The Effects of Interrupting Behavior on Interpersonal Attitude and Engagement in Dyadic Interactions

Angelo Cafaro, Nadine Glas and Catherine Pelachaud
LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay
{angelo.cafaro, nadine.glas, catherine.pelachaud}@telecom-paristech.fr

ABSTRACT

Interruptions frequently occur in dyadic human interaction. In addition to serve as turn-taking mechanism, they may lead to different perceptions of both the interruptee and interrupter's interpersonal attitude, engagement and involvement. We present an empirical study to investigate whether different interruption types (i.e. amount of overlap between speakers and utterance completeness) and strategies (disruptive vs. cooperative) in agent-agent interaction have an impact on perceived agents' interpersonal attitude, engagement and involvement. We found that the interruption type has more influence on the perceived attitudes of both agents, whereas by using a cooperative strategy (as opposed to a disruptive one) an interrupter is perceived as more engaged and more involved in the interaction.

Keywords

Turn-taking; Interruptions; Interpersonal attitude; Engagement; Empirical evaluation; Embodied conversational agents

1. INTRODUCTION

The organization of turn-taking is fundamental in human-human interaction [29]. Interruptions represent a violation of the basic turn-taking rules and a possible way of claiming the turn [4]. An interruption is "a starting up of some intervention by one person while another's turn is in progress" [28]. In order to model socially believable conversational agents in human-agent and/or agent-agent interaction, it is important to consider and to handle phenomena, such as interruptions, that frequently occur between humans [9].

Interruptions have long been associated with interpersonal dominance [35], but more recently a more balanced view has emerged, distinguishing two **strategies** based on the content of the interruption: *disruptive* and *cooperative* [24, 25]. At the same time, different **types** of interruptions can be distinguished from an organizational turn taking point of view, considering the amount of simultaneous speech and utterance completeness [10].

Different interruption types and strategies influence the way in which interactants are perceived with respect to their mutual interpersonal attitude [4] and interaction-related aspects such as engagement and involvement [32]. Human-agent interaction is often modeled on human-human interaction theories [7], therefore interruption mechanisms may have similar effects in human-agent interaction. For example, a user's interrupting behavior while interacting with a conversational agent might lead to different perceptions of **interpersonal attitudes** (e.g. dominance and friendliness in Argyle's representation [1]), **engagement** (i.e. the coordination and cooperation that is necessary to perform the joint activity of interacting [12]) and **involvement** in the interaction ("being captured by the experience" [31]).

Previous findings showed that an agent's turn-taking behavior influences human's impressions of it in terms of interpersonal attitude and personality [21]. However, we believe that interrupting behavior is not always a sign of power, disturbance and/or disengagement, instead the interplay between different interruption types and strategies might reveal effects on the perceived attitudes (e.g. friendliness and dominance), engagement and involvement of an interrupter. For example, a cooperative interruption strategy might be perceived as indicator of increased friendliness and engagement.

The assessment of these socio-emotional qualities in dyadic interaction represents a fundamental aspect for building affective and socio-believable intelligent conversational agents. We questioned whether deploying specific interruption strategies (e.g. cooperative) has greater impact on the above mentioned assessments compared to interruption type (i.e. amount of overlap between the interactants during the interruption and utterance completeness) or vice-versa. Is an interrupter agent by definition always dominant? What is the impact of interrupting behavior on the agents' engagement level?

In this paper, in the context of the European project *ARIA-VALUSPA* on affective information retrieval through a virtual assistant, we present a study aimed at investigating the effects of interruption strategies and types, in agent-agent interactions, on human perception of both agents' interpersonal attitudes towards each other, level of engagement, and involvement in the interaction. We considered a dyadic agent interaction as it allowed us a complete systematic control of both the **interrupter** and the **interruptee**'s behavior. While the main focus is on the interrupter's behavior, we also took into account the user's perception of the interruptee's attitudes and engagement/involvement as exploratory assessments.

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

2. THEORETICAL BACKGROUND

2.1 Interruption Types

An interruption is defined as a starting up of some intervention by one person while another's turn is in progress, often including "not letting them finish" [28]. An overlap refers to the fact of more than one person talking at a time [28], which means that an interruption does not necessarily implies overlap. There are multiple ways in which interruptions can be further subdivided based on the temporal aspects of the speaker's switch (e.g. [18], [10], [28]). We use the taxonomy of attempted speaker-switches by [10] that was later modified by [4]. Figure 1 shows this taxonomy and the types of interest in the present study are marked in boldface (more details about the experimental design are provided in Section 4.1.1). The classification makes a first division

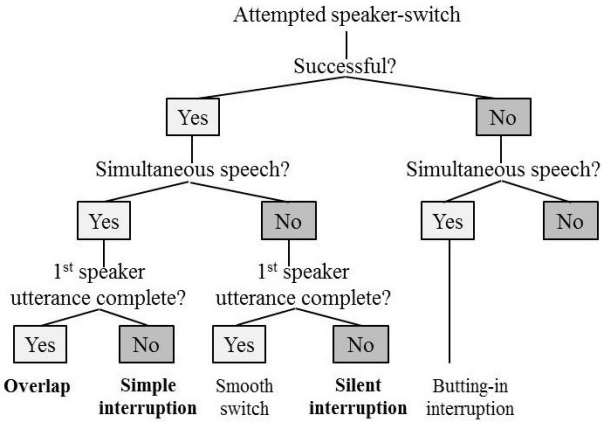

Figure 1: Beattie's taxonomy of interruption types.

according to whether or not the speaker-switch is successful, meaning that the initiator of the attempted interruption takes the floor [4]. Further division is based respectively on the presence of simultaneous speech, and first speaker utterance completeness. In this taxonomy back-channels (such as *yeah, hmm, exactly*) are not considered as interruptions [4].

2.2 Interruption Strategies

Interruptions have been often considered power displays (i.e. dominance), while recently a more balanced view has emerged [18]. This view has led to the identification of two broad types of interruptions, often referred as intrusive and cooperative [18, 24]. In this paper, we name these two types **strategies**. We adopted the terms *disruptive* [25] and *cooperative* as they refer to the interaction, where other terms may already imply an attitude or stance of the participants (i.e. (non) power [14], (dis) confirming [16]).

