"Hmm, Did You Hear What I Just Said?": Development of a Re-Engagement System for Socially Interactive Robots

Hoang-Long Cao [1,2,*,†] , **Paola Cecilia Torrico Moron** [1,†], **Pablo G. Esteban** [1], **Albert De Beir** [1] , **Elahe Bagheri** [1], **Dirk Lefeber** [1] **and Bram Vanderborght** [1]

[1]   Brussels Human Robotics Research Center, Vrije Universiteit Brussel and Flanders Make, 1050 Brussels, Belgium; Paola.Cecilia.Torrico.Moron@vub.be (P.C.T.M.); pablo.gomez.esteban@vub.be (P.G.E.); albert@debeir.com (A.D.B.); Elahe.Bagheri@vub.be (E.B.); dirk.lefeber@vub.be (D.L.); Bram.Vanderborght@vub.be (B.V.)
[2]   College of Engineering Technology, Can Tho University, Can Tho 90000, Vietnam
*   Correspondence: hoang.long.cao@vub.be
†   These authors contributed equally to this work.

check for updates

Abstract: Maintaining engagement is challenging in human–human interaction. When disengagements happen, people try to adapt their behavior with an expectation that engagement will be regained. In human–robot interaction, although socially interactive robots are engaging, people can easily drop engagement while interacting with robots. This paper proposes a multi-layer re-engagement system that applies different strategies through human-like verbal and non-verbal behaviors to regain user engagement, taking into account the user's attention level and affective states. We conducted a usability test in a robot storytelling scenario to demonstrate technical operation of the system as well as to investigate how people react when interacting with a robot with re-engagement ability. Our usability test results reveal that the system has the potential to maintain a user's engagement. Our selected users gave positive comments, through open-ended questions, to the robot with this ability. They also rated the robot with the re-engagement ability higher on several dimensions, i.e., animacy, likability, and perceived intelligence.

Keywords: socially interactive robots; re-engagement; gaze; user's performance; user's perception; Pepper

1. Introduction and Background

Socially interactive robots are designed to interact with people by perceiving the complex surrounding environment and expressing verbal and nonverbal behaviors using speech, facial expressions, paralanguage, and body language [1,2]. These robots are expected to be present in many domestic applications including education [3,4], healthcare [5–9], and museum guidance [10,11] due to their abilities to engage, entertain, and enlighten people [1]. These social abilities enable the robots to be perceived as trusting, helpful, reliable, and importantly, engaging [12], which are essential traits for a harmonic human–robot coexistence and an effective human–robot interaction (HRI) [13].

Although robots with social abilities can provide engaging interactions with people, maintaining engagement with different kinds of users is important but challenging [14–16]. For example, people interacting with robots 'in the wild' can disengage with the robots at any time compared to in-laboratory settings [15]. In child–robot interactions, many authors found that children's social engagement gradually declines as time progresses e.g., [17–19]. There are different factors influencing people's engagement while interacting with robots. Moshkina et al. [20] found that people are engaged with

robots in public places if the robots produce human-like actions and social cues. Ivaldi et al. [21] found that people's personality influences the tendency and the length of human–robot conversation in assembly tasks. Kuno et al. [22] and Sidner et al. [23] discovered that a robot's gaze heightened human–robot engagement. Yamazaki et al. [24] found that coordination of verbal and non-verbal actions in the robot affects visitor engagement at museums and exhibitions. Corrigan et al. [25] suggested that users' perceptions of the robot's characteristics (e.g., friendliness, helpfulness, and attentiveness) might lead to sustained engagement with both the task and robot in task-oriented human–robot interactions. Therefore, human–robot interaction should not follow pre-defined sequences and the robot should apply re-engagement strategies.

In human–human interaction (HHI), people tend to adapt their behavior in different ways when disengagements happen. For example, when students do not focus in class, a teacher might raise their voice or verbally ask to get students' attention. Or in retail, a sale-person may try to emphasize certain words or perform arm gestures to bring customers back to the conversation. Although these strategies obviously do not work all the time and with all people, there is an expectation that engagement will be regained. Some studies in HRI have attempted to apply human strategies to improve people's engagement in human–robot interaction, mainly in maintaining conversation. The first strategy is generating human-like behaviors. Sidner et al. [26] created an engaging robot by mimicking human conversational gaze behavior in collaborative conversation. Bohus and Horvitz [27] explored the use of linguistic hesitation actions (e.g., "uhm", "hmm") to manage conversational engagement in open-world, physically situated dialog systems. The second strategy is adapting robot behaviors to the user's affective states. Ahmad et al. [16] showed that emotion-based adaptation is the most effective way to sustain social engagement during long-term children-robot interaction. Chan and Nejat [28] implemented a method to promote engagement in cognitively stimulating activities taking into account user's affective states. In child–robot interaction, Mubin et al. [29] also suggested using user state adaptation to sustain engagement. Leite et al. [30] found that including empathy is beneficial for children's long-term engagement with robots.

