

Guy Barrier, Jean Caelen, Brigitte Meillon
Laboratoire CLIPS-IMAG, BP 53, Domaine universitaire, 38041 Grenoble Cedex

La visibilité des gestes : paramètres directionnels, intentionnalité du signe et attribution de pertinence

L'étude présente a été menée au CLIPS parallèlement à la conception d'un ACE, l'hôtesse virtuelle Angela dans le cadre du projet Agents Conversationnels Expressifs financé par le RIAM . Elle avait pour but de s'interroger sur la visibilité des gestes du personnage, à partir du film réalisé pour constituer une banque de gestes. Bien que notre étude visuelle se soit déroulée de façon autonome pendant la modélisation de l'agent, l'analyse de scène gestuelle soulève différentes questions transposables aux artefacts humains : que perçoit-on d'une représentation gestuelle et comment peut-elle influencer notre attention ? On sait par exemple que du côté des gestualistes, plusieurs controverses ont été partagées à propos de la visibilité des gestes manuels (et donc leur valeur communicationnelle) . En contraste la pertinence de l'expression faciale n'a jamais été mise en doute et a été largement validée.

Une des fonctions attendue des gestes d'un agent, serait d'être un des ingrédients de la multimodalité expressive et de donner du réalisme à la scène (*réention de l'attention*) . Une autre fonction qui nous intéresse ici est que certains gestes peuvent de manière plus incitative, fonctionner en quelque sorte comme des projecteurs directionnels et guider le regard vers telle direction cardinale de l'espace graphique (*redirection de l'attention*). Une question qu'on peut alors se poser à propos de tout agent est : est-ce qu'il va capturer (voire accaparer) l'attention sur lui, ou est-ce qu'il fonctionne comme transition pour rediriger l'intérêt vers d'autres stimuli ? Cette balance entre les deux foyers d'attention (qui ne s'excluent pas mutuellement), dépend notamment du type de gestes réalisés.

La visibilité des gestes dans les interactions humaines

Dans le cadre de leurs recherches sur la visibilité des gestes Gullberg et Holmqvist (2001, 2003) ont constaté que les gestes étaient peu fixés par les interlocuteurs (moins de 7% de fréquences) tandis que l'attention centrale allait vers le visage . Il indiquent également que les gestes latéraux sont fixés trois fois plus que les gestes autocentrés (cette distinction est morphologique, nous essaierons de la revisiter sur un plan sémantique) . Les auteurs signalent la possibilité d'un traitement parallèle des gestes en périphérie, compétences également remarquées en langage des signes par Muir & al. (2003) : le visage du signeur est regardé par le déchiffreur mais celui-ci traite les gestes entraperçus sur le côté . Les théories de la vision proposent qu'au-delà de 5 degrés de la fovéa, la résolution visuelle quoique plus faible, peut caractériser grossièrement une forme globale. Par exemple, la vision parafovéale d'un mot accélère son déchiffrage lorsqu'il est ensuite fixé au niveau focal . Au niveau des gestes les mouvements pourraient être traités par des routines périphériques en basse résolution visuelle (attention exogène) pendant que l'on regarde le visage. Cette information traitée à un niveau infra-conscient pourrait contribuer à la prise de décision du regard en tant que contrôle d'arrière plan.

Etant supposé que dans toute scène peuvent coopérer ces deux niveaux d'attention (mais sachant la complexité de mesurer l'attention périphérique) notre propos serait de montrer dans quelles situations gestuelles l'attention centrale se place volontairement vers les gestes . Pour cela deux expériences ont été menées : la première à partir de séquences vidéo filmées, la seconde à partir de captures d'images

extraites de ce film . L'utilisation d'images en 2^e partie a été inspirée par plusieurs motivations : les images de personnages sont utilisées par des médias tels que le Web, le pictogramme , la publicité. Nous présumons que si l'image gestuelle démontre une certaine typicalité elle pourrait fonctionner comme une métaphore de la séquence filmée, et générer grâce à ses indices visuo-moteurs des inférences spatio-temporelles (telles que tonicité, amplitude, voire parfois : vitesse, soudaineté ou discontinuité du geste) . En outre les analyses oculométriques de gestes mis en image sont peu courantes . Enfin, les avatars d'agents animés se développent sous des formes très variées.

