
Un Modèle d'Interlocuteur Virtuel avec des Comportements d'Écoute Crédibles

Un modèle d'interlocuteur virtuel crédible

Etienne de Sevin* — Elisabetta Bevacqua** — Sylwia Hyniewska** — Catherine Pelachaud**

* Université Pierre et Marie Curie - LIP6
4 place Jussieu, 75005 Paris - FRANCE
etienne.de-sevin@lip6.fr

** CNRS - Telecom ParisTech
37/39 rue Dareau, 75014 Paris - FRANCE
{bevacqua, hyniewska, pelachaud}@telecom-paristech.fr

RÉSUMÉ. Nous présentons un modèle d'interlocuteur pour un agent virtuel. Il génère des comportements d'écoute (rétroactions) en fonction des comportements verbaux et non-verbaux de l'utilisateur. La personnalité de l'agent virtuel doit influencer le choix de ses comportements d'écoute. Nous proposons que les agents extravertis ont tendance à montrer plus de rétroactions que les introvertis et que la stabilité émotionnelle est liée à la tendance des agents à imiter l'utilisateur. Des études perceptives sur internet ont été réalisées pour évaluer notre modèle dans une situation d'interaction agent-utilisateur. Les participants ont jugé que la synchronisation, ainsi que le choix du type et de la fréquence des rétroactions, générées par notre modèle d'interlocuteur virtuel, étaient plus crédibles et permettaient une meilleure interaction avec les utilisateurs

ABSTRACT. We present a computational model that generates listening behaviour for a virtual agent. It triggers backchannel signals according to the user's visual and acoustic behaviour. The appropriateness of the backchannel algorithm in a user-agent situation of storytelling, has been evaluated by naïve participants, who judged the algorithm-ruled timing of backchannels more positively. The choice of the type and the frequency of the backchannels to be displayed is performed considering the agent's personality traits. We link agents with a higher level of extroversion to a higher tendency to perform more backchannels than introverted ones, while linking neuroticism to less mimicry production and more response and reactive signals sent. We run a perceptive study to test these relations in agent-user interactions, as evaluated by third parties.

MOTS-CLÉS : Agents conversationnels animés, comportements d'écoute, rétroactions, traits de personnalité, sélection de l'action.

KEYWORDS: Embodied Conversational Agents, listener's behaviour, backchannels, personality traits, action selection.

1. Introduction

Pendant les dernières vingtaines d'années, plusieurs recherches, dans le domaine des interfaces homme-machine, ont concentré leurs efforts sur le développement d'Agents Conversationnels Animés (ACAs), qui ont pour but de faciliter l'interaction homme-machine. Grâce à ces agents, les utilisateurs peuvent interagir avec les ordinateurs de la même façon qu'avec des humains en utilisant la parole, les expressions faciales, les gestes, etc... Le principal challenge, lors du développement d'agents conversationnels animés, est de les doter de capacités similaires à l'humain pour leur permettre de maintenir une interaction satisfaisante avec les utilisateurs. Ces agents virtuels doivent donc être crédibles non seulement dans leur aspect physique mais aussi dans leurs comportements. En effet, ils doivent être capables de montrer des comportements appropriés pendant qu'ils parlent et écoutent.

Dans cet article, nous nous focalisons sur les comportements d'écoute, et en particulier, sur les rétroactions que l'interlocuteur peut émettre en écoutant. Dans les communications humain-humain, les interlocuteurs envoient des indications au locuteur pour montrer leur participation, maintenir l'interaction et permettre la continuité du discours du locuteur. Dans les interactions ACA-utilisateur, l'absence de comportements d'écoute appropriés de l'agent détériore la qualité de l'interaction. Dans cet article, nous proposons un modèle d'interlocuteur virtuel qui génère automatiquement des rétroactions en fonction des comportements de l'utilisateur.

Le travail présenté dans cet article fait partie du projet européen SEMAINE¹. Son but est de développer un ACA autonome, ayant des traits de personnalité reconnaissables, qui est capable de montrer, en temps-réel, des comportements d'écoute appropriés dans une conversation avec un utilisateur. Quatre personnages, avec différentes personnalités, invitent l'utilisateur à converser. Le but de ces agents virtuels est de maintenir une interaction réaliste avec un utilisateur humain, malgré une capacité et une reconnaissance verbales limitées. Nous évaluons uniquement la partie non-verbale (interlocuteur) de l'architecture. Notre modèle d'interlocuteur virtuel a été intégré dans le projet SEMAINE (de Sevin *et al.*, 2010a).

Pour introduire la notion de personnalité dans notre modèle, nous avons implémenté un algorithme de sélection de l'action. Il choisit en temps-réel le type et la fréquence des rétroactions que l'agent virtuel doit montrer en fonction de sa personnalité. L'algorithme se base sur deux dimensions de la personnalité : l'extraversion et la stabilité émotionnelle.

La section suivante présente l'état de l'art des concepts se référant à ce travail : personnalité et comportements d'écoute. Dans la section 3, nous présentons l'architecture temps-réel de notre modèle d'interlocuteur virtuel. Les sections 4 et 5 décrivent en détails, respectivement, le module qui génère les comportements d'écoute et le module de sélection de l'action. Des études perceptives, réalisées pour évaluer notre modèle, sont présentées et discutées dans la section 6.

¹ <http://www.semaine-project.eu>

2. État de l'art

2.1. *Personnalité*

Les études montrent que les agents qui exhibent des traits de personnalité sont plus crédibles. En particulier, (Nass *et al.*, 1994) ont montré que les utilisateurs réagissent avec des agents dotés de personnalité, de la même manière qu'avec des humains ayant la même personnalité. De plus, ils sont capables d'identifier la personnalité de l'agent virtuel à partir de signaux verbaux et non-verbaux, et préfèrent interagir avec des agents qui montrent des comportements cohérents. Par exemple, lorsqu'un agent extraverti montre des comportements verbaux et non-verbaux typiques d'extravertis, les utilisateurs savent à quoi s'attendre et la cohérence de l'agent leur donne un sentiment de confiance (Isbister *et al.*, 2000).

Les modèles de traits de personnalité supposent que les traits influencent le comportement, et qu'ils sont des priorités fondamentales d'un individu. Nous basons notre travail sur la perception dimensionnelle de la personnalité (McRorie *et al.*, 2009). Actuellement, deux modèles sont très utilisés dans la communauté scientifique. Le modèle OCEAN (Costa *et al.*, 1992) est basé sur une approche lexicale. Il considère cinq dimensions de la personnalité : ouverture à l'expérience, amabilité, application, extraversion, stabilité émotionnelle. Le modèle d'Eysenck (Eysenck, 1991) est basé sur une approche clinique. Il considère seulement trois dimensions : extraversion, stabilité émotionnelle et psychotisme. Dans ce travail, nous nous focalisons sur les dimensions extraversion et stabilité émotionnelle. Ces deux dimensions sont centrales dans toutes les théories majeures de traits de personnalité. Elles vont nous permettre de formuler des prédictions en terme de comportements, comme la tendance à imiter ou la fréquence des mouvements.

Au niveau des différences individuelles, les individus empathiques imitent plus les postures, les expressions faciales des autres, que les non-empathiques (Sonnby-Borgstrom, 2002). Des chercheurs ont montré que les imitations facilitent en général les interactions et les rendent plus plaisantes, augmentant le sentiment d'empathie (Chartrand *et al.*, 2005). L'empathie est la capacité de partager ou d'interpréter correctement les émotions et les sentiments des autres (Decety *et al.*, 2004). D'après (Eysenck *et al.*, 1978), la stabilité émotionnelle est corrélée positivement à l'empathie, donc à la tendance à imiter. Des études ont aussi montré qu'un haut niveau d'extraversion est associé avec une fréquence élevée de gestes et une augmentation générale de la vitesse des mouvements (Borkeneau *et al.*, 1992).

2.2. *Comportements d'écoute*

Pour avoir des communications réussies, les interlocuteurs doivent donner des indications à propos du discours du locuteur et de la communication elle-même. L'interlocuteur doit montrer sa participation dans l'interaction pour la maintenir et permettre au locuteur de continuer. Lorsque des personnes écoutent quelqu'un, elles n'assimilent pas passivement tous les mots, mais assument un rôle actif dans

l'interaction, montrant avant tout qu'elles sont attentives à l'échange de communication. D'après les comportements de l'interlocuteur, le locuteur peut estimer comment celui-ci réagit et décider comment poursuivre l'interaction. Par exemple, le locuteur peut interrompre la conversation si l'interlocuteur n'est pas intéressé ou répéter si l'interlocuteur montre des signes d'incompréhension.

Plusieurs études ont été réalisées sur le comportement de l'interlocuteur pendant les conversations, dans le but de comprendre et de définir les dynamiques appliquées. Une des premières études sur les comportements expressifs des interlocuteurs en interaction a été présentée par (Yngve, 1970). Son travail portait principalement sur les signaux utilisés pour gérer la prise de parole par le locuteur et l'interlocuteur. Pour décrire ce type de signal, Yngve a introduit le terme « rétroactions » (backchannels). Celles-ci sont définies comme des signaux acoustiques ou visuels non-intrusifs, générés par l'interlocuteur pendant le discours du locuteur. D'après (Allwood *et al.*, 1992) et (Poggi, 2007), les rétroactions donnent de l'information sur les fonctions basiques de communication, comme la perception, l'attention, l'intérêt, la compréhension, l'attitude (comme la croyance, l'appréciation, etc...) et l'acceptation envers ce que dit le locuteur. Par exemple, l'interlocuteur peut montrer qu'il fait attention mais ne comprend pas, ou qu'il croit ce que dit le locuteur mais n'apprécie pas.