In this paper, we present the development of a re-engagement system for socially interactive robots to investigate people's reactions when interacting with a robot with re-engagement ability. The system is developed by combining two re-engagement strategies analyzed above and encapsulating these behaviors in a multi-layer behavior system organization. We aim at a compact implementation so that the system is easy to customize and set up when operating in public spaces—especially for non-technical operators. Therefore, we implemented the system on SoftBank Robotics Pepper humanoid robot, one of the first mass-produced personal and service robots [31]. Moreover, the system uses only built-in sensors to measure the user's engagement and affective states without external sensory devices. The robot control structure is available for researchers to adopt the system framework for other targeted applications.

The system is demonstrated through a usability test in a robot storytelling scenario following a five-user usability engineering approach [32–34]. In storytelling, one side talks significantly more than the other. Therefore, the scenario can trigger a lot of disengagement moments for the re-engagement system to perform its functionalities, especially when the story is unfamiliar with the listeners [35]. We compare user perception, engagement, and performance when people interact with a robot using re-engagement strategies to those interacting with a robot without this ability.

The rest of this paper is organized as follows. Section 2 presents the system development and system implementation on the Pepper robot. The system usability test is demonstrated in Section 3. Our conclusions are given in Section 4.

2. System Development

Our system helps robots produce social and task-based behaviors with a re-engagement ability during human–robot interaction, taking into account the user's engagement and affective states.

This information influences the robot's internal affective state, which is used to trigger different re-engagement strategies.

2.1. Design Principles

Our design principles are derived from the system requirements, i.e., generating human-like behaviors, adapting robot behaviors to the user's affective states, and obtaining a compact implementation. First, the robot control system architecture is organized in layers in order to generate different types of behaviors (task-based and social) [36–39]. Depending on the user's affective states, the robot can switch between behavior layers to maintain an effective human–robot interaction. Second, behavior layers are modular for the ease of prioritizing human-like actions in each layer and increasing the ability to customize or improve the control system. Finally, the implemented robot, selected sensors, and sensory interpretation methods should be selected in favor of obtaining a compact implementation.

2.2. System Architecture

The system architecture was designed following the multi-layer behavior organization approach. The information processing model is shown in Figure 1, in which environmental information gathered by the perceptual system (e.g., touch, sound, and vision) is used to vary the robot's internal affective state and to produce abstract behaviors (social and task-based). These behaviors are executed on the robot platform by the actuation system. System architecture components are explained in the following subsection.

The system architecture is designed for general purpose applications and independent of the implemented robot platform, sensors, and interaction scenario. Therefore, numerical values of behavior parameters are decided and tuned during the implementation process (see Section 2.4).
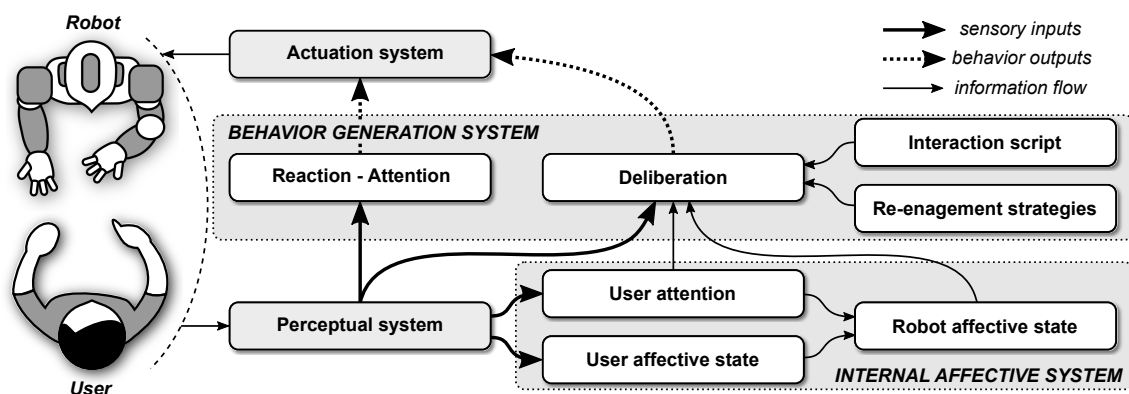

Figure 1. System architecture of the re-engagement system for socially interactive robots. Arrows denote connections between system architecture components.

2.3. System Architecture Components

2.3.1. Internal Affective System

This subsystem computes the robot's internal state allowing the robot to behave as a personal character. The output of this system is used to produce the robot's affective behaviors, e.g., adapting speech and gestures. It is worth mentioning that as a personal character, the robot (through the behavior generation system, Section 2.3.2) decides to express affective behaviors when it is necessary.

