Nonverbal behaviours improving a simulation of small group discussion

Nonverbal Behaviours Improving a Simulation of Small Group Discussion

Emiliano Padilha[†] **Jean Carletta**
Institute for Commmunicating and Collaborative Systems
Cognitive Science Department, University of Edinburgh
2 Buccleuch Place, EH8 9LW Edinburgh, UK
[emilianp,jeanc]@cogsci.ed.ac.uk

Abstract

This paper reports on the development of a multi-agent simulation of small group discussion that focusses on the interaction and the coordination of turn-taking. We describe the addition of nonverbal behaviours, such as gaze, gestures, posture shifts and head and facial expression, to the model; how the agents in the simulation take these behaviours into account in their decisions to speak, to stop, or to give feedback. The simulation is to be evaluated comparing its statistical profile against the statistics generated by a simpler, base model, one without the nonverbal behaviours, to show that it better approximates the statistics of a real group discussion. The properties to be assessed include mean transition intervals, turn lengths, relation of gaze to speaking order, frequency of simultaneous starts, and of feedback.

1 Introduction

Much is known about the organization of talk in conversation, such as its *interactive patterns* and the *coordination of turn-taking*. Research in sociology and psycholinguistics have put forth empirical data and models that explain part of this organization. However, there has been little or no attempt to apply or implement them to a simulated system, either to further investigate and assess such data and models, or simply to show how to make agents that better emulate the behaviours, maybe gaining insight that could help conversational agents or mediated communication.

Simulation is a good method for investigating conversational structure. It allows building hypothesis testing and model evaluation of certain aspects without interference from all the rest of the complexity. Nonetheless, the best known simulation of multi-party interaction (Stasser & Taylor, 1991) only generated patterns of the *order of speaking turns*, based on notions of stable differences in speaking rates amongst participants and transitory tendencies to speak depending on when one last spoke.

More recently, another simulation (Padilha & Carletta, 2002) used a simple multi-agent architecture to emulate the coordination of turn-taking and feedback in *small group discussions*; that is, 3 to 7 participants in informal face-to-face conversations. Coordination was effected by the interaction of the multiple individual behaviours of the various agents, such as starting to talk, talking in turn, finishing an utterance and giving feedback. Agent decisions were also based

---

[†] Supported by a studentship from CNPq, Brazil's Federal Research Council.

on tendencies to speak, and to do other behaviours in a range of modalities including nonverbal ones.

Nonverbal behaviours have been receiving growing interest in the development of conversational agents for human interaction, virtual worlds, and mediated (all-human) multi-party communication (Vertegaal et al., 2000; Rickel & Johnson, 2000; Cassell et al., 2001). They have important roles in face-to-face communication. However, to our knowledge there seems to have been no attempt to recreate their effects in the organization of talk of multi-party conversation.

This paper describes the ongoing work in the development of a multi-agent simulation of small group discussion that incorporates nonverbal behaviours. In particular, it describes how they variously affect the participants' tendencies to speak, to continue talking or to stop, and to give feedback, based on the empirical literature. Of interest is the methodology to be employed in evaluating the simulation, reported in section 6. The basic idea is to show that the simulation containing nonverbal behaviours, described in section 5, yields better approximated patterns verifiable in real data than a simpler base model, without them, which is presented in section 4. These models both extend the work presented in Padilha & Carletta (2002), whose framework outlined in section 3 provides the operational background. But we first lay down in the next section some of the observations and hypotheses from the literature on nonverbal behaviours that our modelling will be based on.

2 Nonverbal behaviours

Nonverbal behaviours play important *regulatory* roles in face-to-face interaction. Apart from complementing the interchanges with emotional content and reflecting interpersonal relationships (Scheflen, 1972), they help coordinate turn-taking and the organization of talk. One of their functions is to show *turn-yielding*, and it has been found that the display of one or more such cues significantly reduces the possibility of simultaneous claimings to the turn (Duncan, 1972), presumably by giving participants better guidance as to when to effect a speaker transition. Gaze, gesture and postural behaviour can show the participants' tendencies not only in yielding the turn, but in continuing to talk (*turn-holding*) or preparing to talk (*turn-requesting*). They thus help to run a more effective interaction without much need of *resources* from the speech signal (which busily carries the contents of the interchanges) to be devoted to it.

