bromberg, M. (1999). Préalables à une classification des actes de parole. *Psychologie Française, 44*, 291-306.

Chapanis, A. (1975). Interactive human communication. *Scientific American, 232*, 36-42. In I. Greif, I. (Ed.) (1988). *Computer-supported cooperative work: A book of readings* (pp. 125-140). San Mateo, CA: Morgan Kaufman.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 15*, 145-182.

Chi, M.T.H. (1997). Quantifying qualitative analyses of verbal data : a practical guide. *The Journal of the Learning Sciences, 6* (3), 271-315.

Chi, M.T.H., Siler, S.A., Jeong, H., Yamauchi, T., & Hausmann, R.G. (2001). Learning from tutoring. *Cognitive Science, 25*, 471-533.

Clancey, W. J. (1985). Heuristic Classification. *Artificial Intelligence, 27*, 289-350.

Clark, H.H. (1996). *Using Language*. Cambridge University Press : Cambridge.

Clark, H.H., & Marshall, C.R. (1981). Definite reference and mutual knowledge. In A.K. Joshi, B.L. Webber, & I.A. Sag (Eds.). *Elements of discourse understanding* (pp.10-63). Cambridge : Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.

Clark, H. H., & Schaefer, E.F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259-294.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition*. (pp. 127 - 149). Washington, DC: American Psychological Association.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Cohen, P. R., & Levesque, H. J. (1995). Communicative actions for artificial agents, *Proceedings of the International Conference on Multi-Agent Systems*. AAAI Press : San Francisco.

Cutler, A., & Pearson, M. (1986). On the analysis of prosodic turn-taking cues. In C. Johns-Lewis (ed.), *Intonation in Discourse* (pp. 139-155). London : Croom Helm.

Daft, R.L., & Lengel, R.H. (1986). Organisational information requirements, media richness and structural design. *Management Science*, 32, 554-571.

Daly-Jones, O., Monk, A.F., & Watts L.A. (1998). Some advantages of video-conferencing over high-quality audio conferencing : fluency and awareness of attentional focus. *International Journal of Human Computer Studies, 49,* 21-58.

Di Paolo, R.E., Graesser, A.C., Hacker, D.J., & White, H.A., (2002). Hints in Human and Computer Tutoring. In M. Rabinowitz (ed.), *The impact of media on technology of instruction*. Mahwah, NJ : Lawrence Erlbaum.

Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face interaction and video mediated communication : a comparison of dialogue structure and co-operative task performance. *Journal of Experimental Psychology : Applied, 3,* 105-125.

Duncan, S., & Fiske, D. W. (1977). *Face-to-face interaction: Research, methods and theory*. Hillsdale, New Jersey : Lawrence Erlbaum.

Durlach, N., & Slater, M. (2000). Presence in shared virtual environments and virtual togetherness. *Presence : Teleoperators and Virtual Environments, 9*(2), 214-217.

Dyck, J., & Gutwin, C. (2002). Groupspace: A 3D Workspace Supporting User Awareness. In *CHI'02 Human factors in computing systems - extended abstracts* (pp.502-503). New York, NY: ACM Press.

Ekman, P. (2003). *Emotions Revealed*. London : Weidenfeld & Nicolson.

Falzon, P. (1994). Les activités métafonctionnelles et leur assistance. *Le travail humain, 57,* 1-23.

Field, A. (2000). *Discovering Statistics using SPSS.* Sage Publications.

Finn, K. E., Sellen, A. J., & Wilbur, S. B. (1997). *Video-mediated communication*. Mahwah, New Jersey: Lawrence Erlbaum.

Foulon-Molenda, S. (2000). A-t-on besoin de se voir pour se comprendre? *Le Travail Humain, 63* (2), 97-120.

Fox, B.A. (1993). *The Human Tutorial Dialogue Project: Issues in the Design of Instructional Systems*. Hillsdale NJ: Lawrence Erlbaum

Fry, R., & Smith, G. F. (1975). The effects of feedback and eye contact on performance of a digit-encoding task. *Journal of Social Psychology, 96,* 145-146.

Fussell, S. R., Kraut, R. E., & Siegel, J. (2000). Coordination of communication: Effects of shared visual context on collaborative work. In *Proceedings of CSCW 2000* (pp. 21-30). NY: ACM Press.