The robot's internal state is influenced by two parameters, i.e., user's attention and user's affective state. User's attention strongly depends on the user's gaze, compared to speech and facial expression. In fact, gaze is the most used factor to access engagement in human–robot communication studies [40,41]. User's affective state includes mood, from negative to positive, and emotional

expressions (e.g., happy, sad, angry, and surprised). The robot's internal affective value (R_{affect}) is calculated as follows,

$$R_{\text{affect}}(t) = \alpha_{attention}U_{\text{attention}} + \alpha_{mood}U_{\text{mood}} + \alpha_{emotion}U_{\text{emotion}} \tag{1}$$

where $\alpha_{attention}$, α_{mood}, $\alpha_{emotion}$ are the influences of the user's attention ($U_{\text{attention}}$), mood (U_{mood}), and emotion (U_{emotion}), on the robot's affective state respectively. These influences depend on the types and intensities of the events. If the user is paying attention to the robot or having a positive affective state (e.g., happy), these events will positively influence the robot's affective state, as the robot is performing well in getting the user's engagement. Consequently, the robot's internal affective state does not simply mimic the user's affective state but allows the robot to behave as an independent character.

2.3.2. Behavior Generation System

This system allows the robot to generate human-like behaviors (social and task-based). The system is designed following the three-layer behavior organization approach widely used in behavioral psychology, i.e., reaction–attention, deliberation, and reflection [36–39].

- The *reaction–attention layer* generates social behaviors, e.g., gazing, eye blinking, and micro-motions. These behaviors are similar to reflexes in neurological perspectives, and they allow the robot to react instantly to external stimuli and create the illusion of the robot being alive [42–44].
- The *deliberation layer* generates task-based behaviors. When the user is engaged with the interaction, this layer generates behaviors following the interaction script, e.g., a story, a lesson, or a guidance. When disengagements occur, the system attempts to apply three levels of re-engagement strategies. We adopted re-engagement strategies from human communication literature as summarized by Richmond et al. [45]). However, since human strategies are highly abstract, previous HRI studies used different ways to (partially) translate these strategies into programmable rules (e.g., [46–50]).
- The *reflection layer* evaluates behaviors decided by the lower layers and might change behavior planning if found not proper, e.g., ethically unacceptable. Due to the required complexity and the scope of the system development, this layer is not implemented.

The core component of the behavior generation system is the behavior selection in the deliberative layer. This whole process is summarized in Algorithm 1. Levels of re-engagement behaviors are selected according to the user's attention level. When the user's attention drops from a fully engaged level to lower ones (i.e., slightly low, moderately low, or significantly low), the noticeability of attention reminders gradually increases by adding different features to script-based behaviors. Specifically, level 1 includes small pausing movements when attention is slightly lost. This can be done by adding a short neutral gesture or a small pause in speech of the current script-based behavior. These pausing movements act as gentle attention reminders. If the user's attention drops to level 2, attention reminders are more noticeable by adapting robot speech with filled pauses (or hesitation markers) e.g., "uhm", "ehem" or with emphasized tone (speaking style). If the attention drops significantly to level 3, the robot pauses the script-based interaction and asks different kinds of questions (e.g., "Did you hear what I just said?", "What did I just say?"). A small off-script chatting creates the chance that the user will be more aware of the ongoing interaction. After engagement is regained, the behavior generation system resumes performing script-based behaviors.

Algorithm 1: Behavior generation mechanism in the deliberation layer. Specific numerical values of parameters (e.g., attention levels, movement speed, and speech) are chosen based on the implemented robot platform and sensors

---

　　Input　: User's attention
　　　　　　　　Robot affective state
　　　　　　　　Interaction script
　　　　　　　　Re-engagement strategies
　　Output: Deliberative behavior

1 **while** *not at end of Interaction script* **do**
2 　　**from** *Interaction script* **get** *current behavior*;
3 　　**switch** *User's attention* **do**
　　　　// Neutral behavior mode
4 　　　　**case** *engaged* **do**
5 　　　　　　**perform** *current behavior*;
6 　　　　**end**
　　　　// Re-engagement mode
7 　　　　**case** *Level 1: slightly low* **do**
8 　　　　　　**adapt** *current behavior* **with** *pause of movement*;
9 　　　　　　**perform** *current behavior* **with** *Robot affect*;
10 　　　　**end**
11 　　　　**case** *Level 2: moderately low* **do**
12 　　　　　　**adapt** *current behavior* **with** *speech connectors and filled pauses*;
13 　　　　　　**adapt** *current behavior* **with** *emphasize while speaking*;
14 　　　　　　**perform** *current behavior* **with** *Robot affect*;
15 　　　　**end**
16 　　　　**case** *Level 3: significantly low* **do**
　　　　　　// Pause Interaction script
17 　　　　　　**while** *User's attention is significantly low* **do**
18 　　　　　　　　**perform** *ask questions* **with** *Robot affect*;
19 　　　　　　**end**
　　　　　　// Resume Interaction script
20 　　　　**end**
21 　　**end**
22 　　**go to** next step in *Interaction script*;
23 **end**

