

Interface faciale émotionnelle : les effets de différentes modalités de présentation.

Auteur 1 : Dupré Damien

LIP, UPMF, BP47, 38040 Grenoble Cedex 9
Email damien.dupre@etu.upmf-grenoble.fr

Auteur 3 : Mandran Nadine

LIG, CNRS, BP 53, 38402 Grenoble Cedex 9
Nadine.Mandran@imag.fr

Auteur 2 : Dubois Michel

LIP, UPMF, BP47, 38040 Grenoble Cedex 9
Michel.Dubois@upmf-grenoble.fr

Auteur 4 : Tcherkassof Anna

LIP, UPMF, BP47, 38040 Grenoble Cedex 9
Anna.Tcherkassof@upmf-grenoble.fr

RESUME

Dans l'aide à la reconnaissance émotionnelle médiatisée, peu de recherches se centrent sur l'analyse de l'effet de la présentation de différentes combinaisons des zones du visage. De plus, dans le champ théorique des émotions, les travaux sur la reconnaissance d'expressions faciales reposent sur des méthodologies utilisant principalement du matériel statique (photos de visage) dans des contextes peu naturels (tâche explicite de reconnaissance d'une émotion).

Nous avons construit et validé du matériel émotionnel présentant des expressions faciales à la fois dynamiques et naturelles et nous avons comparé expérimentalement des modalités de présentation différentes du visage (Yeux + bouche, Visage global + bouche, etc., n = 11) pour connaître leur impact en termes de reconnaissance émotionnelle.

Les résultats permettent de montrer leur pertinence en termes de reconnaissance émotionnelle. L'application de ces travaux est transférable à tout dispositif permettant de présenter le visage d'autrui dès lors que ce dernier est utile pour l'interaction.

MOTS CLES : Emotion, interface faciale, reconnaissance émotionnelle, interaction, modalités de présentation.

ABSTRACT

In supporting emotional recognition in the media, little research has centred on analysing the effect of presenting different combinatorial areas of the face. Moreover, in the theoretical field of emotions, works on the recognition of facial expressions are based on methodologies which use mainly static material (head shots) in unnatural contexts (explicit recognition task of an emotion).

Our research has built and tested material with emotional facial expressions in both dynamic and natural ways and experimentally compared the different methods of presentation of the face (eyes + mouth, face global + mouth, etc., n = 11) to find their impact in terms of emotional recognition.

The results show their relevance in terms of emotional recognition. The application of this work is transferable

to all devices presenting others faces when it is useful for interaction.

KEYWORDS : Emotion, facial interface, emotional recognition, interaction, presentation modality.

INTRODUCTION

Les interfaces permettant la reconnaissance du visage à distance (ou reconnaissance médiatisée) effectuée par un Homme ou par une machine tendent à se développer dans les environnements professionnels, éducatifs ou familiaux. Par exemple, en matière de télécommunication (télé-enseignement, télé-réunion, télétravail, etc.) les plates-formes doivent permettre de supporter les interactions interpersonnelles entre des participants géographiquement séparés. L'enjeu est d'arriver à concevoir des systèmes informatiques de communication permettant de montrer, suivre et comprendre les émotions des différents participants pour pouvoir, en temps réel, réagir et réguler les échanges.

Dans des contextes d'interaction à distance, de nombreuses recherches montrent que le fait de représenter un visage à l'écran peut aider sous certaines conditions la gestion de l'activité des utilisateurs ([28], [29] et [22]). La visualisation des émotions du visage, *via* l'interface informatique, apporterait des éléments pour aider à la gestion de l'interaction car elle permettrait de mieux décoder les attitudes d'autrui et de se repositionner pour interagir ([21] et [2]).

L'étude de la reconnaissance émotionnelle médiatisée pose pourtant quelques problèmes. En effet, dans le champ de la psychologie des émotions, la reconnaissance des expressions faciales est l'objet de multiples débats. La grande majorité des méthodologies employées utilise un matériel statique (photographie) et posé (intentionnellement encodé c'est-à-dire une expression faciale non naturelle et jouée délibérément par un acteur). La validité écologique de ce type de matériel est remise en question ([24], [30] et [31]) ainsi que la possibilité de généraliser ces résultats aux communications interpersonnelles courantes – en face à face ou médiatisées – [20] ou aux expressions faciales recueillies dans des contextes variés (apprentissage, travail, etc.). De plus, ces méthodologies sont réductrices face aux situations naturelles (par essence dynamiques et