Fussell, S. R., Setlock, L. D., & Parker, E. M. (2003). Where do helpers look? Gaze targets during collaborative physical tasks. In *Proceedings of CHI 03 Extended Abstracts* (pp. 768-769). NY: ACM Press

Gentner, D., & Stevens, A.L. (1983). *Mental Models*. Bolt Beranek and Newman Inc.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004). Language efficiency and visual technology: minimizing collaborative effort with visual information. *Journal of Language and Social Psychology, 23,* 491-517.

Ghiglione, R. (1986). *L'homme communiquant*. Paris : Colin.

Giboin, A. (2004). La construction de référentiels communs dans le travail coopératif. In J.-M. Hoc & F. Darses (Eds.) *Psychologie ergonomique : tendances actuelles* (pp. 119-139). Presses Universitaires de France.

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition, 10,* 597-602.

Gnisci, A., Papa, F., & Spedaletti, S. (1999). Usability aspects, socio-relational context and learning performance in the virtual classroom : a laboratory experiment. *Behaviour & Information Technology, 18* (6), 431-443.

Goffman, E. (1963). *Behavior in public places: Notes on the social organization of gatherings.* New York: Free Press.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* New York: Academic Press.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104-137.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring sessions. *Applied Cognitive Psychology, 9*, 1-28.

Grayson, D., & Coventry, L. (1998). The effects of visual proxemic information in video mediated communication. *SIGCHI Bulletin, 30*(3), 30-39.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J.L. Morgan (Eds.), *Syntax and Semantics 3 : Speech Acts* (pp. 41-58). New York : Academic Press.

Hall, E.T. (1971). *La dimension cachée.* Seuil : Paris.

Haviland, S.E., & Clark, H.H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13*, 512-521.

Hirst, G., McRoy, S., Heeman, P., Edmonds, P., & Horton, D. (1994). Repairing conversational misunderstandings and non-understandings. *Speech Communication, 15*, 213-229.

Hoc, J.M. (1996) *Supervision et contrôle de processus : la cognition en situation dynamique.* Presses Universitaires de Grenoble.

Hoc, J-M., & Amalberti, R. (1994) Diagnostic et prise de décision dans les situations dynamiques. *Psychologie Française, 39*(2), 177-192.

Holland, J., Hutchins, E., & Kirsch, D. (2000). Distributed cognition : toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction, 7*, (2), 174-196.

Horaek, H., & Fiedler, A. (2001). Towards understanding the role of hints in tutorial dialogs. In P. Kühnlein, H. Rieser & H. Zeevat (Eds.), *Bi-Dialog 2001* (pp.40-46).

Howell, D.C. (1998). *Méthodes statistiques en sciences humaines*. Bruxelles : De Boeck.

Huang, W., Olson, J.S., & Olson, G.M. (2002). Camera angle affects dominance in video-mediated communication. In *Conference on Human Factors in Computing Systems – CHI 2002* (pp. 716-717). New York, NY : ACM Press.

Hume, G.D., Michael, J.A., Rovick, A.A., & Evens, M.W. (1993). The use of hints as a tutorial tactic. In M.C. Polson (Ed.), *Proceedings of the fifteenth annual conference of the Cognitive Science Society* (pp.563-568). Hillsdale, NJ : Lawrence Erlbaum.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology : General*, 116, 26-37.

Isaacs, E., & Tang, J.C. (1994). What video can and cannot do for collaboration. *Multimedia Systems, 2*, 63-73.

Ishii, H., Kobayashi, M., & Grudin, J. (1993). Integration of interpersonal space and shared workspace: Clearboard design and experiments. *ACM Transactions on Information Systems (TOIS), 11*(4), 349 – 375.

Joiner, R., Scanlon, E., O'Shea, T., Smith, R.B., & Blake, C. (2002). Evidence from a series of experiments on video-mediated collaboration : does eye contact matter ? In G.Stahl

(ed.) *Proceedings of Computer Supported Collaborative Learning* (pp.34-45). Denver, Co. : Lawrence Erlbaum.

Kalyuga, S., Chandler, P., & Sweller, J., (1999). Managing split attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.

