

How turn-taking strategies influence users' impressions of an agent

Mark ter Maat, Khiet P. Truong, Dirk Heylen

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

Abstract. Different turn-taking strategies of an agent influence the impression that people have of it. We recorded conversations of a human with an interviewing agent, controlled by a wizard and using a particular turn-taking strategy. A questionnaire with 27 semantic differential scales concerning personality, emotion, social skills and interviewing skills was used to capture these impressions. We show that it is possible to influence factors such as agreeableness, assertiveness, conversational skill and rapport by varying the agent's turn-taking strategy.

1 Introduction

Turn-taking is a fundamental and universal aspect of conversation that has been described extensively in the literature (for example, in [15, 5, 13]). Although there have been critical reviews on the well-known model by Sacks et al. [15] – “one party speaks at a time”, “occurrences of more than one speaker at a time are brief” – researchers agree that something interesting happens when two or more speakers speak at the same time. Interruptions, for example, have been found to correlate with competitive attitudes such as dominance and power, but also with cooperative attitudes such as attentive listening, backchannel feedback and other rapport-oriented acts [14, 8]. Similarly, pauses in conversation, within or between turns, have various functions and are powerful cues for what is happening in a conversation [3]. Pauses in speech can indicate cognitive processing, but can also be used for grounding, as a politeness marker [2] or as a signal of acceptance or refusal. Endrass et al. [6] showed that there is a cultural component to the perception of pauses.

One application for which turn-taking models are important are Embodied Conversational Agents (ECAs). For ECAs to have humanlike conversations, the ECA needs to know (among other things) when the user has finished or is about to finish his or her turn before it starts speaking. Many turn-taking models have been developed that considered turn-taking theories as described in the literature, in particular with the goal to avoid overlapping speech. For example, Atterer et al. [1] and Schlangen [16] developed algorithms that *predict* turn-endings as soon as possible such that the system can behave quick enough to simulate human-like behavior. Jonsdottir et al. [12, 11] developed a real-time

turn-taking model that is optimized to minimize the silence gap between the human’s speech turn and the system’s speech turn. When evaluating these algorithms they only looked at the performance of the prediction as they were trying to keep the number of overlaps and the average length of the silence gap as low as possible.

But why should the average length of the silence be as short as possible? And why should there be no overlapping speech? In the Semaine project (<http://www.semaine-project.eu>), we aim to create four different ECAs, each with a different personality. We have Poppy, who is cheerful and optimistic, Obadiah, who is gloomy and depressed, Spike, who is aggressive and negative, and Prudence, who is always pragmatic. Several strategies can be applied to evoke different user impressions of the ECA’s personality, for example, one can imagine that the ECA’s appearance or speaking style (e.g. friendly or bored) influences the user’s impressions of the ECA. Here, we are interested in applying different turn-taking strategies (i.e. the management of when to speak) to evoke these user impressions. What we would therefore like to know is how people perceive different turn-taking behaviors, and how we can use this knowledge to assign certain personality-like characteristics to an agent.

Some previous studies have partly addressed these questions. In human-human conversations, interruptions are usually associated with negative personality attributions and assertiveness [14]. In this study, an offline experiment was performed in which participants listened to recorded human-human conversations and judged the participants on several personality and sociability scales. In [7], Fukayama et al. evaluated the impressions conveyed with different gaze models for an embodied agent. Although their research is not related to turn-taking, the concept of ‘impression management’ is very relevant to the ‘perception’ part of the current study. In [17], a similar experiment was carried out, but instead of gaze models, turn-taking strategies were evaluated. Spoken dialogues were simulated with a *Conversation Simulator*, which can simulate conversational agents that use a certain turn-taking strategy, for example, starting to speak before the other agent’s turn ended. The turns were simulated with unrecognizable, mumbling speech. Participants listened to these simulations and rated the ‘speakers’ on scales of personality and affect.

These studies have all used recordings of conversations which were rated by outside observers that were not participating in the conversation. The participants had to rate either audio or video clips which were generated or recorded earlier. Our current study, however, presents a similar agent perception study in which the participants are now *actively involved* in the conversation. We look at how different turn-taking strategies alter the impressions that people have of an agent, and we do this in an *interactive* way. Participants are interacting with a virtual interviewer – using speech only – whose turn-taking strategy is controlled by a Wizard of Oz. After each interview, the participants fill in a questionnaire about the perceived impression of the agent. We describe the setup of this study in Section 2, and the questionnaire used in Section 2.4. The results are presented and discussed in Section 3.

