Noname manuscript No.
(will be inserted by the editor)

A listener model: introducing personality traits

Elisabetta Bevacqua · Etienne de Sevin · Sylwia Julia Hyniewska · Catherine Pelachaud

Received: date / Accepted: date

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 than a random timing. The system can generate different types of backchannels. The choice of the type and the frequency of the backchannels to be displayed is performed considering the agent's personality traits. The personality of the agent is defined in terms of two dimensions, extroversion and neuroticism. We link agents with a higher level of extroversion to a higher tendency to perform more backchannels than introverted ones, and we link neuroticism to less mimicry production and more response and reactive signals sent. We run a perception study to test these relations in agent-user interactions, as evaluated by third parties. We find that the selection of the frequency of backchannels performed by our algorithm contributes to the correct interpretation of the agent's behaviour in terms of personality traits.

Keywords Embodied Conversational Agents · listener's behaviour · backchannels · personality traits · action selection

Elisabetta Bevacqua
CERV - École Nationale d'Ingénieurs de Brest
Parvis Blaise Pascal
Technopôle Brest-Iroise
29280 Plouzané - FRANCE
Tél : +33-2-98058961
Fax : +33-2-98056610
E-mail: bevacqua@enib.fr

Etienne de Sevin
Université Pierre et Marie Curie - LIP6
4 place Jussieu, 75005 Paris - FRANCE
Tél : +33-1-44278743
E-mail: etienne.de-sevin@lip6.fr

Sylwia Julia Hyniewska and Catherine Pelachaud
CNRS - Telecom ParisTech
37/39 rue Dareau, 75014 Paris - FRANCE
Tél : +33-1-45817514
E-mail: hyniewska@telecom-paristech.fr
E-mail: pelachaud@telecom-paristech.fr

1 Introduction

In the past twenty years several researchers in the human-machine interface field have concentrated their efforts in the development of virtual humanoid entities. These agents, which are called Embodied Conversational Agents (ECAs), are a powerful HCI metaphor [35] and help the interaction between human and machine: users enjoy it more, feel more engaged, learn more, etc [34]. Through ECAs users can interact with computers in the same way they interact with their fellows, using channels like speech, facial expressions, gestures (and so on) which they are used to since their birth. To sustain natural and satisfying interactions with users, ECAs must be endowed with human-like capabilities [8]. They must be able to exhibit appropriate behaviour while speaking and while listening.

In this paper we focus on the listener's behaviour and in particular on the signals that an interlocutor can emit while listening. In human-human communications interlocutors provide responses to show their participation in the interaction, to push it forward and make the speaker go on [45, 2, 33]. Similarly, in user-ECA interactions, agents must not freeze while the user is speaking since the absence of the appropriate behaviour would deteriorate the quality of the interaction. We answer the

challenge of improving agent's animations by introducing a new listener model that computes the behaviour of an agent while listening to the user. Its novelty lies in the integration of several modalities (acoustic, hand and face movements) with an on-line computation of behaviour to be generated in accordance with the agent's personality traits.

The work presented in this paper is set within the Sensitive Artificial Listening Agent (SAL) project. It is part of the EU STREP SEMAINE project (http://www.semaine-project.eu). Within SAL, we aim to build an autonomous real-time ECA, endowed with recognisable personality traits, that is able to exhibit appropriate behaviour when it plays the role of the listener in a conversation with a user. Our listener model has been successfully embedded in the SAL system. To encompass the notion of personality, we introduced in our model a listener's action selection algorithm. Such an algorithm works in real-time to choose the type and the frequency of signals to be displayed by the ECA in accordance with its personality. The algorithm is based on the extroversion and neuroticism dimensions of personality.

The next section provides an overview of the background concepts we refer to in this work: personality and listener's behaviour. Section 3 is a brief description of related work. In Section 4 we present the real-time system architecture. Sections 5 and 6 describe in more details respectively the module that generates the listener's behaviour and the action selection algorithm. The perception studies that we performed to evaluate our system are presented and discussed in Section 7.

2 Background

2.1 Personality

Studies have shown that agents that exhibit personality traits are more believable. In particular, Nass et al. [30] showed that people react to agents endowed with personality characteristics in the same manner they would react to humans with similar personalities. Moreover people are able to identify a virtual agent's personality from verbal and non-verbal cues and they prefer to interact with agents that exhibit a consistent behaviour: for example, when an extroverted agent shows typical extroverted traits both in its verbal and non-verbal cues [23]. People know what to expect and the agent's consistency gives them a feeling of confidence. Several psychological models are currently proposed to define human personality. The Big Five [43], based on empirical findings, considers five personality dimensions: Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism. Another model, proposed by Wiggins et al. [44], defines traits based on Affiliation and Dominance, that determine a two-dimensional space where a circular structure can be defined.

Trait models of personality assume that traits influence behaviour, and that they are fundamental properties of an individual. We base our work on a dimensional perception of personality [28].

We focus on the extroversion-introversion and the neuroticism-emotional stability dimensions (as defined by [19,14]), which are central to major trait theories and for which we can formulate concrete predictions in terms of behaviour, such as mimicry or quantity of movement. On the individual differences level it has been shown that empathic individuals exhibit mimicry of postures, mannerisms, and facial expressions of others to a greater extent than not empathic individuals [11]. Similar results were confirmed by [39,38]. Researchers have shown that in general mimicry helps to make the interaction an easier and more pleasant experience improving the feeling of empathy [12]. Empathy is the capability to share or interpret correctly another being's emotions and feelings [15]. As according to Eysenck [20] neuroticism is negatively correlated with empathy, high neuroticism is negatively related to the level of mimicry behaviour. Eisenberg has also shown that characteristics associated with neuroticism have been linked to reduced levels of empathic-responding [16,17]. Researchers have also shown that high extroversion is associated with greater levels of gesturing, more frequent head nods, and a great speed of movement [7].

