Conception et Evaluation d'un Modèle d'Expressivité pour les Gestes des Agents Conversationnels

Stéphanie Buisine* — Björn Hartmann,*** — Maurizio Mancini** — Catherine Pelachaud****

* *Ecole Nationale Supérieure d'Arts et Métiers, LCPI*
151 bd de l'Hôpital, F-75013 Paris
stephanie.buisine@paris.ensam.fr

** *Université de Paris 8, IUT de Montreuil*
140 rue de la Nouvelle France, F-93100 Montreuil
{m.mancini, c.pelachaud}@iut.univ-paris8.fr

*** *Stanford University Computer Science Dept.*
bjoern@stanford.edu

RÉSUMÉ. *Dans cet article nous présentons un modèle permettant de caractériser les gestes à l'aide d'une série de paramètres d'expressivité, puis de générer des gestes expressifs pour les Agents Conversationnels Animés, l'objectif étant d'augmenter la crédibilité des Agents et le naturel de leurs comportements. Nous présentons ensuite deux tests perceptifs destinés à évaluer expérimentalement notre modèle. Les résultats montrent que certains des paramètres sont perçus de manière stable, mais que la qualité de l'animation est parfois insuffisante pour produire des changements subtils d'expressivité. De plus, il apparaît nécessaire d'étudier de manière plus approfondie les effets d'interaction entre paramètres.*

ABSTRACT. *To increase the believability and life-likeness of Embodied Conversational Agents, we introduce a behavior synthesis model for the generation of expressive gesturing. A small set of dimensions of expressivity is used to characterize individual variability of movement. We empirically evaluate our implementation in two separate user studies. The results suggest that our approach works well for a subset of expressive behavior. However, animation fidelity is not high enough to realize subtle changes. Interaction effects between different parameters need to be studied further.*

MOTS-CLÉS : *Agent Conversationnel Animé, Geste, Expressivité, Test Perceptif, Evaluation Expérimentale.*

KEYWORDS: *Embodied Conversational Agents, Gesture, Expressivity, Perceptual Test, Experimental Evaluation.*

1. Introduction

Les Agents Conversationnels Animés sont un type d'interface visant à transférer les propriétés de l'interaction Homme-Homme, avec toute leur richesse, à l'interaction Homme-Machine. Il s'agit de personnages virtuels multimodaux capables de communiquer avec l'utilisateur par de multiples modalités : la voix, les expressions faciales, la direction du regard, les gestes des mains, les postures corporelles, etc. L'efficacité d'un Agent étant en partie liée à sa crédibilité au cours de l'interaction, l'Agent doit exprimer des émotions et des traits de personnalité de manière cohérente (Loyall & Bates, 1997), ce qui est également susceptible d'accroître son acceptabilité auprès des utilisateurs (Lisetti *et al.*, 2004). Or, les individus humains diffèrent entre eux non seulement au niveau de leurs raisonnements, leurs croyances, leurs buts et leurs états émotionnels, mais également au niveau de la manière d'exprimer ces informations à travers leurs comportements. La finalité de cette recherche est de nous rapprocher d'un modèle dans lequel les Agents pourraient produire des comportements individualisés et idiosyncrasiques plutôt que des actions génériques et stéréotypées.

Il reste de nombreux défis à relever au niveau technique pour atteindre un tel objectif, mais il nous faut également intégrer la perception des utilisateurs dans le processus de développement des Agents, afin de guider les choix de conception et d'initier un cycle itératif d'amélioration de notre modèle. L'évaluation des Agents est un domaine de recherche encore émergent (Ruttkay & Pelachaud, 2004), dans lequel on distingue deux axes principaux : la *micro-évaluation*, dans laquelle un point particulier de l'implémentation de l'Agent est testé vis-à-vis du comportement humain qu'il cherche à modéliser, et la *macro-évaluation*, dans laquelle c'est l'application entière qui est évaluée sur les dimensions d'utilité, d'efficacité et de satisfaction. La présente étude se situe dans le champ de la micro-évaluation.

GRETA est un Agent multimodal capable d'interpréter un texte dans lequel des fonctions communicatives ont été annotées (De Carolis *et al.*, 2004) et de générer en retour une séquence comportementale cohérente avec paroles, expressions faciales, direction du regard et gestes appropriés. La synthèse vocale est réalisée à l'aide de Festival (Black *et al.*, 1999). Le rythme de la synthèse vocale sert de base à la synchronisation des autres modalités. Chaque moteur de comportement (expressions faciales/regard et gestes) produit une animation FAP/BAP compatible avec le format MPEG-4 (Taubin, 1998) qui à son tour contrôle le modèle facial et le squelette corporel en OpenGL. Une description détaillée de l'architecture de GRETA peut être trouvée par ailleurs (Hartmann *et al.*, 2002; Pelachaud *et al.*, 2002; Bevacqua & Pelachaud, 2003). Dans cet article, nous décrivons la conception et l'évaluation d'un module de contrôle du contenu expressif des gestes des mains et des bras, qui vient enrichir le système gestuel de GRETA.

La suite de l'article est structurée de la manière suivante : dans la section 2, nous présentons notre modèle du paramétrage des gestes humains et nous décrivons comment nous avons traduit les dimensions d'expressivité théoriques en

spécifications d'animation de bas niveau. Dans la section 3, nous positionnons nos travaux vis-à-vis de l'état de l'art actuel. La section 4 décrit la méthode employée dans les deux expérimentations que nous avons menées, et la section 5 rapporte l'analyse et les résultats de ces expérimentations. Ces résultats sont interprétés et discutés en section 6, puis la section 7 propose une conclusion à cette étude et une ouverture vers de nouvelles recherches.

