
Plateforme expérimentale de capture-restitution croisée pour l'étude de la communication face-à-face

Stephan Raidt, Gérard Bailly & Frédéric Elisei

Institut de la Communication Parlée, 46 av. Félix Viallet, 38031 Grenoble - France
Correspondance: stephan.raidt@icp.inpg.fr

RÉSUMÉ. Nous présentons une plateforme d'expérimentation dédiée à l'analyse du regard mutuel entre deux interlocuteurs lors d'une interaction face-à-face. L'utilisation de deux oculomètres assure une mesure de la direction du regard de haute précision tout en gardant une grande liberté d'action pour les sujets. L'enregistrement des deux interlocuteurs permet d'étudier le bouclage rétroactif qui se développe entre leurs actions mutuelles (regard, parole, expressions faciales, etc.). Les connaissances acquises serviront à améliorer le modèle de gestion du regard de nos agents conversationnels animés.

ABSTRACT. We present here an experimental platform dedicated to analysis of mutual gaze between interlocutors during face-to-face interaction. The unobtrusive gaze trackers used here offer high-precision measurements of gaze with relative free head motion. The aim of this simultaneous monitoring of the two interlocutors is to study the coupling between mutual actions (gaze, speech, facial expressions, etc.). This knowledge will be fed into the module responsible for the control of the eye gaze of our animated conversational agents.

MOTS-CLÉS: interaction face-à-face, direction du regard, tête parlante.

KEYWORDS: face-to-face interaction, eye gaze direction, talking face.

1. Techniques de mesure du regard

Les techniques de la mesure du regard avant le développement des oculomètres impliquaient en général des observateurs qui transcrivaient le regard pendant l'interaction (Argyle and Dean 1965; Kendon 1967; Argyle and Cook 1976).

Des mesures plus précises de la direction du regard sont possibles avec les montages décrits dans (Gibson and Pick 1963; Cline 1967; Anstis, Mayhew et al. 1969). Un premier sujet sert de stimulus pour un deuxième, auquel est demandé de marquer la direction du regard perçue. Cela demande au premier sujet de rester immobile pendant une certaine durée en fixant une cible définie, éventuellement même à l'aide d'un repose-tête. Ces contraintes rendent une réelle communication impossible. La technique appliquée par Yarbus (1967), qui exploite le reflet de petits miroirs sur des lentilles de contact, donne des mesures exactes. En revanche, elle est également trop restrictive et ne permet pas d'interactions libres.

Depuis le développement des oculomètres, les possibilités ont évolué. Vertegaal, Slagter et al. (2001) a étudié la direction du regard d'un sujet vers d'autres sujets dans une conversation avec plusieurs interlocuteurs. Un scénario qui permet l'étude d'une communication face à face a été développé par Gullberg (2001). Un sujet servant de stimulus raconte une courte histoire à un deuxième sujet équipé d'un oculomètre monté sur sa tête. Pendant cette interaction, on réalise un enregistrement vidéo du premier sujet qui servira de stimulus à un troisième sujet. Ce scénario permet d'analyser la direction du regard sur le visage ou d'autres segments du corps et de comparer la situation en direct avec la scrutation différée de la vidéo. La précision de la mesure ne permettait par contre pas de décomposer la direction du regard plus finement de manière à pouvoir analyser séparément les différentes zones du visage regardées. Des données oculométriques plus fines ont été utilisées par Lee, Badler & Badler (2002) pour alimenter un modèle statistique de saccades. Notons que de tels modèles, paramétrés par l'état cognitif du sujet (demande de tour de parole, prise de parole, etc) mais non ancrés sur une analyse de scène ne permettent pas de garantir l'impression d'une réelle co-présence par l'interlocuteur (Garau, Slater et al. 2001).

2. Objectif

La plateforme d'expérimentation décrite ici résulte de notre volonté de doter une tête parlante d'un modèle de gestion du regard et des gestes faciaux associés. Dans la communication face à face, le regard joue un rôle important. Argyle et Cook (1976) donnent un aperçu de multiples rôles que peut jouer le regard. Celui-ci peut avoir des significations élémentaires dans l'enjeu de la dominance, la soumission ou la menace, par exemple. Mais, avant tout, le regard sert au sujet à acquérir des informations sur les intentions de son interlocuteur. De la direction du regard, on

peut déduire l'objet d'attention d'une autre personne, intentionnelle ou non. De même, on peut diriger avec son regard l'attention d'un partenaire vers des objets réels, autant que vers des propos d'un discours. Ainsi le regard joue un rôle crucial dans la gestion du discours, le passage de parole, la mise en relief d'un énoncé, et l'attention mutuelle. D'une manière plus générale, l'analyse du regard d'une personne est un élément important pour estimer ses états mentaux (Baron-Cohen, Baldwin et al. 1997).