The speaker's gaze, in particular, though not essential, is a strong component in floor-apportionment (Beattie, 1981; Kalma, 1992). Speakers tend to look away in the early stages of formulating what to say, gazing back at an interlocutor when formulation is completed and the utterance is being finished so they can monitor uptake (Argyle & Cook, 1976). The person the speaker gazes at when finishing an utterance has an advantage in taking the next turn of talk simply because he can more readily recognise that the speaker is finishing (Steinzor, 1950; Lobb, 1982).

Evidence of gaze both as turn-yielding and turn-holding cues came first from Kendon (1967): he observed that utterances that terminated without speaker gaze more frequently had delayed listener responses. Based on this and other evidence, Beattie (1981) hypothesized that the elimination of gaze should increase speaker transition intervals, at least in situations where gaze levels are low, such as in conversation between strangers. Its importance has also been recognised in multi-party mediated systems: absence of gaze can decrease the turn-taking efficiency of the communications by 25% (Vertegaal et al., 2000).

Gestures and posture shifts, both speaker's and listeners', can also affect turn-taking decisions, although less clearly as with gaze (Beattie, 1981). Changes of posture from as little as a head tilt up to the whole body balance (Kendon, 1972), tend to occur when speakers initiate or finish utterances, or at various levels of discourse junctures (Scheflen, 1972; Cassell et al., 2001). Gestures show the speaker's continuing engagement in talk and can thus indicate turn-holding. Duncan (1972) observed that turn-taking attempts fell virtually to zero when the speaker gesticulated at phonemic junctures.

Listener responses such as head movements (nods) and facial expressions function as or complement backchannel feedback like "yeah", "mm", or "uh?". They help maintain the flow of interaction; without them, the speaker would sooner or later wonder whether the others are listening (Oreström, 1983). Listener responses are given mainly at phonemic junctures, especially by the participant the speaker gazes at (Kendon, 1967; Dittman & Llewellyn, 1968). They can also occur at hesitant phases of speech (filled pauses, etc), or when one has already understood what is being said and might be preparing to speak (Rosenfeld, 1977). With such early responses, the speaker is more likely to emit turn-holding signals like gaze aversion and gesticulation (Duncan, 1974).

Besides early listener responses, behaviours that may act as *turn-requesting* include posture shifts (eg. head forward), noddings and gesture beginnings, like raising a hand or finger (Wiemann & Knapp, 1975). They can delay the speaker in deciding to continue to talk at junctures when he gazes up, and also attract some of the others' attention, preventing them from deciding to speak as well.

3 The simulation

A simulation of small group discussion that contains nonverbal behaviour was presented in Padilha & Carletta (2002). In the current work, we are employing the same framework, and most of the agent behaviours and interactive patterns from that simulation, except for the changes in modelling and decision tendencies to be described in the next sections.

That simulation introduced a general framework that can be used for various kinds of conversational interaction at any level of representation: a simple multi-agent architecture with a *blackboard* where the agents read and write their behaviours. The simulation runs in a loop of *cycles* of a fixed, arbitrarily short length,[1] during which each agent in turn reads the behaviours of the previous cycle from the blackboard, decides what to do and then writes the appropriate behaviours back in a second blackboard. These are collected and read at the next cycle. The agents output multiple behaviours, one of each modality (speech, gaze, gesture, etc). The behaviours of each cycle are treated as simultaneous.

The simulation also had a general way of modelling the agents with respect to the decisions they make. A set of likelihood attributes for each agent was used to govern their various decisions stochastically. They only concerned interaction: whether an agent wants to speak, to give feedback, the confidence or insistence in talking simultaneously with others (that determined who would stop or continue), utterance lengths, and the frequency with which a speaker would continue to talk after each utterance (and thus extend the turn). This set of attributes characterized each agent with regard to their talkativeness, transparency, confidence, interactivity and verbosity in the discussion.