The intensity of re-engagement behaviors depends on the robot's affective state. For example, speech parameters (i.e., speed, pause, and volume) are adapted to the robot's affect value throughout the three levels of re-engagement. When the interaction begins, the robot's affective state is positive. Therefore, it talks with a default style, speed, and volume. During the interaction, if the robot's affect becomes negative, the robot can speak more slowly (i.e., a lower speed and longer delays/pauses between sentences) to passionately raise awareness of disengagement. Depending on which personality the robot is assigned, the speech volume might be increased to passionately insist on attention or decreased to express sympathy. From the different methods of adapting speech parameters (S_i) to the robot's internal state (for a review see [51]), we adopt a linear interpolation approach that was used in previous work due to its intuitiveness and low computational complexity (e.g., [52–54]). Following this approach, speech parameters (S_i) are calculated based on their minimum and maximum values as follows.

$$S_{speed}(R_{affect}) = S_{speed}^{min} + \gamma_{speed}R_{affect}(S_{speed}^{max} - S_{speed}^{min}) \tag{2}$$

$$S_{pause}(R_{affect}) = S_{pause}^{min} + \gamma_{pause}R_{affect}(S_{pause}^{max} - S_{pause}^{min}) \tag{3}$$

$$S_{volume}(R_{affect}) = \begin{cases} S_{volume}^{min} + \gamma_{volume}R_{affect}(S_{volume}^{max} - S_{volume}^{min}) & \text{if } passion \\ S_{volume}^{min} + \gamma_{volume}(1 - R_{affect})(S_{volume}^{max} - S_{volume}^{min}) & \text{if } sympathy \end{cases} \tag{4}$$

where γ_{speed}, γ_{pause}, and γ_{volume} define how strongly each of the speech parameters is influenced by the robot's affect.

2.4. System Implementation

The system was implemented on the Pepper robot of SoftBank Robotics using Python NAOqi SDK and Choregraphe (the software and desktop application that are used to create a robot's behaviors from basic to complex actions). Unlike studies using external devices to access the user's engagement and affective states, our system runs completely on a Pepper robot using the robot's sensors (2D and 3D cameras) with built-in sensory interpretation methods to understand user's behavior. All functions of the system i.e., sensing, decision-making, and executing, are managed by NAOqi API (Application programming interface). e.g., ALGazeAnalysis, ALMood, ALFaceCharacteristics, ALAnimatedSpeech, ALSpeakingMovement, ALAutonomousLife (for the reaction–attention system).

2.4.1. Compact Implementation

Figure 2 shows our developed software containing Choregraphe built-in boxes for sensing and developed Python boxes for different computations. The system is modular since the code of each system component was developed in a separate Choregraphe box. This allows complex algorithms to be developed while keeping the software structure intuitive and understandable. The interaction scenario is scripted in a separate text file and independent from the behavior generation. A log file is created and stored in the robot's internal folder for analysis purposes.
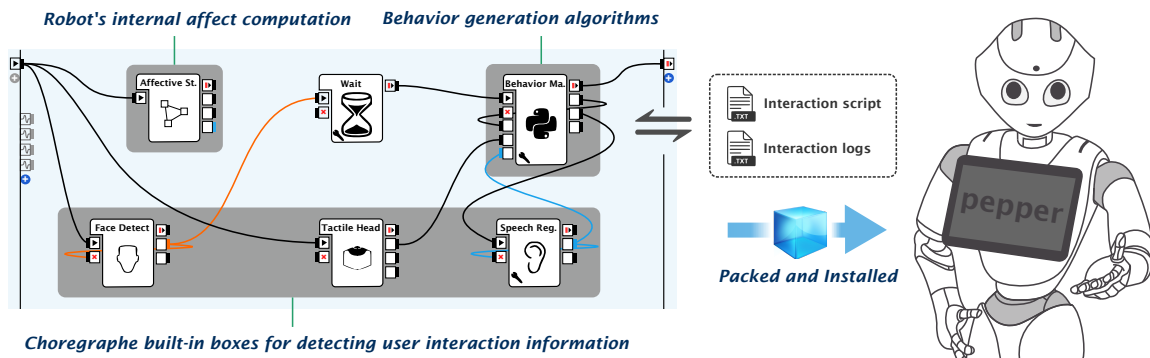

Figure 2. System implementation using Choregraphe built-in boxes and developed Python boxes. The software is packed and run completely on the Pepper robot platform.

The software can be packed and run completely on the robot platform without installing external dependencies or setting up external sensory devices. The robot control structure is accessible through our GitHub project at https://github.com/hoanglongcao/Pepper-Re-engagement (subject to update). Since there are no rigid rules in human communication strategies, developers can adapt the structure or implement more strategies for their targeted applications.