2.2 Listener behaviour

To assure a successful communication, listeners must provide responses about both the content of the speaker's speech and the communication itself. A listener has to show his/her participation in the interaction in order to push it forward and make the speaker go on. In fact, whenever people listen to somebody, they do not assimilate passively all the words, but they assume an active role in the interaction showing before all that they are attending the exchange of communication. According to the listener's behaviour, the speaker can estimate how his/her interlocutor is reacting and can decide how to carry on the interaction. One of the first studies about the expressive behaviours shown by people while interacting has been presented by Yngve [45]. His work focused mainly on those signals used to manage turn-taking, both by the speaker and the listener. To describe this type of signals, Yngve introduced

the term "backchannel". In this conception, backchannels are defined as *non-intrusive acoustic and visual signals provided by the listener during the speaker's turn.* According to Allwood et al. [2] and Poggi [33], acoustic and visual backchannel signals provide information about the basic communicative functions, as perception, attention, interest, understanding, attitude (e.g., belief, liking and so on) and acceptance towards what the speaker is saying. For instance, the interlocutor can show that he is paying attention but not understanding. A particular form of backchannel is the mimicry of the speaker's behavior. By mimicry we mean the behavior displayed by an individual who does what another person does [3]. We are interested in this type of backchannels since studies have shown that mimicry, when not exaggerated to the point of mocking, has several positive influences, making the interaction an easier and more pleasant experience and improving the feeling of engagement [11,9,12]. When fully engaged in an interaction, mimicry of behaviours between interactants may occur [25].

3 State of the art: Listener models for ECAs

First approaches to the implementation of a listener model considered pauses in the speaker's speech as a good timing to provide a backchannel signal. K. R. Thórisson [40] developed a talking head, named Gandalf, capable of interacting with users using acoustic and visual signals. To generate backchannels, the system evaluates the duration of the pauses in the speaker's speech. A backchannel (a short utterance or a head nod) is displayed when a pause, longer than 110 ms, is detected. Gandalf, provided with a face and a hand, has knowledge about the solar system and its interaction with users consists in providing information about the universe. Similarly, Cassell et al. [8] developed a listener model that provides a backchannel signal each time the user makes pause longer than 500 ms. The signal consists in paraverbals (e.g. "m mh"), head nods or a short statements such as "I see". This model has been implemented in the Real Estate Agent (REA). REA is a virtual humanoid whose task consists in showing users the characteristics of houses displayed behind her. Later on, evidences for the assumption that often backchannel signals are provided at pauses were provided by Ward and Tsukahara [42]. Their studies showed that backchannel signals are provided when the speaker talks with a low pitch lasting 110 ms after 700 ms of speech and provided that backchannel has not been displayed within the preceding 800 ms.

Maatman et al. [26] proposed a model that, to determine when a backchannel signal should be displayed, took into account not only acoustic information in the speaker's voice but also visual cues in the speaker's behaviour. From the literature they derived a list of useful rules to predict when a backchannel can occur according to the user's acoustic and visual behaviour. They concluded, for example, that backchannel signals (like head nods or short verbal responses that invite the speaker to go on) appear at a pitch variation in speaker's voice; listener's frowns, body movements and shifts of gaze are produced when the speaker shows uncertainty. Mimicry behaviour is often displayed by the listener during the interaction; for example, the listener mimics posture shifts, gaze shifts, head movements and facial expressions. This model was applied on the Listening Agent [26], developed at the Institute for Creative Technologies in California.

Morency et al. [29] introduced a machine learning method to find the speaker's multimodal features that are important and can affect timing of the agent backchannel. The system uses a sequential probabilistic model for learning to predict and generate real-time backchannel signals. The model is designed to work with two sequential probabilistic models: the Hidden Markov Model and the Conditional Random Field. Backchannels comprehend signals like head nods, head shakes, head rolls and gaze shifts.

Kopp et al. [24] were more interested in a listener model that generates backchannel signals in a pertinent and reasonable way to the statements and the questions asked by a user. Their model is based on reasoning and deliberative processing that plans how and when the agent must react according to its intentions, beliefs and desires. Backchannels are triggered solely according to the written input that the user types on a keyboard. The timing is determined applying the end-of-utterance detection, since listener's signals are often emitted on phrase boundaries. This model has been tested with Max, a virtual human developed at the A.I. Group at Bielefeld University. While interacting with a user, Max is able to display multimodal backchannels (like head nods, shakes, tilts and protrusions with various repetitions and a different quality of movement).

As most of the models presented so far, in this work we propose a listener model to generate backchannels according to the user's acoustic and visual behaviour, however we are particularly interested in the form of backchannel signals and the communicative functions they can transmit. We aim at implementing virtual agents that through their signals show not only that they are listening but also what they are "thinking" of the speaker's speech. Moreover, previous models do not take into account different agents with different person-

ality traits. In this work we propose a first approach to encompass the notion of personality in a listener model.

4 System Architecture

Our system uses the SEMAINE API, a distributed multi-platform component integration framework for real-time interactive systems [36]. The architecture of the whole system is shown in Figure 1. The modules in gray are part of the listener model presented in this work. User's acoustic and visual cues are extracted by analyser modules and then used by the interpreters to derive the system's *current best guess* regarding the state of the user and the dialogue. This information and the user's acoustic and visual cues are used to generate the agent's behaviour both while speaking and listening. The **Dialogue Manager** module determines when the agent should take the turn and which sentence it can utter, whereas the **Listener Intent Planner** module triggers signals while the agent is listening. Then, these signals, called backchannels [45], are filtered by the **Action Selection** module depending on the agent's personality. Then, the **Behaviour Planner** module computes a list of adequate behavioural signals for each communicative function the agent aims to transmit through a backchannel or a sentence. The mapping between a communicative function and the set of behaviours that conveys it is defined in a lexicon. We defined a lexicon for each SAL character partly through perception tests [5] and partly by analyzing videos of human interactions in the SEMAINE database [27]. Afterwards, the behavioural signals are realised by the **Behaviour Realizer** module according to the agent's behavioural characteristics. Finally, the agent's animation is rendered by a 3D character player.

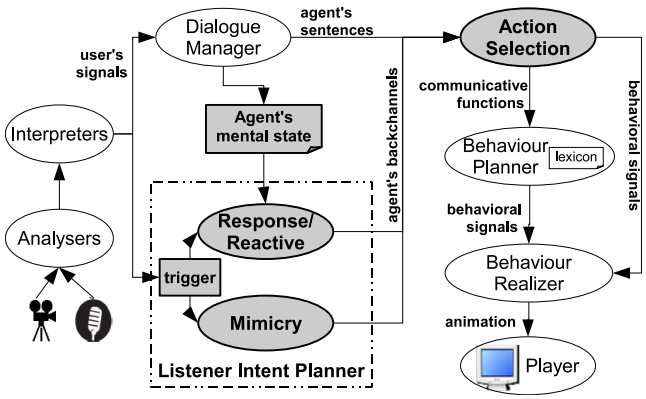

Fig. 1 Architecture of the whole system. The modules in gray are part of the listener model presented in this paper.