2. Une modélisation de l'expressivité

Notre objectif général est de créer un Agent Conversationnel Animé capable de faire preuve d'individualité à la fois au travers (i) du choix de ses signaux non verbaux et (ii) des propriétés expressives de ces signaux. Concernant le choix des comportements non verbaux, nous avons précédemment décrit une méthode permettant de sélectionner des gestes en fonction de critères de haut niveau et de données contextuelles (Maya *et al.*, 2004). De manière complémentaire, la présente étude ne concerne que la manière de réaliser des gestes donnés – et non de sélectionner ceux-ci.

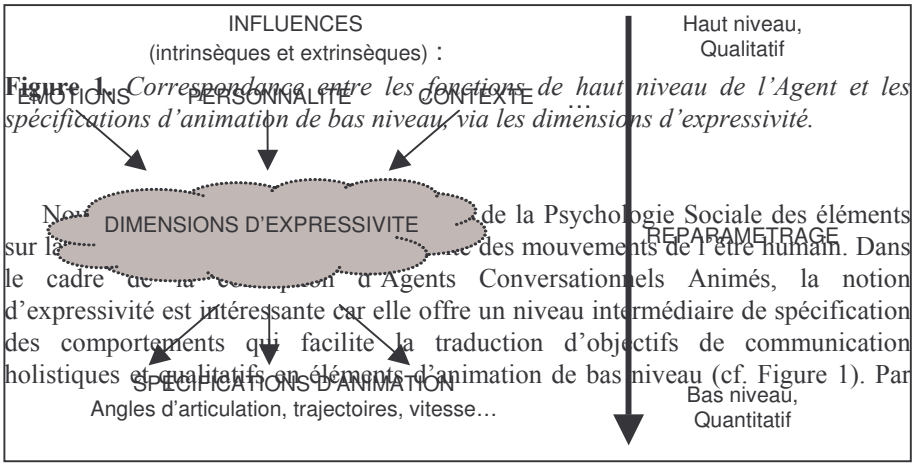


Figure 1. *Correspondance entre les fonctions de haut niveau de l'Agent et les spécifications d'animation de bas niveau via les dimensions d'expressivité.*

Nous nous sommes inspirés de la Psychologie Sociale des éléments sur la perception des mouvements de l'être humain. Dans le cadre de la conception d'Agents Conversationnels Animés, la notion d'expressivité est intéressante car elle offre un niveau intermédiaire de spécification des comportements qui facilite la traduction d'objectifs de communication holistiques et qualitatifs en éléments d'animation de bas niveau (cf. Figure 1). Par

exemple, face à une situation donnée, un individu développera, compte tenu de sa personnalité, une réaction d'excitation. D'un point de vue comportemental, cette excitation pourra se traduire notamment par une quantité accrue de mouvements, des gestes rapides et saccadés, etc. Afin de générer de façon réaliste un comportement d'excitation, il est nécessaire d'identifier les dimensions d'expressivité sur lesquelles l'excitation doit jouer. Notre approche repose sur un point de vue perceptif (comment l'expressivité est-elle perçue par les autres ?), c'est-à-dire que nous cherchons à modéliser les mouvements des surfaces et non l'activation des muscles sous-jacents.

A partir des données rapportées dans la littérature (Wallbott & Scherer, 1986; Gallaher, 1992; Wallbott, 1998), et à partir de notre propre analyse d'un corpus gestuel (Martell *et al.*, 2003), nous avons circonscrit l'expressivité des gestes par un ensemble de six paramètres :

– *Le niveau d'activation général* : quantité de mouvements pendant un tour de parole (par exemple, passif/statique *vs*. animé/impliqué).

– *L'amplitude spatiale* des gestes (par exemple, le volume occupé par le corps).

– *L'amplitude temporelle* : durée des gestes (par exemple, mouvement rapide *vs*. maintenu).

– *La fluidité* : régularité et continuité des mouvements (par exemple, enchaînement doux, harmonieux *vs*. brusque, saccadé).

– *L'intensité* : force, propriétés dynamiques du mouvement (par exemple, faible ou relâché *vs*. fort ou tendu).

– *La répétition* : répétition rythmique de certains mouvements spécifiques.

Pour le moment, les changements de posture ne sont pas intégrés dans ce modèle, bien que leur importance en tant que canal de communication ait été soulignée par ailleurs (Mehrabian, 1969; Harrigan, 1985).

Un Agent GRETA individualisé aura des valeurs par défaut personnelles pour chacun des paramètres ci-dessus. Des indications sur ces paramètres peuvent également être insérées parmi les marqueurs de fonctions communicatives au sein de notre langage balisé. Ces données d'expressivité sont ensuite transformées par notre moteur gestuel en spécifications d'animation de bas niveau. Nous décrivons ci-dessous brièvement l'implémentation de chaque paramètre, mais le lecteur intéressé est renvoyé à (Hartmann *et al.*, 2005) pour une description complète des détails techniques.

Le *niveau d'activation général* est implémenté comme un seuillage au niveau de la planification du geste. Si l'attribut de poids du marqueur communicatif dans le texte d'entrée atteint ou dépasse ce seuil, le marqueur est mis en correspondance avec la base de gestes et un geste est généré. L'*amplitude spatiale* est contrôlée par le dimensionnement des volumes d'atteinte (voir McNeill, 1992) qui déterminent les positions clés des poignets de l'Agent dans l'espace. De plus, l'angle de rotation du coude (angle entre un vecteur pointant vers le bas et la projection du coude sur la

droite reliant l'épaule et le poignet) est ajusté pour modifier l'espace occupé par l'Agent – des coudes déployés vers l'extérieur élargissent la silhouette. L'*amplitude temporelle* contrôle la vitesse du *stroke* [1] en augmentant ou en diminuant la durée du *stroke* par un calcul basé sur la loi de Fitts (1954). La *fluidité* affecte l'enchaînement des gestes – comment la rétraction d'un geste est combinée avec la préparation du geste suivant. Les seuils temporels de rétraction des bras vers une position neutre sont rallongés dans les mouvements fluides. La fluidité affecte également la continuité des courbes Spline de Kochanek-Bartels (1984) utilisées pour l'interpolation de la trajectoire du bras. Pour visualiser l'*intensité* d'un mouvement, nous contrôlons également les paramètres Spline de Kochanek-Bartels et, si la forme de la main n'est pas essentielle au geste (comme c'est le cas pour les gestes de battement), nous y associons une forme de main tendue ou détendue. Enfin, la *répétition* est réalisée par la technique de l'*expansion du stroke* qui a été décrite par ailleurs (Hartmann *et al.*, 2002) : cette technique répète le *stroke* du geste de sorte que les *strokes* successifs coïncident avec les emphases verbales.

