Let Me Finish First - The Effect of Interruption-Handling Strategy on the Perceived Personality of a Social Agent

Ronald Cumbal*
Uppsala University
Sweden
ronald.cumbal@it.uu.se

Reshma Kantharaju
Disney Research
USA
reshma.kantharaju@disney.com

Maike Paetzel-Prüsmann
Disney Research
Switzerland
maike.paetzelprusmann@disney.com

James Kennedy
Disney Research
USA
james.kennedy@disney.com

ABSTRACT

This paper presents an experiment with three artificial agents adopting different strategies when being interrupted by human conversational partners. The agent either ignored the interruption (the most common behavior in conversational engines to date), yielded the turn to the human conversational partner right away, or acknowledged the interruption, finished its thought and then responded to the content of the interruption. Our results show that this change in the agent's conversational behavior had a significant impact on which personality traits people assigned to the agent, as well as how much they enjoyed interacting with it. Moreover, the data also indicates that human interlocutors adapted their own conversational behavior. Our findings suggest that the interactive behavior of an artificial agent should be carefully designed to match its desired personality and the intended conversational dynamics.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in interaction design**; *Empirical studies in HCI*; • **Computing methodologies** → **Discourse, dialogue and pragmatics**.

KEYWORDS

Speech interruption, Barge-in, Overlapping speech, Group interaction, Spoken dialogue system

ACM Reference Format:
Ronald Cumbal, Reshma Kantharaju, Maike Paetzel-Prüsmann, and James Kennedy. 2024. Let Me Finish First - The Effect of Interruption-Handling Strategy on the Perceived Personality of a Social Agent. In *ACM International Conference on Intelligent Virtual Agents (IVA'24), September 16–19, 2024, GLASGOW, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3652988.3673916

*The work was performed while the first author was with The Walt Disney Company

IVA '24, September 16–19, 2024, GLASGOW, United Kingdom

ACM ISBN 979-8-4007-0625-7/24/09…$15.00
https://doi.org/10.1145/3652988.3673916

1 INTRODUCTION

For artificial agents involved in spoken interactions, effective communication strategies go beyond having interesting content to share; they also need an understanding of conversational turn behaviors. In an ideal interaction, individuals take turns speaking, with short pauses between turns facilitating a smooth switch between speakers [44]. Hence, previous work on timing in conversations has mainly focused on predicting when a conversational partner has finished speaking [48]. However, in typical conversations, the turn-switching is not smooth and tends to have a higher frequency of simultaneous speech [24, 47]. No matter how accurate models predicting the end of a speaker's turn become, overlapping speech and interruptions will occur in human conversations as a natural phenomenon that contributes to facilitating the dialogue [8, 60]. With our work, we aim to extend related work by focusing on *different strategies for an agent to react to being interrupted.*

We developed an interactive game where teams of two humans engage with three artificial agents, each equipped with a different strategy for handling interruptions. Most interactive dialogue systems use a turn-by-turn strategy, i.e., the agent does not process any incoming speech while talking, so interruptions are ignored entirely [32]. This strategy constrains the interaction [9], even though research shows that users tend to trust and prefer systems that offer greater interactive capabilities [15]. As such, we designed two different strategies for reacting to interruptions. In the first strategy, the agent responds to the interruption right away and then returns to the previous topic, which mimics the expected conversational behavior if the importance of the human request is considered higher than the ongoing conversational topic. In the second strategy, the agent acknowledges the interruption, finishes the previous conversation thread, and then responds to the interruption. This behavior is more appropriate when the agent considers the current topic to have higher priority than the interruption content.

Our work presents two main contributions. First, we analyze how the three different strategies of responding to being interrupted changes people's *perception of the agent* and how people *adapt their behavior in a conversation accordingly.* Our analysis includes perceptual dimensions covering the personality of the artificial agent as well as the quality of the interaction. Second, we describe how we *integrated a model to classify interruptions into our dialogue framework* and built a content manager to appropriately react for each strategy. While the natural language understanding of the

system is controlled by a human wizard, all decisions around when to speak and what to respond with are taken autonomously by the system. Our work can both *inform researchers on how to best react to an agent being interrupted and help them implement a framework that will work in both fully autonomous and wizard-controlled systems.*

2 RELATED WORK

Instances of simultaneous speech by two or more speakers are commonly referred to as *speech overlaps* and several researchers have proposed taxonomies to identify and categorize them [13, 42, 45]. Two main categories of speech overlaps can be identified: *competitive* and *non-competitive* (commonly referred to as *cooperative*) [7, 48]. *Competitive overlaps* occur when interlocutors compete to claim the floor from the current speaker, with one of them yielding the floor [45]. *Cooperative overlaps* occur when another speaker takes a turn, often to provide feedback or support, with no intention of claiming the floor [14, 19, 58, 61]. An *interruption*, frequently categorized alongside *competitive overlaps*, is defined as an attempt to grab the floor while a speaker's turn is ongoing or during brief pauses [22, 29, 46, 48]. In this study, we focus on interruptions in which participants seek to claim the floor during an agent turn.

2.1 Interruption Detection and Classification

Analyses of corpora annotated with speech overlaps have highlighted audiovisual characteristics differentiating competitive and cooperative overlaps. For example, interruptions are often characterized by a higher pitch and an increase in intensity [22, 23, 45, 47], as well as an increase in speaking rate, glottal stops, or repetitions [45]. These changes in speech provide evidence of speaker conflicts when claiming a turn [45]. Cooperative overlaps, on the other hand, have a lower pitch level [60]. Interruptions are also more likely to occur after certain types of pauses [22], e.g., endpoints of words, including backchannels and disfluencies [47].

Several attempts have been made to automatically classify speech overlaps in conversations. Initial studies focused on audio-based features, e.g., fundamental frequency (F0) [27], energy [28], or a combination of them: prosody, MFCCs, energy, and spectral features [7]. Classifier performance can improve with the addition of visual features e.g., annotated hand motions and disfluencies produced a 21% increase in classification accuracy [28], and slight improvements were observed after including gaze information [56].

In this work, we focus solely on audio-based features to minimize the amount of personal user data recorded. Our interruption detection model prioritizes low latency for use in a real-time dialogue system, in contrast to offline classifiers that tend to underperform in real interactions [30]. Furthermore, our approach aims to enhance model performance, particularly compared to similar systems, emphasizing the importance of facilitating fluent conversations.

2.2 Interruptions in Interactive Systems

Research on interruptions in spoken dialogue systems has primarily focused on evaluating response-strategies used to handle *barge-ins* on task-oriented dialogues, where the user and the system adhere to strict turn-by-turn interactions [11, 40]. The most common method to handle interruptions has been to stop the system's spoken output and continue once the user has finished speaking. For instance, Cassell et al. [5] and Nooraei et al. [36] applied this technique to enhance the communicative abilities for embodied conversational agents, while Matsuyama et al. [33, 34] used this strategy when users barged-in to select articles from a list of news titles read aloud by the system. Crook et al. [10, 11] extended this strategy for agents in social conversations, incorporating a context-sensitive approach to address interruptions and decide whether to continue, abort, or replan the conversation. They also accounted for the user's emotional state to handle different interruptions. In these systems, interruption detection primarily relies on Voice Activity Detection (VAD) [35, 52], Speech Recognition [40] or a combination of both [43]. However, these components are susceptible to detecting false positives, e.g., backchannels, which are often mitigated by allowing user barge-ins only at specific points in the dialogue [11, 40].

