Turn Management or Impression Management?

Mark ter Maat and Dirk Heylen

Human Media Interaction, University of Twente
P.O. Box 217, 7500 AE Enschede, the Netherlands
{maatm,heylen}@ewi.utwente.nl

Abstract. We look at how some basic choices in the management of turns influence the impression that people get from an agent. We look at scales concerning personality, emotion and interpersonal stance. We do this by a person perception study, or rather an agent perception study, using simulated conversations that systematically vary basic turn-taking strategies. We show how we can create different impressions of friendliness, rudeness, arousal and several other dimensions by varying the timing of the start of a turn with respect to the ending of the interlocutor's turn and by varying the strategy of ending or not ending a turn when overlap is detected.

1 Introduction

It has been claimed that in conversations, speaker change occurs and that overwhelmingly one party talks at a time although the occurrences of more than one speaker at a time are common - but brief ([8]). These claims have been contested about as many times at they have been repeated ([2],[5],[6], for instance). However, in this paper we do not want to argue in favor nor against these claims. Instead, we will look at factors that influence this turn taking behaviour.

According to Goffman ([3]) conversation is expected to be effective transmission of talk, which means that it would be helpful to have "norms constraining interruption or simultaneous talk" (system constraints). But conversations are also social encounters, not only governed by requirements that regulate efficient information exchange, but also by what Goffman calls ritual constraints, regarding "how each individual ought to handle himself with respect to each of the others". These constraints should prevent a person from being rude or impolite. Another obvious factor that influences turn taking behaviour is emotions. "A clash of opinions also means a clash of turn-taking" [6]). All these factors influence turn taking behaviour, which means that observing turn taking behaviour could be very informative about the kind of person one is, how one is feeling or how one wants to be perceived. Although turn taking has been studied widely, the connection with these ritual dimensions has been much less so, though investigations into the detection of some interpersonal variables in conversations have shown that turn taking behavior has an impact ([7]).

In the context of the Semaine project (http://www.semaine-project.eu), we are currently exploring how we can express personality, emotion and social attitudes of the agents that we are building through the implementation of their

Zs. Ruttkay et al. (Eds.): IVA 2009, LNAI 5773, pp. 467–473, 2009.

behaviors. In this paper, we focus on the way variations in turn-management strategies can be used as a resource to create different impressions of the characters. Our goal is therefore not so much to have our agents learn to take "smooth" turn transitions as in [4], for instance, in which the authors present a learning algorithm that tries to minimize the gap between two utterances and the periods in which both participants speak. We would not mind if an agent violates such rules. Violations like these happen constantly in every day conversations. What if an agent starts speaking before another has finished? What if the agent detects that the user starts speaking before it is finished? Should it stop speaking, continue speaking normally, or continue using elevated speech (that is speak louder) to try and win the battle for the floor[1]? Maybe it depends on the kind of character that we want to build.

In this paper, we look at how some basic choices in the management of turns influence the impression that people get from an agent. We look at scales concerning personality, emotion and interpersonal stance. We do this by a person perception study, or rather an agent perception study, using simulated conversations that systematically vary basic turn-taking strategies. We describe the study in the next section and discuss the results in Section 3.

2 Agent Perception Study

The goal of this study is to find out how variations in turn management strategies affects what people perceive the personality, emotional state and the interpersonal stance of an agent to be. We generated several conversations where one agent talks to another varying the ways in which one agent deals with overlap and when it starts speaking. These conversations were presented to human raters who judged the agent on various semantic scales. To generate the conversations we build a conversation simulator that allows one to define different turn management rules. Before we specify the turn management strategies and the impression measures that we examined, we will first say more about the conversation simulator.

Conversation Simulator. The conversation simulator allows one to program the behavior of two agents that can communicate with one another by sending each other information on their communicative actions. They indicate whether they are silent, speaking normally or using elevated speech. These signals, however, are subject to a small delay before they are recognized by the other agent.

