Interruptions in Human-Agent Interaction
Liu Yang, Catherine Achard, Catherine Pelachaud

To cite this version:
Liu Yang, Catherine Achard, Catherine Pelachaud. Interruptions in Human-Agent Interaction. ACM International Conference on Intelligent Virtual Agents, 2021, Virtuel, France. 10.1145/3472306.3478340. hal-03428893

HAL Id: hal-03428893
https://hal.science/hal-03428893
Submitted on 15 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Interruptions in Human-Agent Interaction

Liu YANG
yangl@isir.upmc.fr
ISIR,CNRS, Sorbonne University
Paris, France

Catherine ACHARD
catherine.achard@upmc.fr
ISIR,CNRS, Sorbonne University
Paris, France

Catherine PELACHAUD
catherine.pelachaud@upmc.fr
ISIR,CNRS, Sorbonne University
Paris, France

ABSTRACT
Turn management is one of the necessary social interactions skills. In human-human interactions, turn changes are naturally completed by interruption, “cooperatively” or “competitively”. Interruptions are inherent in conversation. They can be considered disruptive at first glance, but can also be cooperative and participate to enriching the interaction. To create natural human-agent interaction, Embodied Conversational Agent (ECA) should be able to communicate autonomously with humans both verbally and nonverbally. A challenge is then to handle interruptions during their interaction. This article presents our ongoing work to endow ECA to manage interruption during the interaction with a human partner. In order to achieve this goal, we start by analyzing human-human interaction data.

CCS CONCEPTS
- Human-centered computing → Human-agent interaction (HAI).

KEYWORDS
Nonverbal behaviour, Conversational interruption, Turn-taking, Embodied conversational agent (ECA)

ACM Reference Format:

1 INTRODUCTION
In face-to-face conversations, interlocutors exchange quickly the role of speaker and listener in turns. During interactions, humans adapt and adjust their behaviour according to their interlocutors. In particular, partners exchange speaking turns which can give rise to interruptions, overlaps or silence.

In most cases speaking turn exchanges smoothly during a conversation with no gap or overlap [11]. The coordination is smooth when the listener waits for his/her turn, or sends signals to specify s/he wants to take the next speaking turn, and the speaker receive perfectly to give the turn. On the opposite, the listener may not want to take the turn signalled by the current speaker that giving rise to silence, or grab the turn before the current speaker finishes that leading to an interruption.

The management of turn-taking during conversation is necessary for ECA development. Giving the agent the capacity to interrupt or respond to an interruption helps to improve and engage the communication [12].

2 BACKGROUND & RELATED WORKS
Interruptions are natural and frequent in real interactions. They can be regarded as a deviation from the simple turn-taking model, to mediate the content and redirection of a conversational exchange. Interruptions can be broadly divided into two strategies: competitive and cooperative interruptions [4]. Both interruption strategies are very similar in their local discourse characteristics, but their global roles in helping interlocutors to exchange information are quite different [8].

Competitive interruption occurs when the listener interrupts to control the interaction, usually disrupting the flow of dialogue between the partners and can be seen as a conflict. A competitive interruption could be [8]:

- Disagreement: the listener disagrees with the speaker and wants to express his opinion immediately.
- Floor taking: the switch does not change the topic of the conversation and aims to develop the topic in the speaker’s place.
- Topic change: to accomplish the task by changing the topic.
- Tangentialization: the listener summing up information from the current speaker to prevent listening to unwanted information.

On the opposite, cooperative interruption helps to complete the current turn [8]:

- Agreement: show agreement, compliance, understanding or support.
- Assistance: the listener provides the current speaker with a word, a phrase or an idea.
- Clarification: to understand the message sent by the speaker. The purpose is to ask the current speaker to clarify or explain information about which the listener is not clear.

Beattie [1] has defined a taxonomy of interruptions as shown in Figure 1. In our work, as a start, we focus on simple, silent, butting-in interruptions.

Lee et al. [7] used both speaker’s acoustic cues and listener’s gestural cues to predict the future interruption in a multimodal dyadic interaction corpus. Chylek and colleagues [3] developed a deep residual network to model acoustic features and predict the timing of interruption.