More information about the whole architecture and the flow of data between modules can be found in [37]. In this work we focus on the Listener Intent Planner and Action Selection modules that are involved in the generation of the backchannel signals while the agent is in the role of the listener. These two modules are detailed in the following two sections.

5 Listener Intent Planner

The Listener Intent Planner (LIP) module computes the agent's behaviour while being a listener conversing with a user. Its task consists in deciding *when* a backchannel signal should be emitted and in determining the types of backchannel the agent could perform. Then it will be up to the Action Selection module to decide *which* backchannel will be actually displayed.

To trigger a backchannel the LIP module needs information about the user's behaviour. Research has shown that there is a strong correlation between the triggering of some backchannel signals and the visual and acoustic behaviours performed by the speaker [26, 42]. Models have been elaborated to predict when a backchannel signal could be triggered based on a statistical analysis of the speaker's behaviours [26,29,42]. We use a similar approach and we have fixed some probabilistic rules based on the literature to prompt a backchannel signal when certain speaker's behaviours are recognized; for example, a raising pitch elicits both vocal and gestural backchannels with a probability higher than 0.9854 [4].

To identify those behaviours of the user that could elicit a backchannel from the agent, the user's acoustic and visual behaviours are continuously tracked through a video camera and a microphone. Audio and visual applications can be connected to our system to provide information about head movements, facial actions, acoustic cues like pauses and pitch variation of the user's voice. In the SEMAINE project the Listener Intent Planner has been connected with video analysis applications [21,41] and with audio analysis applications [18]. The triggering rules have been defined through an XML-based language and are written in an external file uploaded at the beginning of the interaction. So far we have defined rules for head movements (like nods, shakes and tilts), facial actions (like smile, raising eyebrows and frown) and acoustic cues (raising/lowering pitch, silence); however, by using an XML-based language, the set of rules can be easily modified or extended. To take into account the user's signals analysed by new applications, we can add new rules in the external file without modifying the source code. Moreover, we can easily modify the probability associated to those

user's behaviours that can trigger a backchannel signal. The definition of a rule is a triplet:

$$RULE = (name; usersignals; backchannels);$$

in which:

- *name* is the unique name of the rule.
- *usersignals* is the list of the user's signals that must be detected to trigger the rule.
- *backchannels* contains the possible types of backchannels that can be triggered with a certain **probability** when the rule is applied.

The example in Figure 2 shows the rule triggered when a user's head nod is detected.

```
<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>
```

Fig. 2 Example of triggering rule.

Another reason for associating probabilities to the rules is that it allows us to define agents that react differently to a user during an interaction. For example, we can define agents that have high probability to provide a lot of backchannels and that respond especially to the user's acoustic signals. Probabilities could vary according to agent's personality, mood and even culture.

When a user's behavior satisfies one of the rules a backchannel is triggered. The LIP modules can generate three types of backchannels: reactive signals, response signals and mimicry. Our agent can emit reactive backchannels that are signals derived from perception processing: the agent reacts to the speaker's behaviour or speech, generating automatic behaviour. Moreover, our agent can provide response backchannels that are signals generated by cognitive processing: the agent responds to the speaker's behaviour or speech performing a more aware behaviour. These backchannels are a type of attitudinal signals that the agent shows to provide information about what it "thinks" about the user's speech. Previous listener models have mainly considered reactive backchannels, whereas in this work we aim at creating a virtual listener able to transmit its communicative functions through backchannel signals. Response signals are used to show, for example, that the agent agrees or disagrees with the user, or that it believes but at the same time refuses the speaker's message. Another type of signals that our system can generate as backchannel is the mimicry of user's non verbal behaviours. As described previously in this paper, studies have shown that mimicry, when not exaggerated to the point of mocking, has several positive influences on interactions; for such a reason we are interested in this type of behavior.

Response/Reactive sub-module. The Response/Reactive sub-module generates both response and reactive backchannel signals. In order to generate these types of backchannels, information about what the agent "thinks" of the speaker's speech is needed. This information is provided in the *agent's mental state* that describes whether or not the agent agrees or believes and so on. We define the mental state as a set of communicative functions that the agent wishes to transmit during an interaction. We consider twelve communicative functions, a subset chosen from the taxonomies proposed by Allwood et al. [2] and by Poggi [33]: agree, accept, interest, like, believe and their opposites. For each communicative function the value of the importance the agent attributes to it is defined. Such a value is a number between 0 and 1, where 0 represents the minimum importance whereas 1 indicates that the agent gives to the corresponding communicative function the maximum importance.

In this work we provide a representation of the agent's mental state although we do not supply a system that computes it, however we implemented our listener model to be easily connected to this type of systems in order to update the value of the agent's mental state according to the evolution of the interaction. For example we connected our listener module to a *cognitive system* implemented within the SEMAINE Project. When a backchannel is triggered, the Response/Reactive sub-module generates a response backchannel that contains all the communicative functions in the agent's mental state that have a value of importance higher than zero.

It will be up to the Behaviour Planner to select the adequate behaviours to display for each communicative function [6]. The selection is done according to the importance associated to each communicative function and the mapping between the given function and a set of behavioural signals that convey it. Such a mapping has been defined through perception tests that we performed in previous studies [5]: for example, the communicative function "accept" can be mapped in a combination of head nod, smile, raise eyebrows and several paraverbals like *a-ah*, *yeah*, *right* and so on. If

no communicative functions have an importance value higher than zero, this module generates a reactive backchannel: an automatic reaction to the user's behaviour that simply shows contact and perception. This type of backchannel is translated in those typical continuer signals, like head nods and raise eyebrows, that have been studied in the literature [1,10,32]. The agent's mental state could be undefined, for example, when the agent does not want to show any particular attitudinal signal or when no cognitive system is connected to our system and, as a consequence, no information about the agent's reaction towards the interaction can be provided.