Since the focus is interaction and the coordination of turn-taking, the actual contents of utterances were abstracted away in order to concentrate solely on those aspects. Hence, the speech behaviours included: backchannel feedback of two types, "yeah" (representing any kind of positive continuer, such as "mm", "uh-huh", etc) and "uh?" (any negative feedback, such as "sorry?", that induced a response from the speaker); start of talk with timestamp information; talk in turn; utterance near-completion; and utterance completion.

The last one corresponds to arriving at a discourse or phonemic juncture, or in turn-taking terms, a *transition-relevance place* (Sacks et al., 1974), to be referred henceforth as TRP. It is the place where a speaker transition would be expected, but not necessary since the speaker can always continue to talk.[2] It involves at least grammatic (syntax and semantics) and prosodic (intonational contour) completion (Oreström, 1983). These are essential conditions, though usually more are expected for transitions to occur (cf. next section). The behaviour indicating utterance near-completion was then called "pre-TRP" and represented in the simulation these cues from the speech signal.

As a simulation of face-to-face discussion, behaviours also included a range of nonverbal modalities: gaze, gestures, posture shifts, head and facial expression. They were expressed just by indicating their presence or absence or, for gaze, whether one was looking away or at someone (and who). Nods and a puzzled expression were nonverbal versions of "yeah" and "uh?" respectively. The speaker could make gestures on and off throughout the turn up to the pre-TRP, while posture shifts could only occur at the start and end of each utterance. The speaker

[1] Currently defined at 500ms, or half a second, which is a good compromise with the requirements of the interaction and turn-taking. The only situation in which such coarse granularity proves to be a limitation is with simultaneous (or *quasi*-simultaneous) starts. They were resolved by adding a timestamp to the start of talk (timing within the cycle), and by assuming that the first to speak actually did so and the others just intended to (or "inbreathed"), but were cut short by the first.

[2] We are calling *utterance* what Sacks et al. (1974) named a *turn-construction unit* (TCU).

started gazing at the previous one but would soon (randomly) gaze away, gazing back only before the TRP.

In that simulation all these nonverbal behaviours were produced to emulate a general pattern of participant interaction. They were *not* used as input by the agents in their various decisions, so they did not affect the interaction, participation, turn length and the turn-taking coordination. In fact, that simulation was generating too many simultaneous starts: not only starting of talk (and feedback for that matter) could only occur at one position, at the cycle that was tagged TRP, but also there was just one behaviour affecting about as equally everybody's decisions to start: the pre-TRP, or utterance near-completion). In the real data we have, participants are much more careful (and skillful) in avoiding simultaneous starts while creating much more varied speaker transition patterns.

The purpose of the current work is therefore to update and expand the previous simulation by implementing the effect of nonverbal behaviours from the observations of the literature. We need to verify their effect by comparing the profile of the interaction generated by the simulation to the one produced by a base model, without the nonverbal effects. This model is described next.

4 The base model

This model is based on the general turn-taking systematics of Sacks et al. (1974). It described conversations as having a *one-at-a-time* speakership organization, with overlaps and silent gaps occurring mostly at speaker transitions, and with the first participant to start at a TRP getting the turn, or, if no one does, with the current speaker potentially continuing to talk. That is: participants take *turns* at talk during which, overwhelmingly, only one speaks. More-than-one (overlaps) and less-than-one (gaps) are common but brief.[3] Turns correspond to one or more utterances with the TRPs inbetween, at pauses in the speaker's talk. The first to speak at a TRP gets the attention of the others, starting then a new turn. The current speaker readily gives way to the new starter unless another participant (a later starter) reveals anything that could have higher priority, such as a problem of understanding (eg. "uh?").

Participants intending to speak use cues from the speech to *project* (roughly) when it is going to end so as to time their own starts: the grammatic and prosodic completions already mentioned. These are essential but often insufficient for a smooth speaker transition to occur, one not taken as interruption. Usually, one or more additional *turn-yielding* cues are used by participants in deciding to start: drop in pitch and/or loudness towards the end of the utterance, an unfilled pause after completion, the speaker gazing back up to an interlocutor, termination of gesture, and a shift of posture, such as leaning back (Duncan, 1972).[4]

---

[3] Although overlaps are clearly not near as common as gaps between transitions. For instance, Oreström (1983) found it to be as little as 17% in a corpus of dyadic conversation.