According to [24] "cooperative interruptions are intended to help the speaker by coordinating the process and/or content of the ongoing conversation" [18]. "Intrusive interruptions pose threats to the current speaker's territory by disrupting the process and/or content of the ongoing conversation" [18, 24, 14].

Researchers identified different sub-strategies (cooperative/disruptive). For example, in [24] do not mention any cooperative sub-strategy, while [18, 16] mention three: **(1) Agreement**, showing concurrence, compliance, understanding or support. Sometimes also serving as an extension or elaboration of the idea presented by the speaker; **(2) Assistance**, providing the current speaker with a word, phrase, sentence or idea in order to help completing his/her utterance and **(3) Clarification,** having the current speaker clarify or explaining previously elicited piece of information that the listener is unclear about.

In [18] they further combine four disruptive sub-strategies from [24] and [16]: **(1) Disagreement**, disagree with what the current speaker is saying and want to voice his or her opinion immediately; **(2) Floor taking**, developing the topic of the current speaker by taking over the floor from the current speaker; **(3) Topic change**; and **(4) Tangentialization**, reflecting awareness, usually by summarization, of the information being sent by the current speaker. Tangentialization prevents the interrupter from listening to an unwanted piece of information either because it has been already presented previously or because it is already known to the listener.

In our study we needed to consider the sequence organization of the interactions (conversational analysis) [29] as this may (also) influence how an interrupting agent is perceived. More precisely, we needed to consider the adjacency pairs of which the interruption might be a part of. An adjacency pair consists of two speakers' turns where the second part of a pair is responsive to the action of the first part [29]. Adjacency pairs examples are offer-accept/decline and question-answer [29]. The basic rule of operation for a second speaker is to produce a second part of a pair according to the type of the first part. For example, the sequence *assessment-(dis)agreement* [17] may be considered more natural than the sequence *assessment-question*.

In this study we are not interested in the possible different consequences of the second pair part types among each other, but we are aware that they can be perceived differently. We therefore focused on the distinction between cooperative and disruptive strategies for each set of second pair parts separately (e.g. agreement/disagreement) (details in Section 4.1.2).

2.3 Interpersonal Attitudes

Interpersonal attitudes are essentially an individual's conscious or unconscious evaluation of how s/he feels about and relates to another person [1, p. 85]. Several researchers attempted to identify the dimensions that can best represent the different interpersonal attitudes expressed during social interaction. Schutz [30] proposed the dimensions of inclusion, control and affect. Burgoon and Hale [6] identified twelve dimensions defining different communication styles (e.g. dominance, intimacy, confidence). Argyle proposed a two-dimensional representation. A first dimension is the **affiliation**, characterized as the degree of liking or friendliness and ranging from unfriendly to friendly. A second dimension is the **status**, related to power and assertiveness during the interaction and ranging from submissive to dominant [1, p. 86].

We adopted Argyle's model because it has an intrinsic simplicity (i.e. only two dimensions) and it offers the best compromise between explanatory power and parsimony.

2.4 Engagement and Involvement

Engagement is a concept that has many interpretations in the domain of human-agent interaction [12]. All of them,

however, refer more or less directly to the coordination and cooperation that is necessary to perform the joint activity of interacting [12]. Our interpretation of engagement comes from [26] as "the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction".

Involvement refers to "being captured by the experience" [31]. [3] defines five aspects of involvement: (1) Immediacy (physical proximity); (2) Expressiveness (energy, activity, enthusiasm); (3) Altercentrism (focus on the conversation parner); (4) Interaction management (smooth flow); (5) Composure dimesion (body movement, confidence) and (6) Positive affect (good feelings). We used the interpretation of [31], which refers to aspects 2, 3 and 4. While involvement can be closely related to engagement [12, 19, 20], in this paper we differentiate it as it might provide us more fine grained qualities of the dyadic interaction.

We expected that initiating one's turn exactly when the other's turn finishes may be a sign of coordination and thus engagement. Previous research showed that leaving pauses between turns creates the feeling of having more rapport (i.e. a feature of engagement) [21], while other research in human-human interaction has shown that simultaneous speech can be a way to show involvement in an interaction [32].

3. RELATED WORK

Turn-Taking in Perception Studies. Maat et al. [21] showed how different turn-taking strategies (including interruptions) in a user-agent interaction could influence users' impressions of an agent's personality (agreeableness), emotion and social attitudes. However, their analysis did not take into account the speech content nor the interruption strategy. Thórisson and colleagues [33] evaluated a turn-taking model (YTTM) in terms of scalability and believability, however they focused more on timing and prosody features for autonomously assign the turn in a multi-party agent interaction. In [27] a multi-agent model for reflecting interpersonal attitudes in conversing groups has been presented. Their emphasis was on nonverbal behavior (e.g. interpersonal space, gestures) and turn-taking strategies, however they did not consider verbal behavior and different interruption types/strategies.

Interruptions. Crook and colleagues [9, 8] presented a mechanism for handling "barge-in" interruptions from a user interacting with an embodied conversational agent. Their agent was able to detect and respond to user interruptions. The handling process consisted of an address phase — i.e. the agent addressing the particular interruption that occurred — and a resumption phase where the agent implemented context-sensitive strategies for continuing or aborting the current conversational plan. While they modeled an handling mechanism that considered the affective user's perception of the interruptee (i.e. the agent), they did not consider the effects of the interrupter (the user in their case), strategy of interruption on perceived interpersonal attitude, engagement, and involvement in the interaction with the agent.

In [15] they examined interruptions in a corpus of spontaneous task oriented dialogue. They focused on timing and acoustic/prosodic features that predict interruptions, whereas we investigate the effects of different interruption strategies and types (i.e. amount of overlap between speakers and utterance completeness, while keeping constant the timing of the interruption) on perceived interpersonal attitudes, engagement and involvement of the interactants.

