
Adaptation d'une théorie psychologique pour la génération d'expressions faciales synthétiques pour des agents d'interface

Amandine Grizard, Marco Paleari et Christine Lisetti

*Groupe d'informatique affective sociale
Département communications multimedias,
Institut Eurecom
2229 route des crêtes 06904 Sophia-Antipolis*

RÉSUMÉ. Dans cet article, nous décrivons comment les interactions humains-robots peuvent être améliorées en affinant la crédibilité des expressions faciales sur des personnages zoomorphiques ou anthropomorphiques. La crédibilité des émotions exprimées par les agents dépend de l'intensité du lien entre ces expressions (état externe) et le modèle émotionnel (état interne ou processus). La théorie psychologique de Scherer décrit comment les expressions faciales humaines sont liées avec le processus interne d'évaluation cognitif. Dans ce papier, nous montrons comment nous adaptons la théorie de Scherer sur la génération des expressions émotionnelles faciales humaines à partir d'une évaluation cognitive jusqu'à la génération d'expressions émotionnelles artificielles sur deux plateformes différentes : une tête physique de robot et un avatar graphique. Nous présentons les résultats d'études utilisateurs pour évaluer la crédibilité de nos deux personnages.

ABSTRACT. In this article, we describe how human-agent interactions may be improved by refined believability of facial expressions of zoomorphic or anthropomorphic embodiment forms. The believability of agents' emotional expressions depends on the strength of the link between the expressions themselves (external states) and the emotional model (internal states or processes). Scherer's psychological theory describes how human facial expressions are linked with internal cognitive appraisal processes. In this paper we show how we adapted Scherer's theory of human emotional facial expressions generation from cognitive appraisal to the generation of artificial emotional expressions on two different embodiment platforms: a graphical avatar and a physical robotic head. We discuss the results of the user study we conducted to assess the believability of the two embodied forms.

MOTS-CLÉS : Expressions émotionnelles faciales, Informatique affective, Émotion et communication

KEYWORDS: Emotional facial expressions, Affective Computing, Emotion and communication

1. Introduction

Il existe deux objectifs principaux dans le développement d'agents sociaux intelligents pour des interactions avec les humains. Le premier est l'expression des émotions (état externe) et le second est la modélisation de ces émotions (état interne). Le but de notre recherche est de directement lier les processus internes responsables des expressions faciales émotionnelles en appliquant la théorie psychologique de Scherer sur l'évaluation et la génération des émotions. Dans ce papier nous ne parlerons que de la génération d'expressions émotionnelles basée sur cette théorie.

Dans la suite, nous présentons quelques recherches qui ont été faites sur des agents sociaux artificiels sur différentes plateformes. Nous introduisons la théorie psychologique de Scherer que nous utilisons pour générer des expressions faciales artificielles sur deux plateformes : le robot iCat développé par Philips et l'avatar graphique Cherry, et comment nous avons l'adaptée. Pour finir, nous présentons des résultats portant sur la crédibilité de nos expressions sur ces deux plateformes différentes.

2. Agents sociaux

Pour être acceptés dans l'environnement humain, il a été suggéré que les agents ont besoin d'adopter des comportements sociaux comme exprimer leur état interne à travers des expressions émotionnelles faciales ou vocales, ou encore respecter des conventions sociales [FON 02, GRA 05, PIC 97]. Ces agents peuvent être représentés par des avatars graphiques ou des robots sociaux. Les avatars, à l'apparence virtuelle anthropomorphique, sont souvent implémentés par des Agents Conversationnels Animés (ACAs). Un ACA est un agent virtuel qui interagit avec l'utilisateur ou un autre agent à travers un comportement communicatif multimodale [POG 05]. Des études ont déjà été faites impliquant ce type d'agents (figure 1), avec par exemple Greta [POG 05, PEL 03], REA [CAS 99], Grace [AL. 02] ou Cherry [LIS 04]. Cherry est montée sur un robot et aide les personnes à s'orienter à l'université de Floride (UCF). Elle adapte son comportement en fonction de la personne à laquelle elle parle : elle dira "Bonjour Professeur Lisetti" et "Salut" à un étudiant. Cherry possède également une personnalité et des capacités émotives, et est très bien acceptée dans l'environnement humain. En ce qui concerne les robots, deux sortes ont été conçus pour l'interaction sociale : les robots zoomorphiques et les robots anthropomorphiques [FON 02]. Les robots zoomorphiques sont le plus souvent des jouets à l'apparence de chats ou de chiens, par exemple Aibo (Sony) ou iCat (Philips). Nous utilisons cette dernière plateforme pour une partie de nos travaux et nous donnerons plus de détails sur les spécifications de iCat à la section 4. Les robots anthropomorphiques sont beaucoup plus dédiés à l'interaction sociale et ont donc besoin d'exprimer leurs états internes émotionnels, leurs buts et leurs désirs. Des travaux ont été fait pour développer ce genre de robots avec par exemple Kismet [BRE 03] ou bien iCat [van 05].

