A virtual diary companion

Anton Nijholt[1], Ferdi Meijerink[1] and Peter-Paul van Maanen[2]

Abstract. Chatbots and embodied conversational agents show turn based conversation behaviour. In current research we almost always assume that each utterance of a human conversational partner should be followed by an intelligent and/or empathetic reaction of chatbot or embodied agent. They are assumed to be alert, trying to please the user. There are other applications which have not yet received much attention and which require a more patient or relaxed attitude, waiting for the right moment to provide feedback to the human partner. Being able and willing to listen is one of the conditions for being successful. In this paper we have some observations on listening behaviour research and introduce one of our applications, the virtual diary companion.

1 INTRODUCTION

Textual chatbots and embodied conversational agents listen to us. Listening is part of the interaction and listening behaviour should be modelled. Hence, we want chatbots or conversational agents to know how to listen. And, to determine when they have to react. This is an important issue for current and future applications of chatbots and conversational agents, especially when we take into account the convergence we see between traditional text-like chatbots using all kinds of tricks to get a believable conversation and current embodied conversational agents that are equipped with (primitive) models of mind that allow them to show empathy using verbal and nonverbal cues, displayed by animation and speech features. On the one hand we have chatbots like the forty year old Eliza [1] and its successors, on the other hand we have fully embodied agents that have sensors to perceive their conversational partners and have verbal and nonverbal capabilities. But, as everyone knows, those embodied agents may have nice animations, but they are nevertheless hardly able to 'maintain' a believable conversation for more than one utterance. The main obstacle that prevents having believable conversations between a chatbot and its human partner remains our inability to model realistic language use. This is even truer when we ask our bots to understand spoken language since it is less formal than written language and it introduces the additional difficulty of speech recognition.

So, when we ask the question "Do we want our chatbots and embodied agents to know how to listen to us?" and we want to answer it in a confirmative way, then we must have applications in mind where we can somehow circumvent these problems or embed imperfect solutions to these problems in a context where other available information can compensate for this imperfectness. Admittedly, this is not a very original observation, but making it explicit we can understand why so much research is now going on the role of nonverbal signals in human-human interaction and how they can be employed in natural human-computer interaction, where an embodied agent represents an application. Obviously, we prefer embodied agents to be pro-active, act autonomously, but preferably in our interests, act in an intelligent and social way, showing empathy, et cetera, depending on the application. We may even forget about an application and just assume that there is a world in which such agents will live their own life, with or without interacting with human partners.

Obviously, in a world where embodied agents and humans co-exist, communicate with each other, and have other joint activities, embodied agents are active as speaker and as listener. Often these tasks, for example in everyday life conversations, changing the role of being the speaker to being the listener are in balance. But in situations where the interactants have different roles (salesman vs customer, receptionist vs visitor, doctor vs patient, et cetera) the emphasis for one of the conversational partners is on listening and for the other partner on speaking. It is not the case that a partner in a conversation that is (momentarily) listening is not active in the conversation. On the contrary, a listener's facial expressions, head movements and body posture changes, whether they are voluntarily or involuntarily displayed, interact and synchronize with a speaker's verbal and nonverbal activities continuously.

We can think of applications where in a real-life interaction situation one of the participants is replaced by a computer, an embodied agent or a humanoid robot and where we nevertheless want to maintain the same quality of natural interaction. This requires modeling listening behaviour. There are also applications where the main task of an embodied agent or chatbot is to listen, to show empathy and to take care that the conversation continues or that a transaction is completed successfully. Successful performance in such a task very much depends on nonverbal communication abilities that have been designed for such an agent, in particular abilities that relate to human listening behaviour. Obviously, we can look at real-life face-to-face interaction situations where one of the conversational partners is mainly listening and turn such a situation into a computer application. But we can also look at less traditional and not yet existing interaction applications made possible by new multimodal interaction, multimedia access, and multimedia presentation technology. In the former case we can look at embedding 'listening intelligence' in an agent that performs human-like tasks such as being a doctor or being a psychotherapist (as in the case of Eliza). But, not less importantly, we can look at 'new' applications where different than in human-human interaction we can assume explicit design of intelligence and affect that aims at allowing a chatbot or an embodied agent to provide support that is not available from a human partner. One of such applications is the affective diary that will be discussed here.