2.4.2. System Parameters

As mentioned in the design principles and system architecture design in Sections 2.1 and 2.2, the numerical values of system parameters are chosen based on the implemented robot platform (Pepper), sensors (Pepper built-in sensors), and the interaction scenario.

The robot's affect is calculated from the user's attention, mood, and emotion. In this implementation, we consider that these elements contribute equally to the robot's affect ($\alpha_{attention} = \alpha_{mood} = \alpha_{emotion}$). The user's attention is retrieved by checking the ALMemory key PeoplePerception/Person/<ID>/LookingAtRobotScore ($\overline{0,1}$) using ALGazeAnalysis API. The user's mood is retrieved by ALMood, which returns three possible values, i.e., positive, neutral, negative. The user's emotion is calculated based on the user's facial expression (i.e., neutral, happy, surprised, angry, or sad) and its intensity ($\overline{0,1}$). These values are retrieved by reading the ALMemory key PeoplePerception/Person/<ID>/ExpressionProperties using ALFaceCharacteristics API. Since each facial expression (or emotion) influences the robot's affect to a different extent, a correction factor (β_i) is added for each emotion intensity ($\beta_{happy} = 1$, $\beta_{sad} = 1/2$, $\beta_{angry} = \beta_{surprised} = 1/3$). The robot's affect value is normalized ($\overline{0,100\%}$) after adding up values of the contributed elements with their influenced values.

Speech parameters are then adapted to the calculated robot's affect value either positively or negatively. The influences of the robot's affect on numerical speech parameters (i.e., speed, pause, and volume) are selected to be equal ($\gamma_{speed} = \gamma_{pause} = \gamma_{volume} = 1$). Maximum and minimum values of these numerical parameters are selected empirically. For speech style, we selected the 'accented' style (\\emph=2\\) when emphasized voice is needed.

3. System Usability Test

We conducted a usability test in which users interact with the Pepper robot in a storytelling scenario to demonstrate the system technical performance and its ability to regain a user's attention during interaction. For this purpose, we follow a five-user usability engineering approach that can reveal 85% of the usability functionalities and problems [32–34]. We selected a storytelling scenario with a story about social robots and their applications. When storytelling comes with an unfamiliar topic, the story is more difficult to tell and listeners enjoy the story less than a familiar topic [35]. Therefore, this scenario can create more user disengagements and give space for the system to perform its functionalities as well as problems.

3.1. Users

We recruited ten people from different backgrounds including six men and four women. Their ages ranged from 22 to 34 years old ($M = 27.5$, $SD = 3.4$). All users did not have prior experience with robots. Figure 3 shows our system usability setup, in which a user sits in front of a robot. A camera was used to record the interaction.

3.2. Usability Testing Design

We designed a 2×1 between-user usability test, in which our selected users are divided randomly into two groups. Each group consists of five users for usability testing, which aims at showing most of the system functionalities and problems [32–34].

In the first group, users interact with a Pepper robot with the re-engagement mode activated i.e., activated condition. In the second group, the re-engagement mode was deactivated i.e., deactivated condition. The interaction scenarios are the same in the two groups.

Since all users are not familiar with robots, they might have high levels of interest in interacting with the robot at the beginning. However, this issue is consistent between the two groups. More importantly, the selected story about social robots and applications is lengthy and unfamiliar. This story would possibly create disengagements throughout the usability test duration.
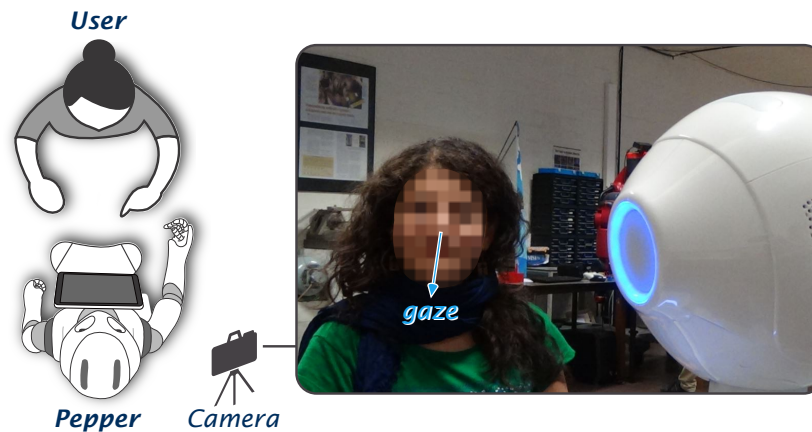
Figure 3. System usability test setup. A user interacting with a Pepper robot with the re-engagement system implemented.

3.3. Interaction Procedure

We asked the selected users to sign a consent form and gave them a brief introduction about the usability test before interacting with the robot. The robot first introduced itself and followed this by giving information about social robots and their applications. Afterward, users were asked to answer a post-test questionnaire about the information given by the robot, their perceptions of the robot, and open questions about their impressions and interaction experience with the robot. Finally, users were compensated with small gifts for their time. The entire session lasted about ten minutes.