spontanées). En effet, dans les situations naturelles, les communications émotionnelles sont apportées par des patterns d'actions complexes. Les mouvements du visage et les changements faciaux plus ou moins rapides permettent de traduire et d'exprimer des expressions à autrui. En d'autres termes, les individus « décodeurs » d'émotions dans des environnements naturels observent surtout les signaux sociaux exprimés par le visage, non pas à partir de stimuli statiques, mais à partir de patterns expressifs complexes et très évolutifs. Ainsi, la catégorisation d'une expression faciale en une émotion s'élabore sur un grand nombre d'informations dynamiques subtiles qu'une photographie statique ne permet ni de présenter, ni de rendre compte ([4], [5], [11] et [13]). L'importance de l'évolution temporelle des mouvements faciaux apparaît comme étant déterminante pour comprendre la reconnaissance des expressions faciales émotionnelles. En effet, les manifestations faciales dynamiques nécessitent de reconnaître et de différencier des patterns faciaux émotionnels peu stabilisés comparativement aux présentations statiques (voir aussi [15]). De plus, [7] a montré que des juges, face à des corpus dynamiques, sont sensibles aux changements subtils des expressions faciales. Ces travaux montrent que la dimension des aspects dynamiques est une question pertinente et utile pour les interactions en face à face et médiatisées. De plus, la précision des reconnaissances, lorsque l'on compare des jugements établis au sujet d'expressions spontanées ou posées montre de meilleurs scores de reconnaissance pour ces dernières [20]. Ces résultats ne sont guère surprenants puisque les expressions posées constituent les meilleurs « exemplaires » de l'émotion exprimée exagérément (comparativement aux expressions spontanées) ce qui les rend plus facilement identifiables ([9] et [14]). Les configurations faciales prototypiques utilisées en laboratoire comme stimuli standards (le plus souvent le *JACFEE set* de [17] ou les photos de [8]) sont rarement rencontrées dans la vie réelle dans cet état de « pureté ». Ces configurations représentent le pattern le plus « abouti et pur » correspondant à l'émotion donnée. Au contraire, il est évident que pour des expressions spontanées cette configuration n'est jamais aussi nette [21]. Dans la vie quotidienne, les expressions spontanées sont moins caricaturales, moins prototypiques et plus ambiguës [25] car nos émotions quotidiennes habituelles sont moins intenses, moins typiques, et moins franches que celles utilisées dans le cadre de photographies exprimant un stéréotype émotionnel [16].

En définitive, bien que les expressions faciales naturelles constituent des signaux sociaux dynamiques (i.e. qui encodent le message émotionnel à travers des patterns d'actions faciales dynamiques), les expressions faciales étudiées dans la littérature ne le sont souvent que sous l'angle statique. Comme le montrent certaines études (même si elles sont encore rares, [1], [13] et [3]), les différences (morphologiques, temporelles, etc.) entre expressions spontanées et posées sont telles que, au

regard de la problématique de la reconnaissance émotionnelle *via* l'expression faciale, il semble difficile de considérer les résultats obtenus avec des expressions prototypiques comme rendant compte de la mise en œuvre des processus cognitifs réellement mobilisés dans les situations naturelles. Les limites méthodologiques du recours à des expressions statiques et posées conduisent actuellement à entreprendre des recherches étudiant spécifiquement des expressions faciales à la fois dynamiques (films) et spontanées (réactions « naturelles » et « spontanées » à des situations).

C'est donc dans cette perspective dynamique et spontanée qu'il convient d'identifier les indices visuels émotionnels pertinents utilisés de façon implicite lors de situation de face à face [26]. Pour [10], lors d'interactions médiatisées c'est l'apparition des expressions faciales et posturales qui permet la reconnaissance des émotions. Ainsi la reconnaissance des émotions d'autrui *via* l'interface faciale est analysée en fonction de caractéristiques globales de la prise de vue du visage (notamment la comparaison absence de visage *vs* visage global *vs* visage et épaule ; [27] et [2]). Cependant les recherches sur la reconnaissance émotionnelle médiatisée se centrent peu sur l'analyse de l'effet de la présentation de différentes combinaisons des zones du visage.

Cette recherche s'intéresse donc à la conception d'interface faciale émotionnelle pour mieux aider, dans des contextes collaboratifs, distants et interactifs, à superviser les actions d'autrui ou à interagir avec lui. Les objectifs de cette étude sont doubles. Dans un premier temps le but est d'élaborer un matériel expressif écologique (i.e. le plus proche des expressions faciales perçus dans les interactions face à face). Notre recherche a donc construit et validé du matériel émotionnel présentant des expressions faciales à la fois dynamiques et spontanées. Dans un second temps, afin d'identifier les indices de la reconnaissance émotionnelle, nous avons comparé l'incidence des différentes modalités de présentation de ces expressions spontanées et dynamiques sur la reconnaissance. Pour ce faire des modalités de présentation différentes (n = 11) du visage (Yeux + bouche, Visage global + bouche, etc.) ont été expérimentalement confrontées pour connaître leur impact en termes de reconnaissance émotionnelle.

EXPERIENCE

L'objectif de cette expérience est de savoir si des modalités de présentation de certaines zones du visage sont plus efficaces pour la reconnaissance émotionnelle médiatisée.

Hypothèses

- La modalité de présentation du visage global permet d'obtenir des taux de reconnaissance

supérieur par rapport aux conditions parcellisées.

- Les modalités de présentation de certaines zones partielles du visage améliorent la reconnaissance d'émotions spécifiques.