Furthermore, to be able to define more refined rules for regulating turns, the agents can also send certain signals that indicate what they are about to do, for instance the intention to start or end the turn (also found in real human conversations, [1]). To simulate that turn taking signals can be ambiguous and misinterpreted by the recipient, a certain error margin is introduced. Signals that are sent can sometimes get lost or signals can be added or changed sometimes.

[1] These are some of the overlap resolution strategies that [9] presents.

The behavior of the two agents can be scripted. In these scripts one can define how an agent reacts to different situations. The core conversation simulator runs the scripts of the agents in parallel and takes care that the rules are executed and the variables are updated accordingly. The conversations of the agents can be visualized (Figure 1) and made audible.

For the speech rendition that was used in the experiment we wanted to output natural incomprehensible speech. To this end we extracted several sentences from the AMI corpus[2] with a clear start and end point. These fragments were then passed through a Pass Hann Band Filter, from 500Hz to 500Hz with a smoothing of 100Hz. With this method the fragments kept their prosodic information but lost their content.

Turn-management strategies/Stimuli. The procedures for turn management that we have considered in this study consist of *start-up* strategies and *overlap resolution* strategies. For each group we have defined three possible strategies, yielding nine different turn taking strategies in total by crossing them. A startup strategy determines when to start a new utterance. In our case, this can be exactly at the moment when the other agent is finished (**At**), with a certain delay (**After**) or before the other agent is finished (**Before**). The overlap resolution strategy determines the strategy to use when overlap is detected. In our case an agent can stop speaking (**Stop**), it can continue normally (**Normally**) or it can raise its voice (**Raised**). Of course in everyday conversations, real life people use a mixture of these strategies for different circumstances, but since the goal is to find out the effect of the single overlap resolution strategies only a single strategy was used in each conversation. The suggestion for these strategies comes from [9].

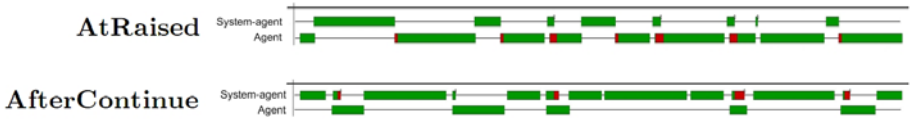

Fig. 1. Example conversations

Using the conversation simulator eight different agents were scripted using the turn taking strategies described. The strategy **Before+Stop** was discarded because it would start while the other person was speaking and immediately stop again, resulting in very unrealistic conversations. The other eight scripts resulted in very different conversations. In Figure 1 two examples of different conversations are shown. The contributions of the agent that varies its strategy (which we will refer to as the agent) are shown on the lower tier. The fixed system-agent (which we will also refer to as the system) was scripted to use different strategies based on chance and is shown on the top tier in each case. Note that the conversations are quite different. The question now is whether these interactions lead to different perceptions of the agent.

[2] http://www.amiproject.org

Experimental Set-Up. The variation in the turn management scripts results in completely different interaction patterns which might change the impression of how the agents relate to each other on an interpersonal scale, or they might change the impression of the personality of the agent. In our study we had ten people rate the eight conversations on each of the dimensions mentioned above. The raters were all between 20 and 30 years old, mainly students, 6 male, 4 female. We asked them to rate them on a five point scale as follows. We used the following scales in our study: *unfriendly-friendly, distant-pleasant, passive-active, cold-warm, negligent-conscientious, disagreeable-agreeable, rude-respectful, unpredictable-stable, unattentive-attentive, submissive-dominant, un-dependable-responsible, negative-positive, not aroused-aroused*. The last 2 scales were added because they are heavily used in the Semaine project.

The raters were seated in front of a PC which ran a powerpoint presentation. On each slide they could click on an audio file that would then play. The audio of the system agent came from the left speaker and the audio from the agent which they had to rate from the right speaker. We made sure that each rater knew which speaker they had to rate. To make the difference even clearer, the agent's speech was somewhat higher in pitch than that of the system agent.