Mimicry sub-module. This sub-module generates the mimicry of the detected user's non-verbal behaviours as backchannel signals. This type of backchannel can be seen as a subset of the reactive and response backchannels: while listening the interlocutor can display signals of mimicry both at perception level, as a reaction of the user's behaviour, and at cognitive level, consciously deciding to imitate the speaker (for example to appear more likeable [3]). However, because of its particular form (that is, the copy of some user's visual behaviours) we decide to compute it in a different submodule. When a backchannel is triggered by a user's visual cue (such as a head nod or a smile and so on), the Mimicry sub-module generates a signal that consists in the mimic of the same visual behaviour. No acoustic mimicry is considered in this model.

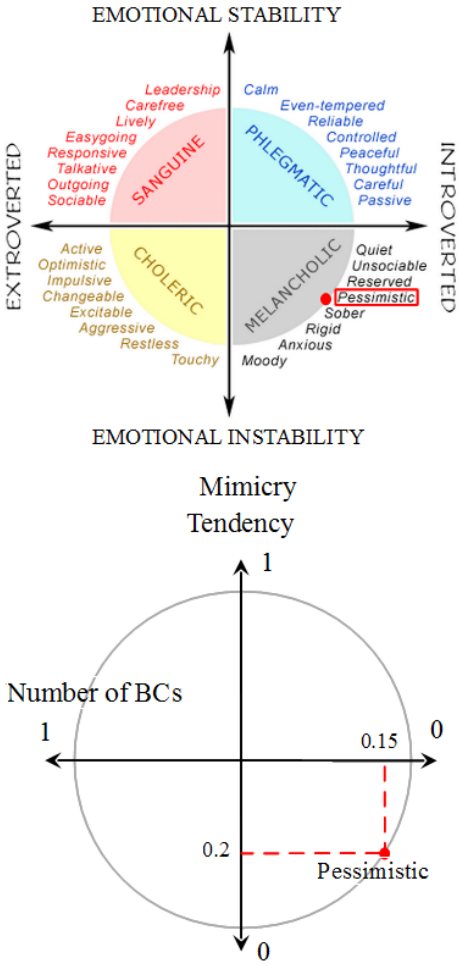

Fig. 3 Eysenck's two dimensional representation and our hypothesis of its implication on tendency to mimicry and number of backchannels. Example of deduction for Obadiah.

6 Action Selection

The Action Selection (AS) module receives all possible actions coming from the Listener Intent Planner and the Dialogue Manager (see Figure 1). The principal role of the Action Selection module is to filter backchannels according to personality of the agent.

In the SEMAINE Project, four SAL characters have been designed with their own personality traits. Poppy is outgoing and cheerful; Spike is aggressive and argumentative; Prudence is reliable and pragmatic; and Obadiah is pessimistic and gloomy. We have defined their respective traits (and associated behaviour tendencies) based on a dimensional approach. We have situated these traits on the dimensions of extroversion and neuroticism (emotional instability). They are important dimensions in all major theories of personality. We use the circle representation validated by Eysenck [19] for the four SAL characters (see Figure 3).

In order to define parameters for the Action Selection module in terms of frequency and type of backchannels according to the two dimensions of the personality, we base our choices on the following assumptions:

- H1: the extroversion dimension is associated to the frequency of the backchannels (mimicry and reactive/response backchannels). Poppy (outgoing) should perform more backchannels and Obadiah (pessimistic) less [7].
- H2: the emotional stability dimension is linked to the type of backchannels displayed by the ECA (mimicry tendency) [16]. Prudence (reliable) should mimic more than Spike (aggressive) [13,38].

	Obadiah	Poppy	Prudence	Spike
BC type	0.2	0.65	0.85	0.1
BC frequency	0.15	0.95	0.2	0.75

Table 1 Setting of BC priority and frequency for the four SAL agents.

We designed a circle equivalent to Eysenck's representation, where the frequency of backchannels axis is similar to Eysenck's extroversion axis and the type of backchannel axis (tendency to perform mimicry over reactive/response backchannels) is similar to the emotion stability axis. Although the parameters of the AS module are not easy to tune, we can easily set frequency and type of backchannels for our listener backchannel selection by following the two hypotheses. On the horizontal axis, 0 corresponds to block all the backchannels coming from the LIP and 1 to let all of them pass to be displayed by the agent. 0.5 corresponds to a moderate frequency of backchannels. Their number depends of the non-verbal behaviours and the voice variation of the users. On the vertical axis, 1 corresponds to favour mimicry over reactive/response backchannels and 0 to favour reactive/response backchannels over mimicry in term of priority for the AS module. 0.5 corresponds to have no preference on the type of backchannels to be displayed by the agent.

We proceed by locating the personality trait on Eysenck's representation and by translating to our graph we obtain values for the frequency and priority of backchannels. For example, Obadiah (pessimistic) performs few backchannels (15% of all backchannels received from the Listener Intent Planner) and more reactive/response backchannels (80%) than mimicry (20%). Poppy who is outgoing, performs a lot of backchannels (95% of all backchannels received) and a little more mimicry (65%) than reactive/response backchannels (35%). We obtain parameters for the Action Selection module according to the four personalities (see table 1). They are coherent with the SEMAINE corpus [27] and the literature [12, 7].

Backchannel types.

Backchannel selection is event-based and is done in real-time. Actions can be a mix of several backchannels if there are no conflicts on the same modality. Only one action can be displayed by the ECA at a given time and the AS module receives continuously candidate backchannels. When the ECA is already displaying an action, no choices are made. The action selection algorithm waits until the display of the current action is over before selecting another one to be displayed. These candidate backchannels received during this time are queued and used during the next selection pass. The choice is made when conflicts appear between modalities of backchannels in the queue. A highly emotionally stable agent shows more mimicry behaviours [11, 39] while a highly emotionally unstable agent shows more reactive/responsive behaviours [31, 17]. The priority value for each backchannel coming from the LIP is modified according to our hypothesis H2. It increases or decreases the priorities for certain type of backchannels (mimicry or reactive/response backchannels) based on the agent's personality (degree of neuroticism). The difficulty lies in the computation of these priorities. Finally backchannels with a high priority have a greater chance to be chosen by the selection algorithm to be displayed by the agent.