In this study, we improve on existing methods by introducing a classifier specifically designed to reduce the aforementioned false positives. Additionally, we extend previous interruption handling techniques, based on basic turn-taking (i.e., stop-continue), by introducing a dialogue framework that dynamically adapts and expands the content of the dialogue flow when interruptions are detected.

2.3 Perception of Interruption Handling

Turn-taking styles in human conversations can affect the perception of social attributes. For example, the occurrence and management of interruptions play a role in shaping perceptions of interpersonal dynamics between conversational participants [3], often associated with perceptions of power, control, or dominance [37, 59].

When analyzing how people perceive different interruption handling strategies with conversational agents, Janowski et al. [25] found that agents that yield to interruptions were linked with introverted and submissive personality traits, while those that interrupt were associated with extroverted personality traits. Cafaro et al. [4] explored interruption types and handling strategies on perceived interpersonal attitude, engagement, and involvement using two conversational agents. The amount of overlap had an effect on users' perception of dominance and friendliness, while a cooperative strategy positively influenced the engagement and involvement levels. Similarly, Gebhard et al. [16] found that an agent was rated as more friendly when overlaps were minimal, and dominant when the agent continued talking after an interruption from the user. Recently, Yang et al. [62] proposed a model to predict interruption initiation timing for virtual agents, finding that randomly chosen interruption times were rated similarly to real and predicted ones.

In contrast to previous work relying on simulated or constrained interactions, our study employs a social conversation to measure perceptual changes. Additionally, we assess the effect of interruption handling strategies on participants' conversational behavior.

3 RESEARCH OBJECTIVES

The main objective of our work is to understand the impact of different interruption-handling strategies on the *personality* of a conversational agent and how they influence its *likability*. We are also interested in studying how these strategies affect the *conversational behavior* of the human interlocutors, measured in terms of user utterances and speech overlaps.

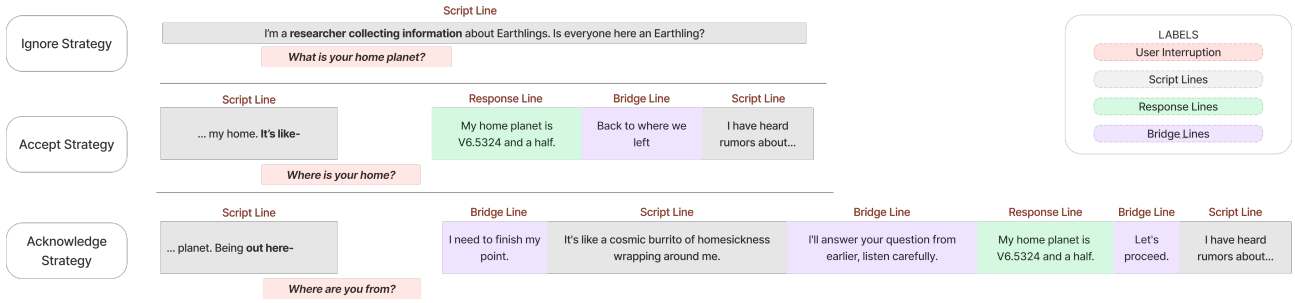

Figure 1: An overview of the three different interruption handling strategies with example dialogues.

3.1 Interruption-Handling Strategies

To address these questions, we created a system that implements three different interruption-handling strategies (Fig. 1) as follows:

Ignore (*IGNR*): This is the standard strategy implemented in most dialog systems, where all of the interruptions are ignored, irrespective of their significance to the conversation. The users must wait for the agent to complete its turn before attempting a turn-switch.

Accept (*ACPT*): In this condition, when an interruption is detected, the agent immediately pauses its speech and yields the turn to the user. Once the user has finished speaking, the agent responds to the user's interruption and resumes the conversation from where it left off. This behavior implicitly signals that the interruption takes precedence over the current conversational topic.

Acknowledge (*ACKN*): When an interruption is detected, the agent again immediately pauses and yields the turn to the user. However, once the user has finished speaking, the agent verbally acknowledges the interruption, finishes its incomplete turn, and finally responds with the answer to the user's interruption. This behavior implicitly signals that the current conversational topic takes precedence over the interruption.

3.2 Hypotheses

Based on the existing literature, we formed the following set of hypotheses about how our strategies will compare:

[H1] Changes in Perception Existing human-agent and human-human studies have shown a strong association between perceived personality traits and interruptive events [4], highlighting that an agent that interrupts is perceived as less agreeable and more assertive [54]. Additionally, studies have shown connections between interruptions, personality traits and likability (Sec. 2.3), indicating that both interrupting and holding on to a turn could significantly reduce an agent's likability [1, 19, 41]. We hypothesize:

[H1A] Personality: Agents that respond immediately to an interruption (*ACPT*) will be perceived as more *open* and *agreeable* compared to agents not responding right away (*ACKN* and *IGNR*). In contrast, agents that do not respond to interruptions at all (*IGNR*) will be perceived as more *extroverted* and *conscientious* compared to agents that respond to them (*ACKN* and *ACPT*).

[H1B] Likability: Agents will be perceived as more *likable* when the human turn is given precedence (*ACPT*), in comparison to the agent finishing its own turn (*ACKN* and *IGNR*).

[H2] Changes in Conversational Behavior In an interaction, interlocutors often adapt to the style of one another [17, 32]. Research indicates that the turn-taking style of an agent has an impact on the user's behavior [55] and, in general, participants accommodate their speech behavior to the agent [51]. We hypothesize:

[H2A] Utterances: As they will wait for clear turn-yielding cues, people will produce fewer utterances with an agent that ignores interruptions (*IGNR*) when compared to an agent that explicitly acknowledges the interruption (*ACKN*, *ACPT*).

[H2B] Overlaps: Conversations will have a higher number of speech overlaps in *ACKN* and *ACPT* because of more interruption attempts than in *IGNR*, where the agent ignores user interruptions.

4 IMPLEMENTATION

4.1 Interruption Detection

We built a dataset by automatically extracting overlapping speech segments based on dialogue act tags from the AMI corpus [26] and adding nine additional labels[1]. We collapsed the labels provided by three annotators to get interruptions (**I**) and non-interruptions (**NI**) with moderate agreement (Cohen's kappa = 0.59). To improve the robustness to instances of vocal sounds e.g., laughter or coughing, we included samples from the VocalSounds dataset [20]. The final dataset contained 9671 samples for training (split 70:30 into train and validation sets) and an additional 322 samples for testing.

Our interruption detection model used concatenated speech features i.e., F0, MFCCs, pitch, and intensity as input. The features were extracted over a 0.96 second context window using *Librosa* with default parameters, except for pitch (fmax=1600, fmin=75). A Feed-forward Neural Network model with four hidden layers, using ReLU activation functions and dropout, was implemented in PyTorch (v1.8.1). Hyperparameters were manually tuned based on precision and recall evaluation, resulting in a batch size of 64, a learning rate of $1e^{-4}$, a hidden layer size of 32 neurons, and a dropout rate of 35%. The model was trained using Cross Entropy Loss and Stochastic Gradient Descent for up to 50 epochs (seed=1), with early stopping based on validation loss (patience=5).