2.3. Agent interlocuteur virtuel

(Thorisson, 1996) a développé une tête pensante, appelée Gandalf, qui est capable d'interagir avec les utilisateurs en utilisant des signaux verbaux et non-verbaux. Gandalf a été conçu avec un visage et des mains. Son but est de donner des informations sur l'univers. Pendant l'interaction avec l'utilisateur, Gandalf est capable de montrer des expressions faciales, de regarder l'objet dont il parle, de gérer le tour de parole, et de générer des rétroactions en temps-réel. Quand les pauses de l'utilisateur sont supérieures à 110 ms, le système génère une rétroaction (une courte phrase ou un acquiescement de la tête).

(Cassel *et al.*, 2001) ont développé un agent immobilier (REA). Les tâches de cet humain virtuel consistent à montrer aux utilisateurs les caractéristiques des maisons à vendre, projetées derrière lui. Il interagit avec les utilisateurs par des gestes et la parole. Le synthétiseur vocal de Rea lui permet de faire varier l'intonation de sa voix. Comme pour Gandalf, les rétroactions de Rea sont générées à partir d'un modèle de temps de pause de l'utilisateur (supérieures à 500 ms). Les rétroactions sont de type paraverbal (« m mh »), hochements de tête ou courtes phrases.

(Ward *et al.*, 2000) ont mis en évidence que les rétroactions sont souvent générées quand le locuteur peut les percevoir. Ils ont proposé un modèle basé sur les signaux acoustiques afin de déterminer le bon moment pour montrer des rétroactions. Celles-ci sont générées quand le locuteur parle avec un ton faible pendant 110 ms après au moins 700 ms de discours, et qu'aucune rétroaction n'a été jouée dans les 800 ms précédentes.

L'agent interlocuteur (Maatman *et al.*, 2005), développé à ICT, produit des rétroactions en fonction du comportement de l'utilisateur. Plusieurs études ont montré que le comportement non-verbal de l'utilisateur peut être utile pour définir quand une rétroaction doit être déclenchée pendant une conversation. Dans certaines conditions, par exemple quand la conversation progresse doucement et de manière satisfaisante, les personnes ont tendance à se synchroniser comme dans une danse. Maatman s'est basé sur la littérature pour produire une liste de règles utiles pour prédire quand l'interlocuteur peut déclencher une rétroaction, en fonction des comportements verbaux et non-verbaux du locuteur. Il a conclu que les rétroactions coïncident avec les variations du ton de la voix du locuteur. Les froncements de sourcils, les mouvements du corps et les changements de regards sont utilisés par l'interlocuteur quand le locuteur montre de l'incertitude. Les comportements d'imitation sont souvent utilisés par l'interlocuteur pendant une interaction. Par exemple, il peut imiter les gestes, les changements de regards, les mouvements de tête et les expressions faciales du locuteur.

(Gratch *et al.*, 2007) ont développé un agent qui génère seulement des rétroactions non-verbales pendant l'écoute. Il a été implémenté pour étudier le niveau d'engagement ressenti par l'utilisateur lorsqu'il interagit avec un interlocuteur virtuel. Le système analyse les comportements non-verbaux et les caractéristiques de la voix de l'utilisateur pour décider quand une rétroaction doit être montrée (mouvements de la tête et du regard). (Morency *et al.*, 2008) ont proposé une amélioration de ce système en introduisant de l'apprentissage pour trouver les caractéristiques multimodales importantes du locuteur, pouvant affecter la génération des rétroactions de l'agent. Le système utilise des modèles probabilistes séquentiels pour apprendre à prédire et générer les rétroactions en temps-réel.

D'après (Kopp *et al.*, 2008) un agent conversationnel doit être capable de répondre d'une façon raisonnable et pertinente aux phrases et aux questions de l'utilisateur. Proposé par Kopp et implémenté sur l'humain virtuel Max, le modèle de rétroactions est basé sur un processus délibératif de raisonnement qui planifie comment et quand l'agent doit réagir, en fonction de ses intentions, ses croyances et ses désirs. Max est capable de montrer des rétroactions multimodales (principalement des mouvements de la tête avec une variation possible sur la qualité et la répétition). Elles sont générées seulement à partir d'une entrée écrite de l'utilisateur. Celle-ci est évaluée de manière incrémentale, en analysant les mots écrits par l'utilisateur pour augmenter continuellement ses connaissances à propos du sujet de la conversation. Pour déterminer quand générer des rétroactions, le système se base sur la ponctuation des phrases.

Tous les systèmes présentés jusqu'ici permettent à un agent virtuel d'exhiber des comportements d'écoute d'un interlocuteur lorsqu'il interagit avec un utilisateur. Cependant, ils ne prennent pas en compte les traits de personnalité des agents. Dans cet article, nous proposons une façon d'inclure la notion de personnalité dans un modèle d'interlocuteur virtuel pour augmenter sa crédibilité.

3. Architecture du modèle d'interlocuteur virtuel

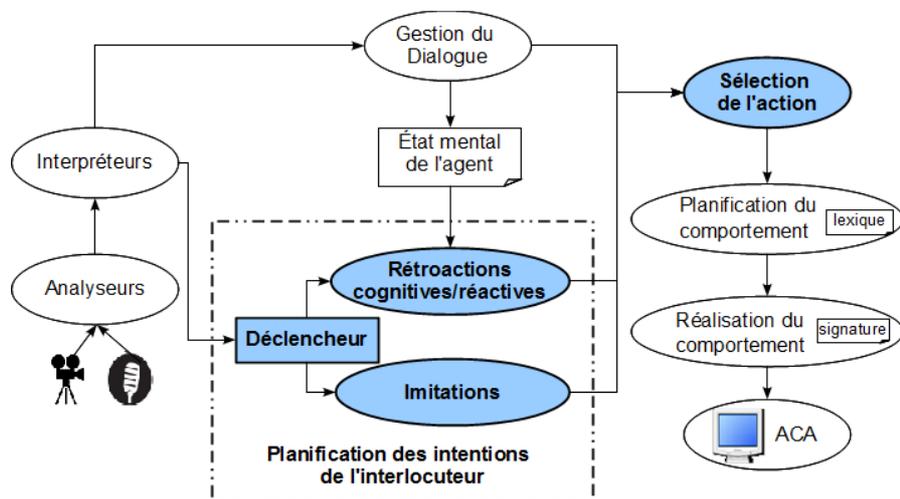


Figure 1. Architecture simplifiée du projet SEMAINE. Les modules en bleu sont impliqués dans notre modèle d'interlocuteur virtuel.

Notre système utilise l'interface de programmation du projet européen SEMAINE qui permet l'intégration multiplateforme et distribuée de modules pour concevoir des systèmes interactifs en temps-réel (Schröder, 2010) (voir figure 1). Les signaux visuels et acoustiques de l'utilisateur sont extraits par les modules d'analyse et ensuite interprétés pour en déduire des estimations sur les émotions, les expressions faciales, le niveau d'intérêt, etc... de l'utilisateur. Ces informations sont utilisées pour générer le comportement de l'agent virtuel lorsqu'il parle ou écoute. Le module de gestion du dialogue détermine quand l'agent doit prendre la parole et quelle phrase il doit dire, alors que le module de planification des intentions de l'interlocuteur génère les rétroactions quand l'agent écoute, en prenant en compte son état mental. Les informations à propos de ce que l'agent « pense » du discours du locuteur sont définies dans l'état mental de l'agent qui décrit si l'agent est d'accord ou pas, s'il croit ou pas, etc... ce que dit l'utilisateur. Les rétroactions et les phrases de l'agent sont filtrées en fonction de sa personnalité par le module de sélection de l'action. Celui-ci devra aussi choisir entre deux rétroactions sur la même modalité.

Ensuite, le module de planification du comportement calcule une liste de comportements appropriés pour chaque intention communicative que l'agent souhaite transmettre à l'utilisateur, à travers des rétroactions ou des phrases (de Sevin *et al.*, 2010c). L'ensemble des comportements possibles pour chaque intention communicative a été défini dans un lexique à partir de tests perceptifs (Heylen *et al.*, 2007 ; Bevacqua *et al.*, 2007), et d'analyse de vidéos d'interactions humain-humain de la base de données SEMAINE (McKeown *et al.*, 2010). Pour créer cette base de

données, quatre acteurs humains, jouant chacun une des quatre personnalités proposées dans le projet SEMAINE, ont été filmés en interaction avec une autre personne. A partir de l'observation de l'interlocuteur, un lexique pour chaque agent virtuel a été défini afin de prendre en compte leur personnalité et leurs moyens d'expression (plus de mouvements de mains ou de tête par exemple).

Une fois le comportement généré, il est calculé par le module de réalisation du comportement en fonction des caractéristiques comportementales de l'agent. Pour les décrire, nous avons adopté le concept de signature de l'agent virtuel proposé par (Mancini *et al.*, 2008). Cette signature est un ensemble de paramètres numériques représentant la préférence de l'agent pour ses modalités d'expression (mains, tête...) et pour l'expressivité de ses comportements (rapide, ample...). Nous déterminons une signature pour chaque agent en fonction de sa personnalité. Par exemple, Poppy qui est extravertie, a tendance à montrer beaucoup de gestes et d'expressions faciales (La France *et al.*, 2004). A chaque fois qu'un agent virtuel est choisi pour interagir avec l'utilisateur, son lexique et sa signature sont automatiquement chargés dans le système. De cette manière, les animations générées varient en fonction de l'agent sélectionné et sont envoyées au niveau du moteur graphique pour être jouées par l'agent conversationnel animé.

Dans cet article nous nous focalisons sur les modules de planification des intentions de l'interlocuteur et de sélection de l'action (en bleu sur la figure 1). Ils sont impliqués sans la génération de rétroactions lorsque l'agent a un rôle d'interlocuteur. Ces deux modules sont détaillés dans les deux prochaines sections.