These conversations were ordered such that conversations in which the system was more talkative than the agent alternated with conversations in which the agent was more talkative. We had five raters listen to this order (A) and five raters listened to an order in which the first four conversations of A changed position with the last four conversations of A. These results were combined for the analysis. The raters were asked to fill in the questionnaire on how they perceived the person from the right loudspeaker after each conversation.

3 Results

In this section we present the results of the ratings and a selection of our analyses.

To get a first impression of the results we give a summary of the ratings in Figure 2. In this table we show the means and the standard deviations for each of the scales and all the conversations. Using boldface we have marked the conversation that received the highest mean rate for a scale and using underlinings we marked the conversation with the lowest mean rate for the scale.

When one looks first at the bold face figures it is immediately obvious that the **AtStop** column, in which the agent will start speaking exactly when the system stops and stops when overlap is detected, attracts most of the high scores. It is the top highest rated version on positivity, friendliness, agreeability, respect, pleasantness, attentiveness, warmth and responsibility. On the other hand, **BeforeRaised**,where the agent starts before the system has ended and raises its voice in case of overlap is the top most rated version on negativity, un-friendliness, disagreeability, rudeness, distance, unpredictability, un-attentiveness, and coldness. It is also rated as the most "aroused" agent. These two strategies appear to be the most extreme.

What to think about the other highest/lowest scores? The **BeforeContinue** and **AfterRaised** agent show the lowest scores on the arousal and responsibility

	Before & Continue		After & Stop		Before & Raised		At & Stop		At & Continue		After & Raised		At & Raised		After & Continue	
Scale	Avg	Sd	Avg	Sd	Avg	Sd	Avg	Sd	Avg	Sd	Avg	Sd	Avg	Sd	Avg	Sd
Negative - Positive	2,6	1,1	2,6	0,5	2,1	0,9	**3,7**	0,7	2,8	0,9	2,4	0,5	2,3	0,7	3,1	0,9
Not aroused - Aroused	3,0	1,2	3,1	1,1	**4,8**	0,4	3,3	0,9	3,9	0,6	3,7	1,1	4,1	0,7	3,0	1,1
Unfriendly - Friendly	3,3	1,1	3,0	0,0	2,2	0,9	**3,8**	0,8	3,5	1,0	3,4	0,5	3,1	1,3	3,7	0,7
Disagreeable - Agreeable	2,4	1,2	3,1	1,5	1,7	0,8	**3,5**	1,1	2,3	0,7	2,2	0,6	2,4	1,1	3,3	0,8
Negligent - Conscientious	3,9	1,2	*3,0*	1,2	*3,2*	1,2	3,7	0,8	3,7	1,1	3,1	0,9	**4,0**	0,7	3,4	0,5
Rude - Respectful	3,2	1,2	3,7	0,8	2,0	1,2	**4,2**	0,8	3,0	1,1	2,9	0,7	2,9	0,9	4,1	0,6
Distant - Pleasant	2,6	0,8	2,5	0,7	2,5	0,8	**4,2**	0,8	2,7	0,8	2,6	1,0	3,0	1,1	3,1	1,0
Unpredictable - Stable	3,7	0,8	3,3	1,2	2,0	0,7	3,6	1,1	3,4	1,0	2,7	1,3	2,8	1,0	**3,9**	0,7
Unattentive - Attentive	3,4	1,2	3,3	0,5	3,0	1,3	**4,3**	0,5	3,5	1,1	3,5	0,7	3,7	0,9	3,6	0,5
Cold - Warm	2,5	1,2	2,9	0,6	2,4	0,5	**3,9**	0,6	2,9	0,9	2,7	0,7	2,7	0,9	3,2	0,8
Passive - Active	4,4	0,5	2,6	1,5	4,7	0,5	3,5	0,5	**4,8**	0,4	3,5	1,0	4,0	0,7	2,8	0,8
Submissive - Dominant	4,4	1,3	1,1	0,3	4,3	0,8	3,3	0,8	**4,5**	0,7	3,3	1,1	3,7	1,3	2,7	0,5
Undependable - Responsible	3,2	1,0	2,9	0,7	2,7	1,2	**3,7**	0,8	3,4	0,7	2,4	0,8	3,3	0,9	3,3	0,7