The best-performing model achieved an overall accuracy of 79% (Interruptions: 69%; Non-interruptions: 87%). The use of a small context window ensures quick detection of interruptions without producing a high rate of false positives, and improves on classifiers deployed in previous work as detailed in Sec. 2.

[1]Feedback [**NI**], Delayed response [**NI**], Side-talk [**NI**], Non-speech [**NI**]; Early Start [**I**], Turn Grab [**I**], Attempt Turn [**I**], Simultaneous Start [**I**]; *Unknown*

4.2 System Overview

A content manager (CM) navigated a dialogue tree and sent requests to a Microsoft Azure[2] Text-to-Speech system (TTS) for audio playback. The system's language understanding was handled by a human wizard who was blind to the experimental condition. We chose to have human control over the language understanding to remove frustration linked to misunderstandings as a confounding factor. The wizard was located in a different room and received an audio stream of the conversation. Speech was detected by a microphone array with on-board VAD. The results from the VAD classification and the audio stream were fed into the interruption detector, which sent outputs to the CM. If a non-interruptive overlap was detected, the CM took no action. If an interruptive event was detected, the CM implemented one of the three strategies explained in Sec. 3.1.

If the strategy selected was *IGNR*, the CM disregarded all messages received while the agent turn was in progress; the user's speech had no impact on the timing or the content uttered by the agent during its turn. In the *ACPT* and *ACKN* conditions, when an interruption was detected, the CM stopped the ongoing utterance and waited for the user to complete their utterance. Afterward, a response to the user was pushed to the queue of upcoming speech. Conversational snippets designed to make the transitions more natural were added to the queue. Depending on the condition, these were either inserted before (*ACPT*) or before and after (*ACKN*) the already planned utterances belonging to the conversational topic (Fig. 1). The wizard needed to input the content of the interruption before a response could be selected. If the wizard failed to provide this input within two seconds after the participant finished speaking, the agent defaulted to asking the human to repeat.

5 EVALUATION

A within-subject design was used for the evaluation of the system and our three interruption-handling strategies. We used a script to generate a balanced random distribution of the conditions to account for potential order effects. Both the wizard and the researcher conducting the experiment were blind to the order of conditions.

Participants were recruited internally through voluntary sign-ups communicated using posters and mailing lists. Forty adults who live and work in the US participated. Each session was conducted with two randomly assigned participants, some of whom did not know each other (n=9) and others who did (n=11). As participants are working in a technology field, most indicated a general familiarity with virtual agents or robots (n=31). However, their involvement with said technologies varied, with some participants working in design or administrative roles, and even among the technologists, most did not work with robots or artificial agents on a daily basis. Due to privacy concerns, we were not permitted to collect additional demographic data. Participants received a non-monetary reward for their participation.

[2]https://azure.microsoft.com/en-us/products/ai-services/text-to-speech

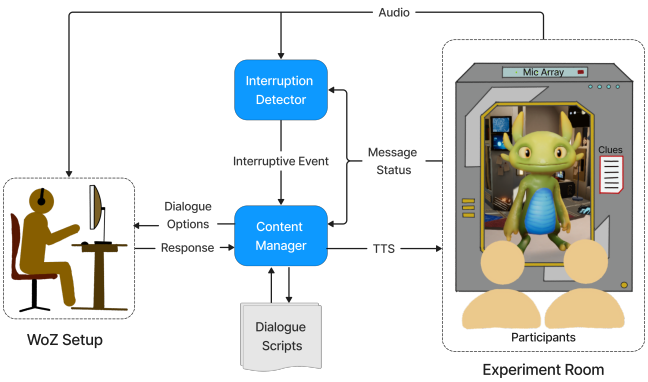

Figure 2: Overall setup of the dialogue system and user study.

5.1 Interaction Design

To measure the impact of interruption-handling, we designed a conversational game with an incentive to interrupt.

(Game) Context: The main aim of designing the scenario was to create an environment in which interruptions need to occur and to create a communicative advantage without explicitly telling participants when to interrupt. Moreover, the game was inspired by real-world entertainment experiences, such as interactive escape rooms and artificial game-playing partners.

The premise of the interaction was based on users collecting information from the artificial agent. Participants were told that an alien species, named Galastrolotls, pretending to be friendly explorers, had been contacting humans to obtain information to organize an invasion. The participants were directed to play the role of investigators and were tasked with collecting answers to questions that helped in sabotaging the Galastrolotl's mission. The agents were designed to speak in long sentences to provide ample time for interruptions. To increase urgency, participants were told they had around 5 minutes to collect information from each agent.

By giving the artificial agent and the human participants competing tasks, the perceived importance of the conversational topic is in contradiction: the humans have a high sense of urgency in getting responses to their requests, while the agent has a high interest in finishing its own topics.

We had two participants per session in order to (i) elicit non-interruptive speech overlaps, e.g., side-conversations and shared laughter, observed frequently in group conversations and verify our system's performance in such conditions; (ii) show that our system can handle multiparty conversations, which is analogous to other real-world scenarios, e.g., in classrooms or museums [18].

Dialogue: Three conversational scripts were designed using the Yarn Spinner scripting tool[3]. Each script consists of three parts: the introduction where the agent presents itself; the discussion in which the agent attempts to ask questions and in which participants can attempt to gather information; and the conclusion, for wrapping-up the conversation. The *discussion* phase was further divided into five alternating on-topic and off-topic dialogue segments. On-topic segments comprise lines related to information a participant needs to gather. They were designed to elicit interruptions in the form of questions and consist of a series of statements by the agent that include keywords matching the information the participants need to collect. Off-topic segments were used as transitions between

[3]https://yarnspinner.dev

different on-topic segments. Generic lines were inserted to bridge between handling interruptions and the main script (see Fig. 1).

Participants were directed to ask questions only when the agent was talking about a relevant topic. If a participant asked a question outside the relevant conversational context, the agent first issued a warning and eventually left the conversation if a threshold was exceeded. This was a preemptive measure to limit random interruptions and control the conversation from being too open-ended.

5.2 Embodiment

The visual scene depicts an agent designed to resemble an extraterrestrial creature in a virtual spaceship implemented in Unity3D. A rendering is shown in Fig. 2 and the supplementary material. The embodiment of the agent is distinctly non-human, but imitates human-like behaviors. The non-verbal behaviors of the agent include body and hand movements (beat gestures), head movements, eye blinks, and lip movements synchronized to speech content. These behaviors were automatically selected based on the speech content and generated using Bolder Games' Nexus engine[4]. The appearance and behaviors of the agents remained constant across the three interactions and only the background was varied to keep the confounding factors to a minimum while creating three distinct interactions. Each background was tied to a specific conversational script. The order of backgrounds was randomly selected and not tied to a specific condition.

5.3 Procedure

Each session included two participants and lasted approximately 45 minutes. Participants were greeted by a researcher who gave a brief overview of the study and the data collected. Participants were reminded that their involvement was voluntary and that they could withdraw and leave at any time. After signing the consent form, participants first listened to a brief audio recording explaining the main objectives of the mission. Next, the researcher gave game instructions that encouraged participants to interrupt the agent and ask relevant questions. The participants were then presented with three interactive scenarios corresponding to the conditions described in Sec. 3.1. Each scenario was designed to be 4-5 minutes long and had five answers that could be collected. Participants had access to a list of potential items during the interaction. After each scenario, participants answered a questionnaire to measure their perception of the agent and received an updated list of potential items for the next interaction. In all sessions, the agent's language understanding was controlled by the same individual who remained hidden from participants throughout the study. At the end of the study, participants answered a final questionnaire about exposure to similar technologies. They were then interviewed by the researcher, who asked whether they noticed any differences between the three sessions and if they had any preferences. During the final debriefing phase, participants received a game score based on the information items collected.