4. Module de planification des intentions de l'interlocuteur

Le module de planification des intentions de l'interlocuteur calcule les comportements d'écoute de l'agent virtuel dans une conversation avec un utilisateur. Sa tâche consiste à décider quand une rétroaction doit être émise et quelle intention communicative l'agent doit transmettre à l'utilisateur à travers ce signal.

Pour déclencher une rétroaction, le module a besoin d'informations à propos du comportement de l'utilisateur. Des recherches ont montré qu'il y a une forte corrélation entre le déclenchement de rétroactions et les comportements verbaux et non-verbaux du locuteur (Maatman *et al.*, 2005 ; Ward *et al.*, 2000). Des modèles ont été élaborés pour prédire quand une rétroaction doit être déclenchée, basés sur une analyse statistique des comportements du locuteur (Morency *et al.*, 2008). Nous utilisons une approche similaire. Nous avons défini des règles probabilistes pour générer des rétroactions quand les systèmes d'analyse ou d'interprétation reconnaissent des comportements spécifiques du locuteur. Les probabilités ont été définies en se basant sur la littérature. Par exemple, une variation du ton de la voix de l'utilisateur peut déclencher une rétroaction avec une probabilité supérieure à 0.9854 (Bertrand *et al.*, 2006). Les règles ont été écrites dans un langage basé sur le XML et sont dans un fichier externe que l'on charge dynamiquement dans le système, au début de l'interaction.

En utilisant ce type de langage, les règles peuvent être facilement modifiées et étendues (voir figure 2). La définition d'une règle est un triplet :

REGLE = (nom, signaux, rétroactions) ;

avec nom (name): le nom unique de la règle, signaux (usersignals): la liste des signaux de l'utilisateur qui peuvent déclencher la règle, rétroactions (backchannels): contenant les types possibles de rétroactions qui peuvent être déclenchées avec la probabilité associée de déclenchement.

```
<rule name="trigger-nod">
  <usersignals>
    <usersignal id="s1" name="nod" modality="head"/>
  </usersignals>
  <backchannels probability="0.85">
    <mimicry>
      <mimicry_signal name="nod" modality="head"/>
    </mimicry>
    <response reactive/>
  </backchannels>
</rule>
```

Figure 2. Exemple de règle de déclenchement

Pour identifier les comportements de l'utilisateur qui provoquent des rétroactions de l'agent, les comportements verbaux et non-verbaux de l'utilisateur sont en permanence capturés grâce à une caméra vidéo et un microphone. Les applications audio et vidéo peuvent être connectées à notre système pour donner des informations sur les mouvements de tête, les expressions faciales, les signaux acoustiques comme les pauses et les variations de ton dans la voix de l'utilisateur (de Sevin *et al.*, 2010a). Toutes ces informations sont ensuite interprétées pour être décrites en termes de comportements grâce à des méthodes d'apprentissage sur des corpus. Par exemple, l'application qui analyse les expressions du visage est capable d'interpréter un mouvement des lèvres pour reconnaître un sourire, et l'application qui analyse le son est capable de reconnaître les émotions dans la voix de l'utilisateur. Dans le projet SEMAINE, le module de planification des intentions de l'interlocuteur est connecté avec les modules de capture et d'interprétation vidéo (Gunes *et al.*, 2010 ; Valstar *et al.*, 2010) et audio (Eyben *et al.*, 2010). Lorsqu'un comportement (contenant un ou plusieurs signaux) du locuteur est détecté, le module de planification des intentions de l'interlocuteur cherche parmi les règles si certaines correspondent aux signaux détectés. Si une telle règle est trouvée, une rétroaction est déclenchée avec la probabilité indiquée par la règle.

Nous prenons en compte trois types de comportements d'écoute : les rétroactions réactives et cognitives ainsi que les imitations. Notre agent peut émettre des

rétroactions réactives directement à partir de ses perceptions. Il peut aussi produire des rétroactions cognitives qui prennent en compte, en plus des perceptions, son propre état mental. Nous définissons l'état mental comme un ensemble d'intentions communicatives que l'agent souhaite transmettre pendant l'interaction. Pour chaque intention communicative, une valeur de l'importance que l'agent lui attribue est définie. Une telle valeur est un nombre entre 0 et 1. Nous considérons douze intentions communicatives choisies à partir des taxonomies proposées par (Allwood *et al.*, 1992) et (Poggi, 2003) : accepter, refuser, approuver, désapprouver, croire, douter, s'intéresser, ignorer, comprendre, mal comprendre. A chaque intention communicative est associé un ensemble de comportements prédéfinis, permettant de transmettre cette intention à l'utilisateur. La représentation de l'état mental de l'agent est stockée dans un fichier externe et écrit dans un langage basé sur le XML pour sa flexibilité et les possibilités d'extension. Ce fichier est chargé dynamiquement dans le système au début de l'interaction. Pour l'instant, il est prédéfini pour chaque agent virtuel en fonction de sa personnalité, mais on souhaite rendre l'état mental de l'agent dynamique dans un futur proche. Il peut donc répondre aux comportements ou au discours de l'utilisateur avec des comportements automatiques d'écoute, ou plus élaborés impliquant des intentions communicatives. Enfin notre agent peut copier certains comportements de l'utilisateur en générant des rétroactions d'imitation.

Le sous-module de rétroactions réactives et cognitives. Ce sous-module utilise les informations de l'état mental de l'agent pour générer des rétroactions appropriées. Quand une rétroaction est déclenchée, le sous-module cherche toutes les intentions communicatives dans l'état mental de l'agent qui ont une importance supérieure à zéro. Si au moins une intention communicative correspond à ce critère, le sous-module génère une rétroaction cognitive. Ensuite, celle-ci sera envoyée au module de planification des comportements qui calculera une liste de comportements appropriés pour l'agent pour chaque intention communicative, en fonction de leur importance. Si aucune information n'est donnée par l'état mental de l'agent, c'est à dire qu'aucune intention communicative n'a de valeur supérieure à zéro, le sous-module génère une rétroaction réactive. Celle-ci correspond à une réaction automatique au comportement de l'utilisateur. Ce type de rétroactions correspond souvent à des hochements de la tête et à des froncements de sourcils qui ont été étudiés dans la littérature (Allwood *et al.*, 2003 ; Cerrato *et al.*, 2003).

Le sous-module d'imitation. Ce sous-module génère des rétroactions sous forme d'imitations de comportements non-verbaux de l'utilisateur. Ce type de rétroactions a aussi deux niveaux : un niveau réactif (généré automatiquement à partir du comportement de l'utilisateur) et un niveau cognitif (imitation volontaire de l'utilisateur) comme par exemple pour paraître plus aimable. Quand une rétroaction est déclenchée, le sous-module d'imitation vérifie s'il possède une règle qui peut générer une imitation du comportement du locuteur à l'origine du déclenchement, comme un hochement de la tête ou un sourire pour augmenter l'engagement de l'agent dans l'interaction avec l'utilisateur (Lakin *et al.*, 2003).

5. Sélection des rétroactions

Le module de sélection de l'action (voir figure 1) reçoit tous les comportements générés par le module de planification des intentions de l'interlocuteur et le module de gestion du dialogue. Nous nous intéresserons à la partie non-verbale de la sélection de l'action : la sélection des rétroactions. Le rôle de ce module est de filtrer les rétroactions qui vont être jouées par l'agent virtuel, en fonction de conflits possibles entre rétroactions, du niveau d'intérêt de l'utilisateur (de Sevin *et al.*, 2009) ainsi que de la personnalité de l'agent (de Sevin *et al.*, 2010b). Nous allons nous focaliser sur l'influence des personnalités des agents sur la sélection de la fréquence et du type des rétroactions qui peuvent être compliqués à déterminer. En effet comment déterminer l'impact d'un agent extraverti ou pessimiste sur le choix des rétroactions ? et comment définir la valeur numérique de cet impact ?

Dans le projet SEMAINE, quatre agents virtuels interlocuteurs ont été définis avec chacun leurs propres traits de personnalité. Poppy est extravertie; Spike est agressif; Prudence est pragmatique; et Obadiah est pessimiste (voir figure 4). Nous avons choisi de représenter ces traits de personnalité sur deux dimensions : l'extraversion et la stabilité émotionnelle (voir figure 3) en utilisant la représentation en cercle validée par (Eysenck, 1991). En effet elles vont nous permettre de formuler des prédictions concrètes en termes de comportements, comme la tendance à imiter ou la fréquence des mouvements.

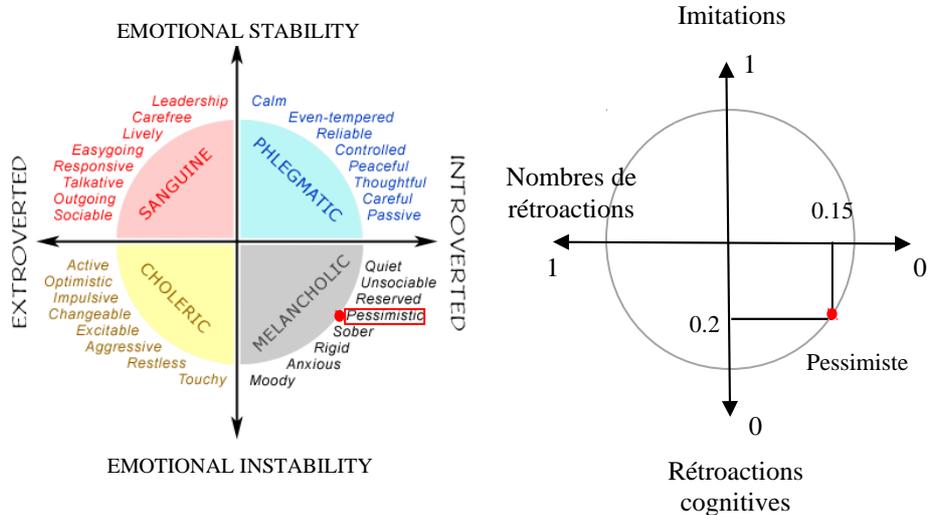


Figure 3. Représentation à deux dimensions d'Eysenck et notre représentation permettant d'en déduire les paramètres d'imitation et du nombre de rétroactions pour les traits de personnalité. Exemple de déduction pour Obadiah.