Fig. 2. Means and Standard Deviations of the ratings for each conversation

dimensions respectively. Interestingly though, the lowest score on arousal is 3,0 (which is shared between **BeforeContinue** and **AfterContinue**, which means that there is no agent that effectively performs low on arousal. Similarly, the current strategies do not yield an agent that is extremely negligent. The opposite, conscientiousness is rated highest when an agent does not wait (long) before the turn of the system has finished and continues (possibly speaking louder) when overlap is detected; i.e. the **BeforeContinue** and **AtRaised** score highest on conscientiousness. The impression of highest stability is reserved for the agent that starts speaking only after the system has finished but does not mind continuing when interrupted (without raising its voice), **AfterContinue**. An agent that waits before the other has finished but raises its voice when interrupted, **AfterRaised**, on the other hand, is not considered to be dependable.

Figure 3 contains the fragments grouped by startup strategy and overlap resolution strategy, with their mean and standard deviation. In these tables significance is indicated with one or more stars. The significance was calculated for every dimension scale by performing a two-paired t-test for all combinations (1+2, 2+3, 1+3). The type of t-test (equal variance or unequal variance) was determined by performing an f-test first. A group was said to be significantly different when both t-tests with the other groups scored an $p < 0.05$. So, for example, the negative value for raising the voice is significant because both the t-test with stop and the t-test with continue resulted in an $p < 0.05$. The left part of figure 3 shows that, for the startup strategy, most significant differences occur with the situation in which the agent starts before the system is finished. Starting before the end is seen as more unfriendly, disagreeable, rude, cold and more active, compared to starting at the end or after the end. The most pleasant person would be a speaker who starts directly at the end of the other person's speech, not sooner or later. The sooner a person starts to talk the more active

	Startup Strategy						Overlap Resolution Strategy					
	Before		At		After		Stop		Continue		Raised	
Scale	Avg	Sd	Avg	Sd	Avg	Sd	Avg	Sd	Avg	Sd	Avg	Sd
Negative - Positive	2,4	1,0	2,9	0,9	2,7	0,7	3,1	0,8	2,8	0,9	*2,3	0,7
Not aroused - Aroused	3,9	1,3	3,8	0,8	3,3	1,1	3,2	1,0	3,3	1,0	***4,2	0,9
Unfriendly - Friendly	*2,8	1,1	3,5	1,0	3,4	0,6	3,4	0,7	3,5	0,9	*2,9	1,1
Disagreeable - Agreeable	*2,1	1,1	2,7	1,1	2,9	1,1	3,3	1,3	2,7	1,0	*2,1	0,9
Negligent - Conscientious	3,6	1,2	3,8	0,9	3,2	0,9	3,4	1,1	3,7	1,0	3,4	1,0
Rude - Respectful	*2,6	1,3	3,4	1,1	3,6	0,9	4,0	0,8	3,4	1,1	*2,6	1,0
Distant - Pleasant	2,6	0,8	3,3	1,1	2,7	0,9	3,4	1,1	2,8	0,9	2,7	1,0
Unpredictable - Stable	2,9	1,1	*3,3	1,0	3,3	1,1	3,5	1,1	3,7	0,8	*2,5	1,0
Unattentive - Attentive	3,2	1,2	3,8	0,9	3,5	0,6	3,8	0,7	3,5	0,9	3,4	1,0
Cold - Warm	*2,5	0,9	3,2	0,9	2,9	0,7	*3,4	0,8	2,9	1,0	2,6	0,7
Passive - Active	*4,6	0,5	***4,1	0,8	*3,0	1,2	**3,1	1,2	4,0	1,1	4,1	0,9
Submissive - Dominant	4,4	1,0	3,8	1,1	***2,4	1,2	***2,2	1,3	3,9	1,2	3,8	1,1
Undependable - Responsible	3,0	1,1	3,5	0,8	2,9	0,8	3,3	0,9	3,3	0,8	2,8	1,0