[4] For example, Oreström (1983) found in his corpus (though *without* any nonverbal behaviours) that 95% of all transitions had grammatic/prosodic completion, and two thirds of these (ie. 63%)

The nonverbal cues are considered in the next section, but the drop in pitch/loudness and the unfilled pause can be represented in this model. Take as an example the following extract from a real group discussion that shows the speaker finishing an utterance grammatically, prosodically, and by gradually reducing loudness.[5] After a brief interval, someone starts to speak:

```
1:...treat all the students fairly, really.
2:                                          You can't really...
```

This transition could be reproduced in the simulation as follows. Utterance contents are not considered, only talk and feedback. Talk is marked by cycles of the behaviour `talk`. When completing an utterance, or arriving at a phonemic juncture, a series of one to three pre-TRP cues are sent: `tlk2`, `tlk1`, `tlk0`. This varying length allows for more patterns of speaker transition overlap and feedback. The sequence acts as a countdown to the `TRP` that follows it. The `TRP` represents the unfilled pause after completion of the utterance. In case of a drop in pitch/loudness, the pre-TRPs become, respectively, `low2`, `low1`, `low0`. So the previous example would be represented as (bold font just for clarity):

```
Agt1: talk talk low2 low1 low0 TRP
Agt2:                              talk talk ...
```

The speaker may decide to continue talking immediately after `tlk0` (not sending out the `TRP`)[6] or continue after the `TRP`, if no one else has started then. Decisions to continue are taken according to one of the likelihood attributes of each agent called *verbosity*. In this next example, the speaker finishes an utterance and receives feedback; after a pause (the TRP), she continues:

```
1:...that he's gonna be, fine wasn't he?       And a second one...
2:                                     yeah
3:                                       yeah
```

This would be represented in the simulation as follows:

```
Agt1:... talk talk tlk2 tlk1 tlk0 TRP  talk talk...
Agt2:                             yeah
Agt3:                             yeah
```

---

also had additional cues: 40% had an unfilled pause, and 44% a reduction in pitch or loudness. He concluded that the more of these (and other) cues, the more likely the transition.

[5] Punctuation is intuitive here. The period represents a falling intonation, as would a question mark represent a raising-falling one. Commas are small prosodic discontinuities, and the space between the two utterances, a short interval. The smaller font conveys the lowering loudness.

[6] This possibility, something like "...`talk tlk1 tlk0 talk`..." represents discourse junctures with no pause, where speaker transitions are not normally expected (no TRP), but where listener responses may be. In the recordings that we have, more than half of them occur far away from the TRPs, at the middle of the utterances, seemingly at lower syntactic-semantic boundaries.

Agents decide to start talk based on the probability expressed by another attribute of each agent called *talkativeness*. They may start at the pre-TRPs (after the first one at least, thus overlapping the finishing speaker), at the `TRP` (when the speaker has just finished), or after it (after the pause). The TRP meant by Sacks et al. (1974) actually comprises all these cycles: a time, not a point, where transitions are expected; our tagging of just one cycle as "`TRP`" is indicative only.

`talk`	`tlk2`	`tlk1`	`tlk0`	`TRP`	1 cycle after	2 after
-	-	T/2-0.2	T/2-0.1	T/2	T/2+0.1	T

Table 1: modifications in the tests against Talkativeness (or Transparency).

The likelihood to talk at each of these times varies though, as it does for each agent. Starting while the speaker is still finishing must be less likley than after it. So, the actual decision is modified by the equation (T/2)-0.#, where T is the agent's talkativeness and # is the `tlk#` or `TRP` cue (with # = -1) just read from the previous cycle. Consider the likelihoods to be real numbers between 0 and 1. Table 1 shows the modifications from this equation in deciding to speak at each of the points of the TRP.[7] It is only two cycles after the `TRP` that the likelihood to start (provided no one has already) comes back to the normal T value. A drop in pitch/loudness (`low#` instead of `tlk#`) further adds +0.1 with no maximum (ie. they can go over T), making it more likely then for agents to start closer to the `TRP` with this additional turn-yielding cue.