[4]https://www.boldergames.com/

5.4 Measures

Perceptual Measurements. To understand the perception of the agent in the three different conditions, we designed a questionnaire to be filled out after each interaction. To measure differences in the perceived *personality of the agent* [H1A], a Ten-Item Personality Inventory (TIPI) questionnaire [21] on a 7-point Likert scale was used. We extracted 3 items from the Godspeed questionnaire's likability scale [2] and measure them on a 5-point Likert scale to capture the perceived *likability* of an agent [H1B].

We added questions to help understand whether the manipulation of the agent's behavior worked as intended. To ensure the perceived *anthropomorphism* of the agent was not impacted by the different behaviors, two items from the respective Godspeed questionnaire sale [2] were added. Next, we measured the agent's *perceived competence* using four items on a 7-point scale from Doyle et al. [12], as we assumed agents showing any reactions to interruptions (*ACKN* and *ACPT*) may be perceived as more competent than agents simply ignoring interruptions (*IGNR*). Moreover, we captured specifics of the communicative behavior related to overlapping speech using 6 items based on Paetzel et al. [38]. These capture questions in relation to the speed and flow of communication, pauses, as well as the ability to get a word in. Finally, we measured the satisfaction with the game and overall enjoyment. The exact questionnaire is available in the supplementary material.

Conversational Behavior Measurements. For comparing conversational behaviors, we consider the utterances of the participants in conversation with the three different agents. SileroVAD [53] was used to extract speech boundaries for the participants, and agent utterance timings were extracted from logs. The number of speech turns and their average length were calculated to verify hypothesis [H2A]. We also extracted the number of overlapping speech segments and their average duration to test hypothesis [H2B].

6 RESULTS

We present the results from a study consisting of 20 sessions with 40 participants. A Shapiro-Wilk test applied individually to all perceptual and behavioral measures revealed that the data is not from a normal distribution ($p < .05$), therefore Kruskal-Wallis and Spearman tests are used as they are designed for non-parametric data. A post-hoc Dunn test with Benjamini-Hochberg (BH) correction is applied and pairwise comparison results are reported along with the epsilon-squared values for effect sizes[5].

6.1 Classifier Performance

There were a total of 1354 instances across conditions where the user overlapped with the agent after the agent started talking, i.e., potential interruptions. Two annotators labeled a randomly selected 10% of these instances as either cases where the user intended to interrupt the agent or not (side talk, laughter, or other). Cohen's Kappa was 0.67, indicating good agreement. Treating the annotator labels as ground truth, the classifier achieves F1 scores of 0.78 and 0.80 from the two annotators when using a weighted average.

Additional annotations by the same annotators were conducted to reach 25% (68) of all system predicted interruptions. Again, the

[5](.0 − .04): Weak, (.04 − .16): Moderate, (.16 − .36): Relatively Strong, (.36 − 1): Strong

annotators labeled whether the user intended to interrupt the agent or not ($\kappa = 0.55$), indicating moderate agreement. Taking this as ground truth, the classifier achieves F1 scores of 0.34 and 0.54. There were many instances where the system would predict interruptions when the user had spoken during a brief pause (on the order of hundreds of milliseconds) in the speech, but the system believed speech was ongoing due to padding in the audio playback files. If the system perspective about the speech status were taken, the classifier's performance would improve to F1 scores of 0.81 and 0.87. The classifier's performance was comparable in all three conditions, with average F1 scores of 84.5% *IGNR*, 81.5% *ACPT*, and 84% *ACKN*. To leverage this potential improvement in future work, the classifier should directly take the agent's audio signal as input, or the speech status reporting should be adjusted to cater to any silence padding. In the works closest to ours, classifiers for interruption detection achieved an F1 score of 69.5 [6] and Equal Error Rate of 32.0% [57], both much worse than the performance of our classifier. Most importantly, both studies conducted their testing offline, while our performance was measured during live interactions.

6.2 Manipulation Checks

In this section, we report on the perceptual measures and observable behaviors used to ensure the manipulations had the intended effect.

Anthropomorphism. The *perceived anthropomorphism* (2 items; $\alpha = 0.79$) is measured to ensure the three different agents remain comparable. Indeed, our results show no significant differences between the agents [$H(2) = 0.94, p = .62, \varepsilon^2 = .007$] and the manipulation can be considered successful.

Interaction Length. While the content scripts were designed to be of equal length, the overall duration of the interaction varied significantly between conditions [$H(2) = 7.82, p = .02, \varepsilon^2 = .13$]. This variation is due to the interruption-handling strategies. In particular, the *ACKN* condition requires many bridge lines to transition between interruptions and other dialogue, evidenced through the ratio between turns with bridge lines and total turns. We observe a significant difference across conditions [$H(2) = 29.78, p < .001, \varepsilon^2 = .49$], with the *ACKN* condition ($M = 0.30, SD = 0.11$) having a higher ratio than *ACPT* ($M = 0.17, SD = 0.10$) and *IGNR* ($M = 0.11, SD = 0.02$). This also results in shorter average turn lengths by the agent when more bridge content is present, as the bridge content is short. As such, there is a significant difference in the average agent turn length [$H(2) = 21.79, p < .001, \varepsilon^2 = .36$], with *IGNR* ($M = 4.55, SD = 0.25$) significantly longer than *ACPT* ($M = 3.86, SD = 0.5, p < .001$) and *ACKN* ($M = 3.74, SD = 0.6, p < .001$). This may have had an impact on the perception of the agent.

Perceived Conversational Abilities. Our goal was to create conditions that move beyond the commonly used *IGNR* strategy of not responding to interruptions. Accordingly, if this were achieved successfully, there would be some expectation that those conditions might be perceived as having greater conversational abilities and communicative competence.

The *perceived competence* (4-items; $\alpha = 0.82$) had observable differences in average ratings across the conditions [$H(2) = 6.00, p = .04, \varepsilon^2 = .05$]. A post-hoc test shows significant differences ($p = .05$) between *ACPT* ($M = 4.46, SD = 1.10$) and *ACKN* ($M = 3.86, SD = 1.34$). Even though *IGNR* was rated higher ($M = 4.47, SD = 1.13$), the differences were not significant ($p = .09$).

The *perceived awkward pauses* varied slightly across conditions [$H(2) = 5.24, p = .07, \varepsilon^2 = .04$]. Both *ACPT* and *ACKN* ($M = 3.78, SD = 1.17$ for both) had more *awkward pauses* than *IGNR* ($M = 3.22, SD = 1.34$). We calculated the *pause* as a ratio between total pauses (within agent turns and between user-agent turns) and total duration of the interaction and found significant differences between conditions [$H(2) = 8.29, p = .01, \varepsilon^2 = .14$] that align with the user perceptions. In addition, participants felt that it was significantly harder to *get a word in edgewise* [$H(2) = 11.04, p = .003, \varepsilon^2 = .09$] in *ACKN* ($M = 3.58, SD = 1.03, p = .003$) and *IGNR* ($M = 3.35, SD = 1.17, p = .02$) in comparison to *ACPT* ($M = 2.78, SD = 1.10$), where the agent responded immediately. There were no significant differences in the ratings for *flow of communication* [$H(2) = 4.63, p = .09, \varepsilon^2 = .03$].