Construction et sélection du matériel émotionnel

Le matériel a été conçu en deux phases. Dans une première phase dite « pilote », quarante trois participants ont été filmés à leur insu pendant qu'ils réalisaient trois tâches émotionnelles. Cette phase a permis de mettre au point les tâches inductrices d'émotions, les conditions expérimentales, techniques, matérielles et éthiques pour recueillir des expressions faciales spontanées et dynamiques (cf. [30] pour plus de détails). La deuxième phase a consisté en l'élaboration du corpus proprement dit [18]. Ce corpus, appelé DynEmo, comporte 358 vidéos d'expressions faciales émotionnelles de participants filmés à leur insu (182 femmes et 176 hommes âgés de 25 à 65 ans ; moyenne d'âge = 48, sd= 9.2) pendant qu'ils réalisaient une tâche émotionnelle ; cf. <https://dynemo.liglab.fr/>.

Induction émotionnelle. Nous avons élaboré des tâches (durée moyenne autour de 10 min) pour induire 10 émotions. Les tâches sont construites sur la base de la présentation de vidéos et/ou de dispositifs type « magicien d'oz ». Un dernier scénario a aussi été mis au point dans l'optique de n'induire aucune expression émotionnelle, le matériel extrait de cette condition étant considéré comme suscitant une « émotion neutre » (cf. [18] pour plus de détails)¹.

Enregistrement des expressions faciales des encodeurs. Les tâches d'induction émotionnelle (dont la durée varie en fonction des participants et des émotions inductrices) étaient présentées sur l'écran d'un ordinateur portable de 17" à 50cm du participant. La dynamique faciale émotionnelle du participant (dit encodeur) était enregistrée à son insu en temps réel par, notamment, une caméra centrée sur son visage.. Dans le cadre de cette étude, nous exploitons les vidéos des visages en plan soit rapproché (face seule) soit éloigné (face + épaules) filmés lors des inductions de gaieté et d'intérêt (états affectifs positifs), de perplexité et d'énervement (états affectifs négatifs) et de la condition contrôle « expression neutre ».

Procédure d'induction. Les participants étaient invités à signer un premier consentement de participation. Ils étaient installés dans la salle d'expérimentation afin de réaliser l'une des 11 tâches d'induction. A l'issue de cette passation, les participants remplissaient différents questionnaires de mesure de *self report* émotionnel destinés à évaluer l'état émotionnel qu'ils avaient ressenti lors de l'accomplissement de la tâche. Ils étaient ensuite débriefés, puis signaient un formulaire de

consentement d'utilisation de leur image dans un cadre scientifique.

Sélection du matériel émotionnel. Un protocole de jugement des expressions faciales émotionnelles a été ensuite mis en œuvre sur une interface informatique pour présenter et mesurer les séquences émotionnelles les plus représentatives de chaque émotion parmi l'ensemble des expressions faciales émotionnelles recueillies auprès des participants encodeurs. Les séquences présentées sont uniquement celles dont le *self report* de l'encodeur correspond à la nature émotionnelle de la tâche d'induction. Pour ce faire, des personnes juges (i.e. « décodeurs » ; n = 22 étudiants par films) visualisaient les visages des encodeurs et devaient en cours de visualisation attribuer des labels émotionnels *via* un procédé (cf. figure 1) « Oudjat » permettant de sélectionner au fur et à mesure le label émotionnel le plus représentatif en regard de la dynamique émotionnelle.

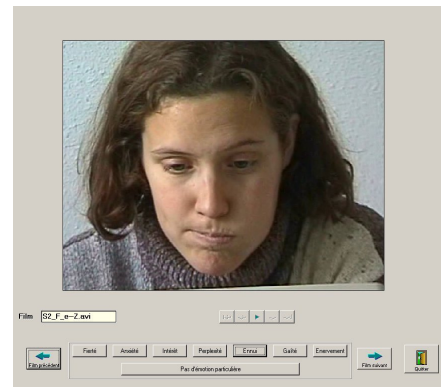


Figure 1 : L'interface de catégorisation émotionnelle.

Les moyens de mesure sont le nombre de clics (changement de catégorisation émotionnelle) et le temps de détection d'une émotion donnée. A l'issue de ce protocole, nous avons pu sélectionner les séquences de films reconnues par les décodeurs comme représentatives des cinq émotions choisies, exprimées à la fois par un encodeur homme et un encodeur femme. Sur un ensemble de films jugés, nous avons retenu au final 10 séquences d'une durée d'une minute chacune.

Participants

Deux cent quarante deux étudiants (215 femmes et 27 hommes) ont participé à cette étude.

Modalités de présentation des émotions

Les séquences de films représentatives d'une émotion donnée ont été travaillées pour être visualisées sous différentes modalités de présentation dynamiques. Notre objectif est de déterminer et de comparer l'efficacité de chaque modalité en termes de reconnaissance émotionnelle. La séquence émotionnelle est diffusée sous différents plans, soit dans des présentations globales (tête + épaules, tête agrandie avec zoom X 4) pour

¹ Nous n'exploitons dans le cadre de ce travail qu'un nombre limité des vidéos d'expressions faciales élaborées lors des programmes de recherche Cognitive (phase pilote) et ANR-06CORP-019-01 (corpus DynEmo).

permettre de visualiser les patterns généraux d'expressions faciales, soit dans des présentations détaillées (zones du visage – des yeux, de la bouche – permettant d'augmenter la visualisation de zones plus expressives et spécialisées dans certaines émotions). Ces présentations sont soit exposées de manière isolée, soit combinées avec des présentations globales (cf. figure 2).