[1] University of Twente, Human Media Interaction (HMI), Enschede, the Netherlands, anijholt@cs.utwente.nl
[2] TNO Human Factors, the Netherlands, peter-paul.vanmaanen@tno.nl

2 HMI RESEARCH ON LISTENING AGENTS

We shortly discuss our previous research (in the Human Media Interaction group of the University of Twente) related to modelling listening behaviour of agents. We looked at various ways to provide our (embodied) agents with intelligence, affect and empathy. In [2] we looked at mechanisms involved in friendship formation and how they can be translated to a human – embodied-agent situation. Short-term and long-term characteristics of a friendship relationship were distinguished, in particular the possibility of adapting to personality characteristics of a human partner. Affective multimodal interaction with an embodied tutor was discussed in [3]. In this research the tutor monitors the performance of a student in a nurse education task. The embodied tutor hardly speaks, but his face shows his appreciation of the student's performance (see Figure 1).

Figure 1 A tutor agent that monitors the student

In [4] we discussed design issues of a virtual coach that was meant to replace a human 'quit smoking coach'. This project was done in close cooperation with the official Dutch organization that aims at supporting people to quit smoking. Therefore an extensive analysis of the practice of individual coaching was possible (see Figure 2 for some characteristic listening expressions of a professional human 'quit smoking coach').

While in these examples of research the emphasis was on embodiment, in particular the possibility to show facial expressions, we also undertook research on textual chatbots. We introduced a chatbot that attempts to employ humour in its conversation with a human partner in [5]. This chatbot tries to generate funny questions by purposely misunderstanding a user's utterance. A chatbot that provides feedback to its human partner that discloses his feelings about emotional events that were experienced was introduced in [6,7]. We will discuss this research in this paper.

Figure 2 Listening expressions of a 'quit smoking coach'

Our research that is explicitly devoted to (embodied) listening agents started in the EU FP6 Network of Excellence Humaine [8] and this research is now continued in the EU FP7 Semaine project [9], and the EU FP7 Network of Excellence SSPNet. Related research takes place in the EU Cost 2102 action in which we are involved. In the Humaine NoE we analysed and annotated listening behaviour with the aim to design a Wizard of Oz environment for research purposes with semi-autonomous listening agents with different personalities interacting with human conversational partners [10]. Personality is shown in verbal and nonverbal behaviour of the listening agents. This research is continued in the Semaine project in which we are concerned with the management of the interaction between user and artificial listener. The NoE SSPNet in which we participate researches social signals: mainly nonverbal signals through which humans communicate, often without conscious awareness, their attitude towards others and social situations.

3 CHATBOTS AND LISTENING BEHAVIOR

Starting with Eliza [1], we can look at the 'listening behaviour' of many chatbots that have been introduced in the past and that sometimes explicitly have been introduced to pass the Turing test. This is ongoing research. In the case of Eliza, Joseph Weizenbaum's program was listening and providing feedback that could be given without any understanding of the contents of the interaction and that was aimed at eliciting more information from its human conversational partner. Mimicking the verbal behaviour of its human conversational partner by continuing and rephrasing verbal content was one of the strategies employed by Eliza. Eliza was not embodied. Eliza had the initiative, the 'user' typed in his or her questions, answers or other utterances, and sometimes tried to take the initiative, Eliza generated textual responses and took the initiative, sometimes by changing the topic, for example, when she was not able to generate adequate feedback to the user. The Eliza textual turn taking approach does not allow continuous and synchronized feedback as is essential in human face-to-face communication [11]. Usually text chatbots are turn-based. Each user utterance is followed by a system utterance. Clearly, when a chatbot is not embodied, we have to accept that all possibly relevant vocal feedback during listening (h'm, aha, yes, go on, really …) and feedback from head nods, facial expressions, gaze and posture can not be used or in a less natural way only.

While on the one hand we see research aiming at introducing 'believable' chatbots, e.g. Alice [12], by modeling more general common sense and domain knowledge, on the other hand we see attempts to have deeper linguistic analysis of dialogue utterances. Such attempts may take the form of general research on natural language dialogue modelling or research guided by chatbot-like applications, such as, affective diaries [13] and empathic buddies [14, 15]. In particular this latter viewpoint on applications where affect plays an important role has received attention. These applications require underlying affect models, e.g. the OCC model [16], and often data (opinion, affect) mining

methods are used to extract affective information from a dialogue and to use this information for affective feedback [17].

4 DIARY COMPANION: MOTIVATION

In our research on 'listening' agents we introduced an agent that plays the role of an interactive affective diary, i.e., an agent that provides emotional feedback based on emotional content analyses of the current and past conversation with the subject [7]. This diary companion is meant to evoke disclosure of emotions and traumatic events for soldiers on peace keeping missions or astronauts on enduring space missions. Obviously, an 'understanding' companion can play a role in many other situations too, including pangs of love and loosing a loved one.