Nous avons déjà implémenté un modèle de gestion du regard pour notre tête parlante en fonction d'une analyse de scène (Picot, Bailly et al. 2006). Pour l'animation cohérente de notre tête parlante dans une situation face à face proche, il sera essentiellement important de savoir exploiter les fonctions du regard dans l'interaction humaine. D'un côté, cela servira à gérer et maintenir l'interaction. D'un autre côté, cela est important pour éviter de provoquer des réactions involontaires chez un interlocuteur par des gestes inattendus.

Cet objectif demande une plateforme d'expérimentation qui permette de mettre des sujets dans une vraie situation de dialogue face à face et de mesurer en même temps leur comportement en termes de gestes faciaux et direction du regard. Comme le modèle va être implémenté sur la tête parlante qui restera soumise aux contraintes d'un affichage sur un écran, on a choisi d'utiliser le même environnement pour notre expérimentation.



Figure 1 : Dispositif de capture croisée et étiquetage des données dans ELAN.

3. Réalisation de la plateforme

Pour construire un scénario qui correspond à l'environnement imposé par la tête parlante, nous avons choisi de mettre en relation les sujets par les mêmes moyens. Nous utilisons un écran et des hauts parleurs pour transmettre au sujet les stimuli audio-visuels. Ceux-ci peuvent être autant des stimuli préenregistrés que l'apparence transmise d'un interlocuteur humain ou virtuel (contrôle autonome ou pilotage depuis mouvements capturés sur un complice humain) en temps réel. Le dernier cas demande des dispositifs de capture qui permettent de saisir les mouvements de l'un des sujets et de transmettre ces signaux à l'écran et aux haut-

parleurs de l'autre, cela dans les deux directions. La plateforme permet ainsi un bouclage d'interaction par un dispositif de capture-restitution croisée. Les actions et réactions de l'un deviennent les stimuli de l'autre et vice versa.

Nous avons étudié différents positionnements des micro caméras Optovision © (cf. Figure 2) utilisées. La procédure de calibration consiste à estimer la direction de regard de l'autre porté sur divers éléments saillant du visage et d'objets virtuels ajoutés à l'écran. Le meilleur compromis trouvé consiste à fixer les caméras sur l'écran en s'assurant que leur orientation mutuelle permet de faire coïncider leur position avec le barycentre des yeux de l'interlocuteur.

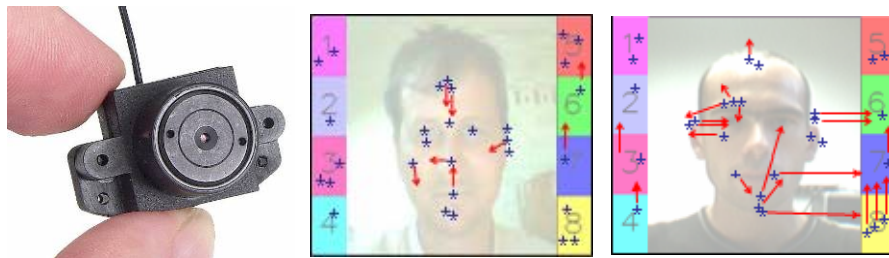


Figure 2 : Caméras utilisées et résultats de calibration de la perception du regard de l'autre en les plaçant de telle manière qu'elles coïncident avec le barycentre des yeux de l'interlocuteur. Les étoiles bleues indiquent la cible du regard de l'observateur, les flèches rouges, si présentes, indiquent la cible perçue par l'observé (personnage de la photo).

Pour documenter l'interaction, nous enregistrons, en plus des signaux audio-visuels échangés, la cible momentanée du regard sur les stimuli visuels. Les écrans doivent alors inclure des moyens de détection de la direction du regard. Cela est réalisé par des oculomètres de type Tobii 1750©.

Une analyse précise, qui permet de révéler les rapports entre geste et fonction, demande des flux de données synchronisés. Nous avons développé à ce propos une application qui rajoute des marqueurs temporels sur les flux enregistrés.

Un outil logiciel étiquette les vidéos de manière automatique pour une visualisation avec ELAN, logiciel d'étiquetage manuel (Hellwig and Uytvanck 2004). Les étiquettes informent sur la cible du regard (yeux, bouche, etc.) et le déroulement du discours. Elles peuvent être corrigées à la main et ensuite être soumises à une analyse statistique.