In sum, previous work focused on turn-taking strategies' impact on perceived agent's personality and attitude [21], handling of interruptions [9, 8] and predicting them [15]. However, to the best of our knowledge no one examined the effects of different interruption strategies and types (according to the definitions provided earlier) on perceived interpersonal attitudes, engagement level and involvement in the interaction between two participants. The next section presents the experimental design of a web study aiming at addressing these aspects.

4. EXPERIMENTAL DESIGN

In order to evaluate the effects of different interruption types and strategies on the human perception of interpersonal attitude (i.e. dominance and friendliness), engagement and involvement in a dyadic agent interaction, we designed an empirical experiment. In particular, we designed a web study in which participants listened to a series of conversational fragments between two agents. The application scenario for our virtual assistant in the ARIA-VALUSPA project is the book *"Alice in Wonderland"* by Lewis Carroll. We therefore adapted the subject of our fragments conforming to this scenario. For each fragment we systematically varied the interruption type (described in Section 2.1) and strategy (introduced in Section 2.2) for a given subcategory (i.e. communicative function of the interruption as described later). After listening to each fragment, participants rated the perceived level of *dominance*, *friendliness*, *engagement* and *involvement* of the two agents which are referred as the **Interruptee** (or **A**) and **Interrupter** (or **B**) in the remainder of the paper.

4.1 Stimuli Design and Preparation

We based the generation of interruptions on existing literature in human-human interaction. However, there are many ways for producing an interruption of a particular type and strategy. For practical reasons, we examined four categories of comparable interruption strategies (later defined as *functions*).

Furthermore, we were aware of possible gender influences on the interactants behavior [36, 4, 35] and the perception of their attitudes [4, 35, 14]. We considered the agents' gender as a blocking factor. Therefore, while we designed our study with this aspect in mind, as a first step and considering that our primary interest was on interruption types and strategy effects, we started with two male gendered agents. In the following sections we describe the design of the different interruption types and strategies.

4.1.1 Interruption Types

We focused on the successful interruptions, as shown in Figure 1, where the interrupter takes the floor. This was necessary in order to make it possible for the interrupter to utter a complete sentence (in the other cases the sentence is partially uttered), and thus effectively deploy an interruption *strategy* (i.e. disruptive vs. cooperative).

In [4] they illustrate each of these attempted speaker-switches with an example, which also gave us some indications about the possibility of pauses in between the turns. We did not consider the smooth speaker switch as it represents an unproblematic "smooth" way of taking the floor and

was not of interest for us in terms of interruption type. According to the classification and examples provided in [4], we generated conversational fragments consisting of sequential turns that imitate each of these interruption types, shown in table 1. Synchronous speech is indicated between squared brackets. The length of the synchronous speech is not only the result of theoretical considerations but depends further on the speech duration of the agents' synthesized speech. The Interruptee is indicated as **A** in the "Agent" column and it always initiates the conversational fragments, whereas the Interrupter is indicated with **B**.

4.1.2 Interruption Strategies

We based the generation of disruptive and cooperative strategies on the categories described in Section 2.2. We examined four categories of comparable strategies, respectively disruptive vs. cooperative, as follows: (1) *Question implying misunderstanding – Clarification question*; (2) *Disagreement – Agreement*; (3) *Tangentialization – Completion*; (4) *New topic introduction – Topic elaboration.*

For practical reasons, we gave a name to each of these categories, that we call the *function* of the interruption and that is based on the commonality of both interruptions in the set: (1) *Understanding question*: both sub-strategies are consequences of the interrupter's (mis) understanding of the interruptee's turn; (2) *Opinion*: both express the interrupter's (dis) agreement with the interruptee; (3) *Partner Communication*: both try to adjust (future) utterances of the interruptee; (4) *Topic*: both strategies contribute to the topic management of the interaction.

In order to verify that for each function the instances of the interruption strategies that we created were indeed perceived as disruptive/cooperative strategies, we performed a manipulation check.

4.1.3 Manipulation Check

Participants rated how disruptive/cooperative they found the interrupter's utterance in the conversation fragments (in written format on a web page), on a 5-points Likert scale (anchors *very disruptive* and *very cooperative*). The second speaker turns were placed directly after the end of the first speaker's turn, corresponding to [4]'s example of smooth speaker-switches (Section 2.1), thereby serving as a baseline.

Results from 11 participants (7 males) indicated that strategies in all functions were correctly recognized, except for the disruptive one in the *Opinion* function (i.e. disagreement). Contrary to our expectations, it was judged, on average, as being closer to cooperative ($M = 3, S.E. = 1.3$). This can be explained by the fact that the disagreement was expressing a mitigated disagreement with respect to the first speaker's evaluation of a story (*Well, that's debatable, it's not my favorite one.*). Therefore, we refined the sub-strategy by creating a direct (non-mitigated) disagreement presented as a factual one. This lead us to use the utterance "*No, it does not tell an amazing story at all*". The final validated list of utterances used in our study is shown in Table 1.

4.1.4 Independent Variables

We split the study in 4 trials according to our **Functions** of interest: *Understanding Question*, *Opinion*, *Partner Communication* and *Topic*. In each trial (i.e for each function), our independent variables (IVs) were **Interruption Type** (with levels *overlap*, *simple* and *silent*) and **Strategy** (*disruptive* vs. *cooperative*). We designed each trial as a 3x2 within-subjects study where the type and strategy were within-subjects factors.

4.1.5 Stimuli Preparation

We produced a total number of 24 conversation fragments (6 for each function/trial) as summarized in Table 1, using a male synthesized voice created with the Cereproc text-to-speech (TTS) tool [2]. We opted for a single synthesized voice as opposed to pre-recorded speech to control for additional biases in the experiment. Indeed, a human speaker might pronounce the same sentence (e.g. A's sentences in our stimuli) with different intonations. We presented our stimuli in the form of videos with subtitles and stereo sound. The subtitles served for compensating the possible lack of clarity due to the synthesized nature of the voice, thus making sure that the participants were correctly exposed to our stimuli. Whereas the stereo sound, in addition to subtitles, allowed us to use the same voice but at the same time clearly discriminate speakers based on their voice sources (i.e. left and right channels). Participants of a pilot test (2 males and 2 females) thought that we used different voices "since the speech came from different audio sources".

