

Multimodal Backchannel Generation for Conversational Agents

Dirk Heylen*

Human Media Interaction - University of Twente
PO BOX 217, 7500 AE Enschede
heylen@ewi.utwente.nl

Abstract

Listeners in face-to-face interactions are not only attending to the communicative signals being emitted by the speakers, but are sending out signals themselves in the various modalities that are available to them: facial expressions, gestures, head movements and speech. These communicative signals, operating in the so-called back-channel, mostly function as feedback on the actions of the speaker; providing information on the reception of the signals; propelling the interaction forward, marking understanding, or providing insight into the attitudes and emotions that the speech gives rise to.

In order to be able to generate appropriate behaviours for a conversational agent in response to the speech of a human interlocutor we need a better understanding of the kinds of behaviours displayed, their timing, determinants, and their effects. A major challenge in generating responsive behaviours, however, is real-time interpretation, as responses in the back-channel are generally very fast. The solution to this problem has been to rely on surface level cues. We discuss on-going work on a sensitive artificial listening agent that tries to accomplish this attentive listening behaviour.

Keywords: listener responses, backchannels, head movements

1 INTRODUCTION

Schegloff (1982)) qualifies face-to-face interaction as an interactional achievement. Communication is constituted by the collaborative action of multiple actors. Gumpertz (1982) states this as follows.

Communication is a social activity requiring the coordinated efforts of two or more individuals. Mere talk to produce sentences, no matter how well formed or elegant the outcome, does not by itself constitute communication. Only when a move has elicited a response can we say communication is taking place.

Responses to talk by recipients can take many forms. They can take the form of a subsequent move in the next turn, but most of the behaviors of the non-talking participant in the conversation displayed during the turn of the speaker can count as some kind of “response”, providing the speaker with feedback on perception, attention, understanding and the way in which the message is received in general: the change in the beliefs, attitudes and affective state of the recipient. These cues and signals enable the synchronization of the communicative actions (for instance turn-taking), grounding and the building of rapport.

Most of the work on the generation of communicative behaviors of embodied conversational agents has been concerned with generating the appropriate non-verbal behaviors that accompany the speech of the embodied agent: the brows, the gestures, or the lip movements. The generation

*The author would like to thank Ruben Kooijman and Boris Reuderink for their work on this project. This research is part of a wider project on listening behaviours for conversational agents in the context of the Humaine Network of Excellence in collaboration mainly with Catherine Pelachaud and Elisabetta Bevacqua (Paris).

of the verbal and non-verbal behaviors to display during the production of speech by another actor, that is the behavior of a listening agent, has received less attention. A major reason for this neglect is the inability of the interpretation modules to construct representations of meaning incrementally and in real-time, that is contingent with the production of the speech of the interlocutor. As many conversational analysts and other researchers of face-to-face interaction have shown, the behaviors displayed by auditors is an essential determinant of the way in which conversations proceed. By showing displays of attention, interest, understanding, compassion, or the reverse, the auditor/listener, determines to an important extent the flow of conversation, providing feedback on several levels.

Besides the fact that most work on embodied conversational agents has focused on speaking behaviors, it also appears that not all expressive behaviors have received the same amount of attention. Language, facial expressions, gestures and gaze are the main kinds of expressive behaviors that have been studied so far. Posture and head movements form another group of nonverbal behaviours, that are very informative about the intentions, attitudes, emotions and the mental state of interlocutors, in particular, “auditors”, but these have been less widely studied.

In our current work on the Sensitive Artificial Listener, the generation of the behaviours that an agent should display while listening are very important. In our first studies we are looking at head movements and gaze in particular. In this paper we describe the general contours of the project, the way we approach the subject and illustrate this with describing the set-up and results of a pilot experiment.