En nous basant sur la littérature, nous avons fait les deux hypothèses suivantes pour associer les deux dimensions (extraversion et stabilité émotionnelle) avec la fréquence et le type (tendance à imiter) des rétroactions :

Hp1 : la dimension d'extraversion est associée avec la fréquence des rétroactions (imitation et rétroactions réactives/cognitives). Poppy (extravertie) devra montrer plus souvent des rétroactions qu'Obadiah (pessimiste) (Smith *et al.*, 1975, Borkeneau *et al.*, 1992).

Hp2 : la dimension de stabilité émotionnelle est associée avec le type de rétroactions à privilégier par l'agent (correspondant à la tendance à imiter). Un agent très stable émotionnellement montre plus d'imitations (Chartrand *et al.*, 2005) alors qu'un agent très instable émotionnellement montre plus de rétroactions cognitives (Noor *et al.*, 2003). Prudence (pragmatique) devra montrer plus d'imitations que Spike (agressif).

A partir de ces deux hypothèses, nous avons construit une représentation en cercle équivalente à celle d'Eysenck en substituant l'axe de l'extraversion par le nombre de rétroactions et celui de la stabilité émotionnelle par la tendance à imiter (voir figure 3). Le nombre de rétroactions représente leur fréquence de haute (1), normale (0,5) à presque nulle (0). Concernant la tendance à imiter, plus la valeur est positive, plus la sélection de l'action va choisir les imitations, alors que, plus la valeur est négative, plus elle va choisir une rétroaction cognitive.

En localisant les personnalités de nos quatre agents virtuels sur la représentation en cercle validée par (Eysenck, 1991), on en déduit, en remplaçant le point au même endroit sur notre représentation (voir figure 3), les valeurs de la fréquence et le type (la tendance à imiter) des rétroactions pour la sélection de l'action de l'interlocuteur, à partir des traits de personnalité de chaque agent. Par exemple, Obadiah qui est pessimiste, montre très peu de rétroactions (fréquence = 0,15) et a tendance à faire plus de rétroactions cognitives que d'imitations (type = 0,2). Poppy, qui est optimiste, montre beaucoup de rétroactions (fréquence = 0,95) et a tendance à faire un peu plus d'imitations que de rétroactions cognitives (type = 0,65). Les valeurs obtenues dans le tableau 1 sont cohérentes avec la littérature (Smith *et al.*, 1975, Borkeneau *et al.*, 1992 ; Chartrand *et al.*, 2005 ; Noor *et al.*, 2003)

Tableau 1. Valeurs obtenues pour la tendance à imiter (type) et la fréquence des rétroactions pour les quatre personnalités de nos agents à partir de la figure 3.

	Obadiah (pessimiste)	Poppy (extravertie)	Prudence (pragmatique)	Spike (agressif)
Type des rétroactions (la tendance à imiter)	0.2	0.65	0.85	0.1
Fréquence des rétroactions	0.15	0.95	0.2	0.75

Pour faire son choix, l'algorithme de sélection de l'action calcule les priorités associées aux rétroactions, en fonction des conflits possibles entre rétroactions et de la personnalité de l'agent. Pour les rétroactions d'imitation, la tendance à imiter (type de rétroactions) va modifier ces priorités en les augmentant ou les diminuant. De ce fait, si les priorités des rétroactions d'imitation sont hautes, la sélection de l'action va les choisir, alors que si les priorités sont faibles, elle va choisir les rétroactions cognitives. La sélection de l'action choisit toujours les priorités les plus hautes. La difficulté réside dans le calcul des priorités des rétroactions (de Sevin *et al.*, 2009). Par exemple, la valeur du type des rétroactions de Poppy (extravertie) est de 0,65 signifiant que l'agent va montrer plus d'imitation que de rétroactions cognitives. A l'opposé, celle d'Obadiah (pessimiste) est de 0,2 signifiant que l'agent va montrer seulement 20% d'imitations.

La fréquence des rétroactions va augmenter ou diminuer la probabilité d'envoyer les rétroactions générées au module de planification du comportement (voir figure 1). Par exemple, la valeur de la fréquence des rétroactions de Poppy (extravertie) est de 0,95, ce qui signifie que 95% des rétroactions générées seront jouées par l'agent virtuel. A l'opposé, celle d'Obadiah (pessimiste) est de 0,15 signifiant que seulement 15% des rétroactions générées seront jouées par l'agent.

6. Évaluations de notre modèle d'interlocuteur

Pour évaluer notre modèle, nous avons réalisé deux études perceptives. La première nous permet d'évaluer le module de planification des intentions de l'interlocuteur alors que la deuxième porte sur le module de sélection de l'action.

6.1. Corpus audiovisuel



Figure 4. Modèles géométriques des quatre agents du projet SEMAINE : Poppy (extravertie), Spike (Agressif), Prudence (pragmatique) et Obadiah (pessimiste).

Chaque agent interlocuteur dans le projet SEMAINE a son propre modèle géométrique et ses propres comportements définis en se basant sur la littérature et sur une analyse du corpus vidéo SEMAINE (McRorie *et al.*, 2009). Dans ces études, nous nous focalisons sur les comportements des agents pour en comprendre leur impact sur la perception des utilisateurs. Nous avons donc utilisé le même modèle géométrique (troisième sur la figure 4) pour notre agent interlocuteur. Il correspond

au visage le plus neutre parmi les quatre afin d'éviter les biais graphiques et de donner des informations sur la personnalité de l'agent interlocuteur.

Pour créer les vidéos, nous avons demandé à une utilisatrice, ne connaissant pas le système, de raconter des histoires à notre agent interlocuteur virtuel (de Sevin *et al.*, 2010b) (voir figure 5). Celui-ci interagit avec l'utilisateur (sans prendre la parole) en montrant des comportements d'écoute générés par le module de planification des intentions de l'interlocuteur à partir des comportements non-verbaux de l'utilisatrice. Les comportements de l'agent sont ensuite filtrés par le module de sélection de l'action en fonction de la personnalité de l'agent.



Figure 5. Capture d'écran d'un clip vidéo utilisé pour les évaluations dans lequel l'agent montre des rétroactions en écoutant l'utilisateur (fenêtre en bas à droite)

En pratique, l'utilisatrice lit deux pages d'une bande dessinée pendant 5 à 10 minutes. Puis quand elle est prête, elle improvise une histoire, à partir des pages de la bande dessinée, pendant trois à six minutes en regardant l'agent interlocuteur. L'intérêt de l'improvisation est que l'utilisatrice regarde en permanence l'agent interlocuteur car elle n'a plus accès à la bande dessinée. Ce qui permet d'avoir une réelle interaction entre les deux. Nous avons vérifié que le sujet de l'histoire et l'intérêt de l'utilisatrice soient identiques afin de limiter les variations de comportements de l'utilisatrice entre les différentes improvisations.

L'utilisatrice interagit plusieurs fois avec les quatre personnalités possibles de l'agent interlocuteur. Pour pouvoir tester l'impact de la fréquence et du type des rétroactions filtrés par notre module de sélection des rétroactions, nous avons défini, pour chacune des quatre personnalités, trois conditions :

- C1 : fréquence des rétroactions filtrée par note module (type par défaut)
- C2 : type des rétroactions filtré par note module (fréquence par défaut)
- C3 : fréquence et type des rétroactions filtrés par note module

La valeur par défaut correspond à la valeur 0,5 sur notre représentation en cercle (voir figure 3) sinon elle correspond à la valeur de la position de la personnalité dans la représentation (Eysenck, 1991) transposée sur notre cercle.

Nous capturons l'image de l'agent interlocuteur et, dans une petite fenêtre en bas à droite, ce que fait l'utilisatrice en interaction avec l'agent (voir figure 5). Au total nous avons enregistré deux heures et demi de vidéos.

6.2. Évaluation du module de planification des intentions de l'interlocuteur

La tâche du module de planification des intentions de l'interlocuteur consiste à déclencher des rétroactions aux moments appropriés en fonction des comportements verbaux et non-verbaux de l'utilisateur (voir section 5). Dans notre évaluation nous avons voulu montrer que la synchronisation des rétroactions générées par notre module permet d'obtenir une meilleure interaction qu'avec des rétroactions désynchronisées. Les participants vont devoir donner leur impression générale sur la réussite de l'interaction, le comportement de l'agent interlocuteur et la synchronisation des rétroactions à partir de vidéos.

Tout d'abord, à partir du corpus audiovisuel, nous avons sélectionné les vidéos où la personnalité de l'agent est pragmatique et où l'agent ne montre que des signaux considérés comme positifs ou neutres (comme l'acquiescement de tête, le sourire...). Ce choix vient du fait qu'on ne voulait pas que la personnalité de l'agent ait une influence sur l'évaluation des participants. Dans cette étude, nous nous sommes intéressés uniquement à la synchronisation des rétroactions.