All videos displayed two still male head silhouettes facing each other on a black background. The silhouettes' color was white grey and they had a label on top respectively showing "Left Agent" and "Right Agent". The subtitles progressively appeared below the corresponding silhouette as the words were audible.

In order to reduce confounding effects, in each conversation fragment (i.e. condition) the interruptee (i.e. first speaker) was uttering the same sentence when it was interrupted. We also kept the onset of the interruptions [28] equal among the different interruption types, meaning that the second speaker (i.e. interrupter) begins speaking **at the same moment** relative to the interruptee's current turn. Only in the silent interruption case we added a small pause in the interruptee's turn following the examples in [4]. For this we added 0.2 seconds of white noise at an amplitude of 0.001 in order to keep the left channel (i.e. of the interruptee) open and to avoid an abrupt termination that might have sounded as a recording error.

Furthermore, in Beattie's examples interruptions occur at possible *completion points* where a speaker's utterance could be produced [23]. There are three aspects that participants in an interaction use to determine possible completion points: syntax, prosody and pragmatics [23]. We generated conversation fragments where the second speaker interrupts at the moment where the first speaker's turn can be judged syntactically and pragmatically complete. However, we did not manipulate prosody features in our stimuli in order to control for possible effects across conditions. Thus, we did not provide the TTS with specific prosody instructions, and we deployed the synthesized output in our videos as it was produced by the tool. In order to compensate for unnatural prosody caused by an imperfect speech synthesis we added punctuation to the subtitles.

4.2 Measurements

For each stimulus, we asked the participants to rate the perceived **attitudes** (**dominance** and **friendliness**) of the agents towards each other, and their level of **engagement** and **involvement** in the interaction.

Table 1: Interruption types and strategies for the function categories investigated in our study. The Interruptee (A) is the first speaker, the Interrupter (B) is the second speaker. In the silent interruption (0.2) refers to a silent pause in seconds.

Interruption Type	Function	Interruption Strategy	Agent	Utterances	
Overlap			A	You know I've read the story Alice in Wonderland. It tells an amazing story [about a little girl.]	(1)
	Understanding Question	Disruptive	B	[When were you in] the Wonderland theme park?	(2)
		Cooperative	B	[Do you mean the book writ]ten by Lewis Carrol?	(3)
	Opinion	Disruptive	B	[No, it does] not tell an amazing story at all.	(4)
		Cooperative	B	[Great! I] love it and I've read it several times.	(5)
	Partner Communication	Disruptive	B	[I've read the book s]o I know the story.	(6)
		Cooperative	B	[About a busy rab]bit and a smiling cat.	(7)
	Topic	Disruptive	B	[I borrowed the book B]eauty and the Beast.	(8)
		Cooperative	B	[It was first rele]ased in 1865.	(9)
Simple Interruption			A	You know I've read the book Alice in Wonderland. It tells an amazing story [about]	(10)
	Understanding Question	Disruptive	B	[When] were you in the Wonderland theme park?	(11)
		Cooperative	B	[Do you] mean the book written by Lewis Carrol?	(12)
	Opinion	Disruptive	B	[No,] it does not tell an amazing story at all.	(13)
		Cooperative	B	[Great]! I love it and I've read it several times.	(14)
	Partner Communication	Disruptive	B	[I've re]ad the book so I know the story.	(15)
		Cooperative	B	[About] a busy rabbit and a smiling cat.	(16)
	Topic	Disruptive	B	[I bo]rrowed the book Beauty and the Beast.	(17)
		Cooperative	B	[It was] first released in 1865.	(18)
Silent Interruption			A	You know I've read the book Alice in Wonderland. It tells an amazing story (0.2)	(19)
	Understanding Question	Disruptive	B	When were you in the Wonderland theme park?	(20)
		Cooperative	B	Do you mean the book written by Lewis Carrol?	(21)
	Opinion	Disruptive	B	No, it does not tell an amazing story at all.	(22)
		Cooperative	B	Great! I love it and I've read it several times.	(23)
	Partner Communication	Disruptive	B	I've read the book so I know the story.	(24)
		Cooperative	B	About a busy rabbit and a smiling cat.	(25)
	Topic	Disruptive	B	I borrowed the book Beauty and the Beast.	(26)
		Cooperative	B	It was first released in 1865.	(27)

For measuring each attitude's dimension we adapted 2 statements from the Riverside Q-Sort inventory [11], using the two most reliable items (i.e. adjectives or key terms) of positive and negative valence as defined in Wiggin's interpersonal circumplex inventory [34]. The items adopted to assess dominance were: "tries to control" and "seems insecure". For friendliness they were: "tries to be likable" and "expresses hostility". For example, a question about the Interrupter dominance (i.e. right agent in our stimuli) towards the Interruptee (i.e. left agent) was: "*The **Right** Agent tries to control the interaction with the Left Agent*". All answers were on a 5-points labeled Likert scale (anchors *1. Completely disagree* and *5. Completely agree*).

For assessing the agents' level of engagement we asked two questions on a 5-points labeled Likert scale (anchors *1. No Value* and *5. A Maximum Value*) that are based on [26, 13]. For instance, the questions for the Interrupter were: (1) *"What value the **Right** Agent attributes to being together with the Left Agent?"* and (2) *"What value does the **Right** Agent attributes to continuing the interaction?"*.

We also assessed the agents' level of involvement as recommended in [31] by asking a question on a 5-points labeled Likert scale (anchors *1. Not at All* and *5. Very Much*). For the Interrupter, for instance, we used: *"How engaging was the interaction for the Right Agent?"*. In [31], more questions are proposed for measuring involvement, however we selected this one as it was more suitable for our design and linked to engagement [19, 20].

A summary of our dependent variables (DVs) is shown in Table 2. As naming convention we preceded the variable names with A and B referring, respectively, to the Interruptee and Interrupter.

4.3 Hypotheses

4.3.1 *Interpersonal Attitudes*

Disruptive interruption strategies in human-human interaction increase the perceived dominance of an interrupter [14, 35], whereas cooperative strategies increase affiliation (i.e. liking or friendliness) [14]. Therefore we expected that the perceived dominance and friendliness of the interrupter agent would be affected by the interruption strategy adopted by the interrupter (i.e. main effect).

