A Listening Agent Exhibiting Variable Behaviour

Elisabetta Bevacqua, Maurizio Mancini, and Catherine Pelachaud

University Paris 8, 140 rue de la Nouvelle France 93100, Montreuil, France
INRIA Rocquencourt, Mirages BP 105, 78153 Le Chesnay Cedex, France

Abstract. Within the Sensitive Artificial Listening Agent project, we propose a system that computes the behaviour of a listening agent. Such an agent must exhibit behaviour variations depending not only on its mental state towards the interaction (e.g., if it agrees or not with the speaker) but also on the agent's characteristics such as its emotional traits and its behaviour style. Our system computes the behaviour of the listening agent in real-time.

1 Introduction

A big challenge that must be faced in the design of virtual agents is the issue of credibility, not only in the agent's aspect but also in its behaviour. Users tend to react as if in a real human-human interaction when the virtual agent behaves in a natural human manner [NST94, RN96]. The work presented in this paper focuses on the listener's behaviour and is set within the Sensitive Artificial Listening Agent (SAL) project, which is part of the EU STREP SEMAINE project (http://www.semaine-project.eu). This project aims to build an autonomous talking head able to exhibit appropriate behaviour when it plays the role of the listener in a conversation with a user. Four characters, with different emotional styles, invite the user to chat trying to induce her/him in a particular emotional state. Within SAL, we aim to build a real-time Embodied Conversational Agents (ECAs) able to automatically generate those verbal and non verbal signals that a human interlocutor displays during an interaction. These signals, called *backchannels*, provide information about the listener's mental state towards the speaker's speech (e.g., if s/he believes or not what the speaker is saying). In our system backchannel signals are emitted not only according to the agent's mental state towards the interaction but also its *behaviour tendencies*, that is the particular way of producing non verbal signals that characterizes the agent. In our work the behaviour tendencies are defined by the preference the agent has in using each available communicative modality (head, gaze, face, gesture and torso) and a set of parameters that affect the qualities of the agent's behaviour (e.g. wide vs. narrow gestures). We call the behaviour tendencies the agent's *baseline*. The proposed work incorporates a pre-existing system for the generation of distinctive behaviour in ECAs [MP07, MP08]. The result is a system capable of computing the verbal and non-verbal behaviours that the agent, in the role of the listener, has to perform on the basis of both its baseline and its mental state.

H. Prendinger, J. Lester, and M. Ishizuka (Eds.): IVA 2008, LNAI 5208, pp. 262–269, 2008.

2 Sensitive Artificial Listener

The Sensitive Artificial Listener (SAL) technique [DCCC+08, WER+08], developed at the Queen's University of Belfast, rises from the need of collecting data about human interactions, where people express various emotions through both verbal and non verbal channels. The SAL idea comes from the observation of chat show hosts who are able to incite people into talking by simply appearing to listen and encouraging now and then with short standard phrases. In the previous SAL a human operator plays the role of the chat show host by selecting, in a Wizard of Oz manner, possible responses from pre-defined scripts. In another room, the user hears the corresponding pre-recorded emotionally coloured statements that not only encourage her/him into talking but also pull her/him towards specific emotional states. To achieve such a goal, SAL provides four characters with different emotional styles: Poppy (who is happy and positive), Obadiah (who is gloomy and sad), Spike (who is argumentative and angry) and Prudence (who pragmatic and sensitive). The user chooses the character s/he wants to talk to and can change it whenever s/he wants. In the other room, the operator chooses the statement to use according to both the selected character and the user's apparent emotional state. For example, if the user is sad, Obadiah approves while Poppy cheers up. Each character is provided with a prefixed script that contains phrases for each phase of a conversation: beginning, maintaining and ending of a conversation. The SAL approach proved successful at provoking sustained and emotionally coloured interactions. Within the SEMAINE project, the SAL system will undergo a substantial transformation: the four characters will be represented by four fully automatic ECAs. Each agent will be able to identify the user's emotional state and select the response; moreover it will provide appropriate verbal and non verbal signals while listening.