Pour analyser le comportement oculaire selon les gestes nous avons auparavant codifié les occurrences en exploitant les typologies de Cosnier (1997) et Mc Neill (1992). Ces derniers distinguent , avec quelques nuances théoriques :

- les gestes référentiels . Cosnier classe ici les **iconiques** : pictographiques (qui dessinent le référent), les spatiographes (relations spatiales), les kinémimiques qui miment une action, et les **déictiques**. Mc Neill met l'accent sur les métaphoriques , dont le contenu pictural permet à la pensée abstraite de se représenter dans l'espace. Par convention nous affecterons ces derniers dans la classe des iconiques.
 - les paraverbaux calqués sur le flux verbal qui incluent les battements prosodiques (bâtons) , ainsi que les gestes cohésifs de Mc Neill associés aux marqueurs grammaticaux .
 - les extra communicatifs (dits aussi régulateurs) qui apparaissent dans les phases d'inconfort verbal tels que les gestes d'autocontact .
- Les quasi linguistiques (*emblems*) et gestes synchronisateurs (coordination interactive avec l'interlocuteur) absents de notre extrait filmé, n'ont pas été analysés.

Nous nous demanderons donc si certaines catégories de gestes sont plus perçues que d'autres, ce qui d'un point de vue pragmatique revient à se demander quels sont les gestes les plus utiles dans la communication humaine ou en IHM .

Première expérience – Gestes en mode vidéo

Les 16 sujets à part égale des deux sexes, recrutés dans un laboratoire partenaire, avaient pour consigne de visionner un casting d'une durée de 5 minutes, filmé quelques semaines auparavant . Ils sont informés qu'ils devront répondre ensuite à une questionnaire. Dans ce film l'actrice jouait le rôle du futur avatar afin de lui prêter ses gestes pour numérisation, un écran blanc avait été choisi comme arrière-plan neutre. Elle devait donc entretenir un dialogue avec un « client en ligne » ; ce dernier était hors du champ visuel mais entendu en voix off, la caméra étant dirigée vers l'actrice seule . Les sujets sont informés qu'ils devront à la fin, répondre à un questionnaire (ce questionnaire nous servira en 2^e expérience). Le film montre l'actrice en train de simuler une commande en ligne de cosmétiques avec une cliente internautes. Les objets de promotion sont absents mais actualisés au niveau narratif. Les fixations oculaires sont enregistrées et extraites au seuil de 150 ms à l'aide d'un eye tracker dont le logiciel de traitement visuel et statistique a été développé par Brigitte Meillon (équipe MultiCom).

Visibilité et catégories gestuelles

Nous remarquons tout d'abord que l'expression faciale recueille 81% des fixations, alors que 12% sont dirigées vers les gestes ou leur prolongement et 7% vers le buste de l'actrice¹ . Nous avons esquissé dans un article précédent le mode d'inspection du visage (Barrier, 2004) et nous intéressons ici plus spécialement aux gestes manuels. Les déictiques sont vus dans 86% des cas où ils apparaissent à l'écran.

¹ Les saccades hors écran ont été éliminées.

Les iconiques sont remarqués dans 78% des situations . Les battements prosodiques et marqueurs intonatifs du discours ont une faible chance d'être vus (14 % en moyenne) mais ils ont plus de succès lorsque le geste est expansif (valeur d'emphase) que lorsqu'il est centré près du buste. Les extra-communicatifs (tortillement des mains, autocontact) qui régulent les tensions discursives du locuteur, ont moins d'une chance sur dix d'être regardés (8%) . Cependant lorsque leur aspect est inattendu ils réactivent l'attention des sujets: par exemple, incidemment l'actrice se pince la lèvre moins d'une seconde mais ce geste atypique déclenche un regain d'intérêt .