T»	.1	.2	.3	.4	.5	.6	.7	.8	.9
at `tlk1`	0	0	0	0	.05	.1	.15	.2	.25
at `tlk0`	0	0	.05	.1	.15	.2	.25	.3	.35
at `TRP`	.05	.1	.15	.2	.25	.3	.35	.4	.45
1 cycle after	.1	.2	.25	.3	.35	.4	.45	.5	.55
2 after (normal)	.1	.2	.3	.4	.5	.6	.7	.8	.9

Table 2: participatory likelihoods at various points of a bare TRP.

Once agents fail in deciding to talk at any of these points, they also decide whether to give feedback according to another attribute, *transparency*, using the same modifications as above (with T now referring to this attribute). If deciding for feedback, the agent stops any further tests of talkativeness for this TRP. We will refer to these two decisions together (either talk or give feedback) later on as

[7] As table 1 shows, reading the `TRP` (the silent pause that is a turn-yielding cue) just adds evenly to the decisions, as if it was another "position" of the pre-TRPs. One might think it ought to have a more salient effect as Duncan (1972) and Oreström (1983) reported. In the simulation, the tendency of gazing back and shifting posture at the same time shall increase its influence.

participatory decisions. Table 2 gives the actual likelihoods of the modifications of table 1 for each of the points of the TRP (values go to a maximum of T).

Other decisions the agents take include whether to continue talking if simultaneously with others (as when more than one starts at the same time), the length of each utterance and of the pre-TRP sequence, whether to reduce pitch/loudness there, and whether the feedback to be given is positive (`yeah`) or negative (`uh?`). If someone starts talk while the speaker is finishing, the others test an *attentivity* attribute to decide whether they continue as normal or shift their attention to the new starter. Failing to shift attention immediately is the cause in the simulation of second (later) starters.

In this model, the agents are using only the speech signal to coordinate their interaction. It is supposed to be like talking on the telephone with various people on the other side. As shown in table 2, starting to speak immediately when one finishes (smoothly, at the `TRP`), then becomes half as likely as the normal likelihood for the agent. Even after a silent pause confirming the TRP, the likelihood is still small for everyone. Only further at the next cycle are the values back to normal. This way, the agents wait a little longer to confirm that the speaker stopped before starting (or doing feedback), increasing transition intervals as is predicted to happen in no-vision conditions.

5 Adding nonverbal behaviours

The base model makes agents extra-careful and less interactive at the TRPs by considerably reducing their probability of doing participatory decisions. Visual input, by providing more regulatory cues, further affects these decisions in this new model through additional, cumulative modifiers. But note that not all of these verbal and nonverbal cues we are arraying here are meant to occur, or actually occur, together at every speaker transition. Most commonly just one or two, or even none are given, which may probably help explain why transitions with some silent interval are much more common than smooth or overlapped transitions.

5.1 The speaker's gaze

Unless the utterance is very short, the speaker soon breaks the mutual gaze obtained when starting to speak. When finishing the utterance, the speaker gazes back at an interlocutor on any cycle up to the `TRP`, even prior to the pre-TRPs. The time of gazing back is decided according to the agent's *interactivity* attribute. The more interactive, the more likely for it to occur sooner. The interlocutor can be any one of three previous speakers, with 0.6, 0.3 and 0.1 probabilities for it to be the latest up to the oldest previous speaker, respectively (agents keep a record of this). Given the effect gaze has on the person being gazed at (below), this predisposition makes it account in part for the *dyadic pattern* of conversation, the ABA tendency (see for instance Stasser & Taylor, 1991).

The effect gaze has on the listeners' participatory decisions is the factor to be modelled here. It is one more turn-yielding cue that increases their likelihood of deciding to speak; although not essential, it is definitely a factor. Moreoever, the effect is greater (or more readily acted upon) on the person being gazed at than on the others. Let's make that the speaker gazing at someone else other than the agent adds +0.1 to all participatory decisions of the base model, with maximum to the T (the agent's talkativeness or transparency). Table 3 shows the probabilities in this case, allowing for reduced intervals in transitions, and more smooth ones.

