Modeling The Impact Of Action Tendency On An Agent Interrupting Behavior

Mathieu Jégou, Céline Clavel, Nicolas Sabouret, Jean-Claude Martin

LIMSI, CNRS (UPR3251), Université Paris-Saclay, Orsay, F-91405, France
{jegou, celine.clavel, nicolas.sabouret, jean-claude.martin}@limsi.fr

Abstract. Expression of emotions and interruption management are two major capabilities that should be displayed by Embodied Conversational Agents. Yet, no work have linked interruption management and emotions. In this paper, we introduce a new model for interruption management that considers one component of emotions: the action tendency. We explain how the action tendency influences the motivation that the agent has to change role (e.g. the motivation to speak if it is currently listening to the user), which impacts its behaviors related to interruptions. We then present an implementation of our model in an existing agent architecture, AgentSlang, and illustrate the behavior of our model with examples of interactions between the agent and one user.

Keywords: Interruptions, Emotion, Action Tendencies

1 Introduction

Embodied Conversational Agents (ECAs) are graphical entities with human-like appearance that are able to dialog naturally and spontaneously with users by recognizing and producing verbal utterances and non-verbal signals [1]. One key ability to ensure natural and spontaneous interactions with users is the ability to manage turn-taking [1]. Turn-taking refers to the ability to coordinate speaking turns during a conversation, so that participants speak mostly one at a time in alternating turns [2]. Human participants actively behave to ensure this coordination [3]. These behaviors are, either taking the turn, when the listener becomes speaker after the end of the previous speaker turn, grabbing the turn when the listener tries to become speaker while the speaker's turn is not finished, yielding the turn to the listener or keeping the turn [3]. Participants explicitly signal their behavior by exchanging verbal and nonverbal signals [2]. Concurrently to this alternation of speaking turns, participants often interrupt each other [4]. Interruptions are either cooperative, showing agreement or helping the speaker complete its turn, or competitive, showing disagreement or disinterest towards what the current speaker is saying [4]. Human participants often vary their behavior related to interruptions. They can choose to continue or stop speaking when detecting that their interlocutor interrupts them, or choose to interrupt or not their partner [4]. These behaviors are partly driven by the participants' dominant or submissive attitudes or their current emotional state [4]. The influence of these factors, especially the emotions, should be taken into account when creating an agent able to interrupt or manage the user's interruption. Indeed, in two perceptual experiments, [7] and [8] have observed that the way the agent handled interruptions influenced the user's judgment about the agent's dominant, submissive attitude or its emotions.

However, very few studies have taken into account dominance or emotions in the way the agent controls its behavior related to its emotions. Past works on interruption management in user-agent interactions encompasses models used to detect user interruption attempts [5], or to determine if the agent should resume its interrupted utterance or plan a new utterance after the user interruption [6]. In most of these models, the agent systematically stops its ongoing turn when it detects the user's utterance and waits the end of the user's interruption before starting to speak again. Moreover, very few models allow the agent to interrupt the user on purpose. To our knowledge, only [9] elaborated a model where the agent behavior varied according to the attitude of the agent towards its interlocutors. Yet, no computational model tried to link interruption management to emotions. This requires finding out which emotional state could lead the agent to interrupt the user, and which emotional states could lead the agent to continue or stop speaking when the user tries to interrupt it.

In this paper, we introduce a new model called EmoTurn (Emotional Turn-Taking). EmoTurn is a computational model for real-time user-agent interactions, where one component of an emotion, the action tendency [10] influences the agent's turn-taking behaviors. According to [10], action tendencies refer to readiness or unreadiness to engage in interaction in some particular fashion.

The proposed model relies on a previous computational model elaborated by Jégou et al. [11]. In Jégou et al.'s work, the agent's behavior towards turn-taking was allowed to vary according to a variable called motivation to change role. We based on this model to elaborate a set of rules determining how the motivation to change role varies according to the agent's current emotional state. After presenting our model, we show the ability of an agent controlled by our model to vary its behavior related to interruptions in real-time interactions with a human partner.