Par regroupement les signes gestuels les plus en vue sont donc les gestes référentiels, qui renvoient à des objets conceptuels par le pointage (déictiques), l'évocation des contours spatiaux du référent (spatiographes) , le mime d'actions sous forme de gestes techniques. En somme, les gestes communicatifs (relatifs à l'objet du discours) intéressent davantage l'attention focale que les extra-communicatifs (relation du locuteur à son énoncé). Ces fréquences « brutes » nous permettent de supposer quels sont les gestes qui suscitent le plus d'attention dans tel contexte. Mais en déroulant les films des trajectoires oculaires, l'observation du mouvement spatio-temporel nous permet de faire davantage d'inférences en matière d'intérêt du spectateur (finesse des inspections, réinspections, durée, etc.).

Ainsi au niveau du décours des tracés oculaires, nous constatons que si les gestes sont centrifuges et si la tête et le buste sont orientés latéralement, l'attention tend alors à se déployer dans la même direction (nous remarquons alors des trajectoires du regard centrifuges). A l'opposé lorsque la posture et le visage font face à la caméra et si les gestes sont de courte amplitude, nous obtenons des trajectoires du regard courtes et centripètes. Nous remarquons également que les déictiques de désignation se traduisent en trajectoires oculaires par une chaîne directionnelle de points de fixation . En comparaison les iconiques montrent plutôt des configurations spatiales analogues à la forme de l'objet référentiel et qui modélisent ses contours et points extrêmes.

Le couplage geste-regard dans les ostensions déictiques

Les psychologues parlent de vision ou attention **conjointe** lorsque quelqu'un attire l'attention d'autrui vers un objet au moyen du regard (intérêt partagé en commun avec le référent). Dans notre série d'observations, la configuration la plus efficace pour rediriger l'attention est, selon les mouvements oculaires, celle qui démontre une projection multimodale vers la cible : buste et visage de profil, geste et œil déictiques, orientés en convergence dans le sens du référent actualisé par l'énoncé .

Dans la situation typique suivante, deux information principales sont inter-connectées par le spectateur : la visée déictique du regard et le geste de pointage qui prolonge cette visée (fig.1). L'axe de la tête du personnage, la rotation de la pupille sont saisis comme indices déictiques conduisant vers le geste ; nous constaterons le même phénomène sur le corpus d'images . Le regard du sujet peut même anticiper le point de chute présumé de la trajectoire balistique (fig.2).



fig.1 connexion pupille-geste



fig.2 visée déictique et pointage

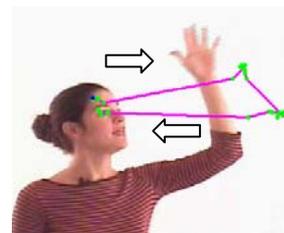


fig.3 aller-retour gestuel et poursuite oculaire

Grâce à cette deixis multimodale, la projection d'intention de l'actrice est modélisée par le regardeur qui se laisse guider vers une cible imaginaire (l'actrice désigne des objets absents mais actualisés au niveau narratif). Dans le contexte représenté fig. 3 : déploiement puis repli du geste, les mouvements des yeux révèlent une forte parenté structurelle avec ces lignes de force directionnelles. L'énergie sensori-motrice est modélisée par le regard du sujet qui en épouse les orientations vectorielles. La synchronisation spatiale et temporelle de la poursuite oculaire, est aussi nette dans la phase de déploiement centrifuge que dans la phase de repli centripète de la posture. En suivant du regard le geste du locuteur qui vise une direction, les spectateurs partagent en miroir son intentionnalité et construisent à partir de repères spatiaux un modèle mental commun.