Les agents décrits ci-dessus utilisent des expressions émotionnelles faciales mais qui ne sont pas modélisées avec une théorie psychologique. De ce fait, ces expressions

ne sont pas liées avec l'état interne de l'avatar causant une perte de la crédibilité et de la compréhension de ces expressions. Nos travaux portent sur la génération d'expressions émotionnelles que nous lierons plus tard avec les états internes des agents.

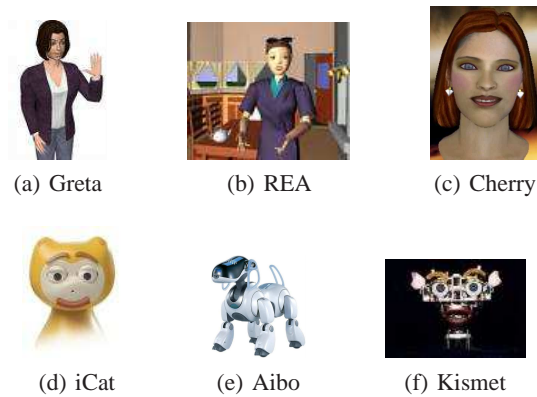


Figure 1. Exemples d'avatars et de robots sociaux : Greta (Pelachaud, 2001), REA (Cassel, 1999), Cherry (Lisetti, 2001), iCat (Philips, 2005), Aibo (Sony, 1993), Kismet (Breazeal, MIT, 2003)

3. La théorie de Scherer pour les prédictions émotionnelles faciales

Scherer [SCH 87, SCH 01] considère les émotions comme un processus avec de constants changements. Sa théorie permet de prédire des réponses émotionnelles avec le visage, le corps, les gestes ou le système nerveux autonome : à partir d'un comportement affectif, il est possible de reconnaître l'émotion, et à partir d'une émotion donnée il est possible de générer l'expression faciale. Dans ce papier nous nous intéressons uniquement aux réponses faciales émotionnelles.

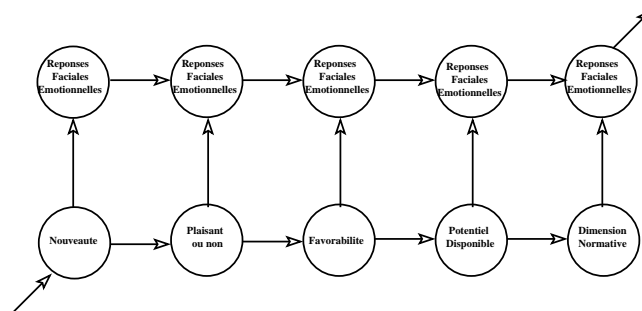


Figure 2. Représentation computationnelle du processus émotionnelle faciale



Figure 3. Exemple d'AUs (AU10, AU15, AU17, combinaison des trois)