The *perceived attentiveness* of the agent was significantly different [$H(2) = 13.39, p = .001, \varepsilon^2 = .11$], with *ACPT* ($M = 3.35, SD = 1.00$) and *IGNR* ($M = 3.42, SD = 1.01$) rated significantly higher ($p = .003$ and $p = .002$, respectively) than *ACKN* ($M = 2.6, SD = 1.10$).

Related to the perceived conversational abilities is the perceived ease of playing the game with the agent. Participants indicated that it was *easier to gather information* in *IGNR* ($M = 3.12, SD = 1.07$) and *ACPT* ($M = 3.02, SD = 1.03$) in comparison to *ACKN* ($M = 2.65, SD = 1.27$), but the differences were not significant [$H(2) = 0.94, p = .62, \varepsilon^2 = .03$]. However, there were significant differences across the conditions [$H(2) = 7.23, p = .02, \varepsilon^2 = .06$] on the reported *satisfaction with the information collected.* A post-hoc test revealed *IGNR* ($M = 3.5, SD = 1.04$) was rated significantly higher ($p = .02$) than *ACKN* ($M = 2.82, SD = 1.26$). Even though *ACPT* ($M = 3.32, SD = 0.97$) received higher ratings, the differences were not significant ($p = .09$). We looked at the differences in the actual number of items collected in each condition (max = 5), and found significant differences across the conditions [$H(2) = 9.25, p = .009, \varepsilon^2 = .15$], where the information obtained was the highest for *ACPT* ($M = 4.85, SD = 0.36$) in comparison to *IGNR* ($M = 4.25, SD = 0.85$) and *ACKN* ($M = 4.15, SD = 0.98$) with ($p = .01$ and $p = .02$, respectively).

These results show that agents using the *IGNR* and *ACPT* strategies were rated higher in conversational abilities compared to the *ACKN* strategy. Participants were more satisfied with the flow of communication in the *IGNR* and *ACPT* conditions, as well as with the attentiveness of the agent and, consequently, with the amount of information collected. This partially aligns with their actual success in collecting information, which was highest in *ACPT*. Our manipulation of the agent's behavior can thus be considered only partially successful: While there are noticeable differences in the agent's conversational abilities, we did not manage to create two new strategies that are perceived as superior communicative strategies in comparison to the *IGNR* baseline.

6.3 [H1] Perception of the Agent

[H1A] Personality: We calculated the five personality traits from the 10 personality items. The Cronbach's alpha on the 2-item sub-scale except Extraversion ($\alpha = 0.74$) and Emotional Stability ($\alpha = 0.62$) was low; this is a commonly observed phenomenon and the ratings are still accepted as a reliable measure for

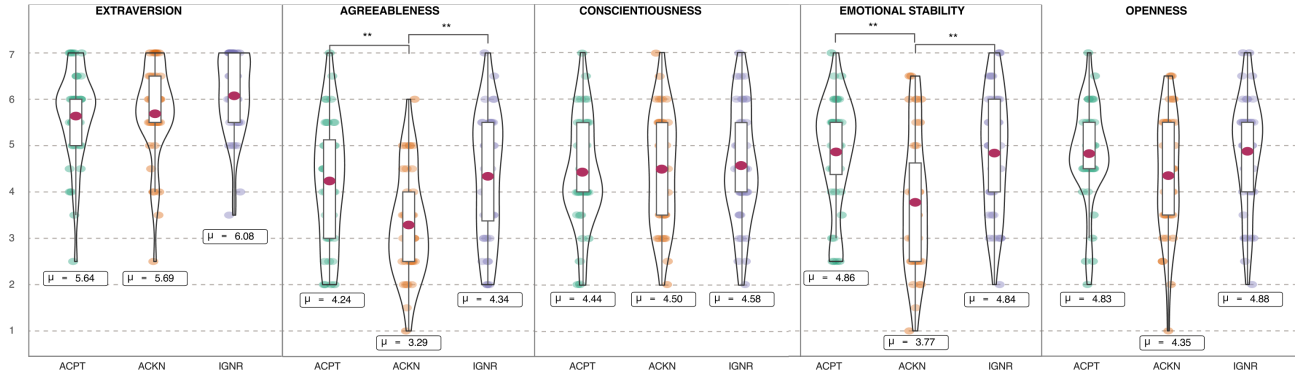

Figure 3: Violin plots of the perceived personality scores for the agent, measured using the 7-point TIPI questionnaire. Vertical boxes indicate the interquartile range and mean values are indicated by a red dot and significant differences between conditions are highlighted by bars for Agreeableness and Emotional Stability (= p < .01)**

personality traits [21]. *Openness* was rated higher for both *ACPT* ($M = 4.82, SD = 1.07$) and *IGNR* ($M = 4.88, SD = 1.21$) than *ACKN* ($M = 4.35, SD = 1.28$) but the differences were not significant, [H(2) $= 4.12, p = .12, \varepsilon^2 = .03$]. *Agreeableness* was rated significantly higher, [H(2)$= 14.37, p < .001, \varepsilon^2 = .12$], for both *ACPT* ($M = 4.82, SD = 1.07$) and *IGNR* ($M = 4.88, SD = 1.21$) than *ACKN* ($M = 3.29, SD = 1.12$). *Extraversion* was rated the highest for *IGNR* ($M = 6.08, SD = 0.86$) when compared to *ACPT* ($M = 5.64, SD = 0.99$) and *ACKN* ($M = 5.69, SD = 1.02$), with [H(2)$= 4.85, p = .08, \varepsilon^2 = .04$]. Although *Conscientiousness* was rated slightly higher for the agent in *IGNR*, there was no significant difference [H(2)$= 0.28, p = .86, \varepsilon^2 = .002$]. Finally, *Emotional Stability* was rated significantly higher, [H(2)$= 14.39, p < .001, \varepsilon^2 = .12$], for both *ACPT* ($M = 4.86, SD = 1.19$) and *IGNR* ($M = 4.84, SD = 1.28$) than *ACKN* ($M = 3.78, SD = 1.45$).

The data is visualized in Fig. 3 and partially confirms H1A. Indeed, the *ACPT* condition was rated as more open and agreeable than *ACKN*, while *IGNR* was considered the most extraverted. Contrary to our expectations, *IGNR* received similar ratings in openness and agreeableness to *ACPT*. Moreover, the data showed no significant difference in perceived conscientiousness between conditions.

[H1B] Likability: The *perceived likability* (3-items, $\alpha = 0.90$) was significantly different across the three conditions [H(2)$= 9.77, p = .007, \varepsilon^2 = .08$], with both *ACPT* ($M = 3.76, SD = 0.85$) and *IGNR* ($M = 3.83, SD = 0.82$) rated significantly higher ($p = .018$ and $p = .012$, respectively) than *ACKN* ($M = 3.11, SD = 1.11$). Additionally, we measured the *conversational enjoyment* and found significant differences across the conditions [H(2)$= 6.10, p = .04, \varepsilon^2 = .05$]. *ACKN* was rated lower ($M = 3.08, SD = 1.19$) than *IGNR* ($M = 3.62, SD = 0.97, p = .08$) and *ACPT* ($M = 3.6, SD = 0.92, p = .05$). Our hypothesis [H1B] is partially confirmed. While an agent following the *ACPT* condition is liked better than one in the *ACKN* condition, there was no significant difference between the *ACPT* and *IGNR* condition.