Les gestes iconiques : analogie de la forme spatiale et de la poursuite oculaire

Comme indiqué les iconiques entraînent des mouvements des yeux analogues à la forme de l'objet référentiel et qui modélisent ses contours et points extrêmes, ce qui procure des indices sur l'imagerie dynamique du receveur. Dans une des séquences narratives, l'actrice décrit un produit cosmétique et mime le geste de saisir ce flacon. La saisie conceptuelle du référent se réalise dans l'espace par la forme qui semble émerger de ses mains. Pour mieux décliner l'offre, elle propose ensuite trois formats de flacon métaphorisés par une partition dans l'espace (petit, moyen, grand format.) Ce découpage spatial est suivi de mouvements oculaires de même amplitude et intervalles que les gestes (fig. 4, 6). Les scanpaths s'organisent selon les propriétés physiques de l'objet virtualisé, le modèle mental du spectateur semble analogue au référent dessiné dans l'espace. De même dans une autre séquence l'actrice décrit un produit cosmétique en forme de stick dont elle mime la forme manuellement (fig. 5) Puis elle explique que cet article peut être glissé facilement dans un sac à main et joint le geste à la parole par une simulation (fig.6). Le regard des spectateurs se pose fidèlement sur les gestes kinémimiques, et modélise le geste renforcé par l'énoncé.



fig. 4 à 6 : expression co-verbale de propriétés référentielles (taille, forme...) modélisées par l'œil du destinataire

Ces observations suggèrent qu'un modèle mental de la tâche ou de l'objet référentiel, est construit par le receveur, en synchronie avec les cognitions de l'émetteur sur la base d'une ressemblance spatiale et motrice. La précision de la prise d'information est révélée par la parenté structurelle entre la forme manuelle et les scanpaths oculaires.

2e expérience – Gestes en capture d'image

En sémiologie visuelle Gandelman (1986) a développé qu'un guidage attentionnel pouvait être exercée par les gestes déictiques représentées sur des oeuvres picturales. Quand les postures, gestes et regards convergent vers la même direction dit-il (et pour peu que s'en mêle un contraste de lumière), un foyer

d'attention est partagé dans le sens de cette visée iconique. Il nous a semblé intéressant de tester sur d'autres supports iconographiques ces théories : le geste en image peut-il capturer l'attention et la déplacer vers telle direction ?

Dans cette seconde expérience, nous avons présenté à 25 étudiants en psychologie (13 F, 12 H), des images fixes retirées du casting précédent sous forme de diapositives présentées trois secondes et séparées par une vue neutre . Les images ont été sélectionnées de façon à représenter au mieux leur classe d'appartenance, afin de suggérer par l'indice le geste dans son mouvement originel . Les sujets sont installés devant un écran 17 pouces incorporé dans l'oculomètre Tobii 1750 . Ils ont pour consigne de visionner les diapositives représentant notre même personnage, et sont informés que des questions leur seront posées ensuite. Il n'y a pas de source sonore mais un sous-titre d'une trentaine de lettres apparaît sous les images, qui prête parole à l'actrice et contextualise les gestes. Son déchiffrage occupe en moyenne 0.8 secondes d'attention. Hormis la première image où le sujet découvre l'actrice, la lecture de ce texte est opérée en entrée immédiate, puis le regard remonte vers le visage ou les gestes . On n'observe pratiquement pas de séquences de régression de l'image vers le texte, simplifié à dessein pour être stocké en mémoire . Nous nous intéressons alors à la manière dont le regard parcourt l'image .

Pour évaluer la visibilité des gestes manuels par paires comparatives nous avons testé les fréquences recueillies à l'aide du Khi deux. En fonction de chacune des deux variables , les modalités sont alors : geste regardé en moins de deux secondes (1) ou geste non regardé (0) dans ce délai.

Effectifs observés			Effectifs théoriques attendus				
	0	1			0	1	
geste de face	305	95	400	72,7%	geste de face	263	137
geste de profil	57	93	150	27,3%	geste de profil	99	51
	362	188	550			362	188

Tableau 1 - Exemple de table des écarts utilisée dans l'expérience 2

Ces variables ont été élaborées en nous appuyant sur les jugements des sujets de l'expérience précédente, à qui nous avons demandé de classer les captures d'image selon des qualificatifs . Les appariements nous ont permis de constituer des critères différentiels, exprimés par les sujets sous l'aspect visuel ou morphologique . Une première catégorisation des réponses nous permettra de différencier :

- les gestes de faible résolution motrice et spatiale, qualifiés souvent de « gestes aléatoires ou parasites », qui « balayent l'espace de manière diffuse »,
- les signes de granularité spatiale plus fine, dont les mouvements et proportions sont contrôlés par l'actrice, ayant l'aspect de « gestes plus précis », « émis volontairement », gestes jugés plus techniques et opératoires , mieux articulés au niveau de la forme .