2 SENSITIVE ARTIFICIAL LISTENER

In the Sensitive Artificial Listening Agent project, we are attempting to build semi-autonomous embodied chat-bots as part of the Sensitive Artificial Listener software. This software, developed in collaboration with Queens University, Belfast (see <http://www.emotion-net.research/>), is used to elicit emotions and accompanying behaviours that occur in conversations. In the original system, a person is sitting in front of a camera and hears the voice of one of the “characters”. The utterances by the characters are selected by an operator who can choose from a collection of pre-recorded phrases. They are indexed by the character they belong to, a pair of emotion dimension labels (positive/negative and active/passive) and by content category. They consist of general moves such as greetings, questions that prompt the persons interacting with SAL to continue speaking, and all kinds of reactions to what the persons are saying. The operator chooses the particular utterances in accordance with the stage of the dialogue and the emotional state of the person. Each character has a different personality expressed through what they are saying and the way they say it and will try to bring the person in a particular emotional state by their comments; cheerful or gloomy, for instance. In the Agent version that we are developing, the voices are replaced by talking heads and the behaviours are partly decided upon automatically.

We are working on the following items.

1. Designing faces and animations that fit the different personalities of the characters
2. Deciding on animations of the nonverbal behaviours that the characters should display when uttering the canned phrases
3. Studing and implementing the behaviours that should be displayed while listening
4. Building a system with some perception and understanding capabilities
5. Building a system that can decide semi-autonomously on which behaviours to display

Our work on building a completely autonomous responsive listening agent is proceeding by making small steps at the time. A detailed data-analysis is needed as the variations and functions of backchannels are highly varied. The real-time detection of features to which the agent can respond is in need of a better understanding of the function of those features in different contexts. Besides the need to know whether a backchannel has to be generated, it is important as well which

kind of backchannel is called for to what effect. Each of these issues can be further investigated in various ways. Besides data analysis, we are also relying on perception experiments in which people have to rate the behaviours of embodied agents in various ways. One of the first experiments that we have carried out in this respect will be discussed in Section 5.

Before we can experiment with fully autonomous agents, some other versions are being tried out. We have started with the off-line simulation of listening behaviours in which we generate behaviours for an agent and show these in combination with pre-recorded fragments of humans to make it appear as if the human interlocutor was talking to the agent. The behaviours are either generated by hand or by rule. In the latter case we are experimenting with probabilistic models and rule based systems. The road to full autonomy takes the following steps:

1. Off-line simulation of listening behaviours
2. Wizard of Oz experiments
3. Experiments with semi-autonomous agents

In this paper we will give an overview of the objectives and the way we approach the project. We discuss, the collection and analysis of data, the implementation of head trackers and modules that drive the behaviour of the agent on the basis of rules and regularities found during data collection and some initial experiments on the evaluation of some model implementations. We start by providing some background on the phenomena under investigation.

3 BACKGROUND

Several research traditions have studied the behaviours that listeners display in conversations. Back-channels, or similar phenomena with a different name such as response tokens, have been studied in the conversational analysis literature, for instance, with the purpose of understanding what role the various contributions of all of the participants play in shaping the conversation. Most relevant in this respect are papers such as Schegloff (1982), Schegloff (1996), Heritage (1984) but there are many others. The literature on turn-taking, both from the CA and other perspectives, also provides useful notes on the behaviours of participants that assume the primary speaker role and the auditors. In the series of papers by Duncan and co-authors¹, for instance, auditor back-channel signal are one of three classes of signals, besides speaker within-turn and speaker continuation signals, that serve to mark units of interaction during speaking turns.

An important issue that comes up with the study of back-channels is the definition of such terms as *speaker*, *hearer* and synonyms. A general assumption behind the concept of back-channel is that all the participants in a face-to-face conversation are both producers and recipients of communicative signals, but that there are different levels on which this occurs. Communicative signals on the primary track, to use the term by Clark (1996), are by the participants that have the floor and the secondary track, ‘in the back’, is constituted by the feedback on the behaviours in the primary track. As Yngve (1970) points out there may be cases of iteration where speakers provide feedback on the back-channels of listeners. To make the definitions of back-channel more precise, one would therefore need a framework that describes the various roles participants take in interaction. We build on the work of Goffman (1981), Levinson (1988), Schegloff (1996), Clark (1992) as a starting point for our theoretical model. However, in this paper, we will be using the terms speaker and hearer without further concerns about the tricky issues that surround them.