Pendant le processus émotionnel (représenté à la figure 2), l'individu évalue séquentiellement un stimulus en fonction d'un ensemble de critères ou SECs (*Sequential Evaluation Checks*) auxquels sont associées des réponses émotionnelles en terme d'Actions Units ou AUs qui ont été définies par Ekman et représentant la position de chaque muscle facial durant les expressions (figure 3). Lorsqu'un événement survient, l'organisme évalue tout d'abord la nouveauté du stimulus pour connaître son importance. Pour cette étape, Scherer propose deux séquences d'AUs possibles : AUs1,2 (haussement des sourcils), AU5 (les paupières qui s'ouvrent) ou bien AUs4,7 (froncement des sourcils), AU26 (la mâchoire qui tombe), en fonction de la nature du stimulus. Le choix d'une d'elles va donner une sous-expression de l'expression globale. Ensuite l'organisme analyse le stimulus afin de savoir si il est plaisant ou non et de cette analyse découle une autre sous-expression. Le processus continue de la même façon avec l'évaluation de l'accomplissement de buts personnels (favorabilité), le potentiel disponible par l'organisme pour atteindre ses buts et la prise en compte de l'évènement en fonction de standards internes ou externes. Pour chaque phase, Scherer donne également le choix entre plusieurs séquences d'AUs. La combinaison et la superposition des sous-expressions permettent d'obtenir l'expression finale. Le début et la fin de chaque sous-expression sont décidés had-hoc de façon à obtenir des expressions le plus crédibles possibles. De plus Scherer donne pour chaque SECs et pour chacune des expressions définies dans [SCH 01] l'intensité à laquelle doit être exprimée les AUs.

4. Deux plateformes différentes : un robot et un avatar

Nous travaillons actuellement sur la création d'expressions faciales crédibles fondée sur l'approche psychologique décrite plus haut. Pour cela nous utilisons deux plateformes différentes : le robot iCat et l'avatar Cherry que nous allons décrire.

iCat est une plateforme robotique développée par Philips [van 05] pour étudier l'interaction humain-robot. iCat a treize moteurs pour les expressions faciales et le mouvement de la tête. Ils permettent de contrôler chacune des parties suivantes : sourcils, paupières, yeux, lèvres, corps et tête. Il a également quatre détecteurs de toucher localisés dans ses pattes et ses oreilles, une webcam placée dans son nez, et des haut-parleurs et un microphone intégrés dans ses pattes. Il possède des lumières (rouges, vertes et bleues) vers ses pattes et ses oreilles qui peuvent traduire l'état interne du robot (sommeil, colère...). Il peut ainsi interagir avec les utilisateurs. Philips a créé



















AUs	Neutre	Moyenne	Très forte
ACTION UNITS POSSIBLES			
AU2 (Sourcil extérieur levé)			
AU5 (Paupière supérieure levée)			
AU12 (Coin des lèvres tiré vers le haut)			
ACTION UNITS EXTRAPOLÉES			
AU4 (Sourcil froncé)			
AU7 (Paupière serrée)			
AU22 (Lèvres en entonnoir)			

Tableau 1. Action Units possibles et extrapolées

des expressions faciales avec iCat en se fondant sur les principes d'animations définis pour les personnages Disney [van 04].

Cherry est un avatar fait avec l'outil commercial Haptik [Hap 05], construit pour pouvoir insérer des avatars dans des applications et dans des pages Web. Il est possible de créer des avatars des deux sexes et de différentes races. La simulation de différents âges est également possible avec l'utilisation de morphologie et de texture de peau. Différents niveaux de contrôle sont disponibles : contrôle des expressions faciales globales, de la morphologie et de la position de l'avatar, ainsi que le contrôle des mouvements faciaux basiques. Nous développons donc des expressions affectives pour cette technologie, en contrôlant les mouvements faciaux basiques de façon à être le plus proche possible des AUs de Ekman [EKM 02].