La saillance des gestes référentiels et articulés

La première catégorie qu'ils décrivent traverse en fait deux classes sémantiques, les extra-communicatifs (dont le contrôle échappe au locuteur) et les prosodiques ou bâtons, de nature semi-intentionnelle. Ces gestes (ex. fig.9) n'expriment pas d'intention de saisie performative ni de visée désignative mais servent comme marqueurs métalinguistiques ou comme indices phatiques. Ils sont identifiés en mode capture d'image par nos sujets, déjà sensibilisés par leur cursus à la communication.

Le test suggère une forte dépendance ($\chi^2 = 58.85$ $p = 0$) entre le fait que les gestes intentionnels attirent spontanément l'intérêt et qu'à l'inverse les gestes aléatoires ne soient pas regardés. Autrement dit : plus un geste en image évoque un signe volontaire et articulé, et plus son intentionnalité est perçue comme liée à une pertinence pragmatique, donc d'une certaine manière, digne d'intérêt.

En ce qui concerne les gestes de battement et phatiques, ils suscitent peu d'attention focale mais ils peuvent être perçus à un niveau infra-conscient. Cette perception périphérique est vraisemblablement plus faible avec des images et en l'absence de contexte vocalique. Néanmoins en condition vidéo nous avons remarqué que les gestes prosodiques relevaient d'une distribution spatiale assez monotone et n'avaient guère de traits distinctifs sauf en cas d'expansivité des bras ; hormis ce trait leurs particularités sont plutôt temporelles. Dans les deux conditions (vidéo et image) ces gestes ne stimulent guère les points d'attention et l'imagerie mentale, sauf en cas de forte amplitude gestuelle.

Pour leur part les gestes référentiels relancent et stimulent l'attention, en créant de la discontinuité dans la chaîne comportementale. Feyeresen (1997) suggère que les gestes iconiques permettent d'ancrer en mémoire le rappel d'une phrase, tandis que les prosodiques ne laissent aucune trace. Néanmoins on peut supposer avec l'auteur, que la perception des gestes prosodiques fonctionne sur un mode analogue à la lecture labiale : il n'est pas obligatoire de voir les mouvements articulatoires de l'interlocuteur pour le comprendre, mais quand cette information est disponible, elle influence plus ou moins notre compréhension.

Les gestes désignatifs montrent leur faculté à drainer de l'attention vers un objet de référence, qui pourrait être un produit phare, un lien à cliquer, un titre de barre de menu. Les sous-titres insérés dans notre 2^e expérience permettaient de contextualiser le geste en pied d'image, sans être entremêlés à celle-ci. A posteriori il nous semble que si les gestes désignatifs pointaient vers une bulle de texte, l'attention se déporterait sans faille vers celle-ci, puisque le personnage parvient déjà par ses gestes à manipuler le regard vers des zones dépourvues de tout stimulus graphique.

L'orientation multimodale geste-regard-posture

Suite aux observations obtenues en 1^{ère} expérience (importance du plan de rotation sagittale), nous avons comparé le geste digital suivant en condition frontale (buste de face, fonction emphatique) puis en condition sagittale (rotation à 90°, fonction désignative). Le texte était pour la figure 7 « je peux vous conseiller ! » et dans l'image suivante « je vous conseille donc une lotion » (l'objet de promotion étant toujours absent de l'image) :



fig 7 geste digital de face

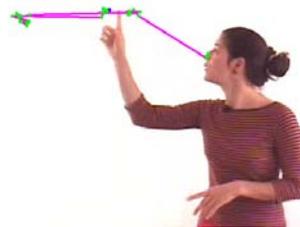


fig 8 geste digital de profil et saillance visuelle de la posture

Les fréquences d'intérêt sont supérieures en faveur du geste déployé latéralement et le test montre un écart significatif ($\chi^2 = 14.49$ $p = 0,00014$). La seconde configuration (fig. 8) semble se détacher plus efficacement comme forme-silhouette saillante dans le rapport entre la figure et le fond. Ici, le fait de

regarder la caméra ramène plutôt de l'attention vers le visage du locuteur que vers sa gestuelle, contrairement à la perspective de profil, qui donne plus de relief ou de contour au geste. Ce résultat qui reste empirique, nous conduira comme dans la première expérience à nous interroger plus largement sur l'importance de l'axe de rotation de tout personnage figuratif :