2 Experimental Setup

The goal of this experiment is to assess how different turn-taking strategies of an agent affect the human interlocutor’s perception of the agent. In a previous experiment [17], this was studied by simulating conversations between two agents using unrecognisable speech. After listening to these simulated conversations, people had to write down how they perceived one of the agents. In each of the conversations the agent was following a predetermined turn-taking strategy.

A disadvantage of that experiment was that the human raters were not subjected to the turn-taking strategies themselves; they were only bystanders overhearing each conversation. When someone is using a certain strategy when talking directly to you, the *effect* on your perception of the other person will probably be stronger than that of a third person, and may even be completely different. For example, when you hear one person interrupting another person, this will have a different (probably weaker) effect than when that person is interrupting you. Therefore, the experiment was adapted in such a way that the human raters had to talk to a computer which was using a certain turn-taking strategy.

2.1 Turn-taking strategies

In our previous study [17], two distinct turn-taking strategies were tested. The first was the *startup* strategy, which determined when to start speaking, relative to the end of the other agent’s turn. The different strategy choices for the startup strategy were ‘before’, ‘at’, and ‘after’. The other strategy was the *overlap resolution* strategy, which determined how the agent would behave when it detected overlapping speech – both agents talking at the same time. The different choices of the overlap resolution strategies were ‘stop’ (stop speaking), ‘normally’ (continuing normally), and ‘raised’ (continuing with a raised voice). Our initial goal was to apply these strategies to the current experiment as well. However, a pilot test showed that overlapping speech was avoided by the human interlocutors; they would immediately stop speaking. Therefore, we decided to simplify the experiment and only assess the startup strategy.

The three startup strategies used in the current study are:

1. EARLY: the system will start its turn just before the end of the interlocutor’s turn
2. DIRECT: the system will start its turn immediately after the interlocutor’s turn has finished
3. LATE: the system will leave a pause (of a few seconds) before it starts its turn after the interlocutor’s turn has finished.

Although the strategies are *conceptually* similar to those used in our previous study [17], the actual *realization* of these strategies is different, so we decided to use other strategy names to stress this difference. In [17], the agent knew exactly when the other agent’s turn would end so the start of turns could be placed at

fixed points before or after the other’s turn (depending on the strategy). In the current study, the strategies are enforced by the human wizard as current automatic algorithms are not advanced enough to accurately predict the end of a turn in the EARLY strategy [4]. In contrast with [17], the human wizard will start turns at variable time intervals before or after the interlocutor’s turn.

2.2 Scenarios

Testing the startup strategy in an experiment in which the participants have to talk to the agent makes the experiment more complex. In the previous experiment [17] the agents used unrecognisable speech to simulate a conversation. However, it does not make much sense to have the participants talk to an agent that uses unintelligible speech. In the current experiment we therefore used an interview setting, with the computer (the agent) in the role of an interviewer. This constrains the flow of the conversation as the initiative is mainly with the agent. It allowed us to limit the number of utterances the agent should be able to say. The agent will ask a question, and independent of the content of the answer of the user, the agent (or Wizard rather) anticipates the user’s turn-end and then asks the next question using one of the three startup strategies.

In such a setup, the agent’s questions are very important. We designed the questions such that they would be easy to answer, since a complex question can disrupt the flow of the conversation. Also, the questions asked by the agent should not be answerable with one or two words only, since it is hard to apply a certain startup strategy when the user only says ‘Yes’.

Another possible problem is that certain questions can influence the results because each question has certain connotations that are perceived differently by each user. Therefore, we decided to create three sets of questions each on a different topic (‘food and drinks’, ‘media’ and ‘school and study’). By making three different topics it was possible to interchange the questions used in each session (a single conversation of a user with the system). This decreases the influence of the questions on the results. Also, by making three sets of related questions, the questions fit in the same context and will not disrupt the flow of the conversation.