Backchannel Frequency. Based on a theoretical model [28], we establish a correlation between the extroversion dimension and the frequency of backchannels [7]. From the video analysis of SEMAINE corpus [27], we computed the backchannel frequency: the highest is Poppy, then Spike, Prudence and then Obadiah. The value of the frequency is deduced from our model. For example, the value for Poppy (extrovert) is 0.95 which means that the largest majority of backchannels will be displayed. In contrast, the value for Obadiah (introvert) is 0.15 which means only 15% of the backchannels will be displayed. When the AS module receives a potential backchannel (mimicry or reactive/response backchannel), it calculates a probability in order to determine if the backchannel will be displayed or not, based on the degree of the agent's extroversion. If not, the backchannel is not queued by the AS module.

7 Evaluation studies

To evaluate our system we conducted two perception studies. The first evaluation allowed us to asses the Listener Intent Planner module while the second one was performed to evaluate the Action Selection module. Both evaluations consisted in showing short videos of interactions between the user and the virtual agent. Participants had to rate them by answering a set of questions.

Fig. 4 Screen shot of the video clip used for the first evaluation study.

To create the corpus of videos we asked a naïve user (a middle-aged woman) to tell stories (improvised from a comic book) to our virtual agent. The agent never took the turn and it just listened to the user displaying backchannel signals automatically generated by our system. We manipulate two variables of the agent's behaviour: the type and the frequency of backchannels according to four personalities (pessimistic, outgoing, reliable and aggressive). To concentrate only on the behaviours and to avoid having to consider extra variables, we used only one facial model: we chose Prudence, one of the virtual agents created within the SEMAINE Project, since she shows the most neutral expression. The resulting videos showed both the agent and the user, as shown in Figure 4.

7.1 Listener Intent Planner evaluation

Since the task of the LIP consists of triggering a backchannel at appropriate times, in our evaluation we aimed at showing that the timing of the backchannels generated by the LIP module allows for better human-agent interactions than random timing. For such a purpose we asked participants to rate a set of user-agent interactions in terms of successfulness, a general impression of the listening agent's behaviour and timing of the signals.

Firstly, from our corpus of videos, we selected those where the personality of the agent was pragmatic and where the agent showed only positive backchannel signals (such as head nod, head tilt, smile, raised eyebrows) to show its participation. Then, from the resulting subset of videos, we extracted nine clips lasting between 40 and 50 seconds. For each clip we generated a new modified clip where the agent was replaced by the same agent performing backchannel signals randomly timed. The random sequences of signals were generated by asking another user to speak to the agent. To avoid the risk that these backchannels were not completely random, we selected the second speaker as more different from the first one as possible. The first speaker was a middle-aged woman who spoke slowly and moved a lot her head. The second speaker was younger and spoke faster. She moved less and her speaking pattern as well as her voice intonation were quite different since her mother tongue was not the same as the one of the first speaker. The agent's behaviour was the same as in the previous interaction in terms of frequency and type of backchannels. Each video contained 8 or 9 backchannel signals. All in all, we prepared eighteen video clips, nine in which the agent's backchannels were triggered by our algorithm and nine in which backchannel signals were given randomly, that is they were not generated according to the user's acoustic and visual behaviour but provided at random timing.

The videos were divided in three groups of six. Each group contained three videos with the backchannels triggered by the LIP module and three videos with the backchannels performed randomly. We hypothesised that when the agent's backchannels are triggered by our algorithm:

- Hp1: the interaction is judged more successful,
- Hp2: the agent's behaviour appears more believable,
- Hp3: the agent is perceived to show less frequently backchannels at inappropriate times,
- Hp4: the agent is perceived to miss less frequently occasions to show a backchannel at appropriate times.

We expected that our algorithm would get higher ratings than the randomly timed backchannels on questions 1 and 2 and lower ratings on questions 3 and 4.

7.1.1 Procedure and participants

Participants accessed the evaluation study through a Web site. The first page introduced them to the evaluation and provided instructions. The second one asked participants to provide some demographic information. Then six video clips were displayed randomly one at a time. Participants could watch a video as many times as they liked before evaluating it through four 7-point Likert-like scales. The four questions were similar to those proposed by Huang et al. [22] in their study.

We asked participants to judge (1) how successful the interaction was (from *not at all* to *absolutely*), (2) how believable the listening agent's behaviour appeared (from *not at all* to *absolutely*), (3) how often the agent performed a backchannel when it should have not (from *very rarely* to *very often*) and (4) how often the agent did not show a backchannel when it should have (from *very rarely* to *very often*).

128 participants (87 women, 41 men) with a mean age of 32.12 years took part in the study. They were mainly from France (75%), and all with a good knowledge of the French language.

7.1.2 Results

The multivariate test of differences between the types of agent's backchanneling (random vs algorithm) on the answers to the four questions, using the Wilks' Lambda criteria, was statistically significant, $F(4, 525)=2.61$, $p<.05$. There was also an effect of the presented video, $F(28,1894)=1.70$, $p<.05$. The interaction between the two was not significant ($p>.05$) using Wilks' Lambda.

Each of the F-ratio transformations of the Wilks criteria were exact.

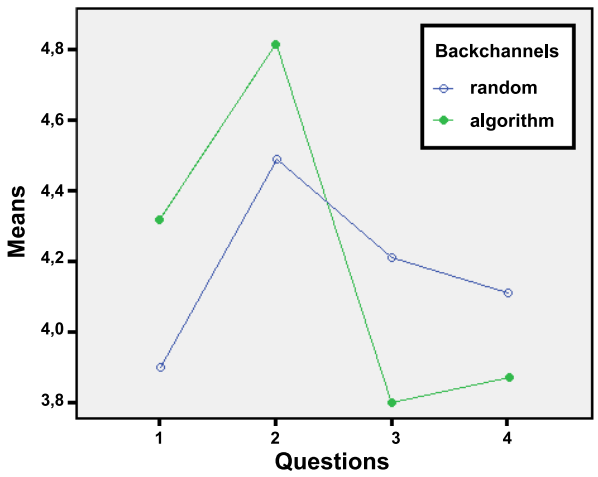

Fig. 5 Algorithm-ruled and random backchannel means for the four evaluated questions.

A test of between participants effects (univariate ANOVA) showed an effect of the presented video for questions 3 ($F(7, 5.8)=2.09$, $p<.05$) and question 4 ($F(7, 5.61)=2.03$, $p<.05$). There was no effect on questions 1 and 2. There was an effect of type of backchannel (algorithm vs random) for all except question 4 ($p>.05$). The means for questions 1 and 2 were higher for the algorithm backchannels than the random ones. It was the opposite for question 3 and 4 respectively, that were lower in the algorithm condition than in the random condition (see Figure 5). The effect of the interaction of agent's backchanneling and video had an effect only on question 2 ($F(7, 4.59)=2.05$, $p<.05$).