Karsenty, L. (1999). Cooperative work and shared visual context : an empirical study of comprehension problems in side-by-side and remote help dialogues. *Human Computer Interaction, 14*, 283-315.

Karsenty, L., & Falzon, P. (1993). L'analyse des dialogues orientés tâche : introduction à des modèles de la communication. In F. Six, X. Vaxevanoglou (Eds). *Les aspects collectifs du travail* (pp.107-118). Actes du XXVIIème Congrès de la Société d'Ergonomie de Langue Française. Toulouse : Octares.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica, 26*, 22-63.

Kerbrat-Orecchioni, C. (1998). *Les interactions verbales.* Paris : Armand Colin.

Kies, J.K. (1997). *Empirical methods for evaluating video-mediated collaborative work.* PhD thesis, University of Virginia, USA.

Kowtko, J., Isard, S., & Doherty-Sneddon, G. (1991). *Conversational games in dialogue*. HCRC Technical Report, University of Edinburgh.

Kraut, R., Miller, M., & Siegel, J. (1996), Collaboration in performance physical tasks : effects on outcomes and communication. In *Proceedings of CSCW'96* (pp.57-66). New York:ACM.

Kraut, R.E., Gergle, D. & Fussell, S.R. (2002). The Use of Visual Information in Shared Visual Spaces : Informing the Development of Virtual Co-Presence. In *Proceedings of CSCW'02* (pp.31-40). New York : ACM Press.

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human Computer Interaction, 18*, 13-49.

Krippendorf, K. (1980). *Content analysis: an introduction to its methodology*. Sage Publications.

Lacoste, M. (1991). Les communications de travail comme interactions. In R. Amalberti, M. de Montmollin, & J. Theureau (Eds.), *Modèles en analyse de travail* (pp.191-227). Liège : Mardaga.

Laurillard, D. (1993). *Rethinking university teaching: A framework for the effective use of educational technology*. London: Routledge.

Le Breton, D. (1998). *Les passions ordinaires. Anthropologie des émotions*. Paris : A. Colin.

Leplat, J. (1991). Organization of activity in collective tasks. In J. Rasmussen, B. Brehemer & J. Leplat (Eds.), *Distributed Decision Making* (pp. 51-74). UK : John Wiley & Sons.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J. L. (1993). Motivational techniques of expert human tutors : Lessons for the design of computer-based tutors. In S. P. Lajoie & S. Derry (Eds.), *Computers as cognitive tools* (pp. 75-105). Hillsdale, NJ: Lawrence Erlbaum.

Lewis, D.K. (1969). *Convention. A Philosophical Study*. Cambridge, MA : Harvard University Press.

Mackay, W. (1999). Media Spaces : Environments for informal multimedia interaction. *Computer Supported Cooperative Work , 1,* 55-81.

Masoodian, M., Apperley, M., & Frederickson, L. (1995). Video support for shared work-space interaction : an empirical study. *Interacting with Computers, 7* (3), 237-253.

Matarazzo, G., & Sellen, A. (2000). The value of video in work at a distance : addition or distraction? *Behavior & Information technology, 19* (5), 339-348.

McCarthy, J.C., & Monk, A.F. (1994). Measuring the quality of computer mediated communication. *Behaviour & Information Technology, 13* (5), 311-319.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

McNeill, D. (1992). *Hand and mind. What gestures reveal about thought*. Chicago : Chicago University Press.

Minneman, S.L., & Bly, S.A. (1991). Managing à trois : a study of a multi-user drawing tool in distributed design work. In *Proceedings of the CHI'91 Conference on Human Factors in Computing Systems* (pp.217-224). NY, Usa : ACM.

Minsky, M. (1980). Telepresence. *Omni*, 45-51.

Moeschler, J. (1985). *Argumentation et conversation. Éléments pour une analyse pragmatique du discours*, Hatier : Crédif.

Monk, A., Nardi, B., Gilbert, N., Mantei, M., and McCarthy, J. (1993): Mixing Oil and Water? Ethnography versus Experimental Psychology in the Study of Computer-Mediated Communication. *Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems* (pp.3-6). ACM Press.

Monk, A., McCarthy, J., Watts, L., & Daly-Jones, O. (1996). Measures of process. In M.MacLeod & D. Murray (Eds.) *Evaluation of CSCW* (pp.125-139). Berlin : Springer-Verlaag.