5. Adaptation des prédictions émotionnelles aux deux plateformes

Toutes les AUs ne peuvent pas être exprimées sur iCat à cause de ses capacités limitées comparées à celles des humains. Nous avons donc ignoré, adapté, ou extrapolé quelques AUs (tableau 1). Par exemple nous pouvons remarquer que les AU2 (le sourcil extérieur qui se lève) et AU4 (froncement du sourcil) sont exprimées de la même façon. Avec notre rapprochement entre les AUs et les mouvements faciaux de iCat, nous sommes capables de créer quelques expressions faciales émotionnelles fondées sur la théorie de Scherer. Une première étude utilisateurs a montré le besoin d'améliorer la crédibilité de nos expressions faciales car les degrés de liberté du visage de iCat sont limités en termes d'AUs comparés à la richesse des expressions des humains [GRI 06]. Nous avons donc adapté les prédictions en utilisant toutes les possibilités d'expressions de iCat : mouvements de tête et des yeux, utilisation des lumières dans les pattes et les oreilles. De plus, nous avons exagéré l'intensité des expressions de iCat comme le fait Philips pour les siennes. Nous donnons quelques exemples de ces prédictions faciales émotionnelles à la figure 5.

Nous avons également modéliser ces expressions sur Cherry en développant un logiciel pour contrôler les mouvements faciaux basiques de l'avatar Haptik en com-

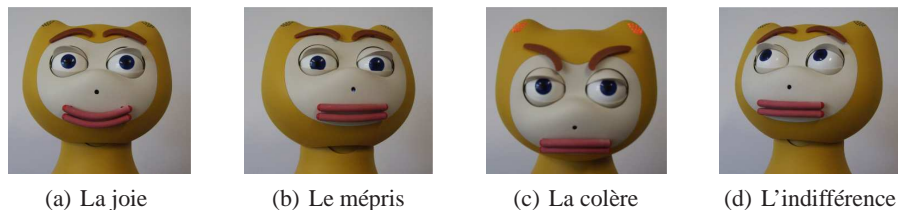


Figure 4. *Quelques exemples d'expressions émotionnelles sur iCat*



Figure 5. *Une évolution possible pour la peur sur Cherry selon la théorie de Scherer*

mençant par les séquences d'AUs, en particulier celles définies dans la théorie de Scherer [SCH 01]. Les avatars Haptik sont très crédibles et nous avons rencontré des problèmes en voulant lier leurs expressions à la théorie que nous utilisons. Nous avons suffisamment de contrôle sur les expressions faciales (comme avec iCat), mais nous manquons de détails pour générer des animations crédibles. En particulier, nous avons besoin d'informations impliquant le temps et l'intensité des AUs concernées pour répondre à des questions telles que : combien de temps dure une partie de l'expression sur le visage ? De plus, certaines prédictions semblent utiliser l'activation des mêmes muscles faciaux par deux SECs en même temps, amenant à se poser des questions sur la crédibilité des expressions. Pour résoudre certains de ces problèmes, nous avons observé des vidéos d'expressions faciales faites par des acteurs et nous en avons extrait des informations temporelles. Nous avons adapté les intensités des AUs pour se concentrer sur les sous-expressions discernables dans les vidéos. Les ambiguïtés et les répétitions des AUs pour l'activation des muscles faciaux ont pu être résolues de cette manière. Finalement, nous avons inclus des mouvements de tête et un contrôle basic sur le regard de l'avatar. Dans [PAL 06], nous détaillons la procédure complète que nous avons développée ainsi que les problèmes que nous avons rencontrés. Un exemple de l'évolution d'une expression est donné à la figure 5.

6. Évaluation de la reconnaissance et de la crédibilité de nos expressions

Nous avons mené des études utilisateurs pour évaluer le taux de reconnaissance et la crédibilité de nos expressions. Pour chacune des études, quinze personnes y ont participé (trois femmes et douze hommes) entre vingt et trente ans. Les expressions leur étaient montrées séparément à partir de la position neutre. Pour la reconnaissance,