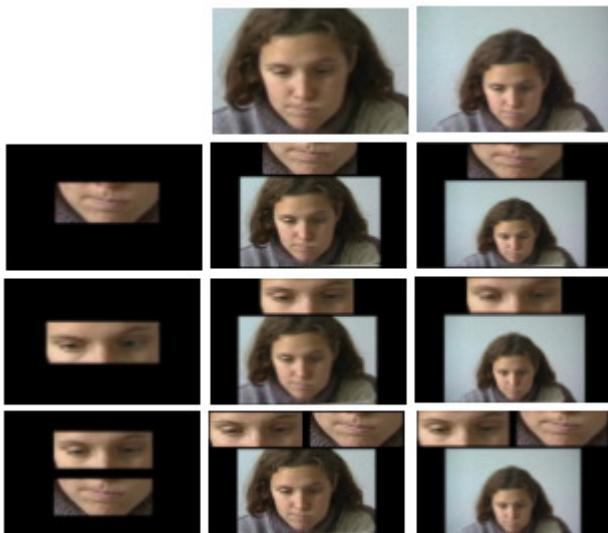


Figure 2 : Disposition des modalités de présentation du visage.

La série de 10 films est présentée sur l'écran d'un ordinateur. Chaque participant ne voit les films présentés qu'avec une seule modalité de disposition. L'ensemble des participants était donc réparti dans un des onze groupes indépendants en fonction des modalités de présentation de l'information des visages. L'équilibre homme/femme était équitable entre les groupes. L'ordre de présentation des films s'effectue aléatoirement. Pour que les participants décodeurs puissent indiquer la ou les émotions perçues au fur et à mesure du film, une interface (basée sur le même principe Oudjat) comportant huit icônes a été construite. Ces icônes représentent les émotions étudiées (gaieté, intérêt, pas d'émotion particulière, énervement et perplexité) et des distracteurs (ennui, fierté et anxiété).

Procédure

Après avoir pris connaissance de la consigne, les sujets commencent l'expérience en s'exerçant sur un film d'essai (ce film n'est pas pris en compte dans l'analyse statistique). Les participants activent les icônes des émotions qu'ils reconnaissent au fur et à mesure de la visualisation des 10 films. Tant qu'ils ne cliquent pas sur une autre icône, l'icône précédente reste enclenchée. L'ordinateur enregistre les réponses des sujets et fournit une mesure automatique des temps de détection des émotions pour les dix films. La durée de la passation s'élève à 30 minutes par sujet. L'analyse des recueils s'effectue sur un intervalle de 0,5 seconde nous permettant ainsi de caractériser précisément les détections spontanées et dynamiques en temps réel. Le

plan d'expérience S_{22} (22 participants par condition) $\langle EF10 \rangle$ (expressions faciales) $\times M_{11}$ (conditions de présentation) est traité par une analyse ANOVA du temps de bonne reconnaissance de l'expression faciale.

PRINCIPAUX RESULTATS

Pour s'assurer de la bonne reconnaissance émotionnelle des différents films, nous avons comparé les temps moyens de reconnaissance de chaque label cible en fonction de chaque film. Ainsi, de manière générale les expressions faciales ont été correctement reconnues quelle que soit la vidéo présentée. Les vidéos d'expressions de faible activation physiologique (intérêt) se révèlent cependant plus ambiguës que les expressions de forte intensité (voir Tableau 1).

Films	Score label			
	S_Gaiete	S_Interet	S_Perpl	S_Energ
gaieté	64,2 (28,11)	24,8 (25,14)	6,7 (12,1)	0,1 (0,5)
intérêt	0,7 (4,16)	36,3 (30,51)	36,1 (28,1)	2,5 (8,2)
neutre	0,1 (0,88)	33,6 (31,80)	35,9 (32,2)	6,1 (16,9)
perplexité	1,1 (6,59)	12,4 (18,29)	44,5 (30,3)	8,8 (17,8)
énervement	0,8 (3,87)	11,1 (18,64)	16,4 (22,3)	33,8 (35,4)

Tableau 1 : Moyenne (en 0.5s) et écart-type de l'utilisation de des labels cibles pour chaque film expressif.

Effet du nombre d'éléments présenté

On pouvait penser que la présentation d'un plus grand nombre d'informations émotionnelles faciliterait la reconnaissance des expressions faciales. Ainsi dans le but de tester notre modèle, nous décomposons notre effet omnibus en deux tests orthogonaux de Helmert : une comparaison visant à tester l'avantage de présenter plusieurs éléments contre un seul et l'autre visant à tester l'avantage de passer de deux à trois éléments.