T»	.1	.2	.3	.4	.5	.6	.7	.8	.9
at `tlk1`	0	0	.05	.1	.15	.2	.25	.3	.35
at `tlk0`	.05	.1	.15	.2	.25	.3	.35	.4	.45
at `TRP`	.1	.2	.25	.3	.35	.4	.45	.5	.55
1 cycle after	.1	.2	.3	.4	.45	.5	.55	.6	.65
2 after (normal)	.1	.2	.3	.4	.5	.6	.7	.8	.9

Table 3: likelihoods when the speaker gazes someone (+0.1, with maximum at T).

The agent being gazed at by the speaker when finishing an utterance, on the other hand, adds not +0.1, but +0.3 without any maximum. This makes participatory decisions to be even more likely at the `TRP` than the normal T (at least at low likelihoods), as shown in table 4, bearing on the strong effect of gaze.

T»	.1	.2	.3	.4	.5	.6	.7	.8	.9
at `tlk1`	.15	.2	.25	.3	.35	.4	.45	.5	.55
at `tlk0`	.25	.3	.35	.4	.45	.5	.55	.6	.65
at `TRP`	.35	.4	.45	.5	.55	.6	.65	.7	.75
1 cycle after	.45	.5	.55	.6	.65	.7	.75	.8	.85
2 after (normal)	.1	.2	.3	.4	.5	.6	.7	.8	.9

Table 4: likelihoods for the agent being gazed at (+0.3 with no maximum).

As described in the base model, the speaker may decide to continue talking after the `TRP`. If so, he might avert gaze depending on how strong the decision to continue was, *before* the completion of the utterance. With the speaker gazing away, the likelihoods remain as that of the base model, as if no visual input existed. This makes the agents less likely to decide to speak and give feedback when the speaker gazes away while finishing an utterance.

5.2 The speaker's gestures and posture shifts

Without some content information, we cannot coordinate gestures to coincide with specific parts of the discourse as conversational agents already do (Cassell et al., 2001). The speaker then just randomly gesticulates on and off while talking, according to the agent's *nonverbal* attribute, allowing different frequencies for each participant.

Gestures seem to convey either a turn-yielding or turn-holding inclination, although it is not as strong as gaze and loudness. While the speaker gesticulates, decisions to start talk are reduced by -0.1 over the other modifiers. Feedback, however, which indicates that one does not wish to speak, goes on as normal.

The speaker, once deciding to continue talk, may keep gesticulating at the pre-TRPs and at the TRP. "Gesticulating" at these times is supposed to represent the speaker not bringing back the arm or hand to a resting position when finishing talk. The decision to continue can be taken at any pre-TRP, so that the agent can go on transmitting the behaviour (and/or gaze away) after finishing the utterace. The likelihood of gesticulating is increased, once one has decided to continue, if the speaker receives early feedback or other turn-requesting behaviours at that time.

Participants may shift posture (or otherwise their body balance) occasionally in the conversation with no particular meaning, just to adjust their seating. But in the simulation, the speaker shifts posture only at the TRP (after finishing the utterance). If the speaker is not gesticulating, it increases the others' participatory decisions by +0.1, *except* for the agent gazed at (if at all). So with the added effect of gaze, a posture shift makes the others' likelihoods almost as good as that of this agent.

5.3 Listener behaviours

Depending on how strong the decision to start talk was, a listener may emit turn-requesting behaviours and only start at the next cycle instead. The choice and number of behaviours are decided according to the *nonverbal* attribute: the higher the value, the more likely that more than one, and more effective behaviours are made. Turn-requesting behaviours are a posture shift, a gesture (the beginning of a gesture), and nods (also representing head movements like those accompanying facial expressions: eg. head tilt with a puzzled expression).

The more behaviours emitted as precursors of the start of talk, the more likely the others will recognise them and *not* decide to start or to continue to talk next. The agents decide whether they perceived the behaviours through the *attentivity* attribute, shifting attention accordingly as explained in the base model. The speaker also tests for whether the behaviours are recognised as well, except if already gazing the requesting participant. If successful in grabbing attention, each behaviour should reduce participatory decisions, or the speaker decision to continue talk, by -0.1. Their effect is to reduce the probability of simultaneous starts by forewarning the others through the visual channel.