7.1.3 Discussion

Our major expectation has been fulfilled, as participants differentiated in their judgement the backchanneling defined by the algorithm and the random backchanneling (i.e. from a different storytelling context than the one presented, with a different user). This effect was significant for three out of four questions. Thus, participants judged the interaction more successful and the agent more believable when the timing of the backchannels was computed by our algorithm than when it was randomly determined. According to the participants' responses, the agent shows also less frequently backchannels at inappropriate times when ruled by the algorithm, however we did not obtain significant results for the fourth question so we cannot affirm that with our algorithm the agent misses less frequently occasions to show a backchannel at appropriate times.

Although we have not formulated any hypothesis regarding the impact of the user's behaviour in itself, we see that it had an impact, the different videos being judged differently by the participants. This was particularly clear for question 3 and 4, that is on backchannels that were more frequent than expected and on the number of missed occasions to backchannel. In some cases the values attributed to the two points were more important than in some others, e.g. question 4 was typically of more value to Poppy, while question 3 of more value to Obadiah. However, there is no interaction between the presented video and the backchannel definition, thus, although there is an impact of the behaviour and the input of the user, we see a general positive effect of the algorithm.

Thus our results show that the timing of the reactions of a listening ECA is important and whether it is contingent or not has a notable impact on the evaluation by a third party. Our algorithm seems to react at appropriate times to the captured and processed audio and visual cues from a user in a storytelling context. We can conclude that the results of our perception study confirm that the Listener Intent Planner allows for better human-agent interactions than random timing.

7.2 Action Selection evaluation

The role of the AS module consists in determining the type of backchannels to favour and adapt the frequency of backchannels according to the personality of the agent. We want to evaluate if the filtration by our Action Selection module allows a better interpretation by the user in terms of the personality of the agent.

From our corpus of videos, we select twelve video clips of twenty seconds. The participants have to evaluate if the frequency and the type of backchannels are appropriate and their impression of the interaction according to a personality of the agent among the four possible. We have formulated the hypotheses that the behaviour of the agent is more appropriate to its personality when our Action Selection module is running:

- H1: the frequency of the backchannels filtered by the AS module should be more appropriate to the personality of the agent.
- H2: the type of the backchannels filtered by the AS module should be more appropriate to the personality of the agent.

Similarly to the first study, this one was accessed through a Web site. Each evaluation page is composed of the description of a virtual agent's personality, a reminder of instructions, and two videos. Before the participants can pass to another page, they have to watch

the two videos and answer two questions for each video concerning our two hypotheses: whether the agent reacts appropriately (BC type) and sufficiently (BC frequency) accordingly to the described personality. Participants use a 5-point Likert scale, from "not at all" to "completely" for the BC type and from "not enough" to "too much" for the BC frequency.

We defined three conditions for each personality to evaluate the effect of the frequency and the type of backchannels filtered by the AS module:

- C1: variation of the backchannel frequency (with BC types baseline)
- C2: variation of the backchannel type (with BC frequency baseline)
- C3: variation of both

Concerning the baseline in C1, the AS module has no preferences in choosing mimicry or response/reactive backchannels. If there is a conflict, it selects the first one. Concerning the baseline in C2, the Action Selection module filters 50% of the backchannels coming from the Listener Intent Planner. These baselines are applied for the four personalities.

The evaluation contains twelve pages (four personalities and three conditions for each) showed randomly. On each page, we have one video with the correct personality (described on the page) in one of the three condition. The second video corresponds to one of the two personalities of the other dimension. For example, if the defined personality is outgoing (extroversion dimension) and the condition is the BC frequency, a video of Poppy with this condition is placed in a random position in the page (up or down). The second video is chosen randomly between Spike and Prudence videos (neuroticism dimension) with the same condition.

The table 2 shows the number of backchannels in the video generated by the Action Selection module according to the four personalities. Their frequencies and types correspond to our assumptions (see table 1).

Video	Obadiah	Poppy	Prudence	Spike
Nb of BC	3	8	4	6
Nb of RBC	3	5	1	5
Nb of M	0	3	3	1

Table 2 Number of backchannels (BC), reponse backchannels (RBC) and mimicries (M) for the four personalities in the video clip (15 sec) generated by the Action Selection module

As we want to evaluate only the Action Selection module, we assume that the backchannels received from the Listener Intent Planner are appropriate and that their timing is correct.

7.2.1 Results

Ninety three participants (57 women, 37 men) mainly from France (80%) took part in the study. Nearly half of the participants has a computer science background (39%), the remaining being mainly from psychology (21%). The majority of the participants (78%) were graduates or post-graduates and declared a good notion of computer use (very good 55%, good 35%). Results showed that Spike's personality was the easiest to recognize (62%), followed by Poppy's and Obadiah's (53%). Prudence's personality appeared to be the most difficult to identify (52%).

To the question on the **frequency** of backchannels, the majority of the participants answered that the agent reacts adequately to the four personalities (see Figure 6). This was not the choice by default so the participants actively chose this response. As the results are homogeneously distributed, we performed an ANOVA and Paired samples tests to verify our hypothesis H1 on the selection of BC frequency. We expected the C3 condition to be evaluated better by participants than the C2 condition.

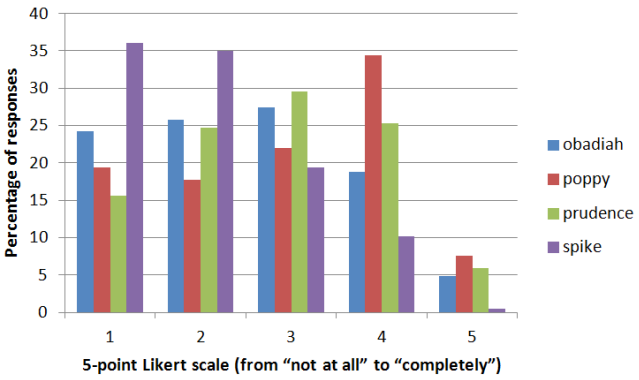

Fig. 6 Answer distribution for the BC frequency concerning the appropriate video for the four personalities (from 1: nor enough, 3: normal to 5: too much).