4. Travaux en cours

Une expérience préliminaire a été conduite avec deux sujets utilisant la plateforme en mode capture-restitution croisée. Nous avons choisi une tâche qui restreint le dialogue pour mieux contrôler l'interaction. On demande aux sujets de lire en alternance des phrases qui doivent être répétées par l'interlocuteur. Ainsi la probabilité d'apparence de gestes de mains et du corps est réduite par rapport à une discussion libre, et l'action de lecture sert à repositionner le sujet régulièrement devant l'écran. En même temps la gestion des tours de parole est contrainte ce qui évite la superposition des énoncés. Les phrases à répéter sont des SUS (Benoît, Grice et al. 1996) qui forcent l'auditeur à être attentif à l'articulation du lecteur. Le Tableau 1 montre les résultats de l'analyse statistique de l'expérience préliminaire. Ceux-ci confirment, par exemple, le regard préphonatoire du lecteur hors du visage de son interlocuteur. Des regards vers la bouche apparaissent plus fréquemment durant la vérification visuelle de l'énoncé. Enfin, le rôle tenu par chaque sujet influe sur la distribution du regard dans les divers états (notamment production/perception de parole). Cette expérience va être répétée avec plus de sujets et avec des stimuli temps réel et préenregistrés. En comparant les résultats respectifs, nous nous attendons à mieux pouvoir comprendre les fonctions du regard et à pouvoir interpréter les réactions à la tête parlante une fois celle-ci équipée d'un modèle amélioré.

Tableau 1: Statistique du regard du Sujet 1 en interaction par le dispositif capture-restitution croisée avec le Sujet 2. La cible du regard est indiquée en pourcentage par rapport à l'action correspondante du dialogue.

Locuteur	Actions de Sujet 1	Cible du regard de Sujet 1 sur le visage de Sujet 2			
		Bouche	Œil gauche	Œil droit	Ailleurs
Sujet 1	Préphonatoire	48,1	0	6,9	45,0
	Parlant	91,6	0	5,1	3,3
	Ecoutant	82,0	0	6,7	11,3
Sujet 2	Ecoutant	64,0	14,6	17,8	3,6
	Parlant	48,4	29,2	19,2	3,2
	Préphonatoire	18,7	10,2	37,6	33,5

Références

Anstis, S. M., J. W. Mayhew, et al. (1969). "The perception of where a face or television "portrait" is looking." *American Journal of Psychology* **82**: 474–489.

- Argyle, M. and M. Cook (1976). Gaze and mutual gaze. London, Cambridge University Press.
- Argyle, M. and J. Dean (1965). "Eye-contact, distance and affiliation." Sociometry **28**: 289-304.
- Baron-Cohen, S., D. A. Baldwin, et al. (1997). "Do children with autism use the speaker's direction of gaze strategy to crack the code of language?" Child Development **68**(1): 48-57.
- Benoît, C., M. Grice, et al. (1996). "The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences." Speech Communication **18**: 381-392.
- Cline, M. G. (1967). "The perception of where a person is looking." American Journal of Psychology **80**: 41-50.
- Garau, M., M. Slater, et al. (2001). The impact of eye gaze on communication using humanoid avatars. SIGCHI conference on Human factors in computing systems, Seattle, Washington: 309-316.
- Gibson, J. J. and A. D. Pick (1963). "Perception of another person's looking behavior." American Journal of Psychology **76**(3): 386-394.
- Gullberg, M. and K. Holmqvist (2001). Visual attention towards gestures in face-to-face interaction vs. on screen. International Gesture Workshop, London, UK: 206-214.
- Hellwig, B. and D. Uytvanck (2004). EUDICO Linguistic Annotator (ELAN) Version 2.0.2 manual. Nijmegen - NL, Max Planck Institute for Psycholinguistics.
- Kendon, A. (1967). "Some functions of gaze-direction in social interaction." Acta Psychologica **26**: 22-63.
- Lee, S. P., J. B. Badler, et al. (2002). "Eyes alive." ACM Transaction on Graphics **21**(3): 637-644.
- Picot, A., G. Bailly, et al. (2006). Scrutation de scènes naturelles par un agent conversationnel animé. Workshop sur les Agents Conversationnels Animés, Toulouse - France.
- Vertegaal, R., R. Slagter, et al. (2001). Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. Conference on Human Factors in Computing Systems, Seattle, USA, ACM Press New York, NY, USA: 301 - 308.
- Yarbus, A. L. (1967). Eye movements during perception of complex objects. Eye Movements and Vision. L. A. Riggs. New York, Plenum Press. **VII**: 171-196.