3.4. Open Questions and Quantitative Measurements

We asked three open questions to have some qualitative insights to understand the user's attitude toward the robot and the interaction. The first question is about the user's impression of the robot itself through verbal and non-verbal behaviors. The second question is related to the user's experience about the interaction, i.e., how it was given by the robot. The last question assessed the user's perception of the role of the robot.

We also performed a quantitative assessment. First, we measured the user's gaze (times, duration) during the interaction. Second, we asked users to fill in a post-interaction questionnaire including eight questions about the information given by the robot to measure the user's performance and 24 items from the Godspeed questionnaire to access the user's perception toward the robot i.e., anthropomorphism, animacy, likability, perceived intelligence, and perceived safety [55].

3.5. Results and Discussion

3.5.1. Technical Performance of the Re-Engagement System

In human–human storytelling, people show different levels of attention depending on many factors, e.g., background, personality, and emotion. Some people are more engaged in the story while some other people are less attentive. To demonstrate the system performance, we present two selected cases representing two main types of users, i.e., lower-engaged users and higher engaged users. In this part, we present two subsets of interactions representing these two types of users with the re-engagement ability activated (activated condition).

In both cases, when the user's attention dropped, the system adapts to the user's attention by applying three levels of human-like re-engagement behaviors trying to bring their attention back. Specifically, the system observes the user's attention level to decide which level of re-engagement should be applied (Figure 4). Level 1 applies slight changes in verbal behaviors. Level 2 increases the intensity of re-engagement behaviors by adding filled pauses and changing to an emphasized ('accented') tone. Level 3 uses direct questions. These actions were not usually performed exactly

at the moments of attention changes. Most of the time, the robot needed to finish its current behavior (completing a speech or a gesture) before applying a certain re-engagement strategy to the upcoming behavior.
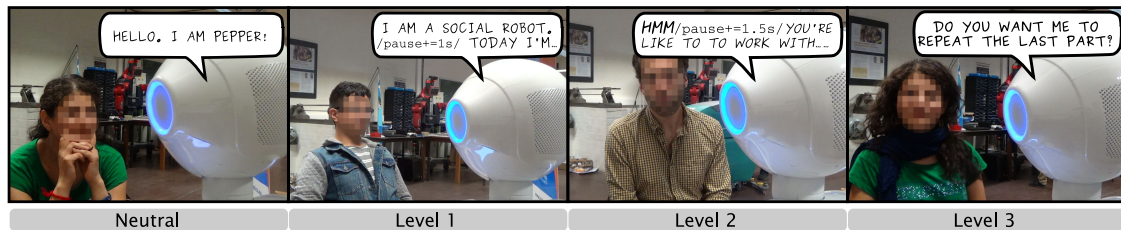

Figure 4. Robot behavior during the interaction. The robot switches from neutral behavior mode to re-engagement behavior mode according to user's attention level.

As mentioned in the system development, the robot's internal affective state does not merely imitate the user's affective state but takes into account the user's mood, emotion, and attention. Therefore, the robot still behaves as an independent and personal character. For example, when the user has a positive affective state but loses attention, the robot affect changes to a negative state to reflect its awareness of the user's disengagement. Or when the user is still engaged in the interaction script, although the robot's internal affective level might vary, this affect is not shown through the robot's behaviors.

The higher-engaged user interaction subset is shown in Figure 5a, in which the user shows a high degree of attention to the story given by the robot. Most of the time, the attention level is above the threshold of level 1. Consequently, the robot mainly performs script-based behavior following the story script. When the attention level slightly dropped (a few times), the system applied the first level of re-engagement behaviors and the user's attention quickly resumed.
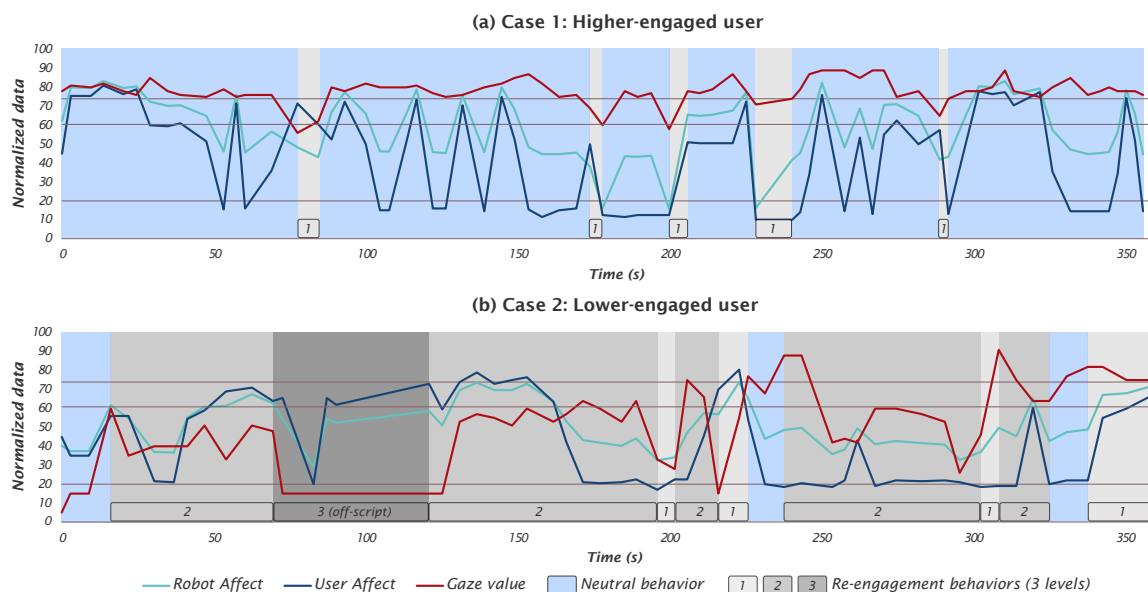