Monk, A. F. & Watts, L. A. (2000). Peripheral participation in video-mediated communication. *International Journal of Human-computer Studies, 52*, 775-960.

Monk, A.F., & Gale, C. (2002). A look is worth a thousand words: full gaze awareness in video-mediated conversation. *Discourse Processes, 33*, 257–278.

Mühlbach, L., Bocker, M. & Prussog, A. (1995). Telepresence in Videocommunications: A Study on Stereoscopy and Individual Eye Contact. *Human factors, 37*, 290-305.

Mullin, J., Smallwood, L., Watson, A., & Wilson, G. M. (2001) New Techniques for Assessing Audio and Video Quality in Real-Time Interactive Communication. In J.Vanderdonckt, A. Blandford & A. Derycke (eds.) *Proceedings of IHM-HCI 2001*, (pp. 221-222). Lille, France.

Munhall, K.G., Gribble, P., Sacco., L. & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics, 58*(3), 351-362.

Nardi, B., Kuchinsky, A., Whittaker, S., Leichner R., & Schwarz, H. (1997). Video-as-data : technical and social apsects of a collaborative multimedia application. In In K.E. Finn, A.J. Sellen, & S.B. Wilbur. (Eds.). *Video-mediated communication* (pp.487-517). New Jersey : Lawrence Erlbaum Associates.

Navarro, C. (1999). Rôle des processus métacognitifs dans l'interaction fonctionnelle. In K. Kostulski & A. Trognon (Eds.) *Communications interactives dans le groupe de travail* (pp.235-258). Nancy : Presses Universitaires de Nancy.

Navarro, C. (2001). Partage de l'information en situation de coopération à distance et nouvelles technologies de la communication : bilan des recherches récentes. *Le Travail Humain,64*(4), 297-319.

Nunnari, F., & Simone, C. (2002). ThreeDness: representing awareness in cooperative applications. In M. Blay-Fornarino, A-M. Pinna-Dery, K. Schmidt & P. Zaraté (Eds.), *Proceedings of the COOP 2002* (pp.7-22). Saint Raphaël, France.

O'Conaill, B., Whittaker, S., & Wilbur, S. (1993). Conversations over videoconferences : an evaluation of videomediated interaction. *Human-Computer interaction, 8*, 382-428.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V. (1996). Comparison of face-to-face and video-mediated communication. *Interacting with computers*, 8, 2, 177-192.

O'Connell, D.C., Kowal, S., & Kaltenbacher, E. (1990). Turn-taking : a critical analysis of the research tradition. *Journal of Psycholinguistic Research, 19*(6), 345-373.

Ochsman, R. B., & Chapanis, A. (1974). The effects of 10 communication modes on the behavior of teams during co-operative problem-solving. *International Journal of Man-Machine Studies, 6*, 579-619.

Olson, G.M., Olson, J.S., & Meader, D.K. (1995). What mix of video and audio is useful for small groups doing remote real-time design work? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 362-368). Denver,CO : ACM Press.

Olson, G.M., & Olson, J.S. (1997). Making sens of the findings : common vocabulary leads to the synthesys necessary for theory building. In K.E. Finn, A.J. Sellen, & S.B. Wilbur. (Eds.). *Video-mediated communication* (pp.75-91). New Jersey : Lawrence Erlbaum.

Ostberg, O., Lindstrom, B., Renhall, P. (1989). Contribution of display size to speech intelligibility in videophone systems. *International Journal of Human-Computer Interaction, 1*(1), 149-159.

Otteson, J.D., & Otteson, C.R., (1980). Effect of teacher's gaze on children's story recall. *Perception & Motor Skill, 50*, 35–42.

Paivio, A. (1986). *Mental representations : a dual coding approach.* New York : Oxford University Press.

Pavard, B. & Karsenty, L. (1997). Différents niveaux d'analyse du contexte dans l'étude ergonomique du travail collectif. *Réseaux, 85*, 5, 73-100.

Pernin, J.P. (2000). *Le projet FORMID (FORMation Interactive à Distance)*. Réunion du GDR I3/IHMC. Université Paris V.