2 The EmoTurn Model

In this section, we present the EmoTurn model. We introduce the different variables used in our model, the motivation to change role coming from Jégou et al.'s model [11], and two action tendencies we selected to elaborate our model. We then explain the rules that combine the action tendencies and the agent's motivation to change role.

2.1 Motivation to change role and action tendencies

The model of Jégou et al. [11] is a theoretical model that controls the agent's verbal and nonverbal signals and the moments when the agent speaks based on its partner non-verbal signals. Two variables directly control the agent's behavior: the *motivation to change role* m and the *degree of certainty* γ towards the perception of the partner's behavior.

The *motivation to change role*, *m*, sets the goals of the agent, namely changing its role (speaker, listener) or keeping its current role. The motivation to change role influences the agent's final behavior, as for a current speaker, a motivation to change role makes it yield the turn, or for a listener, makes it take the turn. Conversely, a motivation to keep role makes the current speaker try to keep its turn and the current listener stay listener. This motivation (*m*) continuously varies between -1 (the agent strongly wants to keep its current role) and 1 (the agent strongly wants to change role).

However, the agent final behavior also depends on its partner's behavior. The agent continuously monitors the signals displayed by its partner and varies accordingly its own behavior. More precisely, the agent computes a degree of certainty γ about the partner's behavior based on the signals displayed by the latter. This degree of certainty informs the agent about whether its partner is currently trying to change role (become speaker for a listener, or become listener for a speaker) or to keep its role (staying listener or speaker). γ is used concurrently with *m* to control the agent's behavior. As a result, γ can potentially make the agent give up its current goal (for example, a listener that gives up its interruption attempt due to the turn keeping signals displayed by the current speaker).

The absolute value of the motivation defines the insistence the agent has in accomplishing its goal. For example, if a listener has a motivation value close to 1, it makes it more insistent to take the turn, even if γ indicates that its partner is trying to keep the turn, leading it to try to interrupt the speaker. Conversely, if the listener has a value close to 0, it will not try to take the turn while γ indicates that the current speaker is willing to keep the turn.

Our goal is to combine this mechanism with action tendencies. We selected several action tendencies from [10] to create the EmoTurn model. In this paper, we will focus on a subset of these action tendencies to illustrate the behavior of the model: the "Excited" action tendency, defined by a tendency to be "excited, restless" [10], and the "Inhibited" tendency; defined as a tendency to feel "paralyzed, or frozen" [10].

2.2 Impact of action tendency on motivation to speak

In our model, the action tendency does not directly influence the agent's behavior but is used to compute the agent's motivation to change role according to the following rules:

```
if the agent has something to say
then  if its current role is speaker
      then m ← motivation to keep turn
      else m ← motivation to take turn
      decrease / increase m according to the agent's
action tendency
else  if its current role is speaker
      then m ← motivation to yield turn
      else m ← motivation to stay listener
```

We consider that the agent has four possible behaviors, keep the turn, take the turn, yield the turn or continue to listen. The factor defining the nature of the agent's behavior is whether the agent has something to say or not. Action tendencies do not influence the nature of the agent's behavior, but impact the strength of the motivation, making it more or less insistent in trying to yield or take the turn.

Depending on the action tendency, the value of the motivation to change role varies as specified in **Table 1**. This mapping is inspired by the semantic descriptions of action tendencies [10].

Table 1. Values for motivation to change role depending on the action tendency

Action Tendency	Role	Motivation to change role
Excited	Speaker	Strongly unwilling
	Listener	Strongly willing
Inhibited	Speaker	Weakly unwilling
	Listener	Weakly willing

`"Strongly Unwilling" means that the agent will insist to keep the turn if the user wants to interrupt it whereas "Strongly Willing" means that the agent will insist to take the turn making him try to interrupt the current user. "Weakly Unwilling" means that the agent will yield the turn when detecting an interruption attempt while "Weakly Willing" means that the agent will not try to take the turn while the agent is still speaking. In section 3, we introduce more in details how, in our implementation, we currently compute the motivation values based on these rules.