As for the interruption type, in the stimuli that we created

Table 2: Summary of the measures. DV names are preceded by A and B referring, respectively, to Interruptee and Interrupter.

MEASURE	DV Names
Attitude: Dominance	A-Dom, B-Dom
Attitude: Friendliness	A-Friend, B-Friend
Engagement	A-Eng, B-Eng
Involvement	A-Inv, B-Inv

the interruption occurs always at the same moment during the interruptee's turn, therefore the difference among the levels of our interruption types (i.e. *overlap*, *simple* and *silent*) lies in the amount of overlap between the two interactants when the interruption occurs (respectively *long* overlap, *short* and *none*). In human-human interaction a speaker holding the turn (in our case resulting in a long overlap) is perceived as more dominant [4, 14] and less likable [14]. Considering that the interrupter agent always completes its utterance (in order to deploy a specific strategy) whereas the interruptee agent varies the amount of produced overlap when the interruption type changes, we predicted that the perceived dominance and friendliness of the interruptee agent would be affected by the interruption type (main effect).

In light of these observations, for each trial (i.e. function examined) we hypothesized the following:

- **H.A-Dom (Interruptee)**: The **Interruption Type** will have a main effect on perceived Interruptee's Dominance, the *higher* the overlap (i.e. from none to long) the *higher* will be **A-Dom**.
- **H.A-Friend (Interruptee)**: The **Interruption Type** will have a main effect on perceived Interruptee's Friendliness, the *higher* the overlap (i.e. from none to long) the *lower* will be **A-Friend**.
- **H.B-Dom (Interrupter)**: The **Interruption Strategy** will have a main effect on perceived Interrupter's Dominance, the disruptive strategy will lead to *higher* **B-Dom** compared to the cooperative one.
- **H.B-Friend (Interrupter)**: The **Interruption Strategy** will have a main effect on perceived Interrupter's Friendliness, the cooperative strategy will lead to *higher* **B-Friend** compared to the disruptive one.

4.3.2 *Engagement and Involvement*

We expected that when an agent is interrupted with a cooperative strategy it would show more engagement and involvement compared to a disruptive one. Cooperative interruptions coordinate the process and/or content of the conversation, while intrusive ones disrupt the process and/or content of the ongoing interaction. Given that engagement is characterized by a coordination between the interaction participants [12], we expect that cooperative interruptions are signs of more interrupter engagement and involvement than disruptive ones (i.e. main effect). Moreover, with respect to the interruption types previous research has shown contradictory effects on engagement related measures regarding the interrupter (Section 2.4) making it unclear to predict a specific outcome.

To the best of our knowledge, previous research did not provide evidence of interruption type/strategy effects on an interruptee's engagement level. We believe that involvement might increase when this agent finishes its turn. However, we considered the interruptee's engagement and involvement as exploratory measures, and thus we did not provide specific hypotheses for those.

In sum, for the interrupter agent in each trial (i.e. function examined) we hypothesized the following:

- **H.B-Eng (Interrupter)**: The **Interruption Strategy** will have a main effect on perceived Interrupter's Engagement, a cooperative strategy will lead to *higher* **B-Eng** compared to a disruptive one.
- **H.B-Inv (Interrupter)**: The **Interruption Strategy** will have a main effect on perceived Interrupter's Involvement, a cooperative strategy will lead to *higher* **B-Inv** compared to a disruptive one.

4.4 Participants and Procedure

We recruited a total of 72 participants via mailing lists (18 in each trial). 46% of the participants were between 18 and 30 years old, 34% between 31 and 40. 57% were male, 39% were female and 4% preferred to not say it. 82% were well educated (above master level). They had different cultural backgrounds[1], the two most prominent groups were France (20.8%) and USA (14%).

We ran this study online. A participant was first presented with a consent page and a sound check ensuring that audio could be played. Then, we showed a tutorial page including the instructions, a video sample and the questionnaires underneath the video player. Once the tutorial was completed, we randomly assigned a participant to a trial (i.e. function), (s)he played the corresponding videos (6) and answered the questions in a within-subjects repeated measures design. To control for first order carryover effects we adopted a counter balanced treatment order for showing the videos according to a latin square design, as recommended in [5].

4.5 Results

We conducted 4 separate statistical analyses for each function that we examined (i.e. trial). An informal comparison among them is discussed in Section 5. For each trial, we conducted 3x2 repeated measures MANOVAs (Multivariate Analysis of Variance) on **Interruptee**'s and, respectively, **Interrupter**'s **Dominance** and **Friendliness** (A-Dom and A-Friend, B-Dom and B-Friend), **Engagement** and **Involvement** (A-Eng and A-Inv, B-Eng and B-Inv). The within-subjects factors were Interruption **Type** (3) and **Strategy** (2).

Prior to conducting the analyses we normalized all DVs in the range $[0-1]$. Due to space constraints we only report significant main effects and interactions of the follow-up UNIVARIATE analyses (sphericity assumption was not violated). We refer to Interruptee (**A**) and Interrupter (**B**) in the reminder of the paper. Simple main effects of interactions between factors were tested using Bonferroni adjustments for multiple comparisons. Effect sizes (η_p^2) for all

[1]As part of the demographic information, we asked participants to indicate the nationality that most represented their cultural identity.

comparisons ranged from **.18** to **.76**. Figure 2 provides a summary of our quantitative findings for all trials.

4.5.1 Function: Question

A's Dominance and Friendliness. The analyses revealed a main effect of Type on A-Dom ($F(2,34) = 9.1$, $p < .001$). As the amount of overlap increased the perceived A's dominance increased (*none*: M=.48, SE=.04; *short*: M=.53 SE=.03; and *long*: M=.64 SE=.04), therefore **H.A-Dom** is **supported**.

Type also had a main effect on A-Friend ($F(2,34) = 3.6$, $p < .05$). In this case an increasing overlap decreased the perceived A's friendliness (*none*: M=.74, SE=.04; *short*: M=.71 SE=.03; and *long*: M=.66 SE=.04), thus the hypothesis **H.A-Friend** is **supported**.