Several studies of nonverbal behaviours have paid attention to the behaviours displayed by listeners. One kind of phenomenon that has received some attention is the way in which behaviours of participants are synchronized and in particular how body movements of listeners are coordinated with the verbal utterances of the speaker. Hadar et al. (1985) showed that about a quarter of the head movements by listeners are in sync with the speaker’s speech. Interactional synchrony in this sense has been studied, amongst others by Kendon (1970), Schefflen (1964), Condon and Ogston (1967). Mirroring is a particular type that has often been commented upon. Schefflen suggests

¹See Duncan (1972), Duncan (1973), Duncan (1974), Duncan (1976), Duncan and Niederehe (1974),.

that this often reflects a shared viewpoint. Also Kendon (1970) hypothesized that the level to which behaviours are synchronized may signal the degree of understanding, agreement or support. These kinds of phenomena show that the behaviours of listeners arise not only from ‘structural concerns’ (e.g. turn-taking signals) but also from ‘ritual concerns’. We take these terms from Goffman (1981) who points out that it is sheer impossible to assign to behaviours a function of only one of these types of concerns (see also Bernieri (1999)).

Besides these synchrony behaviours, listeners display various other nonverbal behaviours as feedback. Chovil (1991), looking in particular at facial expressions, classifies these behaviours in a small set of semantic categories of listener comment displays. These are, besides displays for agreement:

- Back-channel: Displays that were produced by listeners while the speaker was talking or at the end of the speaker’s turn. They take the form of brow raises, mouth corners turned down, eyes closed, lips pressed. In Chovil’s corpus the displays could be accompanied by typical back-channel vocalizations such as “uhuh”, “mhhh”, “yeah”, etc.
- Personal reaction displays: A reaction in response to what the speaker had said rather than just acknowledging the content.
- Motor mimicry displays: displays that might occur in the actual situation that the speaker is talking about (e.g. wincing after hitting ones’ thumb with a hammer, eyes widened and an open mouth in response to a frightening situation). These are interpreted as messages that indicated a sincere appreciation of the situation being described.

Hadar and colleagues have looked in particular at head movements of listeners and how differences in form correspond to functional differences. Several authors writing on head movements have remarked that the precise form of the movements may be informative about the different functions they serve. Kendon (2003), writing on head shakes for instance, states:

Head shakes vary in terms of the amplitude of the head rotations employed, in the number of rotations and in the speed with which they are performed. There is no doubt that these variations in performance intersect with and modify the meaning of the gesture. [...] In this paper, however, necessarily preliminary as it is in many ways, we have made no attempt to subdivide the head shake according to how it may be varied in its performance [...].

In Hadar et al. (1985), such an attempt has been made for a limited number of head movements. They show that kinematic properties such as amplitude, frequency and cyclicity distinguish between signals of ‘yes and no’ (symmetrical, cyclic movements), anticipated claims for speaking (linear, wide movements), synchrony movements occurring in phase with stressed syllables in the other’s speech (narrow, linear) and movements during pauses (wide, linear). As we shall illustrate below, we are performing work along similar lines to reach a better understanding of how the variations in form give rise to variations in meaning.

In the discussion so far, we have mentioned several functions that are served by the behaviours of listeners. They provide feedback to the speaker, acknowledging reception of the signal, possibly its understanding or some kind of comment expressing a particular attitude towards what is being expressed. From its nature as a kind of joint communicative action, conversations require that participants come to react to each other’s actions to ground the actions and provide closure. Feedback is an important part of establishing grounding in the interactional achievement of having a conversation. The variety of functions that feedback serves is partly explained by the various levels on which grounding needs to take place: i.e. levels at which the participants need to have a mutual understanding of each other’s intentions. Clark (1996) suggests that grounding needs to occur on at least four levels with each step a kind of joint action.

1. Joint[A executes behavior t for B to perceive; B attends perceptually to behavior t from A]

2. Joint[A presents signal *s* to B, B identifies signal *s* from A]
3. Joint[A signals to B that *p*, B recognizes that A means that *p*]
4. Joint[A proposes a joint project to B, B takes up the joint project]

As speakers make their utterances, they are usually also monitoring the interlocutors behaviours to find signs of their participatory involvedness on all of these levels.