Figure 5. Subsets from two interactions. (**a**) Re-engagement ability deactivated. (**b**) Re-engagement ability activated to perform different strategies to regain the user's attention. Three horizontal lines represent different levels of user's attention specifically calibrated for the Pepper robot in the usability test.

The lower-engaged user interaction subset is shown in Figure 5b, in which the user lost attention at different levels. Therefore, the system tries to adapt to this by applying all three levels of re-engagement behaviors. With lower-engaged users, the time to finish the story script is longer since the system takes some time to perform re-engagement behaviors.

Although not happening with our users in the usability test, we foresee that we might rarely have very low-engaged users who constantly keep their attention values below level 3. The robot in this case will keep asking direct questions and might annoy the users. The next iteration of system development should be added with a possibility to smoothly end the storytelling instead of continuing to try to regain attention.

3.5.2. Answers to Open Questions

Since all selected users did not have prior experience with robots, most of them were highly interested in interacting with the robot. In both conditions, there was no difference in perception of the robot's role, seven users (three plus four) considered the robot as a friend. Three other users considered the robot as a teacher, a neighbor, and a stranger. However, regarding questions about their impression and interaction experience, answers from our selected users revealed some differences between the two conditions.

Tables 1 and 2 listed all answers about the users' impression of the robot and their interaction experience. In general, users gave quite positive responses in both conditions (e.g., *nice*, *impress(ive)*, and *gesture(s)*) since they did not have prior experience interacting with robots. However, users in the activated condition commented more about human-like behavior from the re-engagement strategies of the robot (even without seeing the other condition).

Table 1. Users' impression of the robot in two conditions: re-engagement deactivated and activated.

	Impression of the Robot
Activated (*n* = 5)	**U01:** *I found it friendly and nice.* **U02:** *It was very lifelike, it was hard to remember it is a robot.* **U03:** *Friendly.* **U04:** *Impressive the way to show emotions, through emotions. Nevertheless, in time the movements become somehow monotonous and predictable.* **U05:** *A lively social robot with nice hand gestures.*
Deactivated (*n* = 5)	**U06:** *First surprised, then a little bit depressed because I have no conversation with him.* **U07:** *It was really trying to impress me and I could feel his try which was wonderful. But his hand motion seems too much and didn't permit me to focus on his speaking which I prefer he pays attention and adjusts it.* **U08:** *I like the hand movements and the way it looks around curiously.* **U09:** *It was nice and good gesture.* **U10:** [No impression]

Table 2. Users' interaction experience with the robot in two conditions: re-engagement deactivated and activated.

	Interaction Experience
Activated (*n* = 5)	**U01:** *I felt comfortable and I wanted to share more, and Pepper identified my reactions super fast which was surprising for me.* **U02:** *He asked questions and understood my responses.* **U03:** *Nice and friendly.* **U04:** *It was a pleasant experience. I would have like to interact longer.* **U05:** *It feels nice to know pepper. It enlightened me about how robots are handling social interactiveness.*
Deactivated (*n* = 5)	**U06:** *Not at all.* **U07:** *It was really unique and wonderful; but a bit monologue! it was much better if i also could speak with him and see his realtime interaction abilities.* **U08:** *Good.* **U09:** *Responsive interaction.* **U10:** *Communication.*

In the activated condition, people mentioned that the robot is *friendly*, *lively*, and *lifelike*. They also realized that the robot could express *emotions*, *identified my*(user's) *reactions*, *understood my* (user's) *response, social interactiveness*. User 04 described the interaction as *pleasant*, and would like to *interact longer* with the robot.