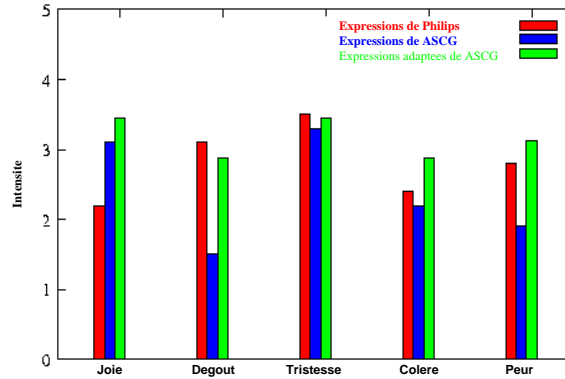


Figure 6. *Crédibilité des émotions pour iCat*

ils devaient choisir quelle était l'expression représentée parmi la joie, le dégoût, le mépris, la tristesse, la fierté, la peur, la colère, l'indifférence et la honte. Nous avons évalué le taux de reconnaissance de ces neuf expressions. Pour la crédibilité, les participants connaissaient l'expression qui leur était montré et devaient choisir entre pas du tout crédible (0) jusqu'à très crédible (5). Les résultats obtenus pour la première expérience montrent que les expressions les mieux reconnues sont la joie, la colère, la peur et l'indifférence pour iCat, la moyenne étant de presque 60% [GRI 06]. Une étude similaire a été faite avec l'avatar. Seules cinq expressions devaient être évaluées (le dégoût, la peur, la tristesse, la colère et la joie). Les résultats ont montré que la moyenne de reconnaissance était de 97% (voir [PAL 06] pour des résultats plus complets).

Les résultats de la seconde expérience sont représentés aux figures 6 et 7. La crédibilité de cinq de nos expressions, notées "Expressions adaptées par ASCG", a été évaluée : le dégoût, la peur, la tristesse, la colère et la joie. Pour iCat (figure 6), nous comparons la crédibilité de nos expressions (adaptées de la théorie de Scherer) avec les expressions de Philips et avec notre première implémentation (sans adaptation). Nous pouvons observer une amélioration de nos résultats en adaptant la théorie. Nos expressions apparaissent plus crédibles que celles de Philips sauf pour la tristesse et le dégoût. Pour Cherry (figure 7), nous comparons la crédibilité de nos expressions avec des expressions humaines et celles de Hapték. Les résultats sont en général satisfaisants, la crédibilité de nos expressions sont toujours proches de celles développées par Hapték.

7. Conclusion

Dans cette étude, nous avons montré comment la théorie psychologique de Scherer peut être utilisée dans le but de représenter des expressions faciales. Nous avons adapté les prédictions de Scherer à un avatar et un robot. Les résultats ont montré le succès de

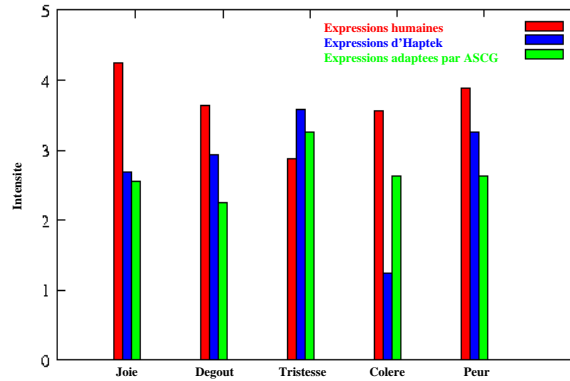


Figure 7. Crédibilité des émotions pour Cherry

l'adaptation de la théorie en terme de reconnaissance et de crédibilité. Notre génération d'émotion peut maintenant créer des expressions sur deux plateformes différentes. Il existe des travaux similaires faits par Kaiser et Wehrle [KAI 01] dans lesquels ils génèrent des expressions faciales créées à partir de la théorie que nous utilisons sur un visage synthétique. Contrairement à nos deux plateformes, ce visage peut exprimer les AUs. Pour pouvoir appliquer les prédictions faites par Scherer, nous avons étudié pour chaque plateforme comment mettre en corrélation leurs façons de bouger certaines parties du visage avec ces AUs.

Nous lierons plus tard ces expressions émotionnelles avec l'état interne de l'agent, généré à partir de la même théorie, avec pour but final d'implémenter une architecture d'agent social intelligent sur des plateformes indépendantes.