BC frequency	n	F	p	
Personality	3	9.737	.000	*
Condition	3	18.032	.000	*
Personality*Condition	6	1.369	.225	

Table 3 ANOVA results for hypothesis H1 on BC frequency.

The answers of the participants to the question on the frequency of the backchannels show an effect (ANOVA, $p<.05$) of the personality (see table 3) and of

BC frequency	n	$\bar{X}$	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	

Table 4 Paired Samples T Test results for hypothesis H1 on BC frequency.

the condition but not of the interaction of personalities and conditions on the judgements (ANOVA, p>.05). The variations of BC frequency (difference between the C2 and C3 conditions) for Poppy (outgoing) and Prudence (reliable) were significant (t-test, p<.05) (see table 4) and not significant for Obadiah (pessimistic) and Spike (aggressive) (t-test, p>.05). The participants consider that the C3 condition for Poppy and Prudence is better than the C2 condition.

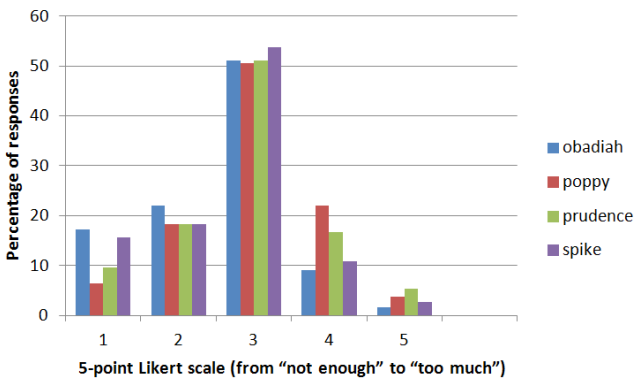

Fig. 7 Answer distribution for the BC type concerning the appropriate video for the four personalities (from 1: not at all to 5: completely)

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

Table 5 Friedman's ANOVA test results for hypothesis H2 on BC type.

BC type	n	$\bar{X}$	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	

Table 6 Wilcoxon Signed Rank Test results for hypothesis H2 on BC type.

Concerning the **type** of backchannels, the backchannels are evaluated as appropriate for Poppy and moderately appropriate for Prudence and Obadiah and not appropriate for Spike (see Figure 7). As the results are not homogeneously distributed, we performed Friedman's ANOVA and Wilcoxon Signed Rank Test to verify our H2 hypotheses concerning the BC type selection. We expected the C3 condition to be evaluated better by participants than the C1 condition.

The answers of the participants to the question about the appropriateness of the backchannels (type) are significant (Friedman test, p<.005) (see table 5). The variation of BC type (difference between conditions C1 and C3) for Obadiah (pessimistic) is significant (p<.005) and not significant for the other personalities (p>.05) (see table 6). The participants consider that only for Obadiah the C3 condition is better than the C1 condition.

7.2.2 Discussion

The main aim of this evaluation study was to check if the variation of the generated BC type and frequency have an impact on participant's perception. The first hypothesis was partially verified: although the attributions were higher with the selection of BC frequency than that of alternatives, the difference was not significant for some personalities. The second hypothesis was verified only for Obadiah (pessimistic).

Concerning hypothesis H1, most of the participants judged the frequency of backchannels as adequate to the personalities, however we obtained significant results only for Poppy (outgoing) and Prudence (pragmatic). Agents who perform a lot of backchannels are associated to extroversion; whereas agents who show a little less backchannels than normal are considered pragmatic. We did not obtained significant results for introversion. Maybe, since Prudence's and Obadiah's backchannel frequency was quite similar (see table 2), participants easily mistook one for the other. As for Spike, who is aggressive, people might have expected a even higher frequency of backchannels (which was already high in our settings).

Concerning hypothesis H2, the results for the hypothesis H2 were not very conclusive for all the personalities (except pessimistic). More evaluations are necessary to validate it. We believe that a part of the problem was the adjective describing the personalities. They might not have been optimal in conveying the meaning we were looking for. For instance, participants said that they did not understand the adjective "pragmatic" and they did not really know how a pragmatic person reacts. If they did not have a clear idea about how the

agent should react, they could not see the difference in the evaluation. Therefore, participants had difficulties in recognizing the video clip associated to the personalty described on the evaluation page. These terms need to be clarified for the next evaluations. Moreover participants also comment on the difficulty to show aggression for Spike or express pessimism for Obadiah only through backchannels. We believe it could explain the judgements that did not meet our expectations.

8 Conclusion

In this paper, we presented a computational model that generates the virtual agent's behaviour while listening to a user, taking into account the agent's personality. The model is composed by two modules: the Listener Intent Planner module that triggers backchannel signals according to the user's visual and acoustic behaviour and the Action Selection module that, according to the agent's personality, chooses the type and the frequency of the backchannels to be displayed by the agent. We evaluated our system through two perception studies. In the first study we evaluated that the timing of the backchannels generated by the Listener Intent Planner module allows for better human-agent interactions than random timing. Participants judged the interaction more successful and the agent more believable when the timing of the backchannels is computed by our algorithm than when it is randomly determined. In the second study we evaluated that behaviour is interpreted as appropriate for a personality when the backchannel frequency is linked with the extroversion dimension and the backchannel type is linked with the neuroticism dimension. The evaluation showed that the selection of frequency of backchannels performed by our Action Selection module does contribute to the correct interpretation of the agent's behaviour in terms of personality traits. Concerning the type of backchannels, more evaluations are necessary to validate our hypothesis.

9 Acknowledgement

This work has been funded by the STREP SEMAINE project IST-211486 (http://www.semaine-project.eu) and Web 2.0 MyPresentingAvatar project.