3 Background

3.1 Personal Tendencies in Behaviour

Argyle [Arg88] states that there are *personal tendencies* that are constantly present in human behaviour: for example people that look more tend to do so in most communicative situations, that is, there is a certain amount of consistency with the personal general tendencies. The reasons behind such personal differences can be due for example to differences in personality, emotional state, mood, sex, age, nationality [WS86]. Also Gallaher [Gal92] found consistencies in the way people behave: she conducted evaluation studies in which subjects' behaviour style was evaluated by friends, and by self-evaluation. Results demonstrated that for example people who are fast when writing have a tendency to be fast while eating; people producing wide gestures also walk with large steps.

3.2 Listener's Behaviour

To assure a successful communication, listeners must provide responses about both the content of the speaker's speech and the communication itself. Through

verbal and non verbal signals, called *backchannel signals*, a listener provides information about basic communicative functions, as perception, attention, interest, understanding, attitude (e.g., belief, liking and so on) and acceptance towards what the speaker is saying [ANA93, Pog05]. For instance, the interlocutor can show that s/he is paying attention but not understanding and, according to the listener's responses, the speaker can decide how to carry on the interaction: for example by re-formulating a sentence.

3.3 Related Work

Previous works on ECAs have provided first approaches to the implementation of a backchannel model. K. R. Thórisson developed a talking head, called Gandalf, able to produce real-time backchannel signals during a conversation with an user [Thó96], while Cassell et al. [CB99] developed the Real Estate Agent (REA) which is able to understand the user requests in real-time. The Listening Agent [MGM05], developed at ICT, produces backchannel signals based on real-time analysis of the speaker's non verbal behaviour (as head motion and body posture) and of acoustic features extracted from the speaker's voice [WT00]. Kopp et al. [KSG07] proposed a model for generating incremental backchannel. The system is based both on a probabilistic model, that defines a set of rules to determine the occurrence of a backchannel, and on a simulation model that perceives, understands and evaluates input through multi-layered processes. All the models described above do not take into account neither the agent's mental state nor its behaviour tendencies.

4 Overall System Description

As explained above, the system we present here is embedded in SAL. A user sits in front of a screen where an ECA, chosen among four different characters, listens to her/him and tries to induce her/him a particular emotional state. A video camera and a microphone record user's movements and voice. This information is used in our system to decide when and how the agent must provide a backchannel signal. The system includes also the concept of user *interest level* based on the Theory of Mind [Pet05]. Figure 1 shows the overall system diagram.

4.1 Agent Definition

Baseline. We define the agent's baseline as a set of numeric parameters that represents the agent's behaviour tendencies. In the baseline we represent two kinds of data: the agent's modality preference and the agent's behaviour expressivity. People can communicate by being more or less expressive in the different modalities: a person can gesture a lot while another one can produce many facial expressions and so on. In our static definition of an agent, we implemented the *modality preference* to represent the agent's degree of preference in using each available modality (face, head, gaze, gesture and torso). If, for example, we want