B's Dominance and Friendliness. We found a main effect of Type on B-Dom ($F(2,34) = 4.0$, $p < .05$), therefore **H.B-Dom** is **rejected**. When the agent interrupted in silence (i.e. no overlap) it was perceived significantly less dominant (M=.70, SE=.04) compared to the two other interruption types (*short*: M=.78, SE=.04; *long*: M=.78, SE=.03).

For friendliness we discovered two main effects of Type ($F(2,34) = 7.9$, $p < .005$) and Strategy ($F(1,17) = 11.9$, $p < .005$) but no significant interaction effects, thus **H.B-Friend** is **partially supported**. We decomposed these effects and discovered that in general a cooperative Strategy accounts for higher friendliness (M=.53, SE=.04) compared to a disruptive one (M=.42, SE=.04). However, when using a disruptive Strategy, the agent was perceived significantly more friendly when interrupting in silence (M=.53, SE=.05) in comparison with the other two Types (*short* overlap: M=.32, SE=.05; *long*: M=.39, SE=.04).

A's Engagement and Involvement. We did not find any significant effect on A-Eng. However, we found a significant interaction effect between Type and Strategy on A-Inv ($F(2,34) = 4.4$, $p < .05$). More specifically, when interrupted in silence (i.e. no overlap), A is considered more involved in the interaction when B's strategy is cooperative (M=.60, SE=.04) compared to disruptive (M=.70, SE=.03).

B's Engagement and Involvement. We found two main effects of Type ($F(2,34) = 5.5$, $p < .05$) and Strategy ($F(1,17) = 17.0$, $p < .005$) on B-Eng, but no significant interaction effects, thus **H.B-Eng** is **partially confirmed**. The decomposition of these effects revealed that, in general, B was more engaged when it used the cooperative Strategy (M=.61, SE=.05) compared to the disruptive one (M=.45, SE=.05). However, when using a disruptive Strategy, it was perceived significantly more engaged when interrupting in silence (M=.54, SE=.05) in comparison with the other two Types (*short* overlap: M=.41, SE=.05; *long*: M=.40, SE=.06).

Strategy had a main effect on B-Inv ($F(1,17) = 12.5$, $p = .003$). Cooperative interruptions made B more involved (M=.59, SE=.04) than disruptive ones (M=.40, SE=.06). Therefore, **H.B-Inv** is **supported**.

4.5.2 Function: Opinion

A's Dominance and Friendliness. We did not find significant effects on A-Dom (**H.A-Dom rejected**). However, we found a significant interaction effect on A-Friend ($F(2,34) = 3.8$, $p < .05$), thus **H.A-Friend** is **partially supported**. In particular, for silent Type of interruptions (i.e. A not overlapping), when the other agent interrupted with a cooperative Strategy, A was perceived less friendly (M=.66, SE=.05) compared to the case when a disruptive strategy was used (M=.74, SE=.03).

B's Dominance and Friendliness. We did not find significant effects on B-Dom (**H.B-Dom rejected**). We found a main effect of Strategy on B-Friend ($F(1,17) = 54.8$, $p < .001$), therefore **H.B-Friend** is **supported**. B was more friendly when doing cooperative interruptions (M=.52, SE=.04) compared to disruptive ones (M=.26, SE=.03).

A's Engagement and Involvement. No significant effects were found.

B's Engagement and Involvement. We discovered a main effect of Strategy on B-Eng ($F(1,17) = 26.0$, $p < .001$) and B-Inv ($F(1,17) = 23.8$, $p < .001$), thus **H.B-Eng** and **H.B-Inv** are **supported**. In particular, interrupting with a cooperative Strategy led to higher engagement (M=.46, SE=.04) and involvement (M=.46, SE=.04) compared disruptive ones (M=.22, SE=.02 and M=.23, SE=.04).

4.5.3 Function: Partner

A's Dominance and Friendliness. We found a main effect of Strategy on A-Dom ($F(1,17) = 12.8$, $p < .005$). When B adopted a cooperative Strategy, A was perceived significantly more dominant (M=.60, SE=.03) compared to when a disruptive Strategy was used (M=.55, SE=.03). Therefore, **H.A-Dom** is **rejected**. We did not find significant main effects or interactions on A-Friend (**H.A-Friend** is **rejected**).

B's Dominance and Friendliness. We found a main effect of Type on B-Dom ($F(2,34) = 5.3$, $p < .05$), therefore **H.B-Dom** is **rejected**. B's dominance was significantly higher when A suddenly stopped its utterance in reaction to the interruption (Type=*simple*: M=.75, SE=.03), compared to the case where A took a pause in its utterance (Type=*silent*: M=.67, SE=.03).

For friendliness (B-Friend) we discovered two main effects of Type ($F(2,34) = 8.6$, $p < .005$) and Strategy ($F(1,17) = 30.8$, $p < .001$) but no significant interaction effects, thus **H.B-Friend** is **partially supported**. We decomposed these effects and discovered that for all interruption Types a cooperative Strategy accounts for higher friendliness (M=.52, SE=.04) compared to a disruptive one (M=.27, SE=.03). However, this effect is more remarked when Type is *silent* (*cooperative*: M=.62, SE=.05; *disruptive*: M=.33, SE=.04).

A's Engagement and Involvement. No significant effects were found.

B's Engagement and Involvement. The analysis revealed a main effect of Strategy on B-Eng ($F(1,17) = 30.7$, $p < .005$) and B-Inv ($F(1,17) = 15.5$, $p < .001$), therefore **H.B-Eng** and **H.B-Inv** are **supported**. In both cases a cooperative Strategy led to higher B's engagement (M=.42, SE=.04) and involvement (M=.43, SE=.03) compared to a disruptive one, respectively (M=.18, SE=.04) and (M=.21, SE=.04).

4.5.4 Function: Topic

For this trial we only found a significant main effect of Type on B-Friend ($F(2,34) = 7.9$, $p < .005$). In particular, B's friendliness was significantly higher when A took a pause in its utterance (Type=*silent*: M=.47, SE=.05), compared to the other two interruption Types (Type=*overlap*: M=.38, SE=.03, Type=*simple*: M=.35, SE=.03).