A partir de ces vidéos, nous avons extrait neuf clips vidéos d'une durée de 40 à 50 secondes. Nous avons ensuite réalisé neuf autres clips vidéos avec le même agent, mais en interaction avec un autre utilisateur. Celui-ci était plus jeune. Sa façon de parler et l'intonation de sa voix étaient très différentes, étant donné que sa langue maternelle était différente. A partir de ces nouveaux clips vidéos, nous avons remplacé la partie de la vidéo correspondant à l'utilisateur (en bas à droite) (voir figure 5) par celle correspondant à l'utilisatrice des clips originaux. Les rétroactions de l'agent dans ces nouveaux clips vidéos sont donc désynchronisées par rapport aux comportements de l'utilisatrice des clips originaux. Les participants vont pouvoir évaluer des clips vidéos où les rétroactions sont synchronisées avec les comportements de l'utilisateur et d'autres avec des rétroactions désynchronisées.

Les paramètres de fréquence et de type de rétroactions sont identiques pour tous les clips vidéos (en moyenne il y a 9 rétroactions par vidéo). Au final, nous avons 18 clips vidéos, 9 obtenus lors de l'interaction avec l'utilisatrice originale et 9 autres générés par l'autre utilisateur mais avec la vidéo originale de l'utilisatrice.

Les clips vidéos sont divisés en trois groupes de six. Chaque groupe contient trois clips avec des rétroactions générées par le module de planification des intentions de l'interlocuteur, et trois désynchronisés. Nous avons fait les hypothèses suivantes lorsque les rétroactions de l'agent sont déclenchées par notre module :

- Hp1 : l'interaction est jugée plus réussie
- Hp2 : le comportement de l'agent apparaît plus crédible
- Hp3 : l'agent montre moins des rétroactions aux moments inappropriés
- Hp4 : l'agent manque moins souvent des occasions de montrer des rétroactions aux moments appropriés.

6.2.1 Procédure et participants

Nous avons réalisé une évaluation sur internet pour atteindre un plus grand nombre de participants de milieux variés. La première page présente l'évaluation et les instructions. La seconde page demande aux participants de donner des informations démographiques. Ensuite six pages, contenant chacune un clip vidéo sélectionné aléatoirement, sont montrées successivement aux participants. Ils peuvent regarder la vidéo autant de fois qu'ils veulent avant de l'évaluer grâce à une échelle de Likert. Les quatre questions sont similaires à celles proposées par (Huang *et al.*, 2010) dans leur étude. Nous demandons aux participants de juger (1) si l'interaction est réussie (de *pas du tout* à *totalelement*), (2) si le comportement de l'agent interlocuteur est crédible (de *pas du tout* à *totalelement*), (3) si l'agent ne fait pas de rétroactions quand il le devrait (de *rarement* à *très souvent*), (4) si l'agent montre des rétroactions quand il ne le devrait pas (de *rarement* à *très souvent*).

128 participants (87 femmes, 41 hommes) avec une moyenne d'âge de 32.12 ans ont pris part à l'étude. Ils provenaient principalement de la France (75%), et avaient tous une bonne connaissance de la langue française.

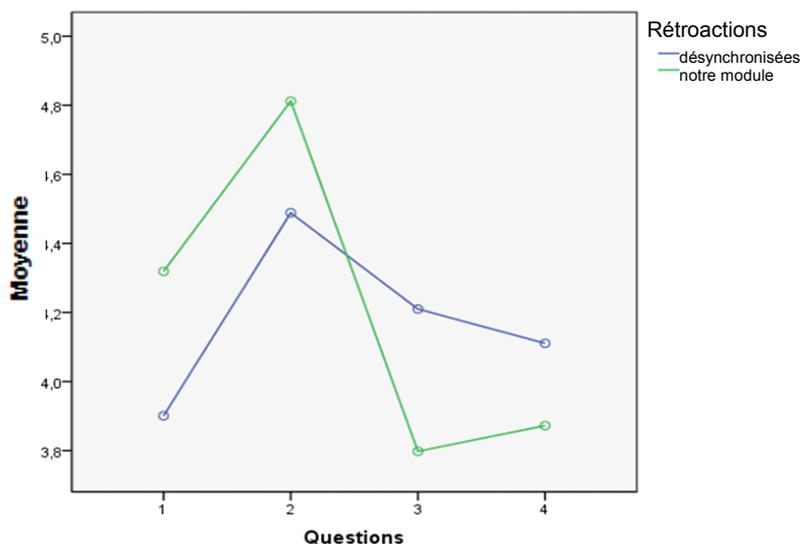


Figure 6. Moyennes des réponses (échelle de Likert) pour l'évaluation des rétroactions générées par notre module et désynchronisées en fonction des quatre questions décrites plus haut.

6.2.2 Résultats

Le test multivarié des différences entre les synchronisations des rétroactions de l'agent (désynchronisées versus par notre module) sur les réponses des quatre questions précédentes est significatif en utilisant le critère Lambda de Wilk, $F(4, 525)=2,61$, $p<.05$. Il y aussi un effet sur le clip vidéo présenté, $F(28,1894)=1.70$, $p<.05$. Par contre leur interaction n'est pas significative ($p>.05$). Chacune des transformations F-ratio du critère de Wilks est exacte.

Un test des effets entre les participants (ANOVA univariée) montre un effet du clip vidéo présenté pour la question 3 ($F(7, 5.8)=2.09$, $p<.05$) et la question 4 ($F(7, 5.61)=2.03$, $p<.05$) mais pas pour les questions 1 et 2. Il y a un effet de la synchronisation des rétroactions (désynchronisées versus par notre module) pour toutes les questions à l'exception de la quatrième ($p>.05$). Les moyennes pour les questions 1 et 2 sont supérieures, pour les rétroactions générées par le module, à celles désynchronisées, alors que c'est l'inverse pour les questions 3 et 4 (voir figure 6). L'interaction des rétroactions de l'agent et des clips vidéos a un effet seulement pour la question 2 ($F(7, 4.59)=2.05$, $p<.05$).

6.2.3 Discussion

Notre principale prédiction a été vérifiée. En effet les participants ont différencié dans leur jugement les rétroactions provenant de notre module de planification des intentions de l'interlocuteur par rapport à celles désynchronisées (c'est à dire à partir d'une autre histoire et d'un utilisateur différent de ceux de la vidéo originale). Cet effet est significatif pour trois des quatre questions. Ainsi les participants ont jugé l'interaction plus réussie et le comportement de l'agent plus crédible quand les rétroactions sont générées par notre module que celles désynchronisées. D'après les réponses des participants, l'agent montre moins fréquemment des rétroactions aux moments inappropriés quand nous utilisons notre module. Cependant nous n'avons pas obtenu de résultats significatifs pour la quatrième question. Nous ne pouvons donc pas affirmer qu'avec notre module, l'agent manque moins fréquemment des occasions de montrer des rétroactions aux moments appropriés.

Bien que nous n'ayons formulé aucune hypothèse concernant l'impact du comportement de l'utilisateur, nous en observons un car les différentes vidéos ont été jugées différemment par les participants. C'est particulièrement probant pour la question 3 et 4 qui portent sur le surnombre de rétroactions et sur le nombre d'occasions manquées de rétroactions. Cependant, il n'y a pas d'interaction entre le clip vidéo présenté et le type de rétroactions. Malgré le fait qu'il y ait un impact du comportement de l'utilisateur, nous observons un effet général positif de notre module.

Nos résultats montrent que la synchronisation des rétroactions de l'agent interlocuteur est importante, et le fait qu'elles soient synchronisées ou pas, a un impact sur l'évaluation par les personnes en interaction. Notre agent interlocuteur semble réagir aux moments appropriés à partir des signaux audio et vidéo de

l'utilisateur dans un contexte où il écoute une histoire de l'utilisateur. Nous pouvons en conclure que les résultats de notre étude perceptive confirment que notre module de planification des intentions de l'interlocuteur permet d'avoir de meilleures interactions humain-agent, qu'avec des rétroactions désynchronisées.

6.3. Évaluation du module de sélection des rétroactions

Le rôle du module de sélection de l'action consiste dans cette évaluation à choisir le type de rétroactions à privilégier, et à adapter leur fréquence en fonction de la personnalité des agents (voir section 5). Dans notre évaluation, nous voulons montrer que le filtrage de la fréquence et du type des rétroactions par notre module de sélection de l'action, en fonction de la personnalité de l'agent, permet une meilleure interaction qu'un filtrage par défaut. En effet si l'utilisateur arrive à identifier la personnalité de l'agent, il aura plus tendance à prendre l'agent pour un véritable interlocuteur.

Nous manipulons deux variables, déduites de nos hypothèses (voir figure 3) : le type et la fréquence des rétroactions en fonction des quatre personnalités possibles de l'agent : extraverti, pessimiste, agressif et pragmatique (définies dans le projet SEMAINE). A partir du corpus audiovisuel, nous avons sélectionné douze clips vidéos, de vingt secondes, représentatifs pour les quatre personnalités et leurs trois conditions (voir section 6.1). Tous les clips vidéos sont générés par notre module de sélection de l'action. Les participants doivent évaluer pour chaque clip vidéo si le nombre et le type de rétroactions sont appropriés, ainsi que leur impression générale de l'interaction par rapport à une personnalité de l'agent définie parmi les quatre possibles.

Nous avons formulé les hypothèses que le comportement de l'agent est plus approprié à sa personnalité lorsque notre module de sélection de l'action est actif :

- Hp1 : la fréquence des rétroactions filtrée par notre module doit être plus appropriée à la personnalité de l'agent qu'une fréquence par défaut
- Hp2 : le type de rétroactions à privilégier par notre module doit être plus approprié à la personnalité de l'agent qu'un choix par défaut

6.3.1 Procédure et participants

Nous avons réalisé une évaluation sur internet pour atteindre un grand nombre de participants de milieux variés. La première page présente l'évaluation et les instructions. La seconde page demande aux participants de donner des informations démographiques. Ensuite douze pages d'évaluation sont montrées aux participants. Chaque page d'évaluation est composée de la description de la personnalité de l'agent virtuel parmi les quatre possibles, d'un résumé des instructions et de deux clips vidéos (voir figure 7).