Another factor to consider is the voice that is used. A male or a female voice can greatly influence the perception of the user, for instance, because the voice sounds more friendly, or because male and female participants may listen differently to male and female voices. To control for this variable we introduced two agents: one with a male and another with a female voice. These voices were changed each session to decrease the influence of the voice on the results.

With these changes the different scenarios were created. A scenario consists of a certain startup strategy (EARLY, DIRECT, or LATE), a certain voice (MALE or FEMALE), and a certain topic (‘food and drinks’, ‘media’, and ‘school and study’). A session is a single conversation of the user with the agent, using a certain scenario. These scenarios were created in such a way that every possible combination and order was used at least once.

2.3 Procedure and participants

22 people participated in the experiment (16 male, 6 female, mean age 27.55, SD 3.41). Prior to the experiment they were told that they would talk with a speech-to-speech dialogue system, with the agent in the role of an interviewer. They were told that we implemented different personalities in different parts of the dialogue system, and that their role was to write down, in a questionnaire, how they perceived each agent. After this introduction they talked with the agent three times (three sessions), each session with a different scenario.

During a session the participant sat in front of a microphone and a set of loudspeakers. The wizard sat behind the participant, out of sight but in the same room. During the interview, the wizard would follow the current startup strategy by clicking on the button to start the next question at the right time. The spoken questions were synthesized with the Loquendo TTS software (<http://www.loquendo.com>). After the interview, the subjects would complete a questionnaire about how they perceived the interviewer.

2.4 Questionnaire

After each interview, the participant received a questionnaire. In order to measure the perceived impression the users had of the agent, we adopted semantic differential scales: pairs of bipolar adjectives were placed at the extremes of 7-point Likert scales. The selection of scales was based on previous experiments by [7, 17, 9, 14]. In general, our goal was to have a set of scales that captures users' impressions of personality-related attributes, social-skills-related attributes, and the interviewer's interviewing capabilities, see Table 1.

Table 1. Semantic differential adjectives used in questionnaire

negative - positive	not aroused - aroused	unfriendly - friendly
disagreeable - agreeable	negligent - conscientious	rude - respectful
distant - close	unpredictable - stable	unattentive - attentive
cold - warm	passive - active	submissive - dominant
competitive - cooperative	impolite - polite	introvert - extravert
inexperienced - experienced	shy - bold	careless - responsible
insecure - confident	tensed - relaxed	disengaged - engaged
aggressive - clam	closed - open	weak - strong
pushy - laidback	arrogant - modest	not socially skilled - socially skilled

3 Results

This section presents the results of the experiment. First, the annotated recordings were checked to confirm that the startup strategies were applied correctly

by the Wizard. Second, a factor analysis was performed to reduce the number of scales to a more manageable number. We will present both steps in the first two subsections below. Finally, the results of the data analysis are presented.

3.1 Strategy validation

During the sessions, it became clear that not just machines have problems predicting the end of the user's turn correctly. Especially with the EARLY strategy it is likely that there are occasions where the user was not almost finished with the turn and wanted to start another sentence, but was interrupted by the interviewer (aka Wizard).

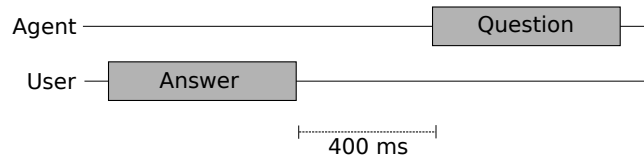


Fig. 1. An example of a pause length of 400 ms

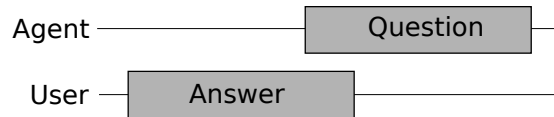


Fig. 2. An example of an instance of overlap

Since applying the correct strategy is error-prone, we need an objective measure to see how consistently each startup strategy was applied. For this we looked at the two objective measures *pause length* and *number of overlaps*. The pause length is the average duration of silence between the end of the user's turn and the start of the agent's next question. Figure 1 illustrates this. We expect that this duration is shortest for the EARLY strategy and longest for the LATE strategy. The number of overlaps is the average number of overlaps per session, where an overlap is defined as an agent that starts the next question while the user is still speaking. For an example, see Figure 2. One should expect that the number of overlaps is highest for the EARLY strategy, and lowest for the LATE strategy. To verify these expectations we annotated the recorded interviews on who was speaking when. With these annotations, we counted the number of overlaps and we measured the average pause length.