Nevertheless, military crew who are on a mission in war zone are often exposed to great stressors. These include threat to life and exposure to grotesque death. Though other factors also contribute to the risk of developing Post Traumatic Stress Disorder (PTSD), war experience is a good predictor for the development of PTSD. Clearly in many cases war experience has a negative effect on one's physical and psychological health.

There have been many attempts as to solve this problem, including training techniques and psychotherapy. Since it is not always possible to provide such support, for example due to lack of resources, it is essential to provide a substitute.

According to Pennebaker [18], the expression of emotions can have a positive effect on psychological and physiological health. This process is commonly referred to as emotional disclosure. This can be facilitated by keeping a diary for example. Since it is important to express yourself emotionally, rather than expressing non-emotional matter, it is important that one is guided during the process of disclosure. This has been our motivation to look at the development of an emotional intelligent agent which should not be regarded as a substitute for psychoanalysis, but rather as an improvement of keeping a diary with the benefit of early intervention.

From, among other things, the literature mentioned above, we derived several requirements. We mention (1) the conversational interface should ensure that the user can express himself freely and is not distracted by the feedback he receives, (2) the agent must behave as expected; this may require (self-)explaining its basic workings, (3) the agent must perform reflections of the user it is talking to, (4) the agent must provide emotional support: expression of esteem and reassurance of worth, and showing affection and attachment, (5) the system has to adapt to the user in order to account for interpersonal differences and preferences.

Obviously, there are other properties of such a system we would like to see, mainly issues that deal with trust and long-term relationships. Until now we have not taken them into account, although it is clear that they are related to the requirements above.

5 DIARY COMPANION: ARCHITECTURE

A global description of the affective diary is as follows. As mentioned earlier, in traditional conversational agents and chatbots, conversation is turn based. Every sentence requires an answer. In emotional disclosure sessions with a therapist, however, it is common that the therapist does not interrupt his patient. Only to guide the conversation to the right direction, will he interrupt the patient. For that reason, and obviously, also the state of the art of natural language processing, we have chosen an interface similar to that of a diary. That is, the user types in text in a text area. He is free to type whatever comes to mind and cannot be interrupted by the system. Therefore, the system communicates with the user through another channel.

In order for the system to communicate sensibly, the text input is analysed. Text input takes place in a text area and no confirmations regarding the end of a discourse unit are supplied. Therefore, the system monitors this field simultaneously. This way, text is analysed in real time and feedback is provided whenever the system feels it is necessary. This feedback consists of emotional support and reflections of the user (i.e. confirmation of correctly interpreted input). The input is pre-processed into complete sentences. 'Part of Speech' processing is applied in order to get more detailed information about the user input.

Providing emotional support requires that the emotional content of the input is analysed. This content is extracted from the word features using WordNet Affect [19] and then, by using information available in the user model, the affective state of the user is calculated. As it is not preferable to provide feedback after every sentence, the Synthetic Partner needs a way to determine if a reaction is necessary. Information from the stored discourse and in the user model is used to determine this and to generate feedback using a template based system. In order to be sure that the system is on the right track with the particular user confirmative questions can be generated regarding the detected emotions. The answers are assessed and the parameters of the emotion detection mechanism are adapted accordingly.

In the feedback process decisions are made whether or not to encourage a user to disclose more of his feelings, to ask the user whether he is (still) content with the amount of feedback provided, and whether the affective state calculated by the diary companion is sufficiently correct. In the next section we have a few more remarks about the emotion detection and the feedback.

6 DIARY COMPANION: IMPLEMENTATION

The virtual diary companion has been implemented in JAVA. There is a designated diary area where a user enters text. The text is processed in the background. The user can concentrate on the disclosure process. System feedback is placed in a separate text field, adjacent to the diary area. This way, the user is not interrupted during his expression of emotions. The system will continuously determine the need to supply feedback to the input from the user. When it determines, based on a set of (adaptable) rules, that it needs to interact with the user, a message will be displayed in the designated area. This message will be displayed for the amount of time specified by this certain feedback type, after which, the message can be overwritten.

Interaction is required to be able to adapt the system according to the user's characteristics. A specific kind of feedback has been implemented for this purpose. The diary companion can supply questions in the designated area, which can be answered using multiple choice answer buttons. These buttons are placed directly under the question. If a button is clicked, they will disappear, indicating that the question has been answered. The answers to these questions are then send to the appropriate handler, which will perform the specified actions.