Pour chaque clip vidéo, les participants répondent sur une échelle de Likert si l'agent interlocuteur réagit de façon appropriée (compris entre *pas du tout* à *complètement*) et suffisamment (compris entre *pas assez* à *trop*) par rapport à la

personnalité décrite. Les participants doivent déterminer quelle vidéo (entre les deux présentées) correspond le plus à la personnalité de l'agent décrite dans l'énoncé.

Personnalité de l'agent: pragmatique

Une histoire est racontée à l'agent virtuel. Celui-ci doit se comporter de façon pragmatique. Veuillez regarder les deux vidéos et répondre aux questions par rapport à la personnalité de l'agent virtuel. Merci.



L'agent réagit-il de façon pragmatique ?

L'agent réagit-il suffisamment ?



L'agent réagit-il de façon pragmatique ?

L'agent réagit-il suffisamment ?

Quelle vidéo correspond le mieux à la personnalité pragmatique de l'agent ?

Figure 7. Exemple d'une page d'évaluation sur internet avec la description de la personnalité de l'agent, un bref résumé des instructions, deux clips vidéos et les questions d'évaluation correspondantes.

Une seule condition (voir section 6.1) sur les trois est évaluée par page, pour les quatre personnalités, correspondant à douze pages d'évaluation au total. Tous les clips vidéos sont générés par notre module de sélection de l'action. Le module de planification des intentions de l'interlocuteur est configuré avec les paramètres par défaut. Sur chaque page, il y a une vidéo dans laquelle les rétroactions de l'agent correspondent à la personnalité décrite sur la page, dans une des trois conditions. L'autre vidéo est choisie aléatoirement parmi celles correspondant à l'autre dimension de la personnalité dans la même condition. Par exemple, si la personnalité de l'agent décrite sur la page d'évaluation est l'extraversion, une des vidéos correspondra à Poppy (extravertie) dans une des trois conditions. L'autre vidéo sera choisie aléatoirement parmi les vidéos de l'autre dimension (stabilité émotionnelle), c'est à dire Spike (agressif) ou Prudence (pragmatique) dans la même condition. La position des deux vidéos dans la page d'évaluation est aléatoire pour empêcher les participants de reconnaître l'ordre des vidéos.

Les participants ne peuvent pas revenir à la page précédente et sont obligés de répondre aux questions de la page courante avant de pouvoir aller à la page suivante. Ils peuvent revoir autant de fois qu'ils le veulent les clips vidéos et n'ont aucune contrainte de temps sur les tâches à réaliser.

93 participants (57 femmes, 37 hommes) avec une moyenne d'âge de 34.56 ans ont pris part à l'étude. Ils provenaient principalement de la France (80%), et avaient tous une bonne connaissance de la langue française.

6.3.2 Résultats

La personnalité la plus reconnue est agressive (62%), puis extravertie (57%), pessimiste (53%), et pragmatique (52%).

Hypothèse Hp1 : La majorité des participants a répondu que la fréquence des rétroactions de l'agent est appropriée pour les quatre personnalités (voir figure 8). Ce n'était pas le choix par défaut (réponse choisie activement par les participants).

Une analyse de la variance a été réalisée (ANOVA), ainsi que des T tests pour échantillons appariés. Nous vérifions que la fréquence des rétroactions filtrées par notre module de sélection de l'action soit plus appropriée à la personnalité de l'agent, qu'une fréquence par défaut (Hp1). Notre prédiction est que la condition C3 (fréquence des rétroactions filtrée par notre module de sélection de l'action) doit être mieux évaluée par les participants que la condition C2 (fréquence des rétroactions par défaut) en particulier pour Poppy (extravertie) et Obadiah (pessimiste).

Les moyennes des réponses des participants à la question « l'agent réagit suffisamment » sont largement supérieures pour toutes les personnalités lorsque la fréquence des rétroactions est filtrée par notre module de sélection (condition C3) par rapport à celles par défaut (condition C2) sauf pour Obadiah (pessimiste) où la moyenne des réponses est légèrement supérieure (voir figure 9).

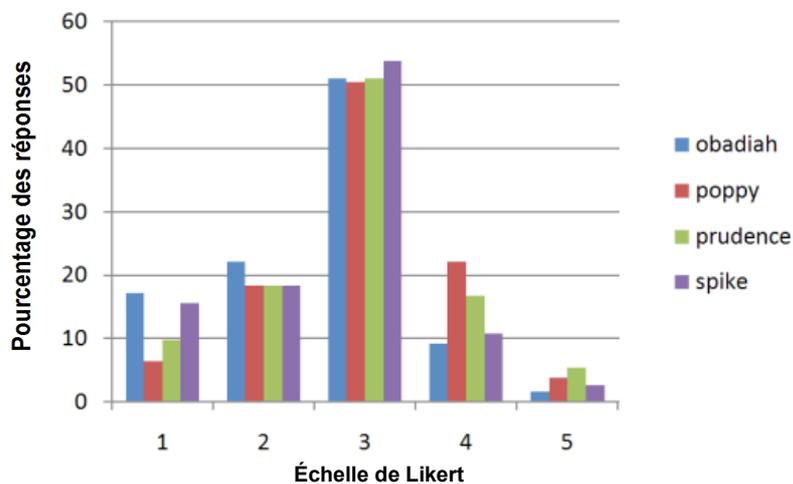


Figure 8. Pourcentage des réponses pour l'évaluation de la fréquence des rétroactions générées par notre module pour les quatre personnalités (de 1 : pas assez, 3 : approprié, 5 : trop)

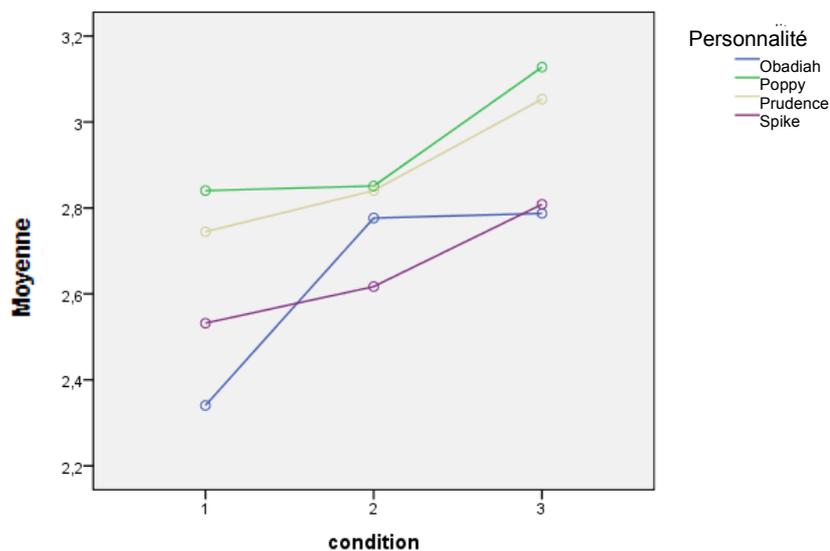


Figure 9. Moyennes des réponses (3 : approprié) pour l'évaluation de la fréquence des rétroactions pour les quatre personnalités en fonction des 3 conditions (C1 : fréquence des rétroactions filtrée par notre module (type par défaut), C2 : type des rétroactions filtré par notre module (fréquence par défaut), C3 : fréquence et type des rétroactions filtrés par notre module).

Tableau 2. Résultats du test ANOVA pour l'hypothèse Hp1

	n	F	p	
Personnalité	3	9.737	.000	*
Condition	3	18.032	.000	*
Personnalité*Condition	6	1.369	.225	

Tableau 3. Résultats des T tests pour échantillons appariés pour l'hypothèse Hp1

Fréquence des rétroactions	n		Test t	p	
Obadiah C2/C3	93	2.78/2.79	-,101	.460	
Poppy C2/C3	93	2.84/3.13	-2.457	.008	*
Prudence C2/C3	93	2.84/3.05	-1.818	.036	*
Spike C2/C3	93	2.62/2.81	-1.592	.057	

Un test ANOVA des réponses à la question « l'agent réagit suffisamment » montre un effet des clips vidéos sur la personnalité et sur la condition ($p < .05$) mais pas sur l'interaction entre les deux ($p > .05$) (voir tableau 2). Il y a un effet de la fréquence des rétroactions pour Poppy (extravertie) et Prudence (pragmatique) ($p < .05$) mais pas pour Obadiah (pessimiste) et Spike (agressif) ($p > .05$) (voir tableau 3).

Les participants considèrent que la condition C3 (fréquence et type filtrés) pour Poppy et Prudence est meilleure que la condition C2 (fréquence par défaut). De plus, les participants considèrent que la condition C3 est meilleure que la condition C1 (type par défaut) pour toutes les personnalités dans le cadre de l'hypothèse Hp1. En effet, nous avons observé qu'il y a un effet du type de rétroactions à privilégier par la sélection de l'action pour les quatre personnalités ($p < .05$).

Hypothèse Hp2 : Concernant le type de rétroactions à privilégier par la sélection de l'action, 66,5% des participants ont évalué supérieur ou égal le clip vidéo correspondant à la personnalité décrite dans l'énoncé, par rapport au clip vidéo aléatoire. Globalement, les participants ont jugé que l'agent virtuel réagit de façon : très appropriée pour Poppy (extravertie), moyennement appropriée Prudence (pragmatique), peu approprié pour Obadiah (pessimiste) et pas approprié pour Spike (agressif) (voir figure 10).

Les moyennes des réponses des participants à la question « l'agent réagit de façon appropriée » sont supérieures pour Obadiah (pessimiste) et pour Spike (agressif) lorsque le type des rétroactions est choisi par notre module de sélection de l'action (condition C3) que celles par défaut (condition C1) alors que celles pour Prudence (pragmatique) et Poppy (extravertie) sont inférieures (voir figure 11).