6.4 [H2] Conversational Behavior

[H2A] Utterances: There were no significant differences between conditions in the amount of user utterances [H(2)$= 0.77, p = .67, \varepsilon^2 = .01$], although they were lower in *IGNR* ($M = 32.7, SD = 11.4$) when compared to *ACPT* ($M = 33.5, SD = 14.5$) and *ACKN* ($M = 36.4, SD = 11.4$). This does not support hypothesis [H2A], that users tend to use fewer utterances when the agent strategically reacts to the user's interruption behavior. However, we also calculated the *average utterance length* for the user speech and found significant differences across conditions [H(2)$= 21.79, p < .001, \varepsilon^2 = .36$]. A post-hoc test shows that the length (in *secs*) was significantly longer in *IGNR* ($M = 4.55, SD = 0.25$) when compared to *ACPT* ($M = 3.86, SD = 0.5, p < .001$) and *ACKN* ($M = 3.74, SD = 0.6, p < .001$).

[H2B] Overlaps: The overall number of speech overlaps were higher for *ACKN* ($M = 30.6, SD = 15.6$) than *IGNR* ($M = 25.2, SD = 13.9$) and *ACPT* ($M = 23.7, SD = 12.6$), but the differences were not significant [$H(2) = 2.47, p = .29, \varepsilon^2 = .04$]. However, the average length (in *secs*) of overlapping speech between the agent and the participants varied significantly between the conditions [$H(2) = 12.67, p = .001, \varepsilon^2 = .21$]. On average, *ACKN* ($M = 0.68, SD = 0.15, p = .001$) and *ACPT* ($M = 0.72, SD = 0.19, p = .02$) had shorter overlapping speech segments than the *IGNR* condition ($M = 0.90, SD = 0.19$). This partially confirms hypothesis [H2B].

7 DISCUSSION

Our results show that the way an artificial agent reacts to being interrupted by human conversational partners influences the *perceived personality* of the agent, as well as how much they *like the agent and enjoy interacting with it*. This finding highlights the importance of consciously choosing conversational strategies to match the desired personality of an artificial agent. Furthermore, our results give an *early indication that the agent's strategy for handling interruptions influences people's conversational behavior*. For example, we found that people's utterances were longer when talking to an agent that ignored interruptions, which calls for caution when designing with a target conversational dynamic in mind.

The original intent of this work was to adapt two strategies commonly occurring in human-human conversations to create a more natural flow of communication for an artificial agent. However, our results show only partial success. The *ACPT* condition was found to be on the same level as the *IGNR* condition in most of the measured conversational dimensions, while the *ACKN* condition

was consistently rated lowest. The ratings of the agent's perceived competence follow the same trend. Given that participants also found it easier to extract information from the agent in *IGNR* and *ACPT* conditions, they may have rated the competence of the agent based on their own ease of extracting information and not on how elaborate the conversational behavior itself was.

Results on the agent's personality [H1A] and its perceived likability [H1B] are in line with the ratings on conversational abilities. Agents adopting the *IGNR* and *ACPT* strategies are liked better, which potentially relates to them being rated as more open and agreeable – generally desirable personality traits. On the contrary, the agent adopting the *ACKN* strategy was rated as more emotionally unstable, less open, and less agreeable. This is in line with previous work: if someone insists on the importance of their own utterances, this person is perceived as more dominant and controlling, and less friendly [4, 16, 25].

While it may be surprising that *IGNR* was preferred over the *ACKN* strategy, we believe that the competitive nature of our game may have influenced participants' perceptions. The conversation was designed such that the agent has a high interest in gathering information from the human. The human participants, however, are attempting to gather information from the agent quickly, which is easier with the *ACPT* strategy as it responds to information requests right away. While both *IGNR* and *ACKN* did not emphasize the importance of the participants' information requests, the agent adopting the *IGNR* strategy did so silently; its behavior could have been attributed to technical limitations. On the contrary, the agent using the *ACKN* strategy clearly communicated that it valued its own conversational content more, which may not be appreciated in the setting we designed. Different, less competitive settings may hence lead to different results.

While we did not explicitly measure whether participants understood the link between each conversational strategy and the implicit hierarchy in content importance, we found that overall, participants were able to identify the differences in the agents' behavior between sessions. Most participant groups ($n = 12$) clearly voiced a preference for the agent in the *ACPT* condition, with the other two conditions being tied with three mentions each. Many groups mentioned the character in the *ACKN* condition as being rude, with some explicitly saying they felt as if they could not extract information from this agent in particular. Those preferring the *ACPT* behavior noted the agent's behavior as being attentive and accommodating and found it easy to get answers from the agent.

The specific way we authored the content acknowledging an interruption in the *ACKN* condition, e.g., "Hold on, let me finish first" may have come across as harsher than anticipated, which could contribute to the feeling of this agent being particularly rude. While the specific utterances authored for *ACKN* may have been a confounding factor in our study, we believe the variety of available lines, as seen in the supplementary material decreases the confound of the authored content. Finally, as ignoring interruptions is the current standard in conversational engines, participants may have noticed behavior deviating from the norm more prominently.

Future Work. We aim to build a platform and develop guidelines for conversations that apply to all agents capable of autonomously communicating with humans [31]. While physical embodiments allow for additional modes of interaction [50, 63], we believe the underlying principles of conversational dynamics we uncovered remain independent of the embodiment type [49]. However, further analysis is needed to confirm this belief. In the future, we would like to understand the conversational dynamics of the three different agent behaviors created in more detail. We also suggest designing an experiment that incorporates different conversational contexts to study the impact of varying levels of importance in the conversational content both on the human and the agent side. Considering different levels of explicit referencing in the language around acknowledgments and interruptions could offer interesting insights into the importance of these conversational bridges. Finally, prior literature suggests that the likability of a conversational partner also depends on the personality of the person being asked to rate the conversational partner [39]. Hence, we suggest asking participants to fill out a personality test and see if certain conversational strategies are preferred by people with specific personality types.

Limitations. This study is limited in the number of participants and the moderate effect sizes we observed for some of our analyses. Even though the order of conditions was balanced, previous conditions may have influenced the behavior and responses of the participants. Moreover, our sample was from a population more familiar with robots and artificial agents than the general public, which could have influenced their perceptions. While the implementation of the conversational behavior was mostly successful, some small malfunctions in the conversation engine could have impacted the rating of the agent. In all conditions, the TTS added pauses between consecutive sentences, which could have impacted the perception of awkward pauses. In the agent adopting the *ACKN* behavior, in 8 pairs the agent would get to a point where it uttered a longer row of content designed to bridge between different parts of the conversation than necessary. In the *ACPT* condition, 3 pairs experienced a situation where they got stuck in a loop of interruptions that led to an early exit of the conversation.

8 CONCLUSION

This paper presented the design and implementation of an experiment devised to understand the impact of different agent interruption strategies on people's perception of the agent. Three conditions were used: one in which the agent would not respond to interruptions, one in which it would respond immediately, and one in which it would acknowledge the interruption but not offer a response until after it had finished its turn. The results of a human-subject experiment revealed that the condition in which the agent would acknowledge but not immediately respond to interruptions was significantly less liked, and perceived as less agreeable and less emotionally stable. These findings highlight the importance of designing appropriate interruption strategies for agents depending on the personality they wish to convey.