1/ est-ce qu'il regarde en face, vers la caméra et réalise des gestes frontaux (i.e. non déportés latéralement, ex. fig. 7)

2/ ou bien la posture est-elle orientée sur le plan sagittal en même temps que les gestes et le regard (fig. 8)

3/ ou encore (contexte mixte) est-ce qu'il entretient le contact par un regard frontal tout en émettant un geste latéral ? (ex. fig. 13)

Ces trois postures pourraient correspondre à des lignes de communications différentes que nous avons testées. Comparons tout d'abord la perspective de face à la perspective de profil (situations 1 et 2). Nous avons considéré l'ensemble des occurrences gestuelles du corpus en condition frontale puis sagittale, soit 550 observations (tableau 1). Le test montre un écart significatif entre effectifs observés et attendus ($\chi^2 = 70.94, p = 0$) : la perspective de face génère moins d'intérêt vers les gestes que la perspective de profil (ce qui ne signifie pas nécessairement que la scène intéresse moins l'utilisateur).

Quelques facteurs peuvent expliquer la différence de perception : les unités faciales sont **plus lisibles** et visibles lorsque le personnage regarde la caméra (perspective de face) , car le visage concurrence alors fortement les gestes. Comparativement le visage de profil apporte moins d'indices émotionnels ou cognitifs, car ces indices faciaux sont partiellement tronqués par l'image (fig. 10) et ceci au bénéfice du geste. D'autre part lorsque le visage et le buste sont en pivot à 90°, ils laissent présager d'un mouvement gestuel centrifuge poussant dans la même direction. L'utilisateur infère alors une intention signalétique dans l'espace .

Elément important : lorsque le visage est de profil, un indice facial recherché systématiquement est l'œil du personnage , comme déjà observé en mode vidéo (fig. 1 à 3): le sujet s'intéresse à **ce qui intéresse l'avatar** (attention conjointe).

<p>perspective frontale</p>  <p>fig. 9</p>	<p>rotation sagittale</p>  <p>fig. 10</p>
syntaxe faciale complète	syntaxe faciale partielle
visage > gestes	gestes > visage
lisibilité optimale des affects	moindre lisibilité des affects
attention utilisateur centripète	attention utilisateur centrifuge
renforcement de la fonction émotionnelle	renforcement de la fonction référentielle
communication de relation	communication de contenu

En liaison avec la même hypothèse , un sous-test a été mené en manipulant la syntaxe de l'image : nous nous demandons alors où se place l'attention du spectateur lorsque l'actrice fixe la caméra de face tandis qu'un geste désignatif est déployé sur le plan latéral (cf. fig. 11). Nous considérons ici uniquement les déictiques de désignation, soit un sous-effectif de 200 observations.

Geste avec regard conjoints ou disjoints ?

Nous comparerons donc les résultats de la catégorie composite 'regard frontal+geste déictique latéral' à la catégorie 'regard latéral +geste déictique latéral' . Les sujets ont vu différentes images appartenant aux deux catégories gestuelles suivantes (aperçu ci-après) :



fig. 11 geste désignatif et regard disjoints



fig. 12 geste désignatif et regard conjoints

L'écart statistique par rapport à la valeur attendue ($\chi^2= 70.94$, $p = 0$) suggère que le geste attire plus nettement l'attention du spectateur lorsque geste et regard sont tournés vers la même direction , en tous cas dans notre contexte d'observation . Le geste déictique acquiert alors une forte valeur centrifuge, phénomène déjà observé en condition vidéo. Comparativement lorsque le regard du personnage est orienté vers une autre direction que son propre geste, l'attention est plus diffuse. En théorie ces deux indices (regard et geste du personnage) peuvent fonctionner soit en synergie, soit en concurrence . Cependant en condition image, le texte ne peut relayer le geste aussi efficacement qu'une phrase sonore offerte par la condition vidéo . En mode animé, l'audition de la phrase « *regardez ce joli flacon* » aura toujours une influence déictique , même si regard et geste de l'actrice sont disjoints.