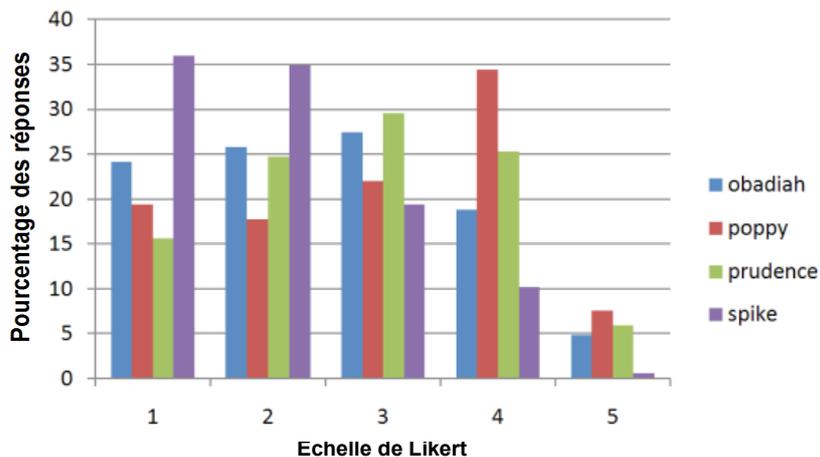


Figure 10. Pourcentage des réponses pour l'évaluation du type des rétroactions sélectionné par notre module de sélection de l'action pour les quatre personnalités (de 1 : pas du tout à 5 : complètement)

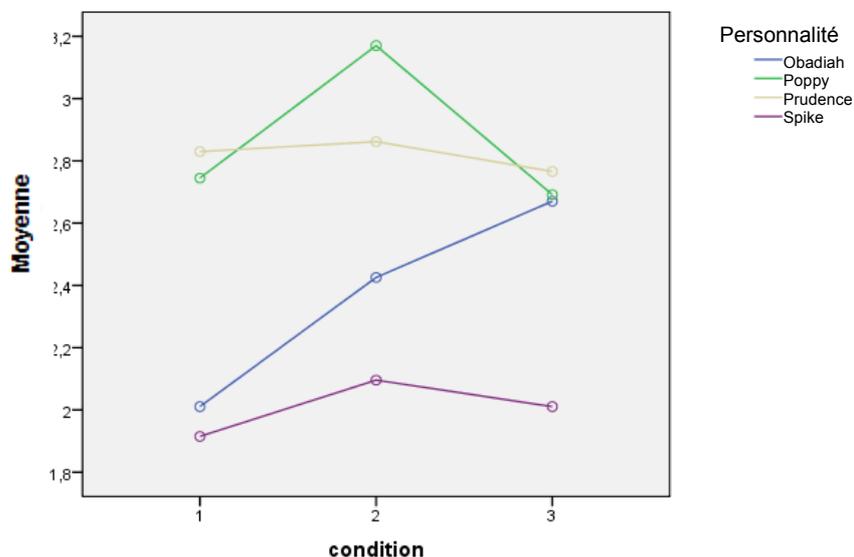


Figure 11. Moyennes des réponses (de 1 : pas du tout à 5 : complètement) pour l'évaluation du type des rétroactions pour les quatre personnalités en fonction des 3 conditions (C1 : fréquence des rétroactions filtrée par notre module (type par défaut), C2 : type des rétroactions filtré par notre module (fréquence par défaut), C3 : fréquence et type des rétroactions filtrés par notre module).

Tableau 4. Résultats du test ANOVA de Friedman pour l'hypothèse Hp2

n	Chi-Square	df	p	
94	141.948	11	.000	*

Tableau 5. Résultats des tests signés des rangs de Wilcoxon pour l'hypothèse Hp2

BC type	n	^s	p	
Obadiah C1/C3	94	2.01/2.67	.000	*
Poppy C1/C3	94	2.75/2.69	.433	
Prudence C1/C3	94	2.83/2.77	.253	
Spike C1/C3	94	1.91/2.01	.182	

Un test ANOVA de Friedman des réponses des participants à la question « l'agent réagit de façon appropriée » montre un effet des clips vidéos sur la personnalité de l'agent décrite dans l'énoncé ($p < .05$) (voir tableau 4). La sélection du type de rétroactions pour Obadiah (pessimiste) a un effet ($p < .05$) mais pas pour les autres personnalités ($p > .05$) (voir tableau 5). Les participants considèrent, seulement pour Obadiah, que la condition C3 est meilleure que la condition C1.

6.3.3 Discussion

Le but de cette évaluation est de vérifier si le filtrage du type et de la fréquence des rétroactions par le module de sélection de l'action a un impact sur la perception de la personnalité des agents par les participants. La première hypothèse Hp1 est partiellement vérifiée. Bien que les attributions soient plus hautes pour les clips vidéos avec la sélection de la fréquence des rétroactions que les alternatives, la différence n'a pas été significative pour certaines personnalités. La seconde hypothèse Hp2 n'a été vérifiée que pour Obadiah (pessimiste).

Concernant l'hypothèse Hp1, la majorité des participants a jugé que la fréquence des rétroactions est plus appropriée à la personnalité de l'agent lorsqu'elle est sélectionnée par notre module de sélection de l'action. Cependant cela n'a été significatif que pour Poppy (extravertie) et Prudence (pragmatique). Notre prédiction pour Poppy est juste mais pas pour Obadiah (pessimiste). En effet quand un agent montre beaucoup de rétroactions, il est associé à l'extraversion. Quand il fait moins de rétroactions que la normale, il n'est pas associé à une personnalité pessimiste. Il faudrait peut être accentuer l'effet pour que cela soit visible. Pour Spike (agressif), les participants s'attendaient à une fréquence importante des rétroactions alors que dans notre prédiction, elle était moyenne. Nous avons aussi vu que le type de rétroactions a un effet sur la perception de la fréquence des rétroactions par les participants. Ceux-ci évaluent mieux la fréquence des rétroactions quand le type de rétroactions est filtré aussi par notre module de sélection de l'action.

Concernant l'hypothèse Hp2, nos prédictions pour Prudence (pragmatique) et Spike (agressif) ne se sont pas révélées justes. Nous pensons que les adjectifs décrivant les personnalités étaient trop ambigus. D'ailleurs plusieurs commentaires de participants ont mentionné qu'ils n'avaient pas compris l'adjectif "pragmatique" et qu'ils ne savaient pas comment des personnes pragmatiques réagissent. Cette incompréhension des termes peut expliquer les mauvaises évaluations sur le type de rétroactions filtré par le module de sélection de l'action. Ces termes doivent être plus explicites dans les prochaines évaluations. De plus, il est parfois difficile de montrer une personnalité agressive ou pessimiste seulement avec des rétroactions. Un travail sur l'expression des rétroactions doit être fait. Le lien entre la stabilité émotionnelle et le type de rétroactions nécessite donc d'autres évaluations pour être validé.

Malgré le fait que nous avons utilisé un modèle neutre de visage de l'agent interlocuteur, pour éviter de donner aux participants des informations sur la personnalité de l'agent, ce visage peut être un biais pour exprimer certaines personnalités, comme celle agressive.

7. Conclusion

Dans cet article, nous avons présenté un modèle d'un interlocuteur virtuel crédible, prenant en compte la personnalité de l'agent. Le modèle est composé de deux modules. Le module de planification des intentions de l'interlocuteur déclenche des rétroactions en fonction des comportements verbaux et non-verbaux de l'utilisateur et génère trois types différents de comportements d'écoute : imitations, rétroactions réactives et cognitives. Le module de sélection des rétroactions filtre les rétroactions en fonction de la personnalité de l'agent. Pour évaluer notre système, nous avons conduit deux études perceptives. Dans la première étude, nous avons évalué si la synchronisation des rétroactions générées par le module de planification des intentions de l'interlocuteur permet d'avoir de meilleures interactions homme-agent. Les participants ont jugé l'interaction plus réussie et l'agent plus crédible quand la synchronisation des rétroactions est calculée par notre module. Dans la seconde étude, nous avons évalué si les comportements de l'agent sont interprétés comme plus appropriés et crédibles, lorsque la fréquence des rétroactions est associée à la dimension de l'extraversion et le type de rétroactions à la dimension de la stabilité émotionnelle. L'évaluation a montré que le choix de la fréquence des rétroactions, par notre module de sélection des rétroactions, contribue à une meilleure interprétation des comportements de l'agent par les utilisateurs en terme de traits de personnalité. Concernant le choix du type de rétroactions par notre module, d'autres évaluations sont nécessaires pour valider notre hypothèse.

Remerciements

Ce travail a été financé par le projet STREP SEMAINE IST-211486². Les auteurs tiennent à remercier l'utilisatrice des enregistrements vidéos pour son aide précieuse.