1. A monitors B for signs of perception activity / B's behaviour provides cues of perception activity
2. A monitors B for signs that B has identified the signal / B indicates that he has identified the signal...

The utterance of speakers and the accompanying behaviours will often be designed to invoke behaviours of interlocutors to ensure this. A typical case of this behaviour is analysed by Goodwin (1981), consisting of hesitations and repetitions of speakers at the beginning of their utterance to evoke gaze behaviours in interlocutors.

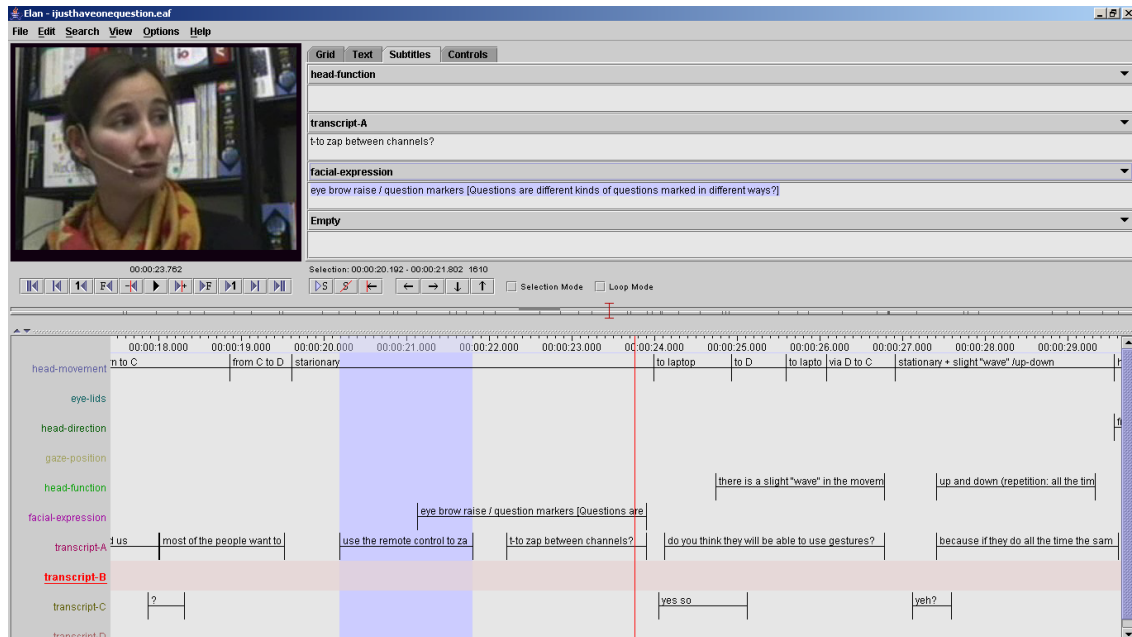
In a similar vein, Allwood et al. (1993) distinguishes four basic communicative functions on which the speaker may require feedback:

1. Contact: is the interlocutor willing and able to continue the interaction
2. Perception: is the interlocutor willing and able to perceive the message
3. Understanding: is the interlocutor willing and able to understand the message
4. Attitude: is the interlocutor willing and able to react and respond to the message (specifically accepting or rejecting it).

The various feedback behaviours are thus not only varied in their form but also in their function. The timing of them is of the essence, as several forms occur in parallel with the utterance of the speaker (synchronous interaction). This poses a big challenge for constructing embodied agents that need to react instantly on the speech produced by speakers. Most of the work on reactive agents has based the reactions on superficial cues that are easy to detect. The listening agent developed at ICT (Maatman et al. (2005) and Gratch et al. (1996)) produces feedback on the basis of head movements of the speaker and a few acoustic features (Ward and Tsukahara (2000)). Similar kind of input will be used in the SAL system.