Autrement dit la posture 'visage de face versus geste latéral' (fig. 11, 13) peut être un compromis intéressant lorsque le message sonore permet une redondance avec le geste désignatif . Le regard vers la caméra procure des indices de repérage labial en interaction humaine et active des canaux émotionnels ou de maintien du contact . En modélisant cette posture l'œil déictique est perdu (l'avatar regarde de face) mais compensé par des mouvements pupillaires phatiques .



fig. 13 contexte regard frontal geste latéral



fig. 14 exemple de point contraire à gauche

Nous avons également effectué un test dans lequel le comptage des fixations s'effectue à partir de l'axe de symétrie délimitant deux zones d'intérêt possibles : à droite et à gauche du personnage (la paume de la main peut être ouverte ou le doigt pointé). Nous dénombrons ensuite des ' points valides' lorsque les

fixations du sujet sont dans le sens de l'hypothèse et des 'points contraires' lorsque le point de fixation est contraire à l'hypothèse. Les images appartenant à la syntaxe : 'regard + geste expansif + posture' orientés au plan sagittal (ex. fig. 14) démontrent 98% de points valides, si l'on exclut les saccades de confort hors écran. Une autre piste à explorer avec cette même bi-partition serait de dénombrer les points contraires à la zone à valoriser lorsque : (a) le personnage est animé, (b) il se présente en perspective disjointe comme indiqué figures 11 ou 13 et (c) il formule un message verbal.

Une dernière configuration gestuelle qui nous a intéressé, se présente comme un continuum entre signe déictique et iconique, où geste et regard référentiels convergent dans la même direction (fig.15, 16).



fig. 15, 16 visée et saisie conjointes : orientation sélective de l'attention

Cette modalité délimite l'attention dans un agrégat de fixations parfaitement localisé, car elle exprime à la fois une visée directionnelle et une saisie métaphorique, soit deux foyers d'intérêt mutuellement reliés. La ligne de projection d'attention du regardeur se négocie sous la forme d'un scanpath entre l'œil directionnel du personnage et le geste iconique. Il en résulte une orientation **sélective** de l'attention avec fermeture des choix possibles donc éventuellement opportunité de guider l'utilisateur en réduisant les points de chute aléatoires du regard. Cette *gestalt* visuelle peut être comparée en quelque sorte à un faisceau de lumière directionnel, qui discrétise la perception et réduit l'exploration diffuse. Gullberg et Holmqvist (op.cit.) remarquent également que lorsque le locuteur regarde vers ses propres gestes, son interlocuteur accorde davantage d'intérêt à ces derniers. Autrement dit ce dernier infère une pertinence communicationnelle sur la base de la multicanalité convergente des indices.

Remarques finales

Durant la perception à l'écran d'un personnage en train de converser, l'expression de son visage demeure l'élément central qui attire l'attention et entretient le canal de la relation. Cet aspect est souvent modélisé par les concepteurs de chatbots, qui donnent de l'importance graphique à la tête du personnage. Pour ce qui est des gestes, nos deux expériences suggèrent qu'ils suscitent de l'intérêt lorsqu'ils témoignent d'une intention pragmatique. Celle-ci est alors perçue par le destinataire, spécialement lorsque les gestes font référence à des objets, en les désignant ou en les traçant dans l'espace. L'œil déictique est un autre signal pertinent, qui engendre du mimétisme attentionnel. Le personnage intéresse parfois en disant quelque chose, mais cette attention peut devenir sélective quand il joint le geste connecté sur cette parole, et tout spécialement quand **montre** ou quand il **fait** quelque chose (geste référentiel). Il génère également de l'intérêt quand il s'**intéresse** lui-même à quelque chose (œil déictique, attention conjointe).