Fig. 3. Strategies, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

he is perceived one could say. This is pointed out by the highly significant result obtained. A final significant result is that the agent that starts after the system's end-of-turn (so with a pause between the utterances) is seen as more submissive.

In the right part of Figure 3 significantly different behaviors occur in the case where the agent stops as soon as overlap is detected and the case when it continues and raises its voice. Stopping when the agent detects overlap is perceived as warmer, more passive and more submissive than continuing. Continuing and talking louder is perceived as more negative, more aroused, less friendly, less agreeable, more rude and more unpredictable than stopping or continuing normally.

These findings to agree with our intuitions which, we think, shows the validity of this methodology. Turn-taking strategies do seem to have an effect on the perception of the agent and the conversations that were generated seem to reflect adequately the principles and parameters that we endow them with.

4 Discussion and Conclusions

In this paper we described a basic conversation simulator that can generate artificial conversations that closely resemble human face-to-face conversations. In particular, the conversation simulator allows one to manipulate several variables related to turn-management, abstracting away from the content of the talk.

We have used the simulator to generate a number of conversations where strategies for timing the beginning of a turn and the decision of continuation when overlap is detected were varied. We showed, through an "agent perception study", how these variations in turn-management changed the impressions that people received from the agent as they listened to the various conversations. The study shows that the manipulation of turn-taking strategies can lead to

different perceptions of an agent on personality scales, interpersonal scales and emotional scales and that therefore these strategies can be used in the repertoire of expressive behaviors of agents reflecting these dimensions.

Future work will continue in several directions. The first concerns a methodological issue which involves the quality of the generated conversations. Another direction is to study other semantic scales than the ones presented here.

The result of this study will be used in actual interactive system, in which an agent will interact in real time with human interactants. This will also allow us to re-introduce contextual features such as topic and semantics that were ignored in this study that focused on the turn-taking tactics as such.

Acknowledgement. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 211486 (SEMAINE).

References

1. Duncan, S., Niederehe, G.: On signalling that it's your turn to speak. Journal of Experimental Social Psychology 10(3), 234–247 (1974)
2. Edelsky, C.: Who's got the floor? Language in Society 10, 383–421 (1981)
3. Goffman, E.: Forms of Talk. Basil Blackwell, Oxford (1981)
4. Jonsdottir, G.R., Thorisson, K.R., Nivel, E.: Learning smooth, Human-Like turntaking in realtime dialogue. In: Proceedings of the 8th international conference on Intelligent Virtual Agents, Tokyo, Japan, pp. 162–175. Springer, Heidelberg (2008)
5. O'Connell, D.C., Kowal, S., Kaltenbacher, E.: Turn-taking: A critical analysis of the research tradition. Journal of Psycholinguistic Research 19(6), 345–373 (1990)
6. Oreström, B.: Turn-taking in English Conversation. Liber, Lund, Sweden (1983)
7. Rienks, R.J., Heylen, D.K.J.: Automatic dominance detection in meetings using easily obtainable features. In: Bourlard, H., Renals, S. (eds.) MLMI 2005. LNCS, vol. 3869, pp. 76–86. Springer, Heidelberg (2006)
8. Sacks, H., Schegloff, E.A., Jefferson, G.: A simplest systematics for the organization of Turn-Taking for conversation. Language 50(4), 696–735 (1974)
9. Schegloff, E.A.: Overlapping talk and the organization of turn-taking for conversation. Language in Society 29(1), 1–63 (2000)