Sur l'ensemble des données disponibles, nos analyses mettent en évidence les différences de reconnaissance en fonction du nombre d'éléments, $F_{(2,2399)} = 6.32$, $p < .05$. Cependant, de manière générale, il existe un avantage à présenter plusieurs éléments plutôt qu'un seul ($t_{(2399)} = 2.95$, $p < .05$) mais les résultats ne montrent pas de différence entre la présentation de deux ou trois éléments ($t_{(2399)} = 1.12$, ns). Ainsi les tests post-hoc de Tukey indiquent, pour certaines expressions, un coût de traitement lorsque trop d'informations sont présentées. Par exemple, la reconnaissance des expressions d'intérêt, de perplexité et d'énervement est facilitée quand on passe de 1 à 2 éléments (intérêt : $M = 32,4$ contre $M = 39,9$; $p < .05$; perplexité : $M = 39,9$ vs $M = 48,5$; $p < .05$; énervement : $M = 28$ vs $M = 37,3$; $p < .05$). Cependant les tests de Tukey révèlent une baisse de performance quand on passe de 2 à 3 éléments pour la reconnaissance de l'intérêt ($M = 35,2$; $p < .05$), de la

perplexité ($M = 44,8$; $p < .05$) et une absence de facilitation pour l'énervement ($M = 37,3$; $p < .05$).

Analyse de la présentation du visage en entier

Les trois types de présentation des plans du visage (visage partiel, rapproché ou éloigné) par émotion ont été traités par analyse de variance. On observe un effet des différents plans sur les émotions neutre ou négative.

La reconnaissance de l'expression d'énervement montre des différences significatives selon les types de présentation du plan ($F_{(2, 477)} = 13,39$; $p < .001$). Les tests post hoc montrent que cette expression est moins bien reconnue sans la présence du visage intégral ($M = 20,5$) et que le rapprochement des zones faciales favorisent les scores de reconnaissance (rapproché : $M = 40,1$; éloigné : $M = 37,3$).

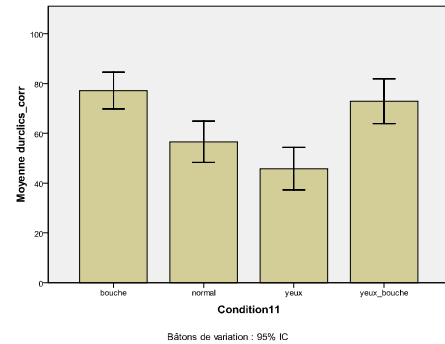
De même, la reconnaissance des expressions de perplexité est affectée par les types de présentation du visage ($F_{(2, 477)} = 5,67$; $p < .05$). La reconnaissance est inhibée lorsque le visage total n'apparaît pas et la visualisation rapprochée favorise la reconnaissance (vues partielles $M = 20,5$ vs visage rapproché $M = 40,1$ vs visage éloigné $M = 37,3$). Néanmoins il n'existe de différence significative d'après le test de Tukey qu'entre le plan rapproché et les autres conditions.

Enfin, la reconnaissance de l'expression « neutre » est, elle aussi, affectée par la présentation du visage entier ($F_{(2, 477)} = 5,95$; $p < .05$). Par contre, celle-ci est mieux reconnue lorsque le visage est présenté en zones éclatées plutôt que dans l'intégralité de la face – rapproché ou éloigné – ($M = 36,6$ vs $M = 25$ et $M = 27,1$).

Effet des conditions parcellisées par émotion

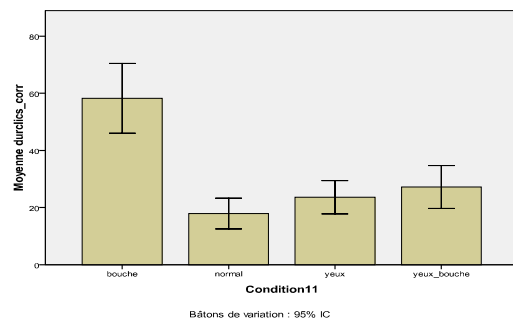
Le test ANOVA de la reconnaissance de la gaieté grâce aux conditions parcellisées comparées à la condition visage entier nous indique des différences significatives dans les modalités de présentation de l'expression faciale ($F_{(3,172)} = 12,62$; $p < .05$). D'après les tests post-hoc de Tukey, la seule présentation de la bouche facilite sa reconnaissance ($M = 77.2$) et la présentation des yeux uniquement entrave la reconnaissance de la gaieté ($M = 45.8$) par rapport à la présentation du visage et des yeux + bouche ($M = 56.6$ et $M = 72.9$; cf. graphique 1).

Les moyennes de reconnaissance de l'expression de l'intérêt ne sont pas significatives.



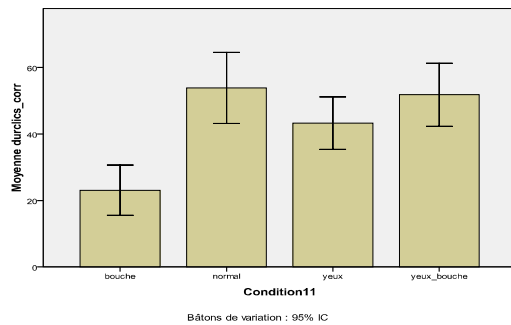
Graphique 1 : Comparaison de la durée moyenne de reconnaissance (en 0.5s avec l'écart-type comme indice de dispersion) des conditions Bouche, Normale, Yeux, et Yeux+Bouche pour la reconnaissance de la Gaieté de l'expression faciale de gaieté.