ACKNOWLEDGMENTS

We would like to thank our character and content team: Laurel, Larry, Bolder Games, Kyna, Michelle and Sam, as well as data annotators Karl, Jake, James and Max. We also thank Naveen Kumar for valuable feedback on the design of the interruption detector.

REFERENCES

[1] D. Aneja, D. McDuff, and M. Czerwinski. 2020. Conversational error analysis in human-agent interaction. In *Proceedings of the 20th ACM international conference on intelligent virtual agents*. 1–8.

[2] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1 (2009), 71–81.

[3] G. W. Beattie. 1981. Interruption in conversational interaction, and its relation to the sex and status of the interactants. *Linguistics: An Interdisciplinary Journal of the Language Sciences* 19, 1-2 (1981), 15–36. https://doi.org/10.1515/ling.1981.19.1-2.15

[4] A. Cafaro, N. Glas, and C. Pelachaud. 2016. The Effects of Interrupting Behavior on Interpersonal Attitude and Engagement in Dyadic Interactions. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems* (Singapore, Singapore) *(AAMAS '16)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 911–920.

[5] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan. 2001. More than just a pretty face: conversational protocols and the affordances of embodiment. *Knowledge-Based Systems* 14, 1 (2001), 55–64. https://doi.org/10.1016/S0950-7051(00)00102-7

[6] S. Chowdhury, M. Danieli, and G. Riccardi. 2015. Annotating and categorizing competition in overlap speech. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5316–5320.

[7] S. A. Chowdhury, M. Danieli, and G. Riccardi. 2015. Annotating and categorizing competition in overlap speech. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (South Brisbane, Queensland, Australia, 2015-04). IEEE, 5316–5320. https://doi.org/10.1109/ICASSP.2015.7178986

[8] J. Coates. 1994. No gap, lots of overlap: Turn-taking patterns in the talk of women friends. *Researching language and literacy in social context* 10, 4 (1994), 177–192.

[9] M. H. Cohen, J. P. Giangola, and J. Balogh. 2004. *Voice user interface design*. Addison-Wesley Professional.

[10] N. Crook, D. Field, C. Smith, S. Harding, S. Pulman, M. Cavazza, D. Charlton, R. Moore, and J. Boye. 2012. Generating context-sensitive ECA responses to user barge-in interruptions. *Journal on Multimodal User Interfaces* 6, 1 (2012), 13–25. https://doi.org/10.1007/s12193-012-0090-z

[11] N. Crook, C. Smith, M. Cavazza, S. Pulman, R. Moore, and J. Boye. 2010. Handling User Interruptions in an Embodied Conversational Agent. In *Proceedings of the AAMAS International Workshop on Interacting with ECAs as Virtual Characters*. 27–33.

[12] P. R. Doyle, L. Clark, and B. R. Cowan. 2021. What Do We See in Them? Identifying Dimensions of Partner Models for Speech Interfaces Using a Psycholexical Approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 244, 14 pages. https://doi.org/10.1145/3411764.3445206

[13] N. Ferguson. 1977. Simultaneous speech, interruptions and dominance. *British Journal of Social and Clinical Psychology* 16, 4 (1977), 295–302. https://doi.org/10.1111/j.2044-8260.1977.tb00235.x

[14] P. French and J. Local. 1983. Turn-competitive incomings. *Journal of Pragmatics* 7, 1 (1983), 17–38.

[15] R. Fusaroli and K. Tylén. 2016. Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive science* 40, 1 (2016), 145–171.

[16] P. Gebhard, T. Schneeberger, G. Mehlmann, T. Baur, and E. André. 2019. Designing the Impression of Social Agents' Real-time Interruption Handling. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents* (Paris France, 2019-07). ACM, 19–21. https://doi.org/10.1145/3308532.3329435

[17] H. Giles, A. Mulac, J. J. Bradac, and P. Johnson. 1987. Speech accommodation theory: The first decade and beyond. *Annals of the International Communication Association* 10, 1 (1987), 13–48. https://doi.org/10.1080/23808985.1987.11678638

[18] S. Gillet, M. Vázquez, C. Peters, F. Yang, and I. Leite. 2022. *Multiparty Interaction Between Humans and Socially Interactive Agents* (1 ed.). Association for Computing Machinery, New York, NY, USA, 113–154. https://doi.org/10.1145/3563659.3563665

[19] J. A. Goldberg. 1990. Interrupting the discourse on interruptions. *Journal of Pragmatics* 14, 6 (1990), 883–903. https://doi.org/10.1016/0378-2166(90)90045-F

[20] Y. Gong, J. Yu, and J. Glass. 2022. Vocalsound: A Dataset for Improving Human Vocal Sounds Recognition. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 151–155. https://doi.org/10.1109/ICASSP43922.2022.9746828

[21] S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.

[22] A. Gravano and J. Hirschberg. 2012. A corpus-based study of interruptions in spoken dialogue. In *Interspeech 2012*. ISCA, 855–858. https://doi.org/10.21437/Interspeech.2012-193

[23] B. Hammarberg, B. Fritzell, J. Gaufin, J. Sundberg, and L. Wedin. 1980. Perceptual and acoustic correlates of abnormal voice qualities. *Acta oto-laryngologica* 90, 1-6 (1980), 441–451.

[24] M. Heldner and J. Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38, 4 (Oct. 2010), 555–568. https://doi.org/10.1016/j.wocn.2010.08.002

[25] K. Janowski and E. André. 2019. What If I Speak Now? A Decision-Theoretic Approach to Personality-Based Turn-Taking. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (Montreal QC, Canada) *(AAMAS '19)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1051–1059.

[26] W. Kraaij, T. Hain, M. Lincoln, and W. Post. 2005. The AMI meeting corpus. In *Proc. International Conference on Methods and Techniques in Behavioral Research*. 1–4.

[27] E. Kurtić, G. J. Brown, and B. Wells. 2009. Fundamental frequency height as a resource for the management of overlap in talk-in-interaction. In *Where prosody meets pragmatics*. Brill, 183–203.

[28] C. Lee, S. Lee, and S. S. Narayanan. 2008. An analysis of multimodal cues of interruption in dyadic spoken interactions. In *Interspeech 2008*. ISCA, 1678–1681. https://doi.org/10.21437/Interspeech.2008-366

[29] H. Z. Li. 2001. Cooperative and Intrusive Interruptions in Inter- and Intracultural Dyadic Discourse. *Journal of Language and Social Psychology* 20, 3 (Sept. 2001), 259–284. https://doi.org/10.1177/0261927X01020003001

[30] T. Lin, Y. Wu, F. Huang, L. Si, J. Sun, and Y. Li. 2022. Duplex conversation: Towards human-like interaction in spoken dialogue systems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3299–3308.