4 DATA

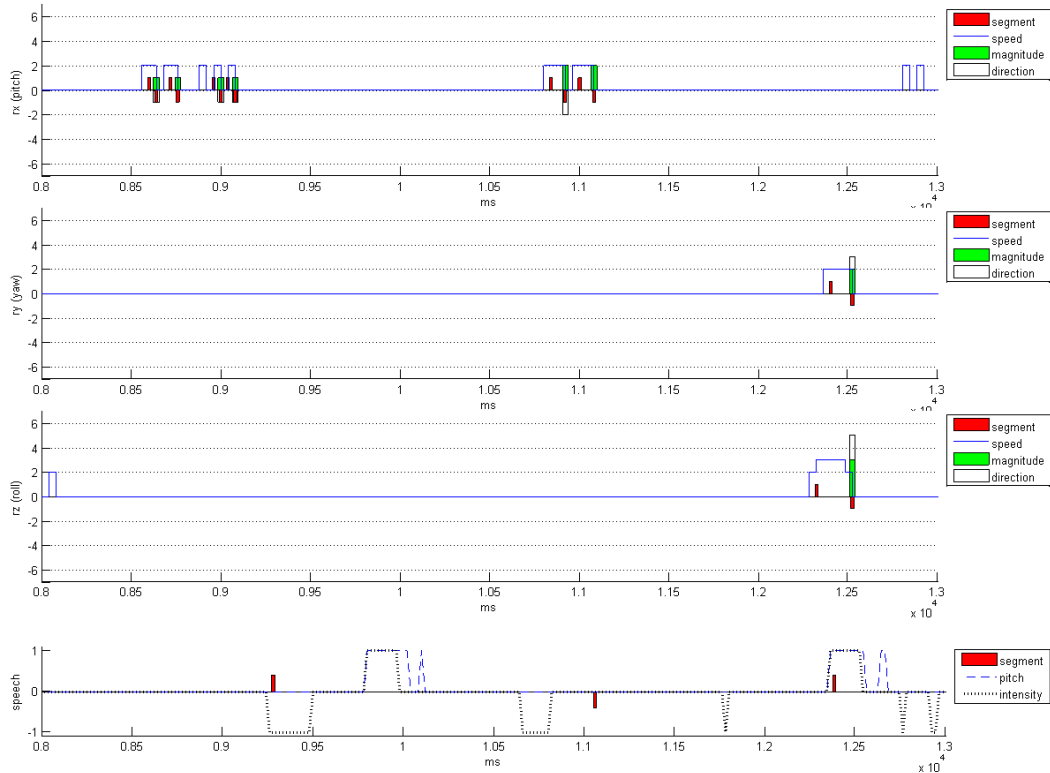
Although there is a fairly rich literature on listener behaviours that be can used to define and implement behaviours of a listening agent, we have also found it useful to look at some of the data that we have available and to collect some new data to learn more about the form, the distribution and the functions of feedback behaviours. Both the corpus collected in the AMI project (<http://www.amiproject.org/>) and the collection of interactions with the Wizard of Oz version of SAL (see <http://www.emotion-research.net/>) are being used in this respect.



The screenshot shows some of the annotations that have been made on the AMI data. In this case, the annotations covered head movements, gaze position, facial expressions, a characterisation of the function of the head movements and the transcripts. The characterisation of the functions was based on the determinants listed in Heylen (2006), which relied in their turn on some of the literature cited above, in particular Chovil (1991) and McClave (2000). The following lists provides the major functions that were used for head movements and back-channels.

1. Cognitive determinants: thinking, remembering, hesitation, correction...
2. Interaction management: turn-regulation...
3. Discourse and information functions: deixis, rhetorical functions (including the narrative functions mentioned in McClave (2000)), question marker, emphasis...
4. Affect and Attitude: epistemic markers (belief, scepticism), surprise, etcetera.

Besides such hand-made annotations, we are using automatic procedures to extract features from the speech and movements (head movements) to be able to list in more detail the distribution of verbal and nonverbal backchannels. The following picture provides a graph representing the head movements of a particular fragment in the SAL data. It shows the various movements of the head on the x , y and z axes in a small fragment of time with an indication of the duration, the velocity and the amplitude.



For this particular fragment the speech runs as “[lip smack]...euhm...well”. Just before the lip smack the head turns slightly upwards as can be read from the first markers on the pitch line. One set of annotations produced for the first half of this fragment is as follows. The form table gives a brief specification of the various behaviours.

