

## THESE

### Modèle de communication émotionnelle pour un agent conversationnel animé

#### Contexte de la thèse :

L'émotion exprimée par les visages, humain ou virtuel est l'une des pierres angulaires de l'**interaction**. Dans le domaine de la formation, et plus particulièrement du *serious game* (formation immersive et interactive), on peut plonger un individu dans un monde virtuel et lui apprendre à manifester les émotions adaptées à sa fonction. Ainsi, une personne souhaitant se perfectionner dans les techniques d'accueil peut utiliser un tel programme de simulation pour apprendre à communiquer avec un client pressé, agressif, mécontent etc. L'interaction se fait par un **système d'immersion 3D** constitué d'un réseau de caméras synchronisées, d'écrans de visualisation, de logiciels permettant de capturer le mouvement d'une personne dans un environnement restreint et de synthétiser en temps réel un clone de la personne immergée dans le monde virtuel.

Le **projet ANR IMMOMO** (Oct 2009 – Sept 2012) vise à concevoir un module pour **capturer et reconnaître les expressions du visage** de la personne immergée (l'apprenant) afin de faciliter son interaction avec un agent conversationnel autonome dont le comportement s'adaptera de façon semi-automatique à celui de l'apprenant. En effet, un formateur pourra décider du type de comportement que l'agent doit adopter (agressif, curieux, insistant, exigeant...) et du message vocal qu'il doit émettre. Comme il lui est impossible de manipuler en temps réel l'agent conversationnel dans toute sa complexité (message vocal émis, dynamique des expressions du visage, gestuelle du corps etc.), un **moteur de comportement multimodal** se chargera de co-manipuler l'agent sur la base des indications dispensées par le formateur et des expressions faciales de l'apprenant, reconnues de façon automatique par le système.

#### Description de la thèse :

Les modèles de communication non verbale pour les Agents Conversationnels Animés, ACAs, traitent généralement ou bien de la modélisation des expressions faciales et gestuelles par rapport aux actes communicatifs ou bien de la modélisation des expressions des émotions. Ces deux modèles sont souvent considérés comme étant indépendants l'un de l'autre. C'est-à-dire que la communication non verbale est l'ensemble (de la juxtaposition) des signaux pour l'émotion et pour les actes communicatifs non-verbaux. Ces modèles ne prennent pas en compte la modulation qu'un état émotionnel apporte à la communication. Ne pas prendre en compte une telle influence ne permet pas de modéliser un comportement cohérent. L'expression des émotions est plaquée aux autres expressions communicatives ; mais elles ne forment pas un tout.

Le sujet de thèse se situe dans ce domaine de recherche. Etant donné un état émotionnel, celui-ci colorie toute la communication, non seulement dans l'expressivité comportementale mais aussi dans la sélection des signaux liés aux actes communicatifs tels que l'emphase ou les ponctuateurs. L'approche envisagée vise d'un côté à prendre les résultats décrits dans la littérature, de l'autre à développer des modèles de communication émotionnelle à partir de modèle d'apprentissage de données comportementales.

Le travail se situe du côté de l'animation de l'agent à partir d'un certain nombre d'informations globales (état émotionnel, ...) et locales (emphase, performatif, etc). Il s'agit d'une part de produire la séquence de haut niveau de l'ensemble de signaux d'animation du mouvement de la tête et du visage à activer en fonction du contexte qui

soit cohérente et véhicule la bonne information. Mais également, à un niveau plus fin de l'animation lié au temps réel, les signaux produits étant des signaux temporels (par exemple des signaux continus pour l'animation de la bouche, des sourcils, de la tête), il faudra produire des séquences temporelles pour chacun des points significatifs du corps de l'agent en visant un rendu le plus naturel possible.

On distinguera deux étapes dans le travail à mener. Dans un premier temps on travaillera dans un cadre plutôt idéal, dans lequel on exploitera des données précises et complètes collectées avec des outils de capture de mouvements sur des personnes (acteurs) équipées. Ces données devront être complètement annotées temporellement par une information contextuelle riche incluant les émotions, l'emphase, les divers signaux communicatifs employés etc.

On cherchera à partir de ces données à apprendre dans un premier temps une « grammaire » de l'usage des signaux communicatifs en contexte. Partant des données de capture de mouvement 3D, les modèles d'apprentissage permettront d'extraire les relations entre les différents composants de l'animation comportementale. D'autre part on cherchera à modéliser finement les signaux temporels naturels correspondant à l'animation d'un personnage dans différents contextes émotionnels. Pour pouvoir les re-synthétiser à volonté.

Un modèle de génération de comportement sera développé. Celui-ci sera une extension d'un moteur déjà existant, Greta (<http://www.tsi.enst.fr/~pelachau/Greta/>)

Dans une seconde étape on s'intéressera au cadre réel plus complexe de l'animation d'un agent virtuel à partir d'informations contextuelles inférées automatiquement à partir de données vidéos (et non plus annotées manuellement). Ces données sont par nature moins précises et moins complètes. Elles seront fournies par des partenaires du projet IMMOMO. Le but sera ici d'interfacer les modèles et outils développés dans la première partie pour permettre leur fonctionnement dans l'environnement de prototypage imaginé dans le projet IMMOMO. Il faudra ici raffiner les modèles précédents pour tenir compte des spécificités des données réelles telles que l'absence de détection d'un marqueur et la prise en compte d'indices de confiance dans les détections.

La validation des modèles se fera par des tests perceptifs permettant d'évaluer la qualité de production de comportements émotionnels communicatifs de notre modèle. Ils permettront aussi de valider si les modèles capturent les éléments essentiels à la transmission d'un état émotionnel.

**Requis :**

Programmation en C/C++, expérience en système interactif et animation 3D fortement souhaitée

Connaissance et intérêt marqué pour l'apprentissage automatique statistique et la reconnaissance des formes.

**Localisation :** Telecom ParisTech + Université de Paris6

**Salaire :** autour de 1400 euros

**Contact:**

Catherine PELACHAUD

[catherine.pelachaud@telecom-paristech.fr](mailto:catherine.pelachaud@telecom-paristech.fr)

Thierry ARTIERES

[thierry.artieres@lip6.fr](mailto:thierry.artieres@lip6.fr)