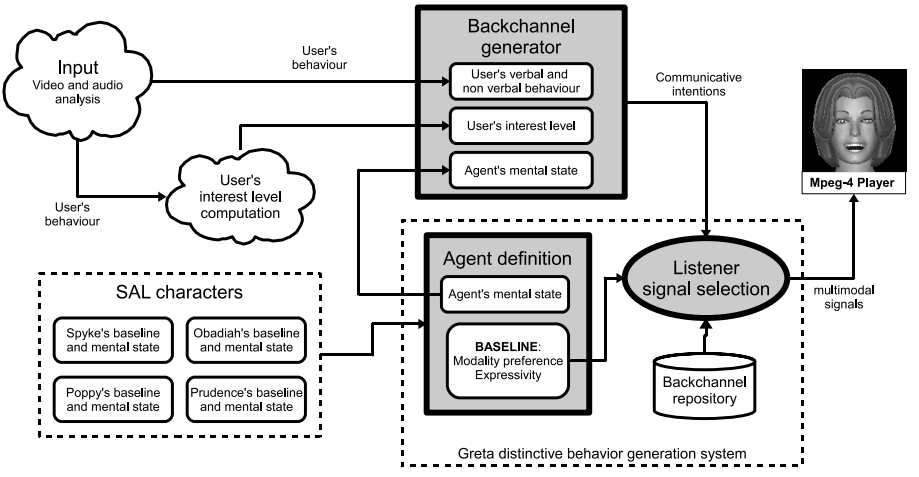

Fig. 1. System diagram

to specify that the agent has the tendency to mainly use hand gestures during communication, we assign a high degree of preference to the gesture modality, if it uses mainly the face, the face modality is set to a higher value, and so on.

To define the behaviour tendencies of an agent we also defined and implemented a set of parameters that allow us to alter the way the agent expresses its actual communicative intention [HMBP05]. The agent's *behaviour expressivity* is defined by a set of 6 parameters that influence the quality of the agent's movements: the frequency (OAC parameter), speed (TMP parameter), spatial volume (SPC parameter), energy (POW parameter), fluidity (FLD parameter), and repetitivity (REP parameter) of the nonverbal signals produced by the agent. These expressivity parameters are defined over each modality in a separate way: a set of parameters for the head movements, another set for the facial expressions, and so on.

To introduce the four SAL characters in our system we defined a baseline for each of them, keeping in mind the expressive behaviour studies showing the existing link between emotional states and behaviour quality [WS86, Gal92]. Thus, each SAL agent is characterized by specific emotional traits. The baselines are determined manually through the observation of videos of real people that exhibit a style of behaviour similar to Poppy, Obadiah, Spike and Prudence's characteristics. For instance, Spyke performs powerful, fast and short movements on all the modalities. Obadiah tends to produce backchannel signals mostly with the head while Poppy prefers to communicate mainly through facial expressions. Prudence performs slow movements mainly on the head and face modalities. When the user chooses a character, its baseline is loaded in the agent definition module and used by the listener signal selection module to computes the agent's distinctive behaviour as described in detail in the Section 4.3.

Agent's mental state. The agent definition module includes also the agent's mental state. In this module we consider solely how the agent reacts towards the interaction, that is how the agent reacts to the user's speech (if it agrees, refuses, understands... what is being said). Having such information the system specifies which *communicative intentions* (agree, refuse, understand,...) it will convey through its backchannel signals. We consider twelve communicative intentions related to backchannels chosen from the literature: agreement, disagreement, acceptance, refusal, belief, disbelief, interest, no interest, liking, disliking, understanding, no understanding [ANA93, Pog05].

The interlocutor's mental state is represented by a set of beliefs and intentions, such as its intentions towards the content of the speaker's speech, its own beliefs, and so on. So far, cognitive modules able to compute the listener's mental state are usually limited to specific domains. At present, such a module is under development within the SAL project [HtM08]. However for sake of simplicity, we link the agent's mental state to the emotional characteristics that differentiate the four SAL agents. Consequently, each SAL agent shows backchannel signals that are compatible with its emotional traits.

Spyke, who is angry and argumentative, conveys negative communicative intentions, in particular dislike, disagreement and not interest. Being gloomy, Obadiah tends to convey negative communicative intentions too, in particular disbelief, refusal and no understanding. Poppy, the happy one, provides backchannel signals that are the expression of positive communicative intentions, as liking, acceptance and interest. Finally, Prudence, who is sensitive and pragmatic, conveys positive communicative intentions, in particular agreement, belief and understanding [SW85, WS86].