3. Etat de l'art

Les ouvrages de Cassell *et al.* (2000) et Prendinger & Ishizuka (2004) offrent un aperçu des développements récents dans le domaine des Agents Conversationnels Animés. De nombreux auteurs étudient spécifiquement la génération des gestes chez les Agents, et leurs travaux peuvent être organisés en deux catégories : ceux qui traitent du problème de la sélection des gestes, et ceux qui étudient l'animation des gestes.

Les travaux sur la sélection des gestes se sont principalement intéressés à la sémantique des gestes chez l'humain, en suivant généralement la classification de McNeill (1992). Cassell *et al*. (2001) sélectionnent des comportements non verbaux adéquats pour accompagner le discours, à partir d'une analyse linguistique. Plus récemment, l'équipe de Cassell a proposé un modèle paramétrique de génération de gestes iconiques, fondé sur l'analyse d'un corpus vidéo (Tepper *et al.*, 2004). Noot & Ruttkay (2004) ont quant à eux proposé GESTYLE pour répondre au besoin de variabilité interindividuelle dans le style des gestes. Leur outil permet de choisir des éléments comportementaux à partir de dictionnaires de style. L'idée d'adapter des gestes existants est également introduite, mais sans être décrite en détail. Par ailleurs, aucune évaluation formelle n'a été menée avec GESTYLE.

Le domaine de recherche sur l'animation des gestes traite de la génération de mouvements réalistes des bras et des mains (Lebourque & Gibet, 1999; Kopp &

[1] D'un point de vue structurel, la réalisation d'un geste est classiquement décomposée de la manière suivante (McNeill, 1992) : le moment principal du geste, la phase la plus énergique est appelée le *stroke* ; une phase de *préparation* précède le *stroke*, elle consiste à amener les bras et les mains jusqu'à la position de *stroke* ; enfin, le geste s'achève avec la phase de *rétraction*, lorsque bras et mains retournent à leur position de repos.

Wachsmuth, 2000; Hartmann *et al.*, 2002). Les systèmes d'animation incluent souvent leur propre langage de représentation pour décrire les gestes. Le système EMOTE (Chi *et al.*, 2000) est étroitement lié à notre travail puisqu'il implémente un modèle modifiant les gestes des Agents pour y ajouter de l'expressivité (en utilisant les principes d'Effort et de Forme de Laban & Lawrence, 1974). Cependant, EMOTE permet uniquement de modifier – pas de générer – des gestes. EMOTE applique une transformation au niveau de l'articulation des membres sans tenir compte du type de geste représenté. A l'inverse, nous utilisons dans notre approche l'information sémantique issue du processus de sélection pour guider et contraindre les adaptations. Un autre point de divergence concerne la méthodologie d'évaluation : l'évaluation d'EMOTE a été menée avec un petit nombre d'utilisateurs entraînés, alors que de notre côté nous avons recouru à un grand nombre d'utilisateurs non entraînés.

De manière générale, les études citées ci-dessus ont été faiblement validées. D'autres études se sont au contraire concentrées uniquement sur l'évaluation des Agents, avec des systèmes parfois spécifiés manuellement : Buisine *et al*. (2004) ont par exemple évalué l'effet des variations des stratégies multimodales et de l'apparence des Agents sur les impressions subjectives et les performances de rappel des utilisateurs. Schröder (2004) a mené une évaluation semblable à la nôtre dans le but de tester l'implémentation de l'expressivité verbale. C'est à partir de son étude que nous avons choisi le type de tâche (ordonner différents items en fonction de leur adéquation à un critère) de notre seconde expérimentation. A notre connaissance, aucune évaluation à grande échelle de l'expressivité des gestes des Agents n'a jamais été rapportée.

4. Protocole expérimental

4.1. *Participants*

106 utilisateurs ont participé à nos deux expérimentations. Tous étaient étudiants de première et deuxième année à l'Université de Paris 8, leur participation étant un exercice obligatoire dans le cadre de leur formation. 80 utilisateurs étaient de sexe masculin, 26 de sexe féminin. Leur âge était compris entre 17 et 25 ans. 56 d'entre eux savaient déjà avant l'expérience ce qu'est un avatar ou un agent autonome. L'ensemble des données a été recueilli de manière anonyme.

Nos deux expériences ont été réalisées par deux groupes différents : 52 utilisateurs ont participé au test 1 et 54 ont participé au test 2.

4.2. *Procédure*

Les passations se sont déroulées par groupes de 5 à 7 utilisateurs dans deux salles de tests. Un ordinateur équipé d'un casque audio a été attribué à chaque utilisateur.

Les deux tests consistaient à visionner de courts clips vidéo dans lesquelles l'Agent GRETA produisait une séquence comportementale unique avec des variations d'expressivité sur une ou plusieurs dimensions. Le test 1 consistait à comparer deux conditions par essai, et le test 2 à comparer quatre conditions par essai (cf. Figures 2 et 3) : l'utilisateur devait répondre à des questions sur ces différentes conditions à l'aide d'une interface graphique classique.