3 Application to real-time user-agent interactions

In this section, we illustrate the behavior of our model in real-time interactions with a human partner. We consider a series of scenarios where the agent computes its motivation to change role according to its action tendency. Moreover, the agent interprets continuously the pitch and the acoustic energy of the user. The agent controls two types of non-verbal signals, its gaze (look towards the user or avert gaze) and its eyebrows (raising or not the eyebrows), which are signals used by humans to coordinate their turns [12], [13]. First, we present how we implemented our model. We then illustrate with four examples, how the agent behaves in real-time interactions with the user.

3.1 Presentation of the implementation

We have implemented EmoTurn in an existing agent architecture, AgentSlang [14]. In this architecture, we have created components dedicated to the real-time coordination of speaking turns with the user. The components are shown on **Fig. 1**.

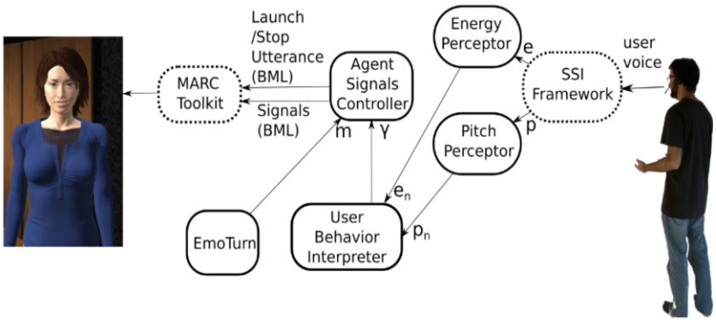

Fig. 1. Implementation of our EmoTurn model in AgentSlang

We used the SSI Framework [15] to extract the energy *e* and pitch *p* from the user's voice. These values, *p* and *e*, are then normalized according to the user's minimum of maximum values of pitch and energy. These values are computed in a previous calibration step. Once the normalized values e_n and p_n determined, they are transmitted to the User Behavior Interpreter component that computes *γ*. The EmoTurn component computes according to the following formula:

$$m = m_u + m_{act} \tag{1}$$

According to the agent's role, the different values of m_u and m_{act} are presented in **Table 1**.

Table 1. Values of m_u and m_{act} used in our implementation

	Listener	Speaker
m_u	**Has something to say :** m_u= 0.5 **Has nothing to say :** m_u=−0.5	**Has something to say :** m_u= 0.5 **Has nothing to say :** m_u=0.5
m_{act}	**Strongly Willing :** m_{act}=0.5 **Weakly Willing :** m_{act}=−0.4	**Strongly Unwilling :** m_{act}=−0.5 **Weakly Unwilling :** m_{act}=0.4

Based on *m* and *γ*, the Agent Signals Controller is in charge to compute gaze and eyebrows variations. It also determines when to launch or stop the agent's utterance according to an internal variable representing the agent acoustic energy. This component takes as inputs, *m* and *γ*. The Agent Signals Controler sends gaze and eyebrows variations commands to the realizer, formulated in BML (Behavior Markup Language) [17]. Based on the theoretical acoustic value, the module decides to launch or stop the agent's utterance. This decision is made such as, when the theoretical energy value is greater than a threshold value (0.2), a launch command is sent to the

realizer, and when the theoretical energy value is less than this threshold, a stop command is sent to the realizer.
We used MARC [16] as the realizer. MARC modulates the agent's nonverbal signals based on the command it receives, launches and stops audio files corresponding to the agent's utterance, and synchronizes the lips of the agent with the audio. We used audio records of human voices rather than a TTS to generate the agent's utterance.