4.2 Backchannel Generator

To display a believable listener behaviour, a virtual agent must be able to decide *when* a backchannel signal should be emitted and select *which* communicative intentions the agent should transmit through the signal. In our system these tasks are performed by the *backchannel generator* module of Figure 1. This module needs three data as input:

- The user's verbal and non verbal behaviour, tracked through a video camera and a microphone;
- The user's estimated interest level, an emotional state linked to the speaker's goal of obtaining new knowledge. Such a level is calculated evaluating the user's gaze, head and torso direction within a temporal window.
- The agent's mental state towards the interaction, as described in 4.1;

Researches have shown that backchannel signals are often emitted according to the verbal and non verbal behaviour performed by the speaker [WT00, MGM05]. On the basis of these results, our system evaluates video and audio data to select user's behaviours that could elicit a backchannel from the agent; for example, a head nod or a variation in the pitch of the user's voice and so on. The probability

that such a behaviour provokes a backchannel signal depends on the user's estimated level of interest. This value is used by the system to vary the backchannel emission frequency: when the interest level decreases the user might want to stop the conversation [SS73], consequently the agent provides less and less backannels. When a backchannel must be emitted the backchannel generator module uses the information about the user's mental state to decide which communicative intentions the agent should convey.

4.3 Listener Signal Selection

In this Section we describe the process of performing the selection of the nonverbal behaviours that the listener has to produce in order to convey its communicative intention. This task is performed by the listener signal selection module of Figure 1, which is an extended version of the corresponding one we implemented for the distinctive behaviour generation system of the Greta agent, presented in [MP07, MP08].

Behaviour sets. In the listener signal selection module, all the listener's communicative intentions, contained in the agent's mental state (see Section 4.1), are associated with the backchannel signals that can be produced by the listener. Each of these associations represents one *entry* of a lexicon, called *backchannel behaviour set*. A backchannel behaviour set is defined by the following parameters:

- The *name* of the corresponding communicative intention. For example *refuse*.
- The set S, containing the name of the signals produced on single modalities that can be used to convey the intention specified by the parameter *name*. For instance, the intention *refuse* can be conveyed by: shaking the head, saying "*no!*" and so on.
- The list of signals that are mandatory to communicate the intention corresponding to the behaviour set; for example to communicate *refuse*, the listener *MUST* shake its head.
- A set of logic rules like *if A then B* where A is a condition involving both the parameters of the agent definition (see Section 4) and the signals contained in the set S and B is a subset of S. For example we could specify a rule in which, if the value of the head Overall Activation parameter (OAC) is higher than a given threshold, the listener has to produce a head nod. In this example we are referencing to a value in the agent's baseline.

The backchannel behaviour sets have been defined in our previous works [BHTP07, HBTP07] and stored in the backchannel repository showed in Figure 1. We performed perceptive tests directed to analyse how users interpret context-free backchannel signals displayed by a virtual agent. From the results we found a many to many mapping between specific signals and most of the meanings proposed in the test.

Performing the listener signal selection. As shown in Figure 1, the listener signal selection process takes as input the agent definition (that is its baseline)

and the agent communicative intention and computes the multimodal behaviour that the agent has to perform. The process consists of some steps of computation: first the system looks for the behaviour set corresponding to the agent's communicative intention and computes all the possible combinations of the signals contained in the set S (see Section 4); it discards the combinations that do not contain the mandatory signals and checks the logic rules contained in the behaviour set, discarding the signal combinations that do not verify these rules; finally it prioritizes the signal combinations depending on the agent modality preference (see Section 4.1) and chooses the signal combination with the highest priority.

5 Conclusion and Future Work

We have proposed a system that computes the behaviour of a listening agent. This system is part of the SAL project that aims at implementing an agent exhibiting realistic behaviour when playing the role of a listener during a conversation. The user can chose among four agents with different styles of behaviour. Each agent provides backchannel signals that are consistent with its emotional traits. In the future, through perceptive tests, we will evaluate our system. We want to verify that we succeed in creating agents that show different behaviours and that these agents are able to sustain an emotionally coloured communication with users.

Acknowledgement

This work has been funded by the STREP SEMAINE project IST-211486 (http://www.semaine-project.eu) and the IP-CALLAS project IST-034800 (http://www.callas-newmedia.eu).