8. Bibliographie

- [AL 02] ET AL. R. S., « GRACE : An Autonomous Robot for the AAI Robot Challenge », SMART W. D., BALCH T., YANCO H., Eds., *AAAI Mobile Robot Competition and Exhibition : Papers from the AAI Workshop*, Edmonton, Alberta, août 2002, AAI Press, p. 1–14, AAI Technical Report WS-02-18.
- [BRE 03] BREAZEAL C., « Towards Sociable robots », *Robotics and Autonomous Systems*, vol. 42, n° 3-4, 2003, p. 167-175.
- [CAS 99] CASSELL J., BICKMORE T., BILLINGHURST M., CAMPBELL L., CHANG K., VILHJáLMSSON H., YAN H., « Embodiment in Conversational Interfaces : Rea », *CHI '99 : Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, 1999, ACM Press, p. 520–527.
- [EKM 02] EKMAN P., FRIESEN W. V., HAGER J. C., *Facial Action Coding System Investigator's Guide*, A Human Face, 2002.
- [FON 02] FONG T., NOURBAKSH I., DAUTENHAHN K., « A Survey of Socially Interactive Robots », *Robotics and Autonomous Systems*, vol. 42, 2002.

- [GRA 05] GRATCH J., MARSELLA S., « Lessons from Emotion Psychology for the Design of Lifelike Characters », *Journal of Applied Artificial Intelligence (special issue on Educational Agents - Beyond Virtual Tutors)*, vol. 19, n° 3-4, 2005, p. 215-233.
- [GRI 06] GRIZARD A., LISETTI C. L., « Generation of Facial Emotional Expressions Based on Psychological Theory », *1rst Workshop on Emotion and Computing at KI 2006, 29th Annual Conference on Artificial Intelligence, June, 14-19, 2006, Bremen, Germany*, Jun 2006.
- [Hap 05] HAPTEK, « <http://www.Haptek.com> », 2005.
- [KAI 01] KAISER S., WEHRLE T., « The Role of Facial Expression in Intra-individual and Inter-individual Emotion Regulation », CANAMERO D., Ed., *Emotional and Intelligent II : The Tangled Knot of Cognition*, Menlo Park, CA, 2001, AAAI Press, p. 61-66.
- [LIS 04] LISETTI C. L., BROWN S., ALVAREZ K., MARPAUNG A., « A Social Informatics Approach to Human-Robot Interaction With a Service Social Robot », *IEEE Transactions on Systems, Man, and Cybernetics - Special Issue on Human Robot Interaction*, vol. 34, n° 2, 2004.
- [PAL 06] PALEARI M., GRIZARD A., LISETTI C., « Phsycologically Grounded Avatar Expressions », *Proceedings of KI06 26th German Annual Conference in Artificial Intelligence*, Bremen, Germany, 2006.
- [PEL 03] PELACHAUD C., BILVI M., « *Communication in Multiagent Systems : Background, Current Trends and Future. Lecture Notes in Computer Science* », vol. 2650, chapitre Computational model of believable conversational agents, p. 300-317, Springer, Berlin, 2003.
- [PIC 97] PICARD R., *Affective Computing*, MIT Press, Cambridge (MA), 1997.
- [POG 05] POGGI I., PELACHAUD C., DE ROSIS F., CAROFIGLIO V., DE CAROLIS B., « *Multimodal Intelligent Information Presentation* », vol. 27, chapitre Great. A Believable Embodied Conversational Agent, Springer, 2005.
- [SCH 87] SCHERER K. R., « Toward a dynamic theory of emotion : The component process model of affective states », *Geneva Studies in Emotion and Communication*, 1(1), 1-98, 1987.
- [SCH 01] SCHERER K. R., « *Appraisal Processes in Emotion : Theory, Methods, Research* », chapitre Appraisal Considered as a Process of Multilevel Sequential Checking, p. 92-120, New York, NY, US : Oxford University Press, 2001.
- [van 04] VAN BREEMEN A., « Binging Robots To Life : Applying Principles of Animation To Robots », *Proceedings of Shapping Human-Robot Interaction workshop held at CHI 2004*, Vienna, Austria, 2004.
- [van 05] VAN BREEMEN A., « iCat : Experimenting with Animabotics », DAUTENHAHN K., NEHANIV C., TE BOEKHORST R., CAÑAMERO L., POLANI D., HEWITT J., Eds., *AISB*, University of Hertfordshire, Hatfield, UK, April 2005, AISB, p. 27-32.