Table 2. Average pause length between user’s turn and following interviewer’s question. *** = $p < 0.001$

	Average pause length			Mean	Sd
	EARLY	DIRECT	LATE		
EARLY	-	***	***	0.72	0.69
DIRECT	***	-	***	1.07	0.58
LATE	***	***	-	1.97	0.57

Table 3. Average number of overlaps (user and interviewer speaking at the same time.) *** = $p < 0.001$

	Average number of overlaps			Mean	Sd
	EARLY	DIRECT	LATE		
EARLY	-	***	***	4.16	2.19
DIRECT	***	-		1.20	1.24
LATE	***		-	0.70	1.13

Table 2 shows the average pause length between the user’s current turn and the following interviewer’s question, grouped by the startup strategy that was used. As expected, the EARLY strategy contains the shortest pauses, and the LATE strategy the longest. Also, the differences between the three strategies are highly significant ($p < 0.001$).

Table 3 shows the average number of overlaps per session – which happened when the agent started speaking before the user was finished – again grouped by the startup strategy that was used. As one should expect, the number of overlaps is highest in the EARLY strategy and lowest in the LATE strategy. The difference between the EARLY strategy and the other two strategies is highly significant ($p < 0.001$), but the difference between the DIRECT and the LATE strategy are not. Because both the DIRECT and the LATE strategy wait for the end of the user’s turn, we did not expect any significant difference between these strategies for overlaps.

These results show that there is indeed a significant difference in the sessions between the different startup strategies in accordance with the desired effect, which means that the different strategies were correctly applied. Therefore, differences in the results are based on significantly different startup strategies.

3.2 Factor analysis

To reduce the number of scales (27 different scales) a factor analysis was performed. We used a Principal Component Analysis, with the rotation method Varimax with Kaiser normalization. From the results we used the items with a correlation > 0.5 , which resulted in four different factors. These four factors, the corresponding scales and the corresponding correlations can be found in Table 4.

The next step is to interpret these factors. The first factor of Table 4 appears to be related to the *agreeableness* trait, one of the five main personality traits as described by Goldberg [9]. A high level of agreeableness corresponds to someone who is cooperative, compassionate, friendly and optimistic. The adjectives strong, dominant, extravert, bold, arrogant, and pushy, are grouped by the second factor and seem to be well described as *assertiveness*, a term that has been used previously by [14] in a similar context. The items of the third factor

do not match with one of the personality traits. It appears that those items say something about the *conversational skill* of the agent. The agent took the role of the interviewer, and a ‘good’ interviewer should be socially skilled, attentive and experienced. The last factor seems to be related to *rapport* [10]. A high level of rapport means the participants are ‘in sync’ or ‘on the same wavelength’, and this often means that the participants are very close and engaged.

In the remaining sections the four factors will be referred to as *agreeableness*, *assertiveness*, *conversational skill* and *rapport*, respectively.

3.3 Data analysis

For the analysis of the data, an ANOVA test was performed on the data with a Bonferroni Posthoc test. The ratings that were used were the four factors found in the previous section, and the four scales that did not fit in these factors: rude-respectful, not aroused-aroused, insecure-confident and passive-active. This section shows the results.

Table 4. Results of Factor Analysis

Factors	Related adjective for low values	Related adjective for high values	Correlation
Factor_1	Cold	Warm	0.86
	Unfriendly	Friendly	0.78
	Tensed	Relaxed	0.72
	Disagreeable	Agreeable	0.70
	Aggressive	Calm	0.63
	Competitive	Cooperative	0.60
	Negative	Positive	0.60
	Impolite	Polite	0.52
Factor_2	Strong	Weak	0.85
	Dominant	Submissive	0.79
	Extravert	Introvert	0.76
	Bold	Shy	0.73
	Arrogant	Modest	0.57
Factor_3	Pushy	Laid back	0.53
	Inexperienced	Experienced	0.72
	Not socially skilled	Socially skilled	0.72
	Unpredictable	Stable	0.69
	Careless	Responsible	0.60
Factor_4	Unattentive	Attentive	0.50
	Closed	Open	0.82
	Disengaged	Engaged	0.71
	Distant	Close	0.62
	Negligent	Conscientious	0.58

Startup strategy Figure 3 shows the results of this analysis for the different startup strategies. This figure only shows the results of the four factors and the rude-respectful scale. The other three scales did not provide any significant results.