3.2 Illustrative examples of potential interactions

We present four examples of interactions between our agent and a human partner. One of the author of this paper played the human partner. These four examples cover four possible qualitative behaviors related to interruption management the agent can have depending on the agent's current role. As a listener, the agent either interrupts the user or take the turn after the user finished its turn. As a speaker, the agent either continues speaking when reacting to the user interruption or stops speaking and let the user become the speaker. The verbal utterances exchanged by the participants, not presented here, were inspired from [18].
The four scenarios are presented in **Fig 2**. For each scenario, we represent the moments when the agent and the user spoke by the waveforms of the audio signals.
In scenario #1, the user speaks at the beginning of the interaction and the agent is the listener. After a few seconds, the dialog manager plan an utterance by setting m_u to 0.5. The agent's action tendency is set to "Excited" in this scenario, thus m_{act}=0.5 . As a result, even if γ indicates that the user is continuing its turn, the agent starts speaking while the user is still speaking. The overlap between the agent and the user lasts 1.6 s, after which the user lets the agent continue its turn.
In the scenario #2, the agent is the listener. Similar to scenario #1, we simulate in the architecture the fact that the agent planned a new utterance ("has something to say" on **Fig 2.**) after few seconds. Here, the agent's action tendency is "Inhibited", thus, m_u=0.5 and m_{act}=−0.5. As a result, the agent does not try to interrupt the user and waits for the end of the user's turn before taking the turn, leaving a gap of 1.3 s.
In scenario #3, the action tendency is set to "Excited", the agent is the current speaker and the user is the current listener. After a while, the user starts to speak, however, since the agent did not finish its utterance, m_u=−0.5 and m_{act}=−0.5. When the user interrupts, γ increases, indicating that the user wants to take the turn. As the acoustic energy remains high during the overlap, the agent continues to speak.
In scenario #4, the agent is also the current speaker. However, its action tendency is "Inhibited". Similar to scenario #3, the user tries to interrupt the agent after several seconds. However, since m_u=−0.5 and m_{act}=0.5 , the agent stops speaking to let the user speak.

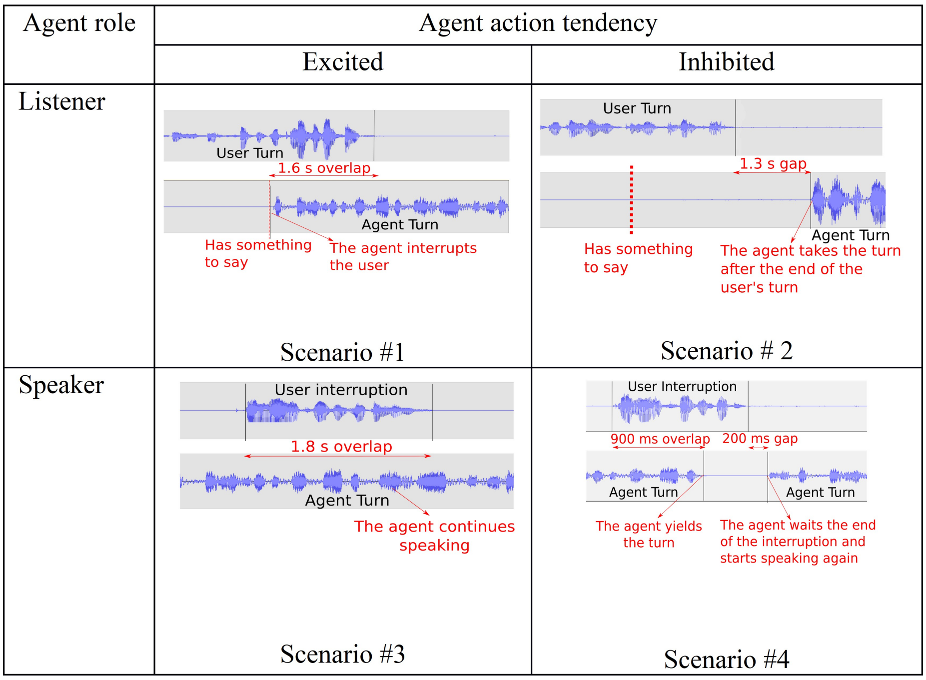

Fig. 2. Four example of interactions between the user and the agent

4 Conclusion

In this paper, we presented EmoTurn, a model designed to make the agent vary the way it manages interruptions with the user according to one component of its emotions, the action tendencies. We then showed how our model integrates in an existing agent architecture, AgentSlang, and how it manages the launch and the stop of the agent's utterance in real-time interactions with a human partner.