La reconnaissance de l'expression neutre est significativement différente selon les zones du visage présentées ($F_{(3,172)} = 20.07$; $p < .05$). Les tests post-hoc de Tukey montrent que la présentation de la bouche seule permet moins de fausse reconnaissance que la présentation des « yeux », des « yeux + bouche » et du plan de l'intégralité du visage ($M = 58.2$ vs $M = 23.6$; $M = 27.2$ et $M = 17.9$; $p < .05$; cf. graphique 2).



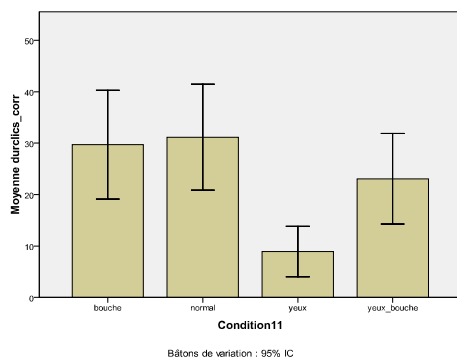
Graphique 2 : Comparaison de la durée moyenne de reconnaissance (en 0.5s avec l'écart-type comme indice de dispersion) des conditions Bouche, Normale, Yeux, et Yeux+Bouche pour la reconnaissance du label Neutre de l'expression faciale neutre.

Au contraire, la reconnaissance de l'expression de perplexité est inversement affectée par les modalités de présentation du visage ($F_{(3,172)} = 20.07$; $p < .05$). La présentation de la bouche seule ne permet pas la reconnaissance alors que la moyenne de reconnaissance de l'expression de perplexité n'est pas significativement différente entre la présentation du visage total, la présentation yeux + bouche et la présentation des yeux seuls ($M = 23.1$ vs $M = 53.9$; $M = 51.8$ et $M = 43.8$; $p < .05$; cf. graphique 3).



Graphique 3 : Comparaison de la durée moyenne de reconnaissance (en 0.5s avec l'écart-type comme indice de dispersion) des conditions Bouche, Normale, Yeux, et Yeux+Bouche pour la reconnaissance de la Perplexité de l'expression faciale de perplexité.

Enfin le test ANOVA de la reconnaissance de l'énerverment met en évidence des différences significatives ($F_{(3,172)} = 5.35$; $p < .05$). La présentation des seuls yeux permet une moins bonne reconnaissance que la présentation de la « bouche », des « yeux+ bouche » et du « visage total » ($M = 8.9$ vs $M = 29.7$; $M = 23.1$ et $M = 31.7$; $p < .05$; cf. graphique 4)



Graphique 4 : Comparaison de la durée moyenne de reconnaissance (en 0.5s avec l'écart-type comme indice de dispersion) des conditions Bouche, Normale, Yeux, et Yeux+Bouche pour la reconnaissance de l'Énerverment de l'expression faciale d'énerverment.

Présentation des yeux en plus d'autres informations

Lorsqu'il s'agit de comparer l'effet de la présentation des yeux en plus du visage, seule la reconnaissance des émotions de forte intensité, quelle que soit leur valence (positive ou négative), sont affectées. Ainsi, la présence des yeux gêne la reconnaissance de la gaieté ($M = 61,9$ contre $M = 67,06$; $F_{(1, 478)} = 4,04$; $p < .05$) et de l'énerverment ($M = 30,8$ contre $M = 37,8$; $F_{(1,478)} = 4,23$; $p < .05$).

Présentation de la bouche en plus d'autres informations

Nos résultats mettent aussi en évidence un effet relatif de la présentation de la bouche en plus du visage entier. Celle-ci facilite la reconnaissances de l'expression de gaieté ($M = 68,9$ contre $M = 58,8$; $F_{(1, 478)} = 16,15$; $p < .001$). De plus, elle permet aussi une

meilleure reconnaissance de l'expression neutre ($M = 31,7$ contre $M = 25,5$; $F_{(1, 478)} = 4,95$; $p < .05$).

DISCUSSION

Cette étude visait à comparer les différentes modalités de présentation d'interfaces émotionnelles. Contrairement à de nombreux travaux disponibles sur les présentations multimodales d'information de contenu de travail ou pédagogique ([19], [28], [12] et [6]), il n'y a pas de travaux sur du contenu émotionnel. La littérature en psychologie des émotions cherche surtout à identifier les processus de reconnaissance émotionnelle et ne s'intéresse pas, d'un point de vue ergonomique, aux modalités de présentation permettant de faciliter une meilleure reconnaissance émotionnelle distante pour supporter les interactions communicationnelles. Or ce travail montre l'intérêt d'un point de vue ergonomique de comprendre comment s'effectue la bonne reconnaissance de ce type de matériel.

Dans un premier temps nos résultats mettent en évidence un effet du nombre d'informations présentées sur la reconnaissance émotionnelle. Il existerait ainsi un nombre optimum de présentation de deux éléments pour la reconnaissance des émotions. L'ajout de ce second élément pour les expressions négatives semble faciliter leur reconnaissance car il permet d'exclure le pattern d'émotions « concurrentes ». En revanche, la présentation de trois éléments entraîne une chute de reconnaissance. On constate ici l'effet de surcharge cognitive déjà démontré sur d'autres matériels non émotionnels [28] et imputable à la surabondance d'éléments informatifs partagés obligeant un traitement informationnel conséquent. Néanmoins la reconnaissance des émotions positives peut s'effectuer à partir d'un seul élément présenté car ces expressions sont suffisamment non ambiguës (cf. gaieté). Un seul élément présenté (mais pertinent) est généralement suffisant pour permettre une bonne reconnaissance. Ces résultats traduisent la nécessité de présenter sur l'interface des informations ou des interactions d'informations strictement pertinentes, et différentes en nombre selon les valences émotionnelles, pour diminuer le coût de traitement tout en augmentant son efficacité.