A propos de la force d'attrait et suggestive démontrée par les gestes référentiels, les déictiques sont naturellement les plus économiques à modéliser de cette classe. On n'imagine pas un agent conversationnel, lequel doit déjà négocier avec l'abstraction lexicale, enchaîner gestuellement des métaphores co-verbales, ou des quasi-linguistiques conventionnels. Dans un autre contexte-développeur, le simple mime de gestes techniques par un artefact non sémantique permet de dynamiser la scène en

distribuant l'attention vers des points de chute (déposer un article dans un caddy, manipuler une calculatrice ou une souris, feuilleter un calepin ...). La limite étant que l'on est alors dans le mode du clip non interfacé sur du guidage intelligent, et les artefacts d'avatars sémantiques ne manqueront pas d'envahir l'e-commerce. Lorsque notre personnage remplit les deux conditions détectées précédemment : 'illustre quelque chose par ses mains' et 'regarde ses propres gestes', il inscrit à l'écran dans le contexte observé, un fort champ de présence visuelle. Certains types de gestes suscitent donc des images mentales et des anticipations de stimuli nouveaux à découvrir à l'écran, étant supposé que les mouvements des yeux enregistrés ne sont pas seulement un processus physique mais un déplacement vers des espaces de sens et de cognitions.

En situation d'interface, cet attrait du geste performatif dépendrait évidemment du rapport de visibilité avec les éléments graphiques environnants : taille de l'agent par rapport aux autres composants, sa distance à l'objet valorisé, son contraste de luminance et chromatique (chaud vs froid ?), sa saturation dans le rapport forme-fond, sa place cardinale à l'écran, l'absence ou présence d'autres représentations humaines dans le champ, sa clôture éventuelle dans une zone, l'aspect statique ou animé des éléments qui l'entourent, etc...

Les gestes désignatifs montrent leur faculté à drainer de l'attention vers un objet de référence, qui pourrait être un produit, un lien à cliquer, un titre de barre de menu, une bulle de texte. Cette compétence directionnelle est relativement autonome mais peut être renforcée par le canal verbal. La configuration sagittale au complet permet de potentialiser l'effet d'un geste vers une cible. Toutefois la posture 'visage de face mais geste latéral' semble un compromis reliant contenu et relation : l'agent désigne de la main mais entretient par le visage, le canal affectif. Le message sonore déictique peut compenser la perte de l'œil déictique, qui devrait être théoriquement tourné vers la cible.

En ce qui concerne les mouvements prosodiques, ils peuvent donner du réalisme à la scène animée et être perçus par des voies infra-conscientes, mais ils semblent regardés de manière plus directe et sortir du lot lorsque l'expressivité manuelle leur confère une certaine emphase. Enfin, au niveau des deux modes expressifs expérimentés, l'oculomètre montre que l'image d'un geste dynamique contient des propriétés cinématiques et peut suggérer par son dispositif fictionnel l'impression de mouvement.

Références

Barrier, G. (2004) Gestes performatifs, expression faciale et partage attentionnel: analyse de la poursuite oculaire à partir d'une scène dialogique, *Sémiotica* 152 (1/4), 217-133 .

Cosnier, J. et Vaysse, J. (1997). *Sémiotique des gestes communicatifs*, Nouveaux Actes Sémiotiques, 52/54, 7-28.

Feyereisen, P. (1997). La compréhension des gestes référentiels. *Nouveaux Actes Sémiotiques*, 52/54, 29-48.

Gandelman, C (1986). The "Scanning" of pictures, *Communication & Cognition*, 19, 1, 3-24.

Gullberg, M., Holmqvist, K.(1999). Keeping an Eye on Gestures: Visual Perception of Gestures in Face-to-Face Communication. *Pragmatics & Cognition* 7 (1999) 35-63.

Gullberg, M. (2003). Eye movements and gestures in human face-to-face interaction. In *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements* (685-703). Oxford: Elsevier Science.

McNeill, D., *Hand and Mind – What Gestures Reveal About Thought*, Chicago Press, 1992.

Muir, L. ; Richardson, I. ; et Leaper, S. (2003). Gaze tracking and its application to video coding for sign language. *Picture Coding Symposium*, 23–25 April 2003, INRIA.