We plan to validate the links between action tendencies and motivation to change role established in section 2. To that purpose users will interact in real-time with an agent varying its turn-taking behavior according to our model. At the end of the interactions, questions about the agent's action tendencies will be asked to the users. Finally, we only covered a subset of the type of interruptions that exist in human interactions. Especially, in order to have a complete model, we should distinguish cooperative and competitive interruptions.

Acknowledgments. This work was funded by the Agence Nationale de la Recherche (ANR), project NARECA (ANR-13-CORD-0015).

References

1. Cassell, J., Bickmore T., Billinghurst, M., Campbell, L., Vilhjálmsson, H., Yan, H.: Embodiment in conversational interfaces, In: CHI'99, pp. 520-527., ACM New York (1999)
2. Sacks, H., Schegloff, E. A., Jefferson, G.: "A Simplest Systematics for the Organization of Turn-Taking for Conversation," Language, 50(4), pp. 696-735 (1974)
3. Bunt, H., Girard, Y.: "Designing an open, multidimensional dialogue act taxonomy," In : DIALOR 2005, pp. 37–44, (2005)
4. Goldberg, J. A.: "Interrupting the discourse on interruptions," *Journal of Pragmatics*, 14(6), pp. 883–903 (1990)
5. Selfridge, E., Arizmendi, I., Heeman, P. Williams, J.: "Continuously predicting and processing barge-in during a live spoken dialogue task," In: SIGDIAL 2013, pp. 384–393 (2013)
6. Crook, N., Smith, C., Cavazza, M., Pulman, S., Moore, R.; Boye, J.: "Handling user interruptions in an embodied conversational agent," In: AAMAS International Workshop on Interacting with ECAs as Virtual Characters, pp. 27–33 (2010)
7. Ter Maat, M., Truong, K. P., Heylen, D.: "How turn-taking strategies influence users' impressions of an agent," in Intelligent Virtual Agents 2016, pp. 441–453 (2010)
8. Cafaro, A., Glas, N., Pelachaud, C.: "The Effects of Interrupting Behavior on Interpersonal Attitude and Engagement in Dyadic Interactions," In: AAMAS 2016, pp. 911–920 (2016)
9. Ravenet, B., Cafaro, A., Biancardi, B., Ochs, M.; Pelachaud, C.: "Conversational behavior reflecting interpersonal attitudes in small group interactions," in IVA 2015, pp. 375–388. (2015)
10. Frijda, N. H., Kuipers, P., ter Schure, E.: "Relations among emotion, appraisal, and emotional action readiness," J. Pers. Soc. Psychol., 57(2), pp. 212–228, (1989)
11. Jégou, M., Lefebvre, L., Chevaillier, P.: "A Continuous Model for the Management of Turn-Taking in User-Agent Spoken Interactions Based on the Variations of Prosodic Signals," in IVA 2015, pp. 389–398 (2015)
12. Oertel, C., Wlodarczak, M., Edlund, J., Wagner, P., Gustafson, J.: "Gaze patterns in turn-taking," in INTERSPEECH 2012, pp 2243-2246, Red Hook, NY (2013)
13. Guaïtella, I., Santi, S., Lagrue, B., Cavé, C.: "Are eyebrow movements linked to voice variations and turn-taking in dialogue? An experimental investigation," Lang Speech, 52(3), pp. 207–222 (2009)
14. Serban, O., Pauchet, A.: "AgentSlang: A New Distributed Interactive System - Current Approaches and Performance:," In: ICAART-2014, pp. 596–603 (2014)
15. Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., André, E.: "The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time," In: ACM Multimedia 2013, pp. 831–834, ACM New York, NY, USA (2013)
16. Courgeon, M., Martin, J.-C., Jacquemin, C.: "Marc: a multimodal affective and reactive character," In: 1st Workshop on AFFective Interaction in Natural Environments, p. 20. (2008)
17. Kopp, S. et al.: "Towards a common framework for multimodal generation: The behavior markup language," In: IVA 2006, pp. 205–217 (2006)
18. Fourati, N. et al.: "Facial Expressions of Appraisals Displayed by a Virtual Storyteller for Children," in IVA 2016, pp. 234–244 (2016)