En s'intéressant plus particulièrement aux conditions de présentation, nos résultats montrent que les expressions faciales négatives sont moins reconnues quand le plan est éclaté (yeux+bouche) qu'entier. Ce résultat inattendu peut s'expliquer par le fait que la reconnaissance des expressions négatives est autant influencée par le mouvement des traits du visage (bouche+yeux) que par l'information configurale de l'ensemble des traits du visage.

Enfin nos résultats mettent en lumière les différences de reconnaissance des différentes émotions en utilisant les modalités de présentation partielles du visage. Ainsi l'expression de gaieté est mieux reconnue en présentant

la bouche seule. Cet indice de la reconnaissance de la gaieté est classiquement retrouvé dans la littérature. Néanmoins on peut se demander pourquoi il n'existe pas d'avantage à présenter les yeux pour différencier les vrais sourires des faux sourires ?² Il semble ainsi que la présentation des yeux dans la reconnaissance de l'expression de *gaieté* n'a pas d'effet quantitatif sur la reconnaissance de cette expression mais plutôt un effet qualitatif de différenciation.

L'expression *d'intérêt* est aussi bien reconnue dans toutes les conditions. Ces résultats posent des questions puisque qu'il s'avère que la présentation des yeux seuls amène une performance de reconnaissance égale à la présentation de la bouche seule et du visage entier. On peut alors imaginer que les yeux portent l'essentiel des informations de l'expression d'intérêt, ces informations quand elles sont disponibles ne sont pas améliorées par d'autres sources informatives.

L'expression *neutre* est mieux reconnue avec la bouche seule alors qu'il n'existe pas de différence entre les autres conditions. La reconnaissance de l'expression neutre est liée à l'absence de reconnaissance d'une autre expression émotionnelle (processus de catégorisation par défaut) et la bouche semble engendrer moins de fausse reconnaissance que les yeux ou que le visage entier. Cela peut s'expliquer par le fait que la bouche est une partie du visage très mobile. Dans cette perspective nous serions entraîné(e)s à percevoir les mouvements de celle-ci, aussi infimes soient ils, contrairement aux yeux qui sont beaucoup moins mobiles et au visage dont l'intégralité des traits rend plus difficile la non reconnaissance d'une expression.

L'expression de *perplexité* est moins bien reconnue avec la présentation de la bouche seule, de même pour l'expression d'*énervement*. Ces résultats indiquent deux effets de présentation de l'interface émotionnelle. D'abord la présentation de la bouche seule n'apporte pas d'indication dans la reconnaissance de la perplexité et de l'énervement. Ensuite la présentation des yeux seuls amènerait une performance égale à la présentation du visage entier pour la reconnaissance de l'expression d'énervement ce qui irait en faveur de l'hypothèse que seuls les yeux sont utilisés pour reconnaître cette expression.

L'ensemble de ces résultats montre que des réflexions doivent s'effectuer à la fois sur le nombre et la nature des éléments émotionnels à présenter sur une interface. D'un point de vue ergonomique, nous devons progresser pour bien identifier les patterns d'informations pertinentes en regard d'une émotion dynamique donnée. Cette recherche exploratoire illustre le fait qu'il n'y a pas une seule modalité de présentation pertinente pour toutes les émotions mais que les modalités doivent s'adapter

aux natures émotionnelles et aux patterns les plus informatifs en regard de telle ou telle émotion.

Grâce à nos résultats, nous sommes en mesure de rendre manifeste des conditions de présentation pertinentes des expressions émotionnelles afin d'obtenir leur meilleure reconnaissance dans des dispositifs de communication interpersonnelle à distance. Mais plus important, nos travaux mettent aussi en lumière le fait que le traitement de certaines zones du visage dans la reconnaissance émotionnelle ont un effet délétère et peuvent même entraîner de fausses reconnaissances. Ces résultats sont absents de la littérature même s'ils représentent un enjeu pour les interfaces d'interaction médiatisée. En effet si différentes études mettent en évidence l'enjeu de la reconnaissance des expressions émotionnelles dans les situations distantes, très peu s'intéressent aux configurations permettant l'identification des « bonnes » et des « mauvaises » reconnaissances émotionnelles.