Person, N. K., & Graesser, A.C. (2003). Fourteen facts about human tutoring : food for thought for ITS developers. In V. Aleven, U. Hoppe, J. Kay, R. Mizoguchi, H. Pain, F. Verdejo, & K. Yacef (Eds.), *AIED2003 Supplemental Proceedings* (pp. 335-344). Sydney, Australia: University of Sydney School of Information Technologies.

Pilkington, R. M. (1992). Question-answering for intelligent on-line help: the process of intelligent responding. *Cognitive Science (4),* 455-491.

Porayska-Pomsta, K., Mellish, C., & Pain, H. (2000). Aspects of Speech Act Categorisation: Towards Generating Teachers' Language. *International Journal of Artificial Intelligence in Education 11*, 254-272.

Prates, R.O., de Souza, C.S., & Barbosa, S.D.J. (2000) A method for evaluating the communicability of user interfaces. *Interactions, 1*, 31-38.

Reboul, A., & Moeschler, J. (1998). *La pragmatique aujourd'hui. Un nouvelle science de la communication*. Paris : Seuil.

Reid, A. (1977). Comparing the telephone with face-to-face interaction. In I. Pool (Ed.). *The social impact of the telephone* (pp. 386-414). Cambridge, MA: MIT Press.

Richard, J.-F. (1990). *Les activités mentales*. Paris : Armand Colin.

Rutter, D.R., & Stephenson, G.M. (1977). The role of visual communication in synchronising conversation. *European Journal of Social Psychology, 2,* 29-37.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking in conversation. *Language, 50*, 696 - 735.

Salber, D. (1995). *De l'interaction homme-machine individuelle aux systèmes multi-utilisateurs. L'exemple de la communication homme-homme médiatisée*. Thèse de Doctorat, Université Joseph Fourier, Grenoble.

Salembier, P. (1993). *Le diagnostic cognitif dans les interactions tutorielles*. Thèse de Doctorat CNAM, Paris.

Sanford, A., Anderson, A.H., & Mullin, J. (2004). Audio channel constraints in video-mediated communication. *Interacting with Computers, 16*, 1069-1094.

Savoyant, A., & Leplat, J. (1983). Statut et fonction des communications dans l'activité des équipes de travail. *Psychologie Française, 28*(3/4), 247-253.

Schiffer, S.R. (1972). *Meaning*. Oxford : Oxford University Press.

Schlegoff, E.A. (1987). Conversation analysis and socially shared cognition. In L. B. Resnick, J. Levine, & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 150-171). Washington, DC : American Psychological Association.

Schmidt, K. (1994). Cooperative work and its articulation : requirements for computer support. *Le Travail humain, (57)*, 4, p. 345-366.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*, 211-232.

Searle, J. (1972). *Les actes de langage*. Paris : Hermann.

Sebillotte, S. (1984). La résolution de problème en situation de diagnostic. Un exemple : le diagnostic medical. *Psychologie Francaise, 29*(3/4), 273-277.

Sellen, A.J. (1995). Remote conversations : the effects of mediating talk with technology. *Human-Computer Interaction, 10*, 401- 444.

Settouti, L.S. (2006). Systèmes à base de trace pour l'apprentissage humain. In *1ères Rencontres Jeunes Chercheurs sur les Environnements Informatiques pour l'Apprentissage Humain* - ATIEF. 11-12 Mai 2006, Evry.

Short, J., Williams, E., & Christie, B. (1976). *The Social Psychology of telecommunications*. London: Wiley.

Siegel, S., & Castellan, J.N. (1988). *Non parametric statistics for the behavioral sciences*. New York : McGraw-Hill.

Slater, M., Sadagic, A., & Schroeder, R. (2000). Small-group behaviour in a virtual and real environment : a comparative study. *Presence : Teleoperators and Virtual Environments, 9*(1), 37-51.

Smith, R., O'Shea, T., O'Malley, C., Scanlon, E., & Taylor, J. (1991). Preliminary experiments with a distributed multimedia problem solving environment. In J.Bowers & S. Benford (Eds.) *Studies in Computer-Supported Cooperative Work* (pp.31-48). Elsevier.

Soury-Lavergne, S. (1998). *Etayage et explication dans le préceptorat distant : le cas de TéléCabri*. Thèse de Doctorat, Université Joseph Fourier, Grenoble.