Form		
Head	tilted left	movement up
Eyes	right corners looking right/up	small shakes
Speech	lip smack + ‘euhm’	
Face	anxious	

The “Intention” values give an indication of the degree to which the head movement was judged to be deliberate or automatic. In many cases both will apply to some extent.

Function	
Intention	deliberate / automatic
Cognition	thinking
Attitude	uncertainty
Social convention	
Emotion	neutral / anxious
Information	
Interaction	stall - responsive - turn-initialisation
Discourse	

With an analysis of some hundred fragments like these we hope to get a better picture of the association between head movements properties and their functions, where the head movements can be quite tiny and the functions more refined than in most analyses currently available in the literature.

5 IMPRESSION MANAGEMENT

The personality of the four characters used in SAL comes out mainly in the kinds of things they say. The character Poppy, for instance, is cheerful and optimistic and will try to cheer up the interlocutors when they are in a negative state and be happy for them when they are in a positive state. Obadiah, on the other hand, is gloomy and passive and will say things with the opposite effect. The voices, created by (amateur) actors are also quite expressive. The choice of talking head should match this, as should their nonverbal behaviors. A good deal of this work might be left to an animator who is skilled in designing and animating characters. An important question with respect to evaluation in this case is what impression the characters generate. So far, we haven't put animators to work on creating particular animations, but we have carried out some experiments varying the gaze behaviour and the head movements of a character and having participants in the experiment judge these behaviours. The basis of our study were the experiments carried out earlier by Fukayama et al. (2002) for gaze and Mignault and Chauduri (2003) for head movements, complemented by many other works on gaze² and head movements.

Similar to the study in Fukayama et al. (2002), a probabilistic model of the behaviours was implemented that determined the gaze of the RUTH talking head (Reuderink (2006)). We limited the variation in movements by fixing the head tilt. Combining some of the outcomes of the two studies we tried to model the behaviours for a happy, friendly, unobtrusive, extrovert agent (A) and for an unhappy, unfriendly and rather dominant agent (B). The combination of two different behaviours together with the fact that different impression variables were attempted to be modeled, raised some interesting issues. The head tilt for A was set to +10° (raised). According to the study by Mignault and Chauduri a head tilted upwards can be perceived as more dominant which is not exactly what we wanted, but it also has an effect on the impression of happiness, which is what we aimed for.

For A, the amount of gaze was set at 75% and a short mean gaze duration which we hoped would create the impression of engagement, friendliness and liking. The mean gaze duration for A was set at 500ms as in the Fukayama et al. (2002) experiment short gaze durations were associated with friendly characters. Gaze aversion for A was downwards, which is associated with submissive rather than dominant personalities.

For B the head tilt was 0°, which may lead, according to Mignault and Chauduri to low scores on happiness. With respect to gaze, we kept the amount of gaze at 75% but changed the mean gaze duration to 2000ms, which results in long periods of gaze, which we hoped would create a rather dominant, unfriendly impression. Gaze aversion for B was to the right.

The settings for both characters are summarised in the following table.

A	
Personality:	happy, friendly, unobtrusive, extrovert
Head tilt	10°
Amount of gaze	75%
Mean gaze duration	500ms
Gaze aversion	down
B	
Personality	unhappy, unfriendly, dominant
Head tilt	010°
Amount of gaze	75%
Mean gaze duration	2000ms
Gaze aversion	to the right

For both A and B we made two animations, one with smaller (A1, B1) and one with larger movements (A2, B2). Each animation we showed in the experiment lasted 40 seconds. We showed the four movies to 21 participants (all students at the University of Twente), divided into three groups for each of which the movies were presented in a different order (A1 B1 A2 B2; B1 A1 B2 A2; A2 B2 A1 B1). The difference in ordering did not show an effect on the result. To rate the

²Argyle and Cook (1976), Cassell and Thórisson (1999), Kendon (1967), and our own work Heylen et al. (2005).

impressions we had the participants fill out a questionnaire for each movie consisting of a rating on a 7-point scale for 39 dutch adjective pairs with the following translations.