Tous les clips ont été générés de la même manière : l'apparence de l'Agent était la même dans tous les cas, de même pour la position de la caméra et l'éclairage. L'Agent jouait toujours la même réplique (en Anglais), issue d'une émission télévisée. Festival (Black *et al.*, 1999) a été utilisé pour la synthèse vocale, et les gestes ont été générés à l'aide de notre propre boîte à outil : les paroles et les gestes étaient synchronisés par notre moteur comportemental.

Pour les deux expériences, les utilisateurs devaient lire la feuille de consignes avant de commencer. Ces consignes incluaient également une courte description de ce qu'est un Agent Conversationnel Animé et présentaient brièvement la finalité de cette recherche (c'est-à-dire créer un Agent expressif). Les utilisateurs avaient la possibilité de poser des questions aux expérimentateurs s'ils le souhaitaient, mais aucune question n'a été posée. Les différents essais constituant l'expérience s'enchaînaient ensuite de manière linéaire – il n'était pas possible de revenir à un essai antérieur pour modifier une réponse déjà validée. Aucune contrainte temporelle n'était imposée aux utilisateurs pour répondre à chaque essai : ils pouvaient rejouer les séquences vidéo autant de fois qu'ils le souhaitaient avant de répondre. Quand tous les utilisateurs d'un groupe avaient fini l'expérience, un nouveau groupe était appelé dans la salle de test.

4.2.1. *Test 1*

Le but de cette expérience était de tester l'hypothèse (H1) : notre implémentation des paramètres d'expressivité (traduction en spécifications d'animation de bas niveau) est appropriée – la modification d'un paramètre sera reconnue par les utilisateurs.

Pour cela, nous avons généré deux séquences vidéo par paramètre : une avec le paramètre à sa valeur maximale, une avec sa valeur minimale. Par exemple pour le paramètre d'*amplitude temporelle*, une des séquences avait des *strokes* très rapides, l'autre des *strokes* très lents. Douze séquences ont ainsi été créées pour notre ensemble de six paramètres. Pour chaque participant, la passation de l'expérience

consistait en 14 essais : une succession aléatoire des douze séquences test et deux supplémentaires (dupliqués arbitrairement parmi les douze autres).

Chacune de ces séquences test était comparée à une séquence contrôle de référence dans laquelle tous les paramètres d'expressivité avaient une valeur neutre. Cette vidéo « neutre » ne correspond pas nécessairement à un comportement « neutre » (si un tel état existe) mais a juste servi de point de référence pour ce test. Les utilisateurs devaient donc identifier par un choix forcé quelle dimension avait été modifiée dans la vidéo test par rapport à la vidéo contrôle, et quel était le sens de la modification (augmentation ou diminution du paramètre identifié).

L'interface utilisée pour ce test est reproduite dans le Figure 2. La vidéo contrôle était toujours située à gauche et la vidéo test à droite. L'utilisateur devait sélectionner le paramètre différant, selon lui, entre les deux vidéos. Il pouvait aussi répondre *Aucune différence* ou *Sans avis*. Une fois qu'il avait sélectionné une réponse, l'utilisateur devait déterminer le sens de la modification (plus intense ou moins intense). Il avançait ensuite vers l'essai suivant.

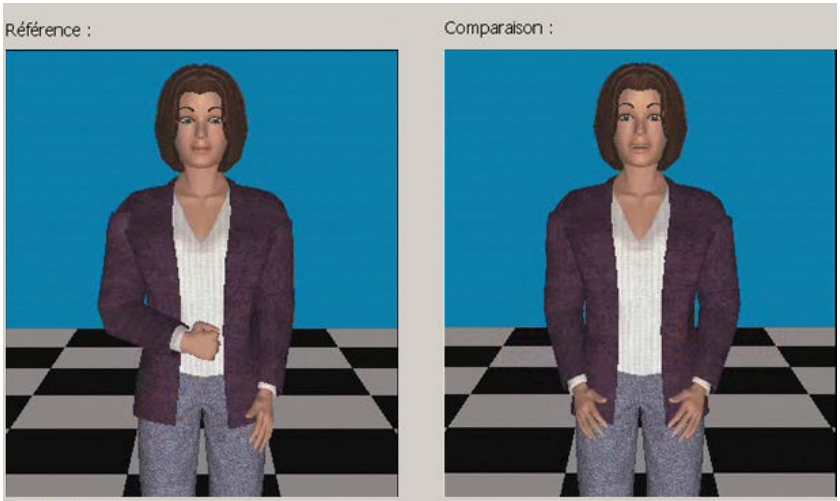


Figure 2. *Interface utilisée pour le test 1.*

4.2.2. *Test 2*

Le second test a été conçu pour tester l'hypothèse (H2) : la combinaison de valeurs appropriées pour les six paramètres crée des comportements cohérents – les utilisateurs pourront reconnaître l'intention communicative sous-jacente.

Pour tester cette hypothèse, nous avons considéré trois types de descripteurs comportementaux : *brusque*, *mou* et *tonique*. Un soin particulier a été porté au choix du vocabulaire utilisé pour ces descripteurs. Nous avons évité les adjectifs impliquant un état émotionnel (par exemple, *joyeux* ou *furieux*) ou un trait de personnalité (par exemple *nerveux* ou *timide*). De tels adjectifs auraient en effet été susceptibles d'engendrer des attentes sur le type de gestes. Or dans cette expérience nous ne nous intéressons pas au choix du type de geste mais à la manière dont les gestes sont produits.