References

[ANA93] Allwood, J., Nivre, J., Ahlsén, E.: On the semantics and pragmatics of linguistic feedback. semantics 9(1) (1993)

[Arg88] Argyle, M.: Bodily Communication, 2nd edn. Methuen & Co., London (1988)

[BHTP07] Bevacqua, E., Heylen, D., Tellier, M., Pelachaud, C.: Facial feedback signals for ecas. In: AISB 2007 Annual convention, workshop "Mindful Environments", Newcastle, UK, April 2007, pp. 147–153 (2007)

[CB99] Cassell, J., Bickmore, T.: Embodiment in conversational interfaces: Rea. In: Conference on Human Factors in Computing Systems, Pittsburgh, PA (1999)

[DCCC+08] Douglas-Cowie, E., Cowie, R., Cox, C., Amir, N., Heylen, D.: The sensitive artificial listener: an induction technique for generating emotionally coloured conversation. In: LREC 2008 (May 2008)

[Gal92] Gallaher, P.E.: Individual differences in nonverbal behavior: Dimensions of style. Journal of Personality and Social Psychology 63(1), 133–145 (1992)

[HBTP07] Heylen, D., Bevacqua, E., Tellier, M., Pelachaud, C.: Searching for prototypical facial feedback signals. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 147–153. Springer, Heidelberg (2007)

[HMBP05] Hartmann, B., Mancini, M., Buisine, S., Pelachaud, C.: Design and evaluation of expressive gesture synthesis for embodied conversational agents. In: 3rd International Joint Conference on Autonomous Agents & Multi-Agent Systems, Utretch (2005)

[HtM08] Heylen, D.K.J., ter Maat, M.: A linguistic view on functional markup languages. In: AAMAS - First Functional Markup Language Workshop, Estoril, Portugal (May 2008)

[KSG07] Kopp, S., Stocksmeier, T., Gibbon, D.: Incremental multimodal feedback for conversational agents. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 139–146. Springer, Heidelberg (2007)

[MGM05] Maatman, R.M., Gratch, J., Marsella, S.: Natural behavior of a listening agent. In: 5th International Conference on Interactive Virtual Agents, Kos, Greece (2005)

[MP07] Mancini, M., Pelachaud, C.: Dynamic behavior qualifiers for conversational agents. In: Intelligent Virtual Agents, pp. 112–124 (2007)

[MP08] Mancini, M., Pelachaud, C.: Distinctiveness in multimodal behaviors. In: Conference on Autonomous Agents and Multiagent System (2008)

[NST94] Nass, C., Steuer, J., Tauber, E.R.: Computers are social actors. In: CHI, pp. 72–78 (1994)

[Pet05] Peters, C.: Direction of attention perception for conversation initiation in virtual environments. In: International Working Conference on Intelligent Virtual Agents, Kos, Greece, pp. 215–228 (September 2005)

[Pog05] Poggi, I.: Backchannel: from humans to embodied agents. In: AISB. University of Hertfordshire, Hatfield (2005)

[RN96] Reeves, B., Nass, C.: The media equation: How people treat computers, television and new media like real people and places. CSLI Publications, Stanford (1996)

[SS73] Schegloff, E.A., Sacks, H.: Opening up closings. Semiotica VIII(4) (1973)

[SW85] Scherer, K.R., Wallbott, H.G.: Analysis of nonverbal behavior. Handbook of discourse analysis 2, 199–230 (1985)

[Thó96] Thórisson, K.R.: Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills. PhD thesis, MIT Media Laboratory (1996)

[WER+08] Wöllmer, M., Eyben, F., Reiter, S., Schuller, B., Cox, C., Douglas-Cowie, E., Cowie, R.: Abandoning emotion classes - towards continuous emotion recognition with modelling of long-range dependencies. In: Interspeech (2008)

[WS86] Wallbott, H.G., Scherer, K.R.: Cues and channels in emotion recognition. Journal of Personality and Social Psychology 51(4), 690–699 (1986)

[WT00] Ward, N., Tsukahara, W.: Prosodic features which cue back-channel responses in english and japanese. Journal of Pragmatics 23, 1177–1207 (2000)