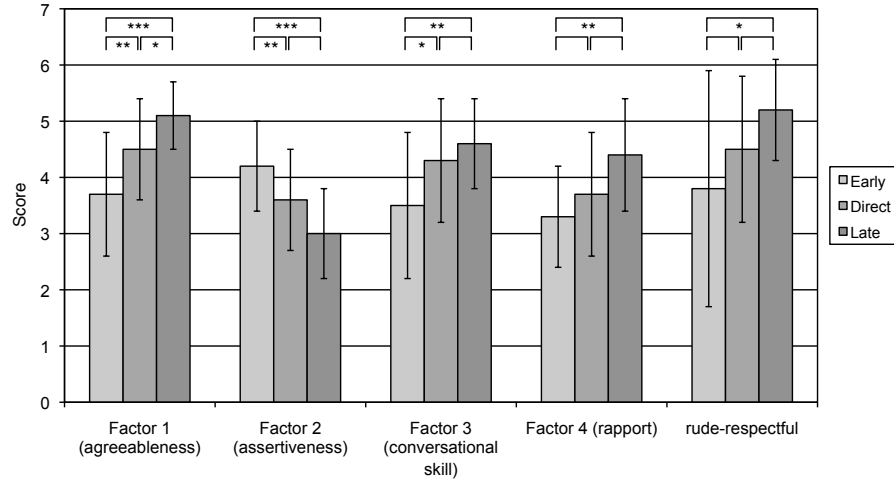


Fig. 3. The results of the different startup strategies. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

The strongest factor clearly is Factor 1 (agreeableness): the ratings for all three strategies are significantly different. Starting EARLY is seen as more unfriendly, tensed, aggressive, competitive and negative, and starting LATE is perceived as more friendly, relaxed, agreeable, cooperative and positive. Factor 2 (assertiveness) is mostly strong in the EARLY strategy. The mean rating is significantly higher compared to the other two strategies, but the DIRECT and the LATE strategy are not significantly different compared to each other. Starting EARLY was rated as more strong, dominant, extravert and bold, and using the DIRECT or LATE strategy was rated as more weak, submissive, introvert and shy.

The same result was found with Factor 3 (conversational skill): the EARLY strategy was rated significantly lower than the other two strategies, but those two strategies were not rated significantly different compared to each other. An agent using the EARLY strategy was rated as less experienced, less socially skilled, less predictable, less careless and less attentive. An agent using the DIRECT or the LATE strategy was seen as more experienced, socially skilled, stable, responsible and attentive.

Agent gender We observed that the voice that was used – male or female – made a big difference in the perception of the user. In the analysis of startup strategy this effect was filtered out by using an equal number of male and female voices. However, the differences between the voices are still interesting, which is why we analyzed the differences between these voices as well. Figure 4 shows the results.

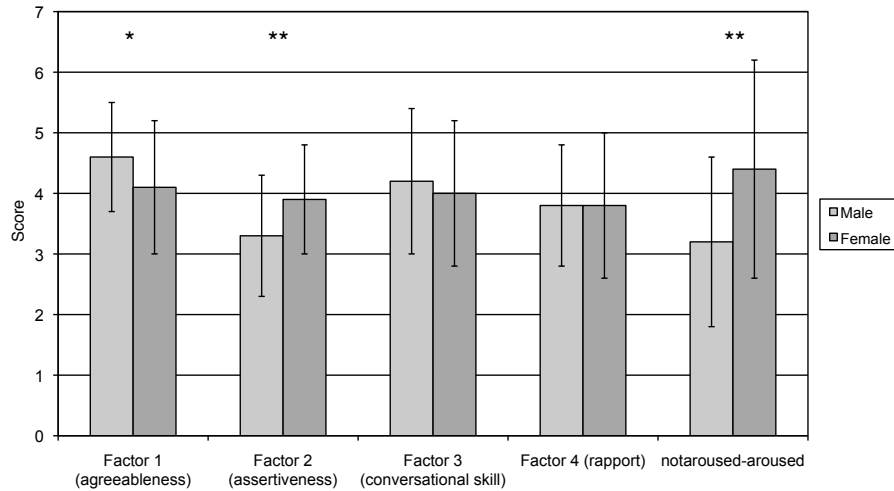


Fig. 4. The results of the different genders. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.005$