Pour chaque descripteur (*brusque*, *mou* et *tonique*), nous avons généré quatre séquences vidéo : une neutre, correspondant à la vidéo contrôle du test 1, deux variantes de l'intention communicative (une séquence fortement expressive, une autre faiblement expressive) et une version antagoniste, avec les valeurs opposées des paramètres (cf. Tableau 1). Les utilisateurs devaient ordonner ces quatre séquences vidéo selon leur adéquation au descripteur comportemental (de la séquence la plus appropriée à la moins appropriée). Deux vidéos ne pouvaient pas avoir le même rang, et les utilisateurs ne pouvaient passer à l'essai suivant que lorsqu'ils avaient ordonné les quatre vidéos. La Figure 3 représente l'interface utilisée pour cette expérience.

	Très brusque	Légèrement brusque	Antagoniste (brusque)
Activation générale	0,6	0,6	0,5
Amplitude spatiale	0	0	0
Amplitude temporelle	1	0,5	-1
Fluidité	-1	-0,5	1
Intensité	1	0,5	-1
Répétition	-1	-0,5	1

Tableau 1. *Valeurs des paramètres d'expressivité pour le descripteur Brusque.*

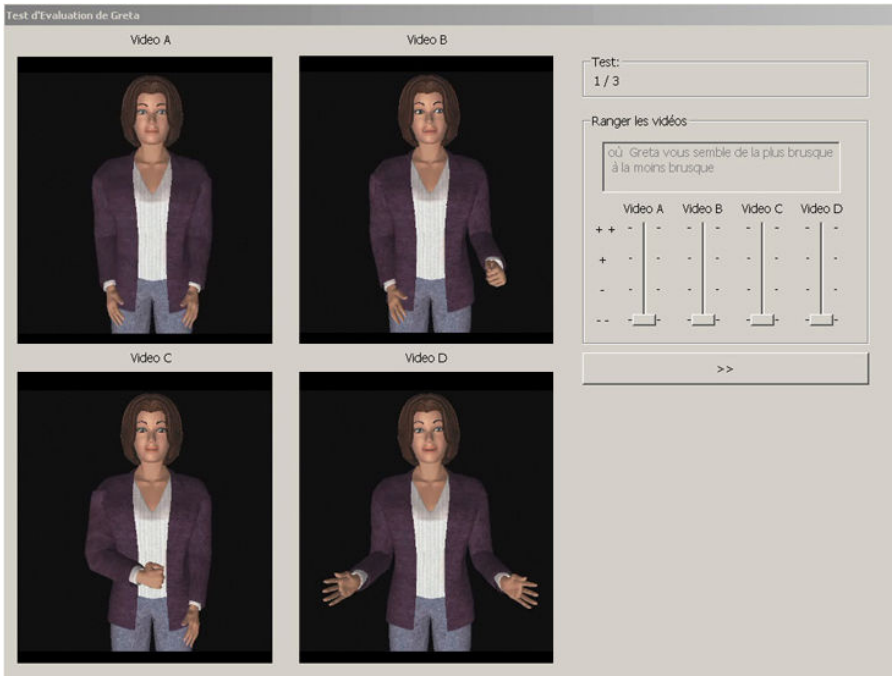


Figure 3. *Interface utilisée pour le test 2.*

5. Résultats

Les analyses que nous avons réalisées ont toutes été destinées à tester la relation entre notre intention d'animation (l'impression que nous avons voulu créer) et la perception des utilisateurs. Pour le test 1, les modifications réelles et perçues étant organisées en une variable catégorisée (les paramètres d'expressivité), nous avons utilisé le calcul du Khi². Pour le test 2, nous avons transformé la variable catégorisée (adéquation au descripteur : ++, +, -, --) en variable numérique (2, 1, 0, -1) de façon à utiliser l'analyse de variance et la corrélation linéaire pour traiter les résultats.

5.1. *Test 1*

Le tableau 2 présente la distribution des réponses des utilisateurs pour chaque paramètre. Les cellules grisées situées sur la diagonale indiquent donc les cas de correspondance entre modifications réelles et modifications perçues : ces cases totalisent 320 réponses, ce qui correspond à 43,1% d'identification correcte des paramètres.

		Modification perçue								Total
		Ampl. Spatiale	Ampl. Temp.	Fluidité	Intensité	Répét.	Niv. Activ.	Aucune diff.	Sans avis	
Paramètre modifié	Amplitude spatiale	77	2	5	5	3	3	3	8	106
	Amplitude temporelle	3	104	7	13	7	1	1	5	141
	Fluidité	2	4	42	10	23	2	34	7	124
	Intensité	7	8	23	42	9	6	27	8	130
	Répétition	18	12	17	20	35	5	10	8	125
	Niveau d'activation	7	7	7	17	6	20	41	11	116
Total		114	137	101	107	83	37	116	47	742

Tableau 2. *Distribution des réponses des utilisateurs en fonction du paramètre modifié.*

Le test du Khi² montre que cette répartition ne peut être attribuée au hasard ($\chi^2(35) = 844{,}16$; $p<0{,}001$). Le tableau 3 présente, pour chacun des paramètres, le taux d'attraction/répulsion entre modification réelle et modification perçue – les attractions/répulsions sont l'équivalent d'un coefficient de corrélation linéaire pour les variables catégorisées. Pour chaque paramètre, l'attraction la plus forte est située sur la diagonale, mais la force des attractions varie beaucoup selon les paramètres (du simple au triple). De plus, il faut souligner que le paramètre *Niveau d'activation* et la réponse *Aucune différence* ont une attraction forte (1,26).