extrovert - introvert, stiff - smooth, static - dynamic, agitated - calm, closed - open, tense - relaxed, sensitive - insensitive, polite - rude, suspicious - trusting, interested - uninterested, credible - incredible, sympathetic - unsympathetic, self-confident - uncertain, cold - warm, weak - strong, selfish - compassionate, formal - informal, winner - loser, thoughtful - reckless, unattractive - attractive, organized - disorganized, unfriendly - friendly, reliable - unreliable, refined - rude, involved - distant, flexible - linear, amusing - boring, attentive - absent, lazy - industrious, inactive - lively, optimistic - pessimistic, happy - depressed, loving - unloving, empathetic - unempathetic, dominant - submissive, aggressive - timid, stubborn - willing, enterprising - passive, realistic - artificial.

Factor analysis reduced the number of dimensions to the following 8 factors.

1. absence, unfriendliness, rudeness
2. submissive, weak, sensitive
3. warm, energetic
4. dull, drained
5. unreliable
6. rigid, static, linear
7. informal
8. attractive

When A and B are compared on these factors, we found that A scores higher on Factors 2 (submissiveness) and 5 (unreliable). B scores significantly higher on Factors 1 (absence, unfriendliness...) and 4 (dullness). There are also some differences between the small and large movements. Large movements create a more unfriendly impression (Factor 1). Small movements score significantly higher on Factor 2 and 8, that is, the smaller movement animations are considered more submissive, but are also more attractive.

All in all we were thus able to generate behaviours that resulted in several impression values that we had designed the agents for. However, this way of designing behaviours by combining functions associated with behaviours as mentioned in the literature poses many interesting problems. As we mentioned before, if one tries to achieve an effect on various impression variables, a particular behaviour may be very well suited for yielding good scores on variable x but mediocre scores on variable y . Also the combination of two behaviours may yield a combined effect that is different from what might be expected from the descriptions of the behaviours independently considered. Adding yet more behaviours - such as speech, for instance - may change the results again.

Furthermore, the precise set of impression categories that one is aiming for may not correspond exactly to the categories used in the studies in the literature. Expressions are ambiguous and fit more than one category.

Context plays an important role as well. The literature reports on functions and impressions, derived from data in a particular context which can be very different from the context of use that we are considering. For the actual design of the style of behaviours for the different personalities we will rely on another methodology. Using actors to collect a corpus of behaviours would be a good option in this case. It will be interesting then to see what gaze behaviours and head movements they display and see how this compares to the literature and the results of this more analytic approach exemplified by the current study. Despite the many obstacles this kind of approach poses, it does produce some useful results as well.

6 CONCLUSION

The SAL context provides us with an interesting set-up to experiment with designing and implementing conversational agents. The fact that the system is in part a wizard of oz set-up makes it possible to have the operator make decisions that need high-level interpretation. Because the agents are primarily designed to “listen” it is important to look in more detail at these less well-studied behaviours.

In this paper, we have presented the way we are proceeding to tackle this project. We have illustrated the process of data collection, analysis and one of the ways in which we are evaluating the generation of multimodal listening behaviours. We believe that a project such as this one can only succeed if many different sources and methodologies are brought into play as the above will have shown.