References

1. Allwood, J., Cerrato, L.: A study of gestural feedback expressions. In: P. Paggio, K. Jokinen, A. Jonsson (eds.) First Nordic Symposium on Multimodal Communication, pp. 7–22. Copenaghen (2003)
2. Allwood, J., Nivre, J., Ahlsén, E.: On the semantics and pragmatics of linguistic feedback. Journal of Semantics pp. 1–26 (1992)
3. van Baaren, R.B., Holland, R.W., Steenaert, B., van Knippenberg, A.: Mimicry for money: Behavioral consequences of imitation. Journal of Experimental Social Psychology **39**, 393398 (2003)
4. 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)
5. 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)
6. Bevacqua, E., Mancini, M., Pelachaud, C.: A listening agent exhibiting variable behaviour. In: H. Prendinger, J.C. Lester, M. Ishizuka (eds.) Proceedings of 8th International Conference on Intelligent Virtual Agents, pp. 262–269. Springer, Tokyo, Japan (2008)
7. Borkenau, P., Liebler, A.: Trait inferences: Sources of validity at zero acquaintance. Journal of Personality and Social Psychology **62**, 645–657 (1992)
8. Cassell, J., Bickmore, T.: Embodiment in conversational interfaces: Rea. In: Human Factors in Computing Systems. Pittsburgh, PA (1999)
9. 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)
10. Cerrato, L., Skhiri, M.: Analysis and measurment of head movements signalling feedback in face-to-face human dialogues. In: P. Paggio, K. Jokinen, A. Jónsson (eds.) First Nordic Symposium on Multimodal Communication, pp. 43–52. Copenaghen (2003)
11. Chartrand, T., Bargh, J.: The Chameleon Effect: The Perception-Behavior Link and Social Interaction. Personality and Social Psychology **76**, 893–910 (1999)
12. 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)
13. Chartrand, T., Maddux, W., Lakin, J.: The new unconscious, chap. Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry, pp. 334–361. New York: Oxford University Press (2005)
14. Costa, P.T., McCrae, R.R.: Four ways five factors are basic. Personality and Individual Differences **13**, 653–665 (1992)
15. Decety, J., Jackson, P.: The functional architecture of human empathy. Behavioral and Cognitive Neuroscience Reviews **3**, 71–100 (2004)
16. Eisenberg, N., Fabes, R.A., Murphy, B., Karbon, M., Smith, M., Maszk, P.: The relations of childrens dispositional empathy-related responding to their emotionality, regulation, and social functioning. Developmental Psychology **32**, 195–209 (1996)
17. Eisenberg, N., Fabes, R.A., Shepard, S.A., Murphy, B.C., Jones, S., Guthrie, I.K.: Contemporaneous and longitudinal prediction of childrens sympathy from dispositional regulation and emotionality. Developmental Psychology **34**, 910–924 (1998)

18. Eyben, F., Wöllmer, M., Schuller, B.: openSMILE - the Munich versatile and fast open-source audio feature extractor. In: ACM Multimedia (MM), pp. 1459–1462. Florence, Italy (2010)
19. Eysenck, H.J.: Dimensions of personality: Criteria for a taxonomic paradigm. Personality and Individual Differences **12**, 773–779 (1991)
20. Eysenck, S., Eysenck, H.: Impulsiveness and venturesomeness - their position in a dimensional system of personality description. Psychol Rep **43**, 1247–1255 (1978)
21. 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)
22. 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)
23. 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)
24. Kopp, S., Allwood, J., Grammer, K., Ahlsen, E., Stocksmeier, T.: Modeling Embodied Feedback with Virtual Humans. Lecture Notes in Computer Science **4930**, 18 (2008)
25. Lakin, J.L., Jefferis, V.A., Cheng, C.M., Chartrand, T.L.: Chameleon effect as social glue: Evidence for the evolutionary significance of nonconsious mimicry. Nonverbal Behavior **27**(3), 145–162 (2003)
26. Maatman, R.M., Gratch, J., Marsella, S.: Natural behavior of a listening agent. In: 5th International Conference on Interactive Virtual Agents. Kos, Greece (2005)
27. 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, Singapore (2010)
28. 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)
29. Morency, L.P., de Kok, I., Gratch, J.: Predicting listener backchannels: A probabilistic multimodal approach. In: 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)
30. Nass, C., Steuer, J., Tauber, E.R.: Computers are social actors. In: CHI, pp. 72–78 (1994)
31. Noor, F., Evans, D.: The effect of facial symmetry on perceptions of personality and attractiveness. Journal of Research in Personality **37**, 339–347 (2003)
32. Poggi, I.: Mind markers. In: N. Trigo, M. Rector, I. Poggi (eds.) Gestures. Meaning and use. University Fernando Pessoa Press, Oporto, Portugal (2003)
33. Poggi, I.: Mind, hands, face and body. A goal and belief view of multimodal communication. Weidler, Berlin (2007)
34. Von der Pütten, A., Krämer, N.C., Gratch, J., Kang, S.: "it doesn't matter what you are!" explaining social effects of agents and avatars. Computers in Human Behavior, in press (2010)
35. Reeves, B., Nass, C.: The media equation: How people treat computers, television and new media like real people and places. CSLI Publications, Stanford, CA (1996)
36. Schröder, M.: The semaine api: Towards a standards-based framework for building emotion-oriented systems. Advances in Human-Computer Interaction **2010** (2010)
37. Schröder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., ter Maat, M., McKeown, G., Pammi, S., Pantic, M., Pelachaud, C., Schuller, B., de Sevin, E., Valstar, M., Wollmer, M.: Building Autonomous Sensitive Artificial Listeners. Journal of Transactions on Affective Computing (to appear)
38. Sonnby-Borgstrm, M., Jnsson, P., Svensson, O.: Emotional empathy as related to mimicry reactions at different levels of information processing. Journal of Nonverbal Behavior **27**, 3–23 (2003). URL http://dx.doi.org/10.1023/A:1023608506243. 10.1023/A:1023608506243
39. Sonnby-Borgstrom, M.: Automatic mimicry reactions as related to differences in emotional empathy. Scandinavian Journal of Psychology **43**, 433–443 (2002)
40. Thórisson, K.R.: Communicative humanoids: A computational model of psychosocial dialogue skills. Ph.D. thesis, MIT Media Laboratory (1996)
41. 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)
42. Ward, N., Tsukahara, W.: Prosodic features which cue back-channel responses in english and japanese. Journal of Pragmatics **23**, 1177–1207 (2000)
43. Watson, D.: Strangers? ratings of the five robust personality factors evidence of a surprising convergence with self-report. Journal of Personality and Social Psychology **57**, 120–128 (1988)
44. Wiggins, J., Trapnell, P., Phillips, N.: Psychometric and geometric characteristics of the revised interpersonal adjectives scale (ias-r). Journal of Multivariate Behavioral Research **23**, 217–305 (1988)
45. Yngve, V.: On getting a word in edgewise. In: Papers from the Sixth Regional Meeting of the Chicago Linguistic Society, pp. 567–577 (1970)