[31] B. Lugrin. 2021. *Introduction to Socially Interactive Agents* (1 ed.). Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3477322.3477324

[32] M. Marge, C. Espy-Wilson, N. G. Ward, A. Alwan, Y. Artzi, M. Bansal, G. Blankenship, J. Chai, H. Daumé, D. Dey, M. Harper, T. Howard, C. Kennington, I. Kruijff-Korbayová, D. Manocha, C. Matuszek, R. Mead, R. Mooney, R. K. Moore, M. Ostendorf, H. Pon-Barry, A. I. Rudnicky, M. Scheutz, R. St. Amant, T. Sun, S. Tellex, D. Traum, and Z. Yu. 2022. Spoken language interaction with robots: Recommendations for future research. *Computer Speech & Language* 71 (Jan. 2022), 101255. https://doi.org/10.1016/j.csl.2021.101255

[33] K. Matsuyama, K. Komatani, T. Ogata, and H. G Okuno. 2009. Enabling a User to Specify an Item at Any Time During System Enumeration — Item Identification for Barge-In-Able Conversational Dialogue Systems. In *Proc. Interspeech 2009*. 252–255. https://doi.org/10.21437/Interspeech.2009-88

[34] K. Matsuyama, K. Komatani, T. Takahashi, T. Ogata, and H. G Okuno. 2010. Improving identification accuracy by extending acceptable utterances in spoken dialogue system using barge-in timing. In *Trends in Applied Intelligent Systems: 23rd International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2010, Cordoba, Spain, June 1-4, 2010, Proceedings, Part II 23*. Springer, 585–594.

[35] P. Murali, L. Ring, H. Trinh, R. Asadi, and T. Bickmore. 2018. Speaker hand-offs in collaborative human-agent oral presentations. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 153–158.

[36] B. Nooraei, C. Rich, and C. L. Sidner. 2014. A real-time architecture for embodied conversational agents: beyond turn-taking. *ACHI* 14 (2014), 381–388. https://api.semanticscholar.org/CorpusID:5981370

[37] J. D. Orcutt and L. K. Harvey. 1985. Deviance, rule-breaking and male dominance in conversation. *Symbolic Interaction* 8, 1 (1985), 15–32.

[38] M. Paetzel, R. Manuvinakurike, and D. DeVault. 2015. "So, which one is it?" The effect of alternative incremental architectures in a high-performance game-playing agent. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 77–86.

[39] E. Park, D. Jin, and A. P del Pobil. 2012. The law of attraction in human-robot interaction. *International Journal of Advanced Robotic Systems* 9, 2 (2012), 35.

[40] A. Raux and M. Eskenazi. 2007. A multi-layer architecture for semi-synchronous event-driven dialogue management. In *2007 IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU)*. IEEE, 514–519.

[41] B. Reeves and C. Nass. 1996. The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK* 10, 10 (1996), 19–36.

[42] D. Roger, P. Bull, and S. Smith. 1988. The Development of a Comprehensive System for Classifying Interruptions. *Journal of Language and Social Psychology* 7, 1 (1988), 27–34. https://doi.org/10.1177/0261927X8800700102

[43] R. C. Rose and H. K. Kim. 2003. A hybrid barge-in procedure for more reliable turn-taking in human-machine dialog systems. In *2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No. 03EX721)*. IEEE, 198–203.

[44] H. Sacks, E. A. Schegloff, and G. Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50, 4 (1974), 696–735. https://doi.org/10.2307/412243

[45] E. A. Schegloff. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in Society* 29, 1 (2000), 1–63. https://doi.org/10.1017/S0047404500001019

[46] E. A. Schegloff and H. Sacks. 1973. Opening up closings. *Semiotica* 8, 4 (1973), 289–327. https://doi.org/10.1515/semi.1973.8.4.289

[47] E. Shriberg, A. Stolcke, and D. Baron. 2001. Observations on overlap: findings and implications for automatic processing of multi-party conversation. In *7th*

European Conference on Speech Communication and Technology (Eurospeech 2001). ISCA, 1359–1362. https://doi.org/10.21437/Eurospeech.2001-352

[48] G. Skantze. 2021. Turn-taking in Conversational Systems and Human-Robot Interaction: A Review. *Computer Speech & Language* 67 (2021), 101178. https://doi.org/10.1016/j.csl.2020.101178

[49] G. Skantze. 2021. Turn-taking in Conversational Systems and Human-Robot Interaction: A Review. *Computer Speech & Language* 67 (2021), 101178. https://doi.org/10.1016/j.csl.2020.101178

[50] M. Skantze, G.and Johansson and J. Beskow. 2015. Exploring Turn-taking Cues in Multi-party Human-robot Discussions about Objects. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (Seattle, Washington, USA) *(ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 67–74. https://doi.org/10.1145/2818346.2820749

[51] L. Staum Casasanto, K. Jasmin, and D. Casasanto. 2010. Virtually accommodating: Speech rate accommodation to a virtual interlocutor. In *32nd Annual Meeting of the Cognitive Science Society (CogSci 2010)*. Cognitive Science Society, 127–132.

[52] N. Ström and S. Seneff. 2000. Intelligent barge-in in conversational systems. In *6th International Conference on Spoken Language Processing (ICSLP 2000)* (2000-10-16). ISCA, vol. 2, 652–655–0. https://doi.org/10.21437/ICSLP.2000-353

[53] Silero Team. 2021. Silero VAD: pre-trained enterprise-grade Voice Activity Detector (VAD), Number Detector and Language Classifier. https://github.com/snakers4/silero-vad.

[54] M. Ter Maat, K. P Truong, and D. Heylen. 2010. How turn-taking strategies influence users' impressions of an agent. In *Intelligent Virtual Agents: 10th International Conference, IVA 2010, Philadelphia, PA, USA, September 20-22, 2010. Proceedings 10*. Springer, 441–453.

[55] M. Ter Maat, K. P. Truong, and D. Heylen. 2011. How agents' turn-taking strategies influence impressions and response behaviors. *Presence: Teleoperators and Virtual Environments* 20, 5 (2011), 412–430.

[56] K. P. Truong. 2013. Classification of cooperative and competitive overlaps in speech using cues from the context, overlapper, and overlappee. In *Interspeech 2013* (2013-08-25). ISCA, 1404–1408. https://doi.org/10.21437/Interspeech.2013-368

[57] K. P. Truong. 2013. Classification of cooperative and competitive overlaps in speech using cues from the context, overlapper, and overlappee. In *Proc. Interspeech 2013*. 1404–1408. https://doi.org/10.21437/Interspeech.2013-368

[58] B. Wells and S. Macfarlane. 1998. Prosody as an Interactional Resource: Turn-projection and Overlap. *Language and Speech* 41, 3-4 (July 1998), 265–294. https://doi.org/10.1177/002383099804100403

[59] C. West. 1979. Against our will: Male interruptions of females in cross-sex conversation. *Annals of the New York Academy of Sciences* 327 (1979), 81–97. https://doi.org/10.1111/j.1749-6632.1979.tb17755.x

[60] L. Yang. 2001. Visualizing spoken discourse: Prosodic form and discourse functions of interruptions. In *Current and New Directions in Discourse and Dialogue*. Springer, 355–381.

[61] L. Yang, C. Achard, and C. Pelachaud. 2022. Multimodal Analysis of Interruptions. In *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Anthropometry, Human Behavior, and Communication*, Vincent G. Duffy (Ed.). Vol. 13319. Springer International Publishing, 306–325. https://doi.org/10.1007/978-3-031-05890-5_24 Series Title: Lecture Notes in Computer Science.

[62] L. Yang, C. Achard, and C. Pelachaud. 2023. Now or When? Interruption timing prediction in dyadic interaction. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*. 1–4.

[63] M. Żarkowski. 2019. Multi-party turn-taking in repeated human–robot interactions: an interdisciplinary evaluation. *International Journal of Social Robotics* 11, 5 (2019), 693–707.