REFERENCES

- Allwood, J., Nivre, J., and Ahlsén, E. (1993). On the semantics and pragmatics of linguistic feedback. *Semantics*, 9(1).
- Argyle, M. and Cook, M. (1976). *Gaze and Mutual gaze*. Cambridge University Press.
- Bernieri, J. (1999). The importance of nonverbal cues in judging rapport. *Journal of Nonverbal behavior*, 23(4):253–269.
- Cassell, J. and Thórisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4-5):519–538.
- Chovil, N. (1991). Social determinants of facial displays. *Journal of Nonverbal Behavior*, 15(3):141–154.
- Clark, H. (1996). *Using Language*. Cambridge University Press, Cambridge.
- Clark, H. H. (1992). *Arenas of Language Use*. The University of Chicago Press, Chicago, London.
- Condon, W. and Ogston, W. (1967). A segmentation of behavior. *Journal of Psychiatry*, 5:221–235.
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23:283–92.
- Duncan, S. (1973). Towards a grammar for dyadic conversations. *Semiotica*, pages 29–46.
- Duncan, S. (1974). On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, 2:161–180.
- Duncan, S. (1976). Language, paralanguage, and body motion in the structure of conversations. In McCormack, W. and Wurm, S., editors, *Language and Man. Anthropological Issues*, pages 239–268. Mouton, The Hague.
- Duncan, S. and Niederehe, G. (1974). On signalling that its your turn to speak. *Journal of Experimental Social Psychology*, 10:234–47.
- Fukayama, A., Ohno, T., Mukawa, N., Sawaki, M., and Hagita, N. (2002). Messages embedded in gaze of interface agents - impression management with agent’s gaze. In *Proceedings of CHI 2002*, pages 41–48. ACM.
- Goffman, E. (1981). *Forms of Talk*. Oxford University Press, Oxford.
- Goodwin, C. (1981). *Conversational Organization: Interaction between Speakers and Hearers*. Academic Press, New York.

- Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R., and Morency, L.-P. (1996). Virtual rapport. In Gratch, J., Young, M., Aylett, R., Ballin, D., and Olivier, P., editors, *Intelligent Virtual Agents*, pages 14–27.
- Gumpertz, J. (1982). *Discourse Strategies*. Cambridge University Press, Cambridge.
- Hadar, U., Steiner, T., and Rose, C. F. (1985). Head movement during listening turns in conversation. *Journal of Nonverbal Behavior*, 9(4):214–228.
- Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In Atkinson, J. M. and Heritage, J., editors, *Structures of Social Action*. Cambridge University Press, Cambridge.
- Heylen, D. (2006). Head gestures, gaze and the principles of conversational structure. *International Journal of Humanoid Robotics*, 3(3):241–267.
- Heylen, D., van Es, I., van Dijk, B., and Nijholt, A. (2005). Experimenting with the gaze of a conversational agent. In van Kuppevelt, J., Dybkjaer, L., and Bernsen, N. O., editors, *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*. Kluwer Academic Publishers.
- Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, 26:22–63.
- Kendon, A. (1970). Movement coordination in social interaction: some examples described. *Acta Psychologica*, 32:100–125.
- Kendon, A. (2003). Some uses of head shake. *Gesture*, 2:147–182.
- Levinson, S. C. (1988). Putting linguistics on a proper footing: explorations in goffman’s concept of participation. In Drew, P. and Wootton, A., editors, *Erving Goffman. Exploring the Interaction Order*, pages 161–227. Polity Press, Cambridge.
- Maatman, R., Gratch, J., and Marsella, S. (2005). Natural behavior of a listening agent. In *5th International Conference on Interactive Virtual Agents*. Kos, Greece.
- McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, 32:855–878.
- Mignault, A. and Chaudhuri, A. (2003). The many faces of a neutral face: Head tilt and perception of dominance and emotion. *Journal of Nonverbal Behavior*, 27(2):111–132.
- Reuderink, B. (2006). The influence of gaze and head tilt on the impression of listening agents.
- Schefflen, A. (1964). The significance of posture in communication systems. *Psychiatry*, 27:316–331.
- Schegloff, E. A. (1982). Discourse as interactional achievement: Some uses of ”uh huh” and other things that come between sentences. In Tannen, D., editor, *Analyzing discourse, text, and talk*, pages 71–93. Georgetown University Press, Washington, DC.
- Schegloff, E. A. (1996). Issues of relevance for discourse analysis: Contingency in action, interaction and co-participant context. In Hovy, E. H. and Scott, D. R., editors, *Computational and Conversational Discourse. Burning issues - An interdisciplinary account*, pages 3–35. Springer.
- Ward, N. and Tsukahara, W. (2000). Prosodic features which cue back-channel responses in english and japanes. *Journal of Pragmatics*, 23:1177–1207.
- Yngve, V. (1970). On getting a word in edgewise. In *Papers from the sixth regional meeting of the Chicago Linguistic Society*, pages 567–77, Chicago: Chicago Linguistic Society.