We noticed that the two predictive models focus on predicting the timing of an interruption, leaving aside to recognize its type.

To automatically distinguish different types of interruptions, Lee et al. [6] analysed the differences in speech intensity, hand motion, and disfluency between cooperative and competitive interruptions. Yang et al. [14] also mentioned acoustic and prosodic differences. Competitive interruptions have typically higher pitch and louder amplitude to gain attention while cooperative interruptions often occur at low or medium pitch levels because of their non-competitive nature.

Most of the classification models are based on features estimated on a long temporal window and thus do not allow to generate non-verbal features in real-time human agent interaction.

3 OBJECTIVE & WORK IN PROGRESS

We plan to develop an ECA as a tool for Social Skill Training (SST) for a large variety of population facing difficulties when interacting with others. Our model will be integrated within the GRETA [10] platform that allows real-time human-agent interaction. We will consider two main situations:

- Interruptions raised by the ECA: when and how to interrupt the human user, including the decision of interruption timing, interruption type and the decision after interruption, whether to grab the turn or to abandon the interruption, depending on the human user’s reaction. Reinforcement learning model has been successfully used to model decision making process. Our RL based decision model takes multimodal input comes from both user and ECA: facial expressions (Action Units), body movement (hand position, body rotation, head movement), gaze direction and also acoustic features (F0, Energy, MFCC). It will also take dialog acts of human user’s speech.

- ECA interrupted by human user: how to respond to human interruptions, whether to ignore the interruption to continue with the current turn, or to stop and yield the current speaking turn to the human user. We propose to develop a multimodal learning model based on Transformer [13] network, taking as input the nonverbal behavior of the human user (facial expression, body movement, gaze direction, acoustic features). Speech content of both the agent and the human will also serve as input to the model with dialog act and keywords. For these last features we will rely on incremental dialog processing technology [5].

We also consider the agent’s behavioural reaction. That is, we also aim to develop a behavioural generation model to compute the agent’s verbal and non-verbal behaviour following different interruptions.

4 CORPUS ANALYSIS & ANNOTATION

As a first step toward developing computational model for the virtual agent, we started studying human-human interaction. We use two corpora, IEMOCAP [9] and NoXi [2]. IEMOCAP is constituted of 80 transcripted, emotion annotated (happiness, anger, sadness, frustration and neutral state) American dyadic spontaneous conversations, while NoXi consists of free conversation about 45 given topics. We chose to use the French dyads for our research.

We manually annotated the interruptions for both IEMOCAP and NoXi databases with the annotation schema in Figure 2.

At each change in the vocal track, we first classify it into interruption, back-channel, and smooth turn-exchange. We can note that interruption may not always consists of overlaps and smooth turn-exchange of gaps. Back-channels occur when the listener interjects phatic responses to the speaker, for example 'uh-huh', 'hmm' or 'yeah'; they are short response messages not aiming at taking the floor. Turn-exchange corresponds always to successful exchanges of speaking turn. On the other hand, interruption corresponds to a successful or failed attempt to grab the turn [1]. In both cases, an interruption can either be cooperative or competitive. We mark it "other" when it is a failed interruption and the interrupter stops grabbing the turn too fast to understand its type (i.e., there is not enough content).

We annotated 953 interruptions in IEMOCAP corpus and 1367 in NoXi. After a first analysis, we found that in both corpora, successful interruptions are significantly more than failed interruptions (IEMOCAP: 92.4% vs. 7.6%, NoXi: 87.52% vs. 12.48%), and most interruptions come up with overlaps (87.6% vs. 12.4%). "Floor taking" takes the largest place of competitive interruptions (43.3% in IEMOCAP, 63.6% in NoXi), and "Agreement" takes the largest place in cooperative interruptions (63.7% in IEMOCAP, 79.02% in NoXi).

5 CONCLUSION

The objective of our research is to improve the capacity of ECA to handle interaction, in particular interruptions. We aim to provide the ECA with the capacity to interrupt, but also to react to human’s interruptions.
ACKNOWLEDGMENTS
This work was performed as a part of TAPAS and PANORAMA project.

REFERENCES