As described in the base model, to give verbal or nonverbal feedback is only considered once the agent has already decided not to talk. The two decisions are based on different attributes of each agent, allowing for their independent frequency variation; say, we can have a non-talkative participant that gives frequent feedback. But it also implies that the more talkative the agent is, the less feedback he tends to give relatively, which seems to be the general tendency in the discussions we have seen.

Feedback (termed *continuers* by many) also invites the speaker to continue to talk by indicating that that listener is not intending to do so ("go on, I understand you"). In a sense, it is the inverse of turn-requesting. Therefore, let's make in the simulation that the likelihood of the speaker to continue talking increases by +0.1 after receiving *only* feedback (no matter how much) *and no* turn-requesting behaviour. This makes situations where no feedback is given less likely to make speakers to continue talking as normal. It then begins to exert an influence in the organization of talk as well.

6 Evaluation

The modelling of this simulation of nonverbal behaviours is still being assessed, and requiring evaluation. We can only report here how we intend to do this, with the aid of a small corpus of audio- and video-taped group discussions.

We have a 5-person and an 8-person discussion, each totalling around 16 minutes. The 5-person discussion, for example, has over 80 speaker transitions (not counting simultaneous talk) plus over 50 clearly recognizable in-turn TRPs (TRPs *without* speaker transition). They were recorded in an experiment where the participants were told to discuss a hypothetical case of plagiarism. Except for occasional minor interferences by the microphones attached to record one channel per participant, the conversations seem to have flowed naturally enough, being representative, we think, of typical discussions between strangers or colleagues. The video was recorded from the ceiling and the participants wore hats with big arrows painted over them that clearly show head movement and direction. Unfortunately occasional uncertainty still exist where it is impossible to tell sometimes who they were gazing at, or if at all, since eye movement cannot be seen. Gestures and posture shifts, though, can be recognized easily.

The methodology for evaluation is similar to the one used by Stasser & Taylor (1991). We will first adjust the attributes of the agents so that a series of runs of the base model yields as good an approximation as possible to the statistics of the real data, the frequencies and mean lengths of the emergent events we are modelling. Then we run both the base model and the nonverbal behaviour model with the same attributes for a series of hundreds of simulations of the same length of the real data (say, 80 speaker transitions like the 5-person recorded discussion). With the statistics of the two sets of simulations calculated, we are able to compare both models and verify whether the second fits or approximates the real data better than the first, the base model. All other things being equal, the differences should be due to the effect of the implemented behaviours on the interactive decisions of the agents.

The properties to be analyzed for the evaluation may include some or all of these: mean speaker transition intervals and no-transition TRP intervals, relationship of the speaker transitions to the TRP total (frequency with which a

TRP actually resulted in a transition), turn lengths, the relation of gaze to the speaking order, frequency of simultaneous starts, overlaps and gapped transitions, and frequency and location of feedback.

7 Concluding remarks

We presented a simulation of the organization of talk and interaction in small group discussions that accounts for the effects of nonverbal behaviours to the participants' decisions in a stochastical way. The various cases do not always ever or never happen; they have different likelihoods to happen, depending on the context and the interchange of behaviours. The described model synthesizes some of the more relevant observations available on the subject. In this sense, it is intended to be a descriptive model.

This is a new research area. Available corpora of multi-party conversation such as group discussion is as yet practically inexistent. Moreover, not only quantitative analyses, but also empirical and descriptive research have yet to be carried up more extensively. Many of the evidence regarding the effect of nonverbal behaviours is still contradictory (Beattie, 1981). At such a preliminary stage, even the synthesis of the empirical data and evaluation of a first-pass model against some real data is a useful research aim.

However, the method of modelling a process using purely simulated means limits the representativeness of it. Since the simulation is necessarily simplified, ignoring many aspects of the process, arguably the "good" results cannot always be considered proof that the concepts and definitions employed to produce them are good. They emulate empirical data, but are the underlying means adequate or acceptable for the visible ends? The same reservations might certainly apply to other (maybe all) stochastic methods as well. In sum, the model may be descriptive, but is it explanatory? The reader is left to judge for himself.