The answers can for example be used in the adaptation of the system's emotion detection parameters. This type of feedback is also displayed for the amount of time specified. However, if for some questions the answer might be critical to the functioning of the system it can also be displayed indefinitely, until the user decides to answer the question.

WordNet Affect contains a list of word senses, which are related to a label. The labels are hierarchically classified in 312 affective categories, the largest being the emotion category. Because WordNet Affect is categorised hierarchically, it allows us to view a word from an arbitrary emotion level. We have chosen to take a top-down approach for integrating WordNet Affect. In the current implementation, only positive emotion, negative emotion, and neutral emotion from the first level are taken into account when detecting emotions.

The emotion extraction processed is summarised as follows. Because WordNet Affect contains only nouns, adjectives, verbs, and adverbs, these are the only words that will be checked for emotional category.

We have modified the WordNet Affect database so it contains lists of words (grouped by part of speech) linked with one or more emotion categories. So looking up a token in the WordNet Affect database results in a list with zero or more emotion categories. For each of these categories, the system will determine whether they are part of the positive, negative or neutral emotion category in the emotion hierarchy. Tokens are then scored for the number of references they have to each emotion category. For example, ‘happy’ has three references to an emotion category: ‘contentment’, ‘euphoria’, and ‘happiness’. These are all in the positive group. As a result the positive score for the adjective ‘happy’ is 3. There are no negative and neutral references, so they both score nil. Then a vector (in this case [3,0,0]) will be associated with the token.

After having determined the emotion vector for each token in a sentence, the vectors are summed, which will then be regarded as an affective state, associated with the sentence. This affective state is then used to update the user model. The current affective state is updated after every processed sentence. For updating the state we have chosen to implement a method similar to the bell analogy mentioned in [19]. Some thresholds in the emotion detection are user-dependent and based on the performance of the system they are updated.

Various visual aids are available for analysis. Figure 3, for example, shows the user model window in which three charts represent the detected positive, negative, and neutral emotions over a period of 50 sentences.

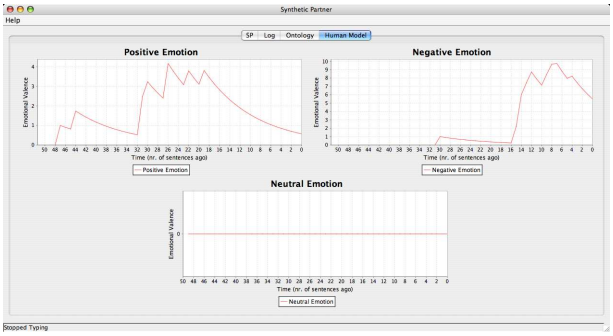

Figure 3 The user model view, displaying the affective states

The feedback process is essential for achieving the diary companion's goal. It is used to keep the user on the right track, gather extra information from the user, to confirm that the diary companion is doing its job well, and to support the user emotionally as well. As we have noted earlier, this feedback takes place in a template based manner. These templates represent several types of feedback. Hence, there is feedback that is used to confirm a user's affective state, feedback that is used to advice users about their monologue (types of emotions that are disclosed, the amount of emotional matter that is disclosed), there are confirmation questions regarding the cause of the disclosed emotion, and there is general feedback for social support (reflective behaviour, emotional support).

7 CONCLUSIONS

Our aim was to build a fully working system that could be seen and tested as a diary companion. Until now we had a limited number of test persons (four) that were asked to use the system for 15 minutes and take the role of a very optimistic or a very pessimistic person. The system performed well, but obviously this was a far from natural situation and the subjects exaggerated their roles. One thing that immediately became clear that there was too much repetition in the way the template feedback was provided to the users. In the implementation of the feedback mechanism and in particular the feedback that supplies reflections, concessions have been made on the part of natural language understanding. The system cannot interpret the text on a semantic level and therefore lacks the ability to respond ‘intelligently’ to the user input. Instead, it reuses only parts of the sentences, for which it has calculated the affective state. This can lead to responses that may seem unusual, since the system has no sense of knowing whether a particular response is coherent. The way we used WordNet Affect is also a very crude approach compared with a more refined linguistic approach combined with an OCC-like approach. Nevertheless we think it has been a useful exercise to aim at a fully working prototype and from there, made possible by its modular design, incrementally improve the various components of the architecture.