		Modification perçue							
		Ampl. Spatiale	Ampl. Temp.	Fluidité	Intensité	Répét.	Niv. Activ.	Aucune diff.	Sans avis
Paramètre modifié	Amplitude spatiale	3.73	-0.90	-0.65	-0.67	-0.75	-0.43	-0.82	0.19
	Amplitude temporelle	-0.86	2.99	-0.64	-0.36	-0.56	-0.86	-0.95	-0.44
	Fluidité	-0.90	-0.83	1.49	-0.44	0.66	-0.68	0.75	-0.11
	Intensité	-0.65	-0.67	0.30	1.24	-0.38	-0.07	0.33	-0.03
	Répétition	-0.06	-0.48	0.00	0.11	1.50	-0.20	-0.49	0.01
	Niveau d'activation	-0.61	-0.67	-0.56	0.02	-0.54	2.46	1.26	0.50

Tableau 3. *Coefficients d'attraction/répulsion.*

Nous avons également étudié la perception que les utilisateurs ont eue du sens de la modification. Pour cela, nous n'avons retenu de l'échantillon initial que les « bonnes réponses », c'est-à-dire les cas où le paramètre modifié avait été correctement identifié (soit 320 essais sur 742). Le tableau 4 présente la distribution des réponses pour chacun des deux sens de modification. Les identifications correctes (cellules grisées) représentent 63,75% des réponses.

		Sens de modification perçu			Total
		-1	0	+1	
Sens de la modification	-1	110	19	13	142
	+1	50	34	94	178
Total		160	53	107	320

Tableau 4. *Distribution des réponses des utilisateurs en fonction du sens de la modification.*

Le Khi² ($\chi^2(2) = 85,09$; $p<0,001$), ainsi que les taux d'attraction/répulsion (cf. Tableau 5) confirment que le sens de modification réel et perçu sont reliés positivement.

		Sens de modification perçu		
		-1	0	+1
Sens de la modification	-1	0.55	-0.19	-0.73
	+1	-0.44	0.15	0.58

Tableau 5. *Coefficients d'attraction/répulsion.*

5.2. *Test 2*

Pour le descripteur *brusque*, l'effet de nos stimuli est significatif ($F(3/153) = 31,23$; $p<0,001$), ce qui signifie que les utilisateurs ont réussi à discriminer les différents degrés de brusquerie que nous avons cherché à créer (cf. Figure 4). La relation entre le réglage de nos paramètres et la perception des utilisateurs peut aussi s'exprimer par un coefficient de corrélation linéaire, qui atteint ici +0,655.

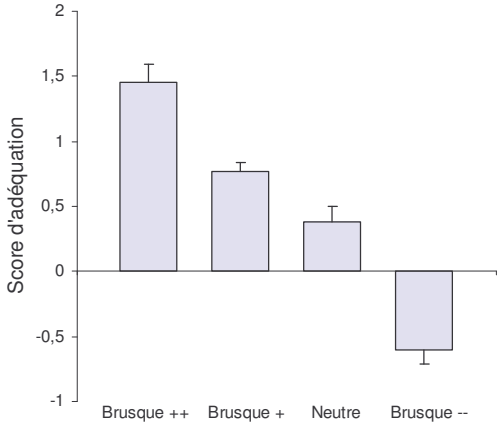
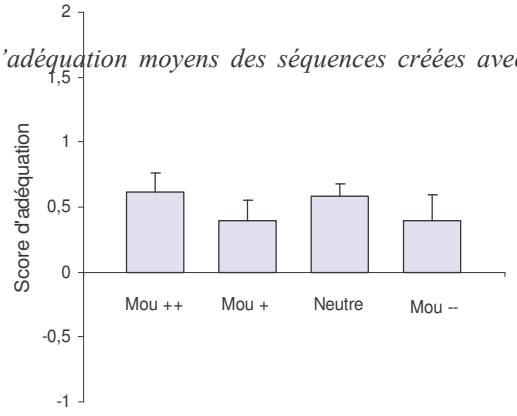


Figure 4. *Scores d'adéquation moyens des séquences créées avec le descripteur Brusque.*

A l'inverse pour le descripteur *mou*, l'effet de nos stimuli n'est pas significatif (F(3/153) = 0,71 ; NS) : comme le montre la figure 5, le taux de reconnaissance des stimuli est du niveau du hasard et le coefficient de corrélation linéaire est presque nul (+0,047).

Figure 5. *Scores d'adéquation moyens des séquences créées avec le descripteur Mou.*



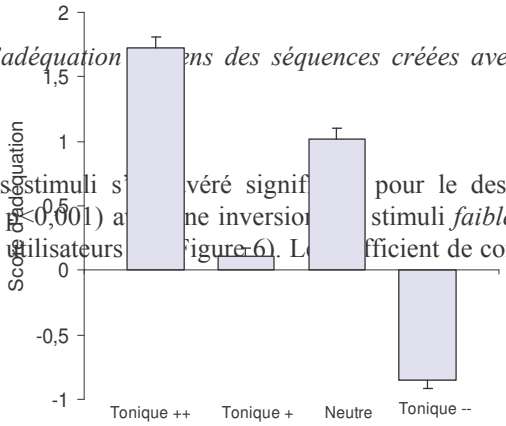


Figure 6. *Scores d'adéquation moyens des séquences créées avec le descripteur Tonique.*

Enfin, l'effet des stimuli s'est révélé significatif pour le descripteur *tonique* ($F(3/153) = 104,86$; $p < 0,001$) avec une inversion des stimuli *faiblement tonique* et *neutre* de la part des utilisateurs (Figure 6). Le coefficient de corrélation linéaire s'élève ici à +0,684.

6. Discussion

Dans cette section, nous proposons de discuter les résultats précédents afin de répondre à notre problématique de la pertinence de notre modèle pour la génération de comportements expressifs crédibles chez un Agent Conversationnel.

6.1. *Test 1*

En testant l'hypothèse (H1), nous avons cherché à déterminer si notre traduction des paramètres d'expressivité en spécifications d'animation était appropriée – si une modification d'un paramètre pouvait être reconnue et correctement interprétée par les utilisateurs.