Function	INTERRUPTEE AGENT (A) Type (↗ *overlap*)	Strategy (*disr.* to *coop.*)	Type AND Strategy (*disr.* to *coop.*)	INTERRUPTER AGENT (B) Type (↗ *overlap*)	Strategy (*disr.* to *coop.*)	Type OR Strategy (*disr.* to *coop.*)
QUESTION	↗ Dominance ↘ Friendliness	*n.s.*	↗ Involvement (when type: *no overlap*)	↗ Dominance	↗ Involvement	↗ Friendliness ↗ Engagement (particular cases: see *Section 4.5.1*)
OPINION	*n.s.*	*n.s.*	↗ Friendliness (when type: *no overlap*)	*n.s.*	↗ Friendliness ↗ Engagement ↗ Involvement	*n.s.*
PARTNER	*n.s.*	↗ Dominance	*n.s.*	↗ Dominance	↗ Engagement ↗ Involvement	↗ Friendliness (more remarked when type: *no overlap*)
TOPIC	*n.s.*	*n.s.*	*n.s.*	↘ Friendliness	*n.s.*	*n.s.*

Figure 2: Summary of results for all functions examined. IVs are listed in the header. Inclined arrows indicate effects directions for the DVs in the cells. In the QUESTION trial, for example, changes in interruption Type (an increasing ↗ *overlap*) decreased A's perceived Friendliness ↘. For OPINION, by changing interruption Strategy from *disruptive* to *cooperative* B's perceived Engagement increased ↗.

5. DISCUSSION AND FUTURE WORK

Type and strategy of interruption had main and interaction effects on the mutual attitudes of the agents. Contrary to our expectations, type had greater importance compared to strategy on user's perceived dominance and friendliness of both A (interruptee) and B (interrupter).

In the "Question" trial, A's dominance increased (and friendliness decreased) as the amount of overlap with B increased. B was perceived less dominant during silent interruption types. This would suggest that no matter how disruptive/cooperative is a question, the amount of overlap between interactants has higher impact on users' perception of their dominance and friendliness.

For "Opinions" (i.e. disagreement vs. agreement), strategy had greater impact but only on friendliness. In particular, a cooperative strategy increases B's friendliness. However, for A this effect also depended on the type of interruption (i.e. *silent*). When B used a cooperative strategy in silent type interruptions, A' friendliness decreased and vice-versa for disruptive ones. We think that B has been perceived very hostile while placing a disruptive strategy (strong disagreement), making A looking more friendly in comparison. The lack of differences in dominance levels can be explained by the tendency of opinions (i.e. A's utterance when interrupted was "It tells an amazing story. . . ") to elicit a (dis) agreement, thus both strategies seemed appropriate.

In the "Partner Communication" trial, B's strategy had effects on A's dominance (B's disruptive strategy led to lower A's dominance compared to the cooperative one), whereas the type had impact on B's dominance as found in other trials. The first outcome can be explained by the notion introduced in [22] about complementarity of interpersonal attitudes in dyads, and stating that dominant behavior induces submissive responses.

Strategy had an effect on B's level of friendliness (more friendly when cooperative). In particular, this effect was more pronounced when the interruption happened after A's silence.

As for the engagement and involvement assessments, A's engagement was never influenced by B's interrupting behavior. Even if A stopped/continued its utterance in reaction. A's involvement was influenced by type and strategy only in the "Question" trial. A possible explanation is that a misunderstanding/clarification question is the only interruption in this study that can reveal something about A's quality of interacting. In the "Question", "Opinion", and "Partner Communication" trials cooperative strategies led to higher B's engagement and involvement levels compared to disruptive ones. This reflects the nature of cooperative interruptions as less face-threatening compared to disruptive ones [18], and displays of joint involvement [14].

The overall lack of significant effects in the "Topic" trial (except for a main effect of type on B's friendliness) can be explained by difficulties in capturing differences among the stimuli, as also reported by participants' feedback (6 out of 18).

In sum, the **turn-taking** mechanism (i.e. **type** of **interruption**) had more impact on the users' perception of **interpersonal attitudes** of both agents, though changing from a disruptive to a cooperative strategy increased interrupter's friendliness and reduced its dominance. We believe that for silent interruption types, when no overlap occurred, the content of the interrupter's utterances became more noticeable, thus increasing the effect of the two different strategies. On the other hand, the **strategy** had important main effects on **engagement** and **involvement** of the interrupter. These results have implications for the design of autonomous conversational agents that can be interrupted and can interrupt human users in a mixed-initiative incremental dialogue. When a cooperative interruption occurs, for example, the agent's mental model of the user (i.e. *Theory of Mind*) can be updated with the perceived user's intention of being friendly and subsequent agent's communicative intents can be (re)planned and adapted accordingly.

Future work should be considered. In human-human interaction personal characteristics such as the gender and status of the interactants can play a role in their interruption behavior [36, 4]. While we kept such variables stable in our study, they form interesting concepts to be considered in future work on interruptions in human-agent interaction. We also kept important factors, such as the timing of the interruption, constant across conditions to avoid further biases. However, we believe that manipulating the moment when the interruption occurs during the interruptee's turn might reveal interesting outcomes on the perceived interrupter's attitude. Finally, we are working on an agent able to manage interruptions (i.e. implementing different handling strategies in response to a user's interruption) or proactively interrupt the user by deploying a specific interruption strategy.

Acknowledgments

This work was supported by the European project H2020 ARIA-VALUSPA and the French ANR project MOCA. It was partially performed within the Labex SMART (ANR-11-LABX-65) supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02.

REFERENCES

[1] M. Argyle. *Bodily communication*. Methuen, New York, 2nd ed. edition, 1988.

[2] M. Aylett and C. Pidcock. The cerevoice characterful speech synthesiser sdk. In C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé, editors, *Intelligent Virtual Agents*, volume 4722 of *Lecture Notes in Computer Science*, pages 413–414. Springer Berlin Heidelberg, 2007.