The decisions to talk, to stop talk and to give feedback are *social* decisions after all, involving a range of other issues besides the pure structure of auditory and visual signals. They involve the *contents* of communication, of course, but more than that, the social relationships between participants (status, their roles, acquaintanceship, etc), their expectations (politeness, typical procedures, etc) and emotional states, among others. In allowing the abstraction of these whole levels of complexity by simulating through probabilities, we are able at least to begin outlining what the systematics that produces this complex process of conversational interaction may look like. The results could be used to inform, for example, better modelling of embodied conversational agents or mediated communication.

References

M. Argyle & M. Cook, 1976. *Gaze and mutual gaze*, Cambridge University Press, Cambridge, Mass., US.

G. Beattie, 1981. The regulation of speaker turns in face-to-face conversation: some implications for conversation in sound-only communication channels, in: *Semiotica, 34*, pp.55-70.

J. Cassell, Y. Nakano, T. W. Bickmore, C. Sidner & C. Rich, 2001. Non-Verbal Cues for Discourse Structure, in: *Proceedings of the 41st Annual Meeting of the ACL*, Toulouse, France, pp.106-115.

A. T. Dittman & L. G. Llewellyn, 1968. Relationship between vocalizations and head nods as listener responses, in: *Journal of personality and social psychology, 9*, pp.79-84.

S. Duncan, 1972. Some signals and rules for taking speaking turns in conversation, in: *Journal of Personality and Social Psychology, 23(2)*, pp.283-293.

S. Duncan, 1974. On the structure of speaker-auditor interaction during speaking turns, in: *Language in Society, 3(2)*, pp.161-180.

A. Kalma, 1992. Gazing in triads - a powerful signal in floor apportionment, in: *British Journal of Social Psychology, 31*, pp.21-39.

A. Kendon, 1967. Some functions of gaze-direction in social interaction, in: *Acta Psychologica, 26*, pp.22-63.

A. Kendon, 1972. Some relationships between body motion and speech, in: A. W. Siegman and B. Pope (eds.), *Studies in Dyadic Communication*, Pergamon Press, Elmsford-NY, US, pp.177-210.

M. O. H. Lobb, 1982. Seating arrangement as a predictor of small group interaction, in: *Journal of Advanced Nursing, 7*, pp.163-166.

B. Oreström, 1983. *Turn-taking in English Conversation*, LiberForlag Lund, Malmö, Sweden.

E. Padilha & J. Carletta, 2002. A simulation of small group discussion, in: *Proceedings of the EDILOG'02, Sixth workshop on the semantics and pragmatics of dialogue*, pp. 117-124, `http://www.ltg.ed.ac.uk/edilog/papers/117.pdf`

J. Rickel & W. Johnson, 2000. Task-Oriented Collaboration with Embodied Agents in Virtual Worlds, in: J. Cassell, J. Sullivan, S. Prevost et al. (eds.) *Virtual Worlds, in Embodied Conversational Agents*, MIT press, Boston.

H. M. Rosenfeld, 1977. Convervational control functions of nonverbal behavior, in: A. W. Siegman & S. Feldstein (eds.), *Nonverbal Behavior and Communication*, Lawrence Erlbaum, Hillsdale, US, pp.291-328.

H. Sacks, E. A. Schegloff & G. Jefferson, 1974. A Simplest Systematics for the Organization of Turn Taking for Conversation, in: *Language, 50(4)*, pp.696-735.

A. E. Scheflen, 1972. *Body language and the social order*, Prentice Hall, Englewood Cliffs, US.

G. Stasser & L. A. Taylor, 1991. Speaking Turns in Face-to-face Discussions, in: *Journal of Personality and Social Psychology, 60(5)*, pp.675-684.

B. Steinzor, 1950. The spatial factor in face to face discussion groups, in: *Journal of Abnormal and Social Psychology, 45*, pp.552-555.

R. Vertegaal, G. van der Veer & H. Vons, 2000. Effects of Gaze on Multiparty Mediated Communication, in: *Graphics Interface*, pp.95-102.

J. Wiemann & M. Knapp, 1975. Turn-taking in conversation, in: *Journal of Communication, 25*, pp.75-92.