Acknowledgements

This work has been supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie).

REFERENCES

[1] J. Weizenbaum. ELIZA--A Computer Program for the Study of Natural Language Communication between Man and Machine. Communications of the ACM, Volume 9, Number 1 (January 1966): 36-35.

[2] B. Stronks, A. Nijholt, P. van der Vet & D. Heylen. Designing for friendship: Becoming friends with your ECA. In: Proceedings Embodied conversational agents - let's specify and evaluate them. A. Marriott et al. (eds.), Bologna, Italy, 2002, 91-97.

[3] D. Heylen, A. Nijholt & R. op den Akker. Affect in Tutoring Dialogues. Journal of Applied Artificial Intelligence (special

issue on Educational Agents - Beyond Virtual Tutors), Vol. 19, No. 3-4, Taylor & Francis, 2005, 287-311.

[4] J. Grolleman, B. van Dijk, A. Nijholt & A. van Emst. Break the habit! Designing an e-therapy intervention using a virtual coach in aid of smoking cessation. In: Proceedings Persuasive 2006. First International Conference on Persuasive Technology for Human Well-being. Lecture Notes in Computer Science 3962, W. IJsselsteijn et al. (eds.), Springer-Verlag, Berlin Heidelberg, 2006, 133-141.

[5] H.W. Tinholt & A. Nijholt. Computational Humour: Utilizing Cross-Reference Ambiguity for Conversational Jokes. In: 7th International Workshop on Fuzzy Logic and Applications (WILF 2007), 2007, Camogli (Genova), Italy, Lecture Notes in Artificial Intelligence 4578, F. Masulli, S. Mitra & G. Pasi (eds.), Springer-Verlag, Berlin, 477-483.

[6] F. Meijerink, P.-P. van Maanen, A.J. van Vliet & A. Nijholt. Disclosure with an emotional intelligent synthetic partner. Proc. workshop on Tools for Psychological Support during Exploration Missions to Mars and Moon. European Space Agency, ESTEC, Noordwijk, The Netherlands, I. Solodilova-Whiteley (ed.), SEA (Group) Ltd, UK, 2007, 31-32.

[7] F. Meijerink. Synthetic Partner: The Design of a Relational Affective Diary. Master's Thesis, Human media Interaction, University of Twente, 2008.

[8] http://emotion-research.net/

[9] http://www.semaine-project.eu/

[10] D. Heylen, A. Nijholt & M. Poel. Generating Nonverbal and Para-verbal Signals for a Sensitive Artificial Listener. Verbal and Nonverbal Communication Behaviours. A. Esposito et al. (eds.), LNCS 4775, Springer, Berlin, 2007, 29-31.

[11] A. Nijholt, D. Reidsma, H. van Welbergen, H.J.A. op den Akker, and Z.M. Ruttkay. Mutually Coordinated Anticipatory Multimodal Interaction. In: Nonverbal Features of Human-Human and Human-Machine Interaction, Patras, Greece. LNCS 5042, Springer Verlag, Berlin, 2008, 73-93.

[12] http://alicebot.blogspot.com/

[13] M. Lindström, A. Ståhl, K. Höök, P. Sundström, J. Laaksolathi, M. Combetto, A. Taylor, R. Bresin. Affective Diary – Designing for Bodily Expressiveness and Self-Reflection. CHI 2006, Montréal, Québec, Canada.

[14] H. Liu, H. Lieberman, and T. Selker. A Model of Textual Affect Sensing using Real-World Knowledge. Proceedings of the Seventh International Conference on Intelligent User Interfaces, Miami, FL), ACM, 2003, 125-132.

[15] M. Al Masum Shaikh, H. Prendinger, and M. Ishizuka. An Analytical Approach to Assess Sentiment of Text. Proceedings ICCIT 2007.

[16] A. Ortony, G. Clore, & A. Collins. The Cognitive Structure of Emotions. Cambridge: Cambridge University Press, 1988.

[17] Li Zhang, J.A. Barnden, R.J. Hendley, and A.M. Wallington. Exploitation in Affect Detection in Open-Ended Improvisational Text. In Proceedings of the Workshop on Sentiment and Subjectivity in Text, Sydney, Australia, July 2006. Association for Computational Linguistics, 47-54.

[18] J.W. Pennebaker. Writing about emotional experiences as a therapeutic process. American Psychological Society, 8(3), 1997, 162–166.

[19] R.W. Picard. Affective Computing. MIT Press, Cambridge, MA, USA, 1997.