Les résultats montrent que de nombreux participants ont perçu les modifications des paramètres d'expressivité et ont correctement attribué ces modifications aux paramètres de notre modèle. Les taux de reconnaissance les plus élevés ont été observés pour les paramètres d'*amplitude spatiale* et d'*amplitude temporelle*. Les variations de *fluidité* et d'*intensité* ont parfois été interprétées de façon inappropriée, mais les « bonnes réponses » ont tout de même obtenu les scores les plus élevés. Le

paramètre de *répétition* a parfois été confondu avec l'*intensité*. Enfin, le *niveau d'activation général*, qui détermine la quantité de mouvement, a été globalement mal reconnu. Lorsque les paramètres ont été correctement reconnus, la plupart des utilisateurs ont également reconnu le sens de modification. Ce résultat suggère que notre implémentation aboutit à des comportements cohérents. De manière générale, les résultats du test 1 semblent indiquer que nous avons implémenté de manière appropriée les dimensions d'*amplitude spatiale* et *temporelle*, mais que notre modèle nécessite d'être amélioré pour les autres paramètres.

Cette conclusion reflète les limitations du présent système : la *fluidité* n'est pas encore entièrement implémentée puisque seul l'enchaînement des gestes est modifié. Or, d'autres facteurs comme la courbure de la trajectoire du poignet ou l'accélération peuvent être considérés pour traduire la *fluidité*. L'*intensité* est un concept ambigu qui implique à la fois des notions d'énergie et de tension, qui peuvent parfois s'exprimer de façon contradictoire : un mouvement exécuté avec tension peut être caractérisé par une restriction de son amplitude, alors qu'un surplus d'énergie peut au contraire se traduire par une amplitude exagérée. Nous envisageons donc notamment de décomposer l'*intensité* en deux paramètres indépendants. Par ailleurs, le fait que la *répétition* ait parfois été interprétée comme de l'*intensité* révèle que ces deux concepts sont liés : une manière d'augmenter l'énergie d'un geste est de répéter son *stroke*. Alors que la séparation conceptuelle de l'*intensité* et de la *répétition* est pertinente du point de vue du programmeur, ces termes sont parfois confondus du point de vue des observateurs. Une solution pourrait alors consister à intégrer la *répétition* dans l'implémentation de notre futur paramètre d'énergie. Enfin, la mauvaise reconnaissance du *niveau d'activation général* peut être due au fait que ce paramètre intervient à un niveau de granularité différent : il ne modifie pas les gestes de manière continue mais au contraire détermine s'ils sont produits ou pas. Les utilisateurs ont peut-être été davantage attentifs aux modifications dans l'exécution des gestes plutôt qu'à la présence ou l'absence de gestes.

6.2. *Test 2*

Par l'hypothèse (H2), nous avons cherché à savoir si la combinaison de nos paramètres permettait de transmettre une intention communicative crédible.

Les participants du test 2 ont effectivement perçu la dimension de brusquerie comme nous l'avions espéré. Ils ont également perçu la tonicité, même si le comportement *faiblement tonique* a été inversé avec le comportement *neutre*. Il est en effet possible que la combinaison des paramètres dans le comportement *faiblement tonique* ait produit une animation peu réaliste et que les utilisateurs aient trouvé le comportement *neutre* plus naturel, lui accordant alors un score d'adéquation plus élevé. Pour le descripteur *mou*, nous n'avons pas réussi à faire varier l'expressivité en modifiant le réglage de nos paramètres. Ceci est peut-être dû

en partie aux types de gestes qui accompagnaient la réplique verbale de notre exemple : la parole était chargée négativement, ce qui se ressentait également dans les gestes de l'Agent. La séquence montrait notamment l'Agent fendre l'air horizontalement de ses mains, d'un geste net et tranchant. Or une personne molle aurait probablement choisi d'autres gestes, avec des formes plus vagues. Nos paramètres d'expressivité affectent seulement la manière dont les gestes sont exécutés, et n'agissent pas sur leur morphologie. Nous supposons que c'est cette contradiction qui a créé une ambiguïté dans la reconnaissance du descripteur *mou*. Ce résultat montre que, pour améliorer l'expressivité des gestes, il est important de considérer à la fois la sélection et la modification des gestes.

Dans le test 2, nous avons pris soin d'éviter les descripteurs liés à la personnalité ou aux émotions. La finalité de notre modèle est bien de simuler de tels traits et de tels états mentaux, mais le lien entre ces concepts de haut niveau et les dimensions d'expressivité n'est pas encore clair à l'heure actuelle. Ce constat n'est pas valable uniquement pour le domaine des Agents Conversationnels – la littérature en Psychologie Sociale est également lacunaire sur ce point.

7. Conclusion et perspectives

Dans cet article nous avons présenté l'évaluation, par un grand nombre d'utilisateurs non entraînés, d'un module de contrôle de l'expressivité des gestes, dont l'objectif est d'enrichir le comportement des Agents Conversationnels Animés. Les résultats nous encouragent à poursuivre cette approche, même si une partie seulement des paramètres d'expressivité et des comportements créés ont été reconnus. Cette étude ouvre plusieurs perspectives à notre travail.

Nous envisageons tout d'abord d'améliorer l'implémentation technique des paramètres afin d'obtenir une meilleure qualité d'animation et de meilleurs résultats sur la perception des modifications dans le réglage des paramètres. Nous souhaitons également étudier de manière plus approfondie l'interdépendance entre les paramètres. Les deux dimensions qui ont été le mieux reconnues sont aussi celles qui sont clairement indépendantes des autres : l'utilisation de l'espace et du temps. Les autres paramètres impliquent des modifications plus subtiles de ces deux dimensions fondamentales. Par ailleurs, nous projetons d'implémenter la possibilité d'intégrer la sélection et la modification des gestes. Enfin, notre protocole expérimental pourra être amélioré pour de prochaines expériences : dans les deux tests décrits dans cet article, une seule réplique verbale a été utilisée et déclinée avec différentes animations. Pour éviter de biaiser nos résultats par le choix et la séquence des gestes, il nous faudra utiliser un échantillon varié de situations et de répliques verbales.