This figure shows that the male voice was rated higher in Factor 1 (agreeableness), lower in Factor 2 (assertiveness) and lower in the arousal scale. This means the male voice was perceived, among others, as more friendly, positive, polite, submissive, shy, and less aroused. The female voice was perceived, among others, as more cold, aggressive, negative, dominant, bold, and aroused. This may appear strange, but the results probably say more about the voices used than about gender in general.

Along the same line, we also did some other gender comparisons in the data which showed only minor differences. Robinson and Reis [14] explain that an important factor could be the difference between the gender – same-sex or opposite-sex. To study this we compared the gender of the user with the gender of the agent. However, only two minor results were found. Male users rated male agents significantly lower ($p < 0.05$) in the notaroused-aroused scale than they rated female agents. Also, female agents were rated significantly lower ($p < 0.05$) in Factor 4 (rapport) by male users than they were rated by female users.

We were also interested in the combination of the startup strategy and the gender of the user. For example, to check whether a male user perceives an agent using the EARLY strategy differently than a female user. However, no significant results were found. Another interesting combination is the startup strategy and the gender of the agent. A male agent using the EARLY strategy might be perceived differently than a female agent using the same strategy. Only one result came out of this. A male agent using the DIRECT strategy is perceived significantly higher in Factor 1 (agreeableness) than a female agent using the same strategy.

The final thing we checked was the combination of user gender, agent gender and startup strategy, but we only found one significant result here. A male user rates a male agent using the LATE strategy significantly lower ($p < 0.05$) in the notaroused-aroused scale than a female agent using the same strategy.

Comparison with previous study In a previous study [17], we also studied the effects of different turn-taking strategies, but instead of actively involving the user in the conversation we used recordings of simulated conversations. The questionnaire that was used contained 13 scales, of which 11 were also used in this experiment. Eight of those scales can be put in a factor (see section 3.2), and most of them (four) belong in Factor 1 (agreeableness).

When comparing the results of the two studies, the results are about equal. However, an interesting difference can be found in the scales that belong to Factor 1. As can be seen in Figure 3, the DIRECT strategy scores significantly higher in agreeableness than the EARLY strategy, and the LATE strategy scores significantly than the DIRECT strategy. Looking at the results of the previous experiment, to the scores of the four scales that belong in Factor 1, the EARLY strategy (on average) scores lower than the other two strategies, but the DIRECT and LATE strategy do not have a significantly different score. On average, the score for the LATE strategy is even slightly *lower* than that of the EARLY strategy. While this is not a sufficient statistical conclusion, this could point to the fact that humans are influenced more by LATE strategies when being actively involved than when being a passive bystander. However, this difference could also be caused by the interview scenario that was used.

4 Conclusions and Discussion

In this paper we studied how three different turn-taking strategies affect how people perceive the actor of those strategies. With a Wizard-of-Oz setup we simulated a conversational interviewing agent, and the start-time of the next question of the agent was determined by the turn-taking strategies we were testing: EARLY, DIRECT or LATE. Although it is not easy to guarantee that a Wizard uses a certain strategy consistently, the analysis of the recordings revealed that the three different strategies were applied accordingly in this experiment.

Based on the results we found, we can conclude that an agent that uses a certain turn-taking strategy can indeed influence the impression that a user has

of this agent. Starting too early (that is, interrupting) is mostly associated with negative and strong personality attributes: agents are perceived as less agreeable and more assertive. Leaving pauses between turns has contrary associations: it is perceived as more agreeable, less assertive, and creates the feeling of having more rapport. The agent’s voice did play a role in the results too. In general, we can say that the male voice was perceived as more agreeable and less aroused than the female voice. However, this effect could be more related to the quality of the synthesized voices than to the gender of the agent. Since we only used two different voices for each gender (one Dutch and one English) it is very hard to generalize these results to gender. Previous studies also report on relations between gender and interruptions and interpersonal perceptions of interlocutors – for example, females who interrupt would be penalized more than male interrupters – but we did not find such effects in our data. This was mainly because our prime interest was the turn-taking strategy. Also, we have to keep in mind that the results are specific to this interviewing domain, and that some findings might probably not generalize to a ‘free-talk’ conversation in which dialogue partners can talk about anything they like, or in a setting in which both dialogue partners can ask questions to each other.