Sperber, D., & Wilson, D. (1989). *La pertinence. Communication et cognition*. Paris : Les Editions de Minuit.

Stefik, M., Foster, G., Bobrow, D., Kahn, K., Lanning, S., & Suchman, L. (1987). Beyond the chalkboard: computer support for collaboration and problem solving in meetings. *Communications of the ACM, (30)*, 1, pp. 32-47.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition & Instruction, 12,* 185-233.

Tajariol, F., Adam, J.-M. & Dubois, M. (2003). A study on the effect of communication and monitoring tools in web-based tutoring. *International Journal of Computers and Applications*, 25 (3), 206-211.

Tajariol, F., Dubois, M., & Adam, J.-M. (2002). Mediated-communication in tutoring : When the face isn't enough. In *Proceedigs of the 5th IASTED International Conference on Computer and Advanced Technology in Education (CATE 2002)*, Cancun, Mexico.

Tang, J. C. (1991). Findings From Observational Studies of Collaborative Work. *International Journal of Man-Machine Studies, 34,* 143-160.

Tang, J.C., & Isaacs, E. (1993). Why do users like video video ? Studies of multi-media supported collaboration. *Computer supported cooperative work, 1,* 3, 163-196.

Tcherkassof, A., Bollon, T., Dubois, M., Pansu, P., Paignon, A., & Adam, J.-M. (2004). Reconnaissance des expressions émotionnelles faciales spontanées et dynamiques. *Cinquième Congrès International de Psychologie Sociale en Langue Française*. Lausanne, Suisse, 1-4 Septembre 2004.

Traum, D. R. (1998). On Clark and Schaefer's *contribution* model and its applicability to human-computer collaboration. In *Proceedings of COOP'98 Workshop*. Sophia Antipolis : France.

Trevino, L.K., Lengel, R.H., & Daft, R.L. (1987). Media symbolism, media richness, and media choice in organizations. *Communications Research, 14*(5), 553-574.

Van Lehn, K., Siler, S., Murray, C., & Bagget, W.B. (2003). Human tutoring : why do only some events cause learning ? *Cognition and Instruction, 21*(3), 209-284.

Velichkovsky, B.M., (1995). Communicating attention: gaze position transfer in cooperative problem solving. *Pragmatics & Cognition, 3*, 199–222.

Vertegaal, R., van der Veer, G., & Vons, H. (2000). Effects of gaze on multiparty mediated communication. In *Proceedings of Graphics Interface* (pp.95-102). Montreal, Canada: Morgan Kaufmann Publishers.

Vertegaal, R., Slagter, R., Van der Veer, G.C., & Nijholt, A. (2001). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *ACM Conference on Human Factors in Computing Systems.* Seattle, WA : ACM Press.

Vitkovitch, M., & Barber, P. (1994). Effect of Video Frame Rate on Subjects' ability to Shadow One of Two competing verbal Passages. *Journal of Speech and Hearing, 37,* 1204-1210.

Wagemann, L., & Percier, M. (1995). *Contribution à l'étude de la formation à la gestion des processus continus : le cas de l'entraînement sur simulateur machine des élèves officiers de la marine marchande*. Thèse de Doctorat, EHESS Paris.

Watson, A. (1999). *Assessing the quality of audio and video components in desktop multimedia conferencing*. Ph.D. Thesis, Department of Computer Science, University College London.

Watts, L.A., Monk, A.F., & Daly-Jones, O. (1996). Interpersonal awareness and synchronization : assessing the value of communication technologies. *International Journal of Human-Computer Studies, 44,* 849-875.

Wenger, E. (1987). *Artificial Intelligence and Tutoring systems*. Los Altos, CA : Morgan Kaufmann Publishers.

Whittaker, S. (2002). Theories and methods in mediated communication. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123-164). Hillsdale, NJ: Lawrence Erlbaum.

Whittaker, S., Brennan, S.E., & Clark, H.H. (1991). Co-ordinating actvity : an analysis of interaction in computer-supported co-operative work. In S. Robertson, J. Olson & G. Olson (Eds.) *Proceedings of CHI'91 Human Factors in Computing Systems* (pp. 360-367). New Orleans, USA: ACM Press.