8. Références bibliographiques

Bevacqua, E., Pelachaud, C., "Modelling an italian talking head", *Proceedings of Auditory-Visual Speech Processing AVSP'03*, 2003.

Black, A., Taylor, P., Caley, R., The Festival Speech Synthesis System. System documentation edition 1.4, 1999. http://www.cstr.ed.ac.uk/projects/festival/

Buisine, S., Abrilian, S., Martin, J.C., Evaluation of multimodal behaviour of embodied agents, In Z. Ruttkay & C. Pelachaud (Eds.), *From Brows to Trust: Evaluating Embodied Conversational Agents*, p. 217-238, Kluwer Academic Publishers, 2004.

Cassell, J., Sullivan, J., Prevost, S., Churchill, E., *Embodied Conversational Agents*. Cambridge, MIT Press, 2000.

Cassell, J., Vilhjálmsson, H., Bickmore, T., "BEAT: the Behavior Expression Animation Toolkit", *Proceedings of SIGGRAPH '01*, 2001, p. 477-486.

Chi, D., Costa, M., Zhao, L., Badler, N., "The EMOTE model for effort and shape", *Proceedings of SIGGRAPH'2000*, 2000, p. 173-182.

De Carolis, B., Pelachaud, C., Poggi, I., Steedman, M., APML, a mark-up language for believable behavior generation, In H. Prendinger & M. Ishizuka (Eds.), *Life-Like Characters*, Cognitive Technologies, Springer, 2004.

Fitts, P.M., "The information capacity of the human motor system in controlling the amplitude of movement", *Journal of Experimental Psychology*, 47, 1954, p. 381-391.

Gallaher, P., "Individual differences in nonverbal behavior: Dimensions of style", *Journal of Personality and Social Psychology*, 63, 1992, p. 133-145.

Harrigan, J., "Listener's body movements and speaking turns", *Communication Research*, 12, 1985, p. 233-250.

Hartmann, B., Mancini, M., Pelachaud, C., "Formational parameters and adaptive prototype instantiation for MPEG-4 compliant gesture synthesis", *Proceedings of Computer Animation'2002*, 2002.

Hartmann, B., Mancini, M., Pelachaud, C., "Implementing expressive gesture synthesis for Embodied Conversational Agents", *Proceedings of GestureWorkshop 2005*, 2005.

Kochanek, D.H.U., Bartels, R.H., "Interpolating splines with local tension, continuity, and bias control", *Proceedings of SIGGRAPH '84*, 1984, p. 33-41.

Kopp, S., Wachsmuth, I., "A knowledge-based approach for lifelike gesture animation", *Proceedings of ECAI'2000*, 2000.

Laban, R., Lawrence, F., *Effort: Economy in body movement*. Boston, Plays Inc, 1974.

Lebourque, T., Gibet, S., "High level specification and control of communication gestures: The gessyca system", *Proceedings of Computer Animation*, 1999, p. 24.

Lisetti, C.L., Brown, S., Alvarez, K., Marpaung, A., "A social informatics approach to human-robot interaction with an office service robot", *IEEE Transactions on Systems, Man, and Cybernetics*, 34, n°2, 2004, p. 195-209.

Loyall, A.B., Bates, J., "Personality-rich believable agents that use language", *Proceedings of the First International Conference on Autonomous Agents (Agents'97)*, 1997, p. 106-113.

Martell, C., Howard, P., Osborn, C., Britt, L., Myers, K. Form2 kinematic gesture corpus - Video recording and annotation, 2003.

Maya, V., Lamolle, M., Pelachaud, C., "Influences on embodied conversational agent's expressivity: Towards an individualization of the ECAs." *Proceedings of AISB'04*, 2004.

McNeill, D., *Hand and Mind*, University of Chigago Press, 1992.

Mehrabian, A., "Significance of posture and position in the communication of attitude and status relationships", *Psychological Bulletin*, 71, 1969, p. 359-372.

Noot, H., Ruttkay, Z., Gesture in style, In A. Camurri & G. Volpe (Eds.), *Gesture-Based Communication in Human-Computer Interaction - GW'03*, Vol. LNAI #2915, p. 324, Springer, 2004.

Pelachaud, C., Carofiglio, V., De Carolis, B., de Rosis, F., Poggi, I., "Embodied contextual agent in information delivering application", *Proceedings of AAMAS'2002*, 2002, p. 758-765.

Prendinger, H., Ishizuka, M., *Life-Like Characters*, Cognitive Technologies, Springer, 2004.

Ruttkay, Z., Pelachaud, C., *From brows to trust: Evaluating Embodied Conversational Agents*, Kluwer Academic Publishers, 2004.

Schröder, M. Speech and Emotion Research: An overview of research frameworks and a dimensional approach to emotional speech synthesis (PhD thesis), Institute of Phonetics, Saarland University, 2004.

Taubin, G. SNHC verification model 7.0. - MPEG-4 (Technical report), 1998.

Tepper, P., Kopp, S., Cassell, J., "Content in context: Generating language and iconic gesture without a gestionary", *Proceedings of AAMAS' 04 Workshop on Balanced Perception and Action in ECAs*, 2004.

Wallbott, H.G., "Bodily expression of emotion", *European Journal of Social Psychology*, 28, 1998, p. 879-896.

Wallbott, H.G., Scherer, K.R., "Cues and channels in emotion recognition", *Journal of Personality and Social Psychology*, 51, 1986, p. 690-699.