[3] B. S. Bamoallem, A. J. Wodehouse, G. M. Mair, and G. A. Vasantha. The impact of head movements on user involvement in mediated interaction. *Computers in Human Behavior*, 55, Part A:424 – 431, 2016.

[4] G. W. Beattie. Interruption in conversational interaction, and its relation to the sex and status of the interactants*. *Linguistics*, 19(1-2):15–36, 1981.

[5] J. V. Bradley. Complete counterbalancing of immediate sequential effects in a latin square design. *Journal of the American Statistical Association*, 53(282):525–528, 1958.

[6] J. K. Burgoon and J. L. Hale. The fundamental topoi of relational communication. *Communication Monographs*, 51(3):193–214, 1984.

[7] J. Cassell. *Embodied conversational agents*. MIT press, 2000.

[8] N. Crook, D. Field, C. Smith, S. Harding, S. Pulman, M. Cavazza, D. Charlton, R. Moore, and J. Boye. Generating context-sensitive eca responses to user barge-in interruptions. *Journal on Multimodal User Interfaces*, 6(1-2):13–25, 2012.

[9] N. Crook, C. Smith, M. Cavazza, S. Pulman, R. Moore, and J. Boye. Handling user interruptions in an embodied conversational agent. In *Proceedings of the AAMAS International Workshop on Interacting with ECAs as Virtual Characters, Toronto*, pages 27–33, 2010.

[10] N. Ferguson. *Interruptions: speaker-switch nonfluency in spontaneous conversation*. PhD thesis, University of Edinburgh, 1975.

[11] D. C. Funder, R. M. Furr, and C. R. Colvin. The riverside behavioral q-sort: A tool for the description of social behavior. *Journal of Personality*, 68(3):451–489, 2000.

[12] N. Glas and C. Pelachaud. Definitions of engagement in human-agent interaction. In *Workshop on Engagement in Human Computer Interaction (ENHANCE) at ACII 2015*, pages 944–949, 2015.

[13] N. Glas and C. Pelachaud. User engagement and preferences in information-giving chat with virtual agents. In *Workshop on Engagement in Social Intelligent Virtual Agents (ESIVA) at IVA 2015*, pages 33–40, 2015.

[14] J. A. Goldberg. Interrupting the discourse on interruptions: An analysis in terms of relationally neutral, power-and rapport-oriented acts. *Journal of Pragmatics*, 14(6):883–903, 1990.

[15] A. Gravano and J. Hirschberg. A corpus-based study of interruptions in spoken dialogue. In *INTERSPEECH, 13th Annual Conference of the International Speech Communication Association*, pages 855–858, 2012.

[16] C. W. Kennedy and C. T. Camden. A new look at interruptions. *Western Journal of Communication (includes Communication Reports)*, 47(1):45–58, 1983.

[17] S. C. Levinson. *Pragmatics (Cambridge textbooks in linguistics)*. Cambridge University Press, 1983.

[18] H. Z. Li. Cooperative and intrusive interruptions in inter-and intracultural dyadic discourse. *Journal of Language and Social Psychology*, 20(3):259–284, 2001.

[19] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2):0–0, 1997.

[20] M. Lombard, T. B. Ditton, D. Crane, B. Davis, G. Gil-Egui, K. Horvath, J. Rossman, and S. Park. Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. In *Third international workshop on presence, delft, the netherlands*, volume 240, 2000.

[21] M. T. Maat, K. P. Truong, and D. Heylen. How turn-taking strategies influence users' impressions of an agent. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova, editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, pages 441–453. Springer Berlin Heidelberg, 2010.

[22] P. M. Markey, D. C. Funder, and D. J. Ozer. Complementarity of interpersonal behaviors in dyadic interactions. *Personality and Social Psychology Bulletin*, 29(9):1082–1090, 2003.

[23] H. J. Mazeland. *Inleiding in de conversatieanalyse*. Coutinho, 2003.

[24] K. Murata. Intrusive or co-operative? a cross-cultural study of interruption. *Journal of Pragmatics*, 21(4):385–400, 1994.

[25] S. H. Ng, M. Brooke, and M. Dunne. Interruption and influence in discussion groups. *Journal of Language and Social Psychology*, 14(4):369–381, 1995.

[26] I. Poggi. *Mind, hands, face and body: a goal and belief view of multimodal communication*. Weidler, 2007.

[27] B. Ravenet, A. Cafaro, B. Biancardi, M. Ochs, and C. Pelachaud. Conversational behavior reflecting interpersonal attitudes in small group interactions. In W.-P. Brinkman, J. Broekens, and D. Heylen, editors, *Intelligent Virtual Agents*, volume 9238 of *Lecture Notes in Computer Science*, pages 375–388. Springer International Publishing, 2015.

[28] E. A. Schegloff. Accounts of conduct in interaction: Interruption, overlap, and turn-taking. In *Handbook of sociological theory*, pages 287–321. Springer, 2001.

[29] E. A. Schegloff. *Sequence organization in interaction: Volume 1: A primer in conversation analysis*, volume 1. Cambridge University Press, 2007.

[30] W. C. Schutz. *FIRO: A Three-dimensional Theory of Interpersonal Behaviour*. Rinehart books in the assessment of personality. Holt, Rinehart and Winston, 1958.

[31] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1):140–164, 2005.

[32] D. Tannen. Gender differences in topical coherence: Creating involvement in best friends' talk. *Discourse Processes*, 13(1):73–90, 1990.

[33] K. R. Thórisson, O. Gislason, G. R. Jonsdottir, and H. T. Thorisson. A multiparty multimodal architecture for realtime turntaking. In *Intelligent Virtual Agents*, pages 350–356, 2010.

[34] J. S. Wiggins, P. Trapnell, and N. Phillips. Psychometric and geometric characteristics of the revised interpersonal adjective scales (ias-r). *Multivariate Behavioral Research*, 23(4):517–530, 1988.

[35] J. Youngquist. The effect of interruptions and dyad gender combination on perceptions of interpersonal dominance. *Communication Studies*, 60(2):147–163, 2009.

[36] D. H. Zimmermann and C. West. Sex roles, interruptions and silences in conversation. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, pages 211–236, 1996.