Whittaker, S., Geelhoed, E., & Robinson, E. (1993). Shared workspaces: how they work and when are they useful? *International Journal of Man-Machine Studies, 39,* 813-842.

Whittaker, S., & O'Conaill, B. (1997). The role of vision in face-to face and medited communication. In K.E. Finn, A.J. Sellen, & S.B. Wilbur. (Eds.). *Video-mediated communication* (pp. 23-49). New Jersey : Lawrence Erlbaum.

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication. *Psychological Bulletin, 84,* 963-976.

Wynnikamen, F. (1998). Approche Psychologique de la tutelle. In Dumas Carré, A., & Weil-Barais, A. (Eds.) *Tutelle et médiation dans l'éducation scientifique* (pp.29-58). Bern: Peter Lang.

Yngve, V. (1970). On getting a word in edgewise. *Papers from the Sixth Regional Meeting Chicago Linguistic Circle*. Chicago, IL : Chicago Linguistic Circle.

Zhang, J., & Norman, D.A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18* (1), 87-122.

Liste des tableaux

Tableau 1-1 : Les fonctions principales des indices non-verbaux en situation de communication face-à-face. 23
Tableau 2-1 : Les effets de l'image-vidéo du partenaire distant sur les interactions communicatives et les impressions subjectives. 47
Tableau 2-2 : Les effets de l'image-vidéo relative au visage *vs.* au visage & buste. 53
Tableau 2-3 : Les effets du regard mutuel et de la connaissance de la direction du regard du partenaire distant. 66
Tableau 2-4 - Synthèse des effets des facteurs expérimentaux relatifs à l'image-vidéo du partenaire distant. 67
Tableau 3-1 : Les effets de l'espace visuel commun sur les interactions communicatives et sur la réalisation de la tâche. 91
Tableau 3-2 : Les effets de l'espace de travail commun. 97
Tableau 3-3 : Synthèse des effets de l'espace visuel commun et de travail commun. 98
Tableau 4-1 : Items du pré-test pour les apprenants 106
Tableau 4-2 : Tableau récapitulatif des indices verbaux utilisés par l'apprenant. 115
Tableau 4-3 : Tableau récapitulatif des indices verbaux utilisés par le tuteur. 118
Tableau 4-4 : Les items relatifs à l'appréciation du tuteur sur la qualité de la contribution de l'apprenant 122
Tableau 4-5 : Les items relatifs à l'appréciation de l'apprenant sur la qualité de la contribution du tuteur 122

Tableau 4-6 : Les items relatifs à l'appréciation du tuteur sur la qualité de sa propre contribution. 123
Tableau 4-7 : Les items relatifs à l'appréciation de l'apprenant sur la qualité de sa propre contribution. 123
Tableau 4-8 : Matrice des correlations entre les items relatifs au jugement du tuteur sur l'apprenant.. 135
Tableau 4-9 : Matrice des correlations inter-items des impressions centripètes des tuteurs. ...136
Tableau 5-1: Matrice des corrélations inter-items des impressions centrifuges des tuteurs 159
Tableau 5-2 : Matrice des corrélations inter-items des impressions centripètes des tuteurs. ...162

Liste des figures

Figure 2-1 : Le regard latéral vertical (à gauche) et le regard latérak horizontal (à droite). 54
Figure 2-2 : L'écran du sujet-volontaire en Vertegaal & al. (2000). 62
Figure 3-1 : Dispositif expérimental (Kraut & al., 2003) 80
Figure 3-2 : L'espace visuel commun en Gergle & al. (2004). 84
Figure 3-3 : A gauche : le novice en train de construire le robot. A droite : l'écran de l'expert avec l'espace visuel commun fixe (en haut à droite) et mouvant (en bas à droite) (Fussell & al., 2003) 87
Figure 4-1 : Schéma technique de la situation expérimentale 108
Figure 5-1 : Image relative à l'écran de l'ordinateur du tuteur : la partie supérieure est réservée à l'apprenant A, la partie inférieure à l'apprenant B. 145
Figure 6-1 : La forme de visualisation analogique utilisée dans le projet Topase 172
Figure 6-2 : La forme de visualisation symbolique de FORMID 172

This document was created with Win2PDF available at http://www.win2pdf.com.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.
This page will not be added after purchasing Win2PDF.