BIBLIOGRAPHIE

1. Ambadar, Z., Schooler, J., & Cohn, J. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, 16, 403-410.
2. Anderson A.H., Smallwood L., MacDonald R., Mullin J., Fleming A.M., & O'Malley C. (1999). Video data and video links in mediated communication: What do users value? *International journal of human-computer studies*, 2, 165-187.
3. Back, E., Jordan, T.R., & Thomas, S.M. (2009). The recognition of mental states from dynamic and static facial expressions. *Visual cognition*, 17, 1271-1286.
4. Bassili, J.N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 373-379.
5. Bassili, J.N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, 37, 2049-2058.
6. Bétrancourt, M., Dillenbourg, P., & Clavier, L. (2008). Reducing cognitive load in learning from animations: Impacts of delivery features. In J.-F. Rouet, W. Schnotz and R. Lowe (Eds.), *Understanding Multimedia Documents* (pp 61-78).
7. Ekman, P. & Friesen, W.V. (1976). *Pictures of facial affect*. Palo Alto, C.A.: Consulting Psychologists Press.
8. Fernández-Dols, J.-M., Sánchez, F., Carrera, P., & Ruiz-Belda, M.-A. (1997). Are spontaneous expressions and emotions linked? An experimental test of coherence. *Journal of Nonverbal Behavior*, 21, 163-177.
9. Heath C. & Luff, P. (1991). Disembodied conduct: Communication through video in multi-media office environment. *Proceedings of CHI'91* (pp. 99-103), New Orleans: ACM press.
10. Hess, U. & Kleck, R.E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology*, 20, 369-385.
11. Jamet, E., Bétrancourt, M., & Rouet, J.-F. (2008). La compréhension des documents complexes. In A. Chevalier & A. Tricot (Dir.) *Ergonomie cognitive des documents électroniques*. Paris : Dunod.

² Les vrais et faux sourires font référence aux sourires types « Duchenne » avec plissement des yeux et « non-Duchenne » sans plissement des yeux.

12. Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception*, 30, 875-887.
13. Kanade, T., Cohn, J.F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)* (pp. 46-53), March 2000, Grenoble, France.
14. Lemay, G., Kirouac, G., & Lacouture, Y. (1995). Expressions faciales émotionnelles spontanées, dynamiques et statiques : comparaison d'études de jugement catégoriel et dimensionnel. [Dynamic and static spontaneous emotional facial expressions: A comparison of categorical and dimensional judgment studies]. *Canadian Journal of Behavioral Science*, 27, 125-139.
15. Manstead, A.S.R., Fisher, A.H., & Jakobs, E.B. (1999). The social and emotional functions of facial displays. In P. Philippot, R.S. Feldman & E.J. Coats (Eds.), *The social context of nonverbal behavior* (pp. 287-313). Cambridge, UK: Cambridge University Press.
16. Matsumoto, D. & Ekman, P. (1988). *Japanese and Caucasian Facial Expressions of Emotion and Neutral Faces (JACFEE and JACNeuF)*. Available from Human Interaction Laboratory, University of California, San Francisco, 401 Parnassus Avenue, San Francisco, CA, 94143.
17. Mayer, R.E. & Anderson, R.B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84, 444-452.
18. Meillon, B., Tcherkassof, A., Mandran, N., Adam, J.-M., Dubois, M., Dupré, D., Benoît, A.-M., Guérin-Dugué, A., Caplier, A. (2010). DYNEMO: A Corpus of dynamic and spontaneous emotional facial expressions. Communication orale à paraître dans *Proceedings of the International Workshop Series on Multimodal Corpora, Tools and Resources: Advances in Capturing, Coding and Analyzing Multimodality (LREC)*, Malte, 18 Mai.
19. Motley, M.T. & Camden, C.T. (1988). Facial expression of emotion: A comparison of posed versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication*, 52, 1-22.
20. Nummenmaa, T. (1992). *Pure and blended emotion in the face*. Helsinki, Finland : Academia Scientiarum Fennica.
21. O'Connell B., Whittaker S., & Wilbur S., (1994). Making the case for face-to-face interaction: Empirical studies of video-mediated interaction. *Human Computer Interaction*, 8, 389-428.
22. O'Malley C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V. (1996). Comparison of face-to-face and video-mediated interaction. *Interacting with computers*, 8, 177-192.
23. Russell, J.A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102-141.
24. Russell, J.A. (1997). Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective. In J.A. Russell & J.M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 295-320). Cambridge, UK: Cambridge University Press.
25. Rutter, D. & Stephenson G., (1977). The role of visual communication in synchronising conversation. *European Journal of Social Psychology*, 7, 29-37
26. Short, J., Williams, E., & Christi, B. (1976). *The social psychology of telecommunications*. New York: Wiley.
27. Sweller, J. & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12, 185-233.
28. Tajariol F., Adam, Michel, J.M., & Dubois, M. (2008). *Seeing the face and observing the actions: The effects of nonverbal cues on mediated tutoring dialogue*. In E. Aimeur, & B. Woolf (Eds.) *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS 2008)* (pp.480-489). Berlin, Germany: Springer Verlag. ITS: Montréal, Canada..
29. Tajariol, F., Adam, J.M., & Dubois, M. (2009). The impact of nonverbal cues on mediated tutoring interaction: an experimental study. In press.
30. Tcherkassof A., Bollon T., Dubois M., Pansu P., & Adam J.-M. (2007). Facial expressions of emotions: A methodological contribution to the study of spontaneous and dynamic emotional faces. *European Journal of Social Psychology*, 37, 1325-1345.
31. Wallbott, H.G. & Scherer, K.R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, 51, 690-699.