² <http://www.semaine-project.eu>

8. Bibliographie

- Allwood, J., Cerrato, L., « A study of gestural feedback expressions », P. Paggio, K. Jokinen, A. Jonsson (eds.), *First Nordic Symposium on Multimodal Communication*, pp. 7–22, Copenhagen, 2003.
- Allwood, J., Nivre, J., Ahlsen, E., « On the semantics and pragmatics of linguistic feedback », *Journal of Semantics*, pp. 1–26, 1992.
- Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., Rauzy, S.: Le cid - corpus of interactional data: protocoles, conventions, annotations. In: Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA), vol. 25, pp. 25–55, 2006.
- Bevacqua, E., Heylen, D., Tellier, M., Pelachaud, C., « Facial feedback signals for ECAs », *In: AISB'07 Annual convention, workshop Mindful Environments*, pp. 147–153, Newcastle upon Tyne, UK, 2007.
- Borkenau, P., Liebler, A., « Trait inferences: Sources of validity at zero acquaintance », *Journal of Personality and Social Psychology*, 62, 645–657, 1992.
- Cassell, J., Bickmore, T., « Embodiment in conversational interfaces: Rea », *In Human Factors in Computing Systems*, Pittsburgh, PA, 1999.
- Cassell, J., Nakano, Y., Bickmore, T., Sidner, C., Rich, C., « Non-verbal cues for discourse structure », *In: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pp. 114–123, Association for Computational Linguistics Morristown, NJ, USA, 2001.
- Cerrato, L., Skhiri, M., « Analysis and measurement of head movements signalling feedback in face-to-face human dialogues », P. Paggio, K. Jokinen, A. Jonsson (eds.), *First Nordic Symposium on Multimodal Communication*, pp. 43–52, Copenhagen, 2003.
- Chartrand, T., Bargh, J., « The Chameleon Effect: The Perception-Behavior Link and Social Interaction », *Personality and Social Psychology* 76, 893–910, 1999.
- Chartrand, T., Maddux, W., Lakin, J., « Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry », *The new unconscious*, pp. 334–361, 2005.
- Chartrand, T., Maddux, W., Lakin, J., « The new unconscious », *chap. Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of non-conscious mimicry*, pp. 334–361, New York: Oxford University Press, 2005.
- Costa P. T., J., McCrae, R.R., « Four ways five factors are basic », *Personality and Individual Differences*, 13, 653–665, 1992.
- Decety, J., Jackson, P., « The functional architecture of human empathy », *Behavioral and Cognitive Neuroscience Reviews*, 3, 71–100, 2004.
- de Sevin, E., Bevacqua, E., Pammi, S., Pelachaud, C., Schröder, M., Schuller, B., « Multimodal Behaviour Listener Driven by Audio Input », *Workshop : In Interacting with ECAs as Virtual Characters, In the Ninth International Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'2010*, Toronto, Canada, 2010a.

- de Sevin, E., Hyniewska, S., Pelachaud, C., « Influence of Personality Traits on Backchannel Selection », In *Proceedings of Intelligent Virtual Agents 2010, IVA'10*, Philadelphia, USA, 2010b.
- de Sevin, E., Niewiadomski, R., Bevacqua, E., Pez, A.M., Mancini, M., Pelachaud, C., « Greta, Une Plateforme d'Agent Conversationnel Expressif et Interactif », In *Journal: Technique et science informatiques*, 29(7), 751-776, 2010c.
- de Sevin, E., Pelachaud, C., « Real-time Backchannel Selection for ECAs according to User's Level of Interest », In *Proceedings of Intelligent Virtual Agents 2009, IVA'09*, Amsterdam, Holland, 2009.
- Eyben, F., Willmer, M., Schuller, B., « opensmile - the munich versatile and fast open-source audio feature extractor », In: *ACM Multimedia (MM)*, pp. 1459–1462, Florence, Italy, 2010.
- Eysenck, H.J., « Dimensions of personality: Criteria for a taxonomic paradigm » *Personality and Individual Differences*, 12, 773–779, 1991.
- Eysenck, S., Eysenck, H., « Impulsiveness and venturesomeness - their position in a dimensional system of personality description » *Psychol Rep*, 43, 1247–1255, 1978.
- Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R., « Creating rapport with virtual agents », *7th International Conference on Intelligent Virtual Agents*, Paris, France, 2007.
- Gunes, H., Pantic, M., « Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners » In: *International Conference on Intelligent Virtual Agents*, pp. 371–377, Philadelphia, USA, 2010.
- Heylen, D., Bevacqua, E., Tellier, M., Pelachaud, C., « Searching for prototypical facial feedback signals » In: *7th International Conference on Intelligent Virtual Agents*, pp. 147–153, Paris, France, 2007.
- Huang, L., Morency, L.P., Gratch, J., « Parasocial consensus sampling: Combining multiple perspectives to learn virtual human behavior » In: *The 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, Toronto, Canada, 2010.
- Isbister, K., Nass, C., « Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics » *Int. J. Hum-Comput. Stud*, 53(2), 251–267, 2000.
- Kopp, S., Allwood, J., Grammer, K., Ahlsen, E., Stocksmeier, T., « Modeling Embodied Feedback with Virtual Humans » *Lecture Notes in Computer Science*, 4930, 18, 2008.
- La France, B., Heisel, A., Beatty, M., « Is there empirical evidence for a nonverbal profile of extraversion? a meta-analysis and critique of the literature » *Communication Monographs*, 71, 28–48, 2004.
- Lakin, J.L., Jefferis, V.A., Cheng, C.M., Chartrand, T.L., « Chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry » *Nonverbal Behavior*, 27(3), 145–162, 2003.
- Maatman, R.M., Gratch, J., Marsella, S., « Natural behavior of a listening agent » In: *5th International Conference on Interactive Virtual Agents*, Kos, Greece, 2005.

- Mancini, M., Pelachaud, C., « Distinctiveness in multimodal behaviors » In: *Conference on Autonomous Agents and MultiAgent System*, 2008.
- McCroskey, J., Heisel, A., Richmond, V., « Eysencks big three and communication traits: three correlational studies » *Communication Monographs*, 68, 360–386, 2001.
- McKeown, G., Valstar, M., Cowie, R., Pantic, M., « The semaine corpus of emotionally coloured character interactions » In: *IEEE Int'l Conf. Multimedia & Expo*, pp. 1079–1084, Singapore, 2010.
- McRorie, M., Sneddon, I., de Sevin, E., Bevacqua, E., Pelachaud, C., « A model of personality and emotional traits » In: *Intelligent Virtual Agents 2009, IVA'09*, Amsterdam, Holland, 2009.
- Morency, L.P., de Kok, I., Gratch, J., « Predicting listener backchannels: A probabilistic multimodal approach », H. Prendinger, J.C. Lester, M. Ishizuka (eds.), *Proceedings of 8th International Conference on Intelligent Virtual Agents, Lecture Notes in Computer Science*, vol. 5208, Springer, Tokyo, Japan, 2008.
- Nass, C., Steuer, J., Tauber, E.R., « Computers are social actors » In: *CHI*, pp. 72–78, 1994.
- Nass, C.I., Moon, Y., Morkes, J., Kim, E.Y., Fogg, B.J., « Computers are social actors: a review of current research », pp. 137–162, 1997.
- Noor, F., Evans, D., « The effect of facial symmetry on perceptions of personality and attractiveness » *Journal of Research in Personality*, 37, 339–347, 2003.
- Poggi, I., « Mind markers », N. Trigo, M. Rector, I. Poggi (eds.), *Gestures. Meaning and use*, University Fernando Pessoa Press, Oporto, Portugal, 2003.
- Poggi, I., *Mind, hands, face and body. A goal and belief view of multimodal communication*. Weidler, Berlin, 2007.
- Schröder, M., « The semaine api: Towards a standards-based framework for building emotion-oriented systems » *Advances in Human-Computer Interaction*, 2010.
- Smith, B., Brown, B., Strong, W., Rencher, A., « Effects of speech rate on personality perception » *Language and Speech*, 18, 145–152, 1975.
- Sonnby-Borgstrom, M., « Automatic mimicry reactions as related to differences in emotional empathy » *Scandinavian Journal of Psychology*, 43, 433–443, 2002.
- Thorisson, K.R., *Communicative humanoids: A computational model of psychosocial dialogue skills*, Ph.D. thesis, MIT Media Laboratory, 1996.
- Valstar, M., Martinez, B., Binefa, X., Pantic, M., « Facial point detection using boosted regression and graph models », In: *IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 2729–2736, San Francisco, USA, 2010.
- Ward, N., Tsukahara, W., « Prosodic features which cue back-channel responses in english and japanese », *Journal of Pragmatics*, 23, 1177–1207, 2000.
- Yngve, V., « On getting a word in edgewise » In: *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pp. 567–577, 1970.

Etienne de Sevin est chercheur en informatique au LIP6, UPMC. Il a obtenu son doctorat en informatique au VRLab EPFL suisse en 2006. Il s'intéresse plus particulièrement à la prise de décision en temps réel des agents virtuels en fonction des facteurs internes (motivations, émotions, personnalités, raisonnement) et externes (la perception de l'environnement, des autres agents et des utilisateurs).

Elisabetta Bevacqua, Maître de Conférences à l'École Nationale d'Ingénieurs de Brest (ENIB). Elle a obtenu son Doctorat en Informatique à l'Université Paris VIII en 2009. Depuis 2001, elle travaille sur les Agents Conversationnels Animés et son domaine de recherche inclut la communication verbale et non-verbale, l'interaction homme-machine et l'implémentation de modèles pour la simulation du comportement humain pour des agents virtuels, en particulier lorsqu'ils jouent le rôle de l'interlocuteur."

Sylwia Hyniewska est doctorante à l'Université de Genève (Suisse) sous la direction de Susanne Kaiser et à Télécom ParisTech (France) avec Catherine Pelachaud. Elle a obtenu en 2008 un master en psychologie affective et clinique à l'Université de Genève. Depuis 2008 elle contribue à plusieurs projets européens (EU HUMAINE, CALLAS, SEMAINE, SSPNet). Ses travaux portent sur la perception des comportements affectifs et des différences individuelles chez les humains et agents virtuels humanoïdes .

Catherine Pelachaud est Directrice de Recherche CNRS au LTCI, Télécom ParisTech. Elle a travaillé à l'élaboration du premier système d'agent conversationnel animé, *Gesture Jack*, avec Justine Cassell, Norman Badler et Mark Steedman. Ses intérêts de recherche portent sur les agents conversationnels animés, la modélisation du comportement non verbal et émotionnel, les langages de représentation et les interfaces multimodales.