In future work, we will implement several of these turn-taking strategies to convey different personalities and agent impressions (in the Semaine project). It would be interesting to design more finely-grained turn-taking strategies and to look more locally, for example, at what meaning or impressions pauses can convey. Rather than leaving pauses between *each* turn, it would be better to adapt to the conversational context and leave a rightly-timed pause that may convey a certain meaning.

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), and by the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 231287 (SSPNet).

References

1. M. Atterer, T. Baumann, and D. Schlangen. Towards Incremental End-of-Utterance Detection in Dialogue Systems. In *Proceedings of the 22nd International Conference on Computational Linguistics*, number August, pages 11–14, 2008.
2. P. Brown and S. C. Levinson. *Politeness: Some universals in language use*. Cambridge University Press, 1987.
3. H. H. Clark. *Using Language*. Cambridge University Press, 1996.
4. I. de Kok and D. Heylen. Multimodal End-of-Turn Prediction in Multi-Party Meetings. In *ICMIMLM 2009 Proceedings*, pages 91–98. ACM Press, 2009.
5. C. Edelsky. Who’s got the floor? *Language and Speech*, 10:383–421, 1981.
6. B. Endrass, M. Rehm, E. André, and Y. I. Nakano. Talk is silver, silence is golden: A cross cultural study on the usage of pauses in speech. In *Proceedings of the UII-Workshop on Enculturating Interfaces (ECI 2008)*, 2008.

7. A. Fukayama, T. Ohno, N. Mukawa, M. Sawaki, and N. Hagita. Messages embedded in gaze of interface agents: impression management with agent's gaze. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI'02)*, pages 41–48, 2002.
8. J. A. Goldberg. Interrupting the Discourse on Interruptions: an Analysis in Terms of Relationally Neutral, Power- and Rapport-Oriented Acts. *Journal of Pragmatics*, 14:883–903, 1990.
9. L. R. Goldberg. The structure of phenotypic personality traits. *American Psychologist*, 48(1):26–34, januari 1993.
10. J. Gratch, S. Marsella, A. Okhmatovskaia, F. Lamothe, M. Morales, R.J. Werf, and L.P. Morency. Virtual rapport. In *6th International Conference on Intelligent Virtual Agents*, pages 14–27. Springer, 2006.
11. G. R. Jonsdottir and K. R. Thorisson. Teaching Computers to Conduct Spoken Interviews: Breaking the Realtime Barrier with Learning. In *Proceedings of the International Conference on Interactive Virtual Agents (IVA 2009)*, pages 446–459, 2009.
12. G. R. Jonsdottir, K. R. Thorisson, and E. Nivel. Learning Smooth, Human-Like Turntaking in Realtime Dialogue. In *Proceedings of the 8th international conference on Intelligent Virtual Agents (IVA2008)*, pages 162–175, 2008.
13. D. C. O'Connell, S. Kowal, and E. Kaltenbacher. Turn-Taking: A Critical Analysis of the Research Tradition. *Journal of Psycholinguistic Research*, 19(6):345–373, 1990.
14. L. F. Robinson and H. T. Reis. The Effects of Interruption, Gender, and Status on Interpersonal Perceptions. *Journal of Nonverbal Behavior*, 13(3):141–153, 1989.
15. H. Sacks, E. A. Schegloff, and G. Jefferson. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4):696–735, 1974.
16. D. Schlangen. From Reaction To Prediction Experiments with Computational Models of Turn-Taking. In *Proceedings of Interspeech 2006*, pages 2010–2013, 2006.
17. M. ter Maat and D. Heylen. Turn Management or Impression Management? In Z. M. Ruttkay, M. Kipp, A. Nijholt, and Hannes Högni Vilhjálmsson, editors, *Proceedings of the 9th international conference on Intelligent Virtual Agents (IVA2009)*, pages 467–473, Amsterdam, The Netherlands, 2009.