In the deactivated condition, the users showed less impressions and commented mainly about the robot's *gestures* and *hand movements*. User 07 mentioned that the robot gestures influenced their focus on the story, possibly because there was no awareness of the user's attention and affective state. User 10 did not show any impression. Users also gave short and less positive answers about the interaction experience. User 06 eventually did not appreciate the interaction experience.

Since we chose a novel story to our selected users in a storytelling scenario to trigger disengagements, users felt that the robot talked more than listened in both cases. However it was mentioned more in the deactivated condition.

3.5.3. Quantitative Assessment: User's Engagement, Performance, and Perception

We performed a between-user comparison to compare user's engagement, performance and perception in two conditions. Table 3 summarized our results, which show some differences between the two groups of our selected users in the usability test. Results of the quantitative measures are in line with the answers of our users to the open questions (see Section 3.5.2).

Table 3. Results of quantitative measurements between two conditions: re-engagement deactivated and activated. M, SD, and Mdn represent mean, standard deviation, and median, respectively.

	Deactivated (*n* = 5)			Activated (*n* = 5)		
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*
Age	28	4.90	30	27	2.00	27
Engagement						
Time gazing at the robot (%)	43.04	29.50	56.28	63.13	18.30	57.95
Each gaze duration (s)	17.25	13.67	13.02	22.99	1.84	23.36
User's perception						
Anthropomorphism	3.72	0.58	3.60	3.76	0.62	3.60
Animacy	3.47	0.27	3.33	4.03	0.43	4.00
Likability	4.24	0.33	4.20	4.44	0.43	4.40
Perceived intelligence	3.88	0.67	4.0	4.04	0.59	4.20
Perceived safety	3.87	0.73	3.67	3.80	0.84	3.67

When the re-engagement ability was activated, users seemed to be more engaged in the interaction compared to when this ability was deactivated by having a higher mean time gazing at the robot (63.13%/43.04%) and a higher mean duration of each gaze (22.99 s/17.25 s). In this condition, users also recalled correctly more information given by the robot during interaction (45.0%/32.5%).

The Godspeed questionnaire results show that when the re-engagement ability was activated, users perceived the robot with higher scores in animacy (4.03/3.47), likability (4.44/4.24), and perceived intelligence (4.04/3.88). This can be explained by the human-like behaviors and the awareness of user's attention the robot expressed during the interaction. There was no clear difference in the other scales, i.e., anthropomorphism (3.76/3.72) and perceived safety (3.80/3.87), since the robot appearance (Pepper) and the interaction script were the same in both conditions.

Although the five-user usability engineering approach only aims at revealing system functionalities and problems, we attempted to run a non-parametric Wilcoxon rank-sum test with continuity correction to compare the medians of our measurements. Results showed that the difference in animacy (W = 2.5, $p < 0.05$) between the two groups of our selected users is significant. Due to the small sample size of this type of usability engineering approach, a quantitative usability test with more users is required to generalize these results to a bigger population.

4. Conclusions

We proposed the development of a re-engagement system for socially interactive robots to maintain a user's attention during human–robot interactions. The system was developed and implemented on a Pepper robot following three design principles, i.e., generating human-like behaviors, adapting robot behaviors to the user's affective states, and obtaining a compact implementation.

We conducted a usability test with a robot storytelling scenario with two groups of selected users to investigate how people react to the system. One group interacted with the robot when the re-engagement ability was activated, and in the other group this ability was deactivated. Results of the usability test show that the system has the potential to help users achieve higher engagement and performance. As in HHI, a user's attention in HRI strongly depends on the user's interest and willingness to interact with the robot. Other possible influencing factors include the interaction scenario, the user's background, and the quality of robot behavior realization. Therefore, applying different re-engagement strategies might not guarantee success in bringing the attention back to a fully engaged level, however, it will still enhance the robot's human-likeness. Our selected users gave more positive comments for the robot with re-engagement ability on impression and interaction experience. They also gave higher scores in animacy, likability, and perceived intelligence. Our future works will focus on improving the system in iterative usability tests and applying the system on different types of users (e.g., children, and students) with various interaction scenarios.

References

1. Fong, T.; Nourbakhsh, I.; Dautenhahn, K. A survey of socially interactive robots. *Robot. Auton. Syst.* **2003**, *42*, 143–166. [CrossRef]
2. McColl, D.; Nejat, G. Recognizing emotional body language displayed by a human-like social robot. *Int. J. Soc. Robot.* **2014**, *6*, 261–280. [CrossRef]
3. Belpaeme, T.; Kennedy, J.; Baxter, P.; Vogt, P.; Krahmer, E.E.; Kopp, S.; Bergmann, K.; Leseman, P.; Küntay, A.C.; Göksun, T.; et al. L2TOR-second language tutoring using social robots. In Proceedings of the ICSR 2015 WONDER Workshop, Paris, France, 26–30 October 2015.
4. Vogt, P.; De Haas, M.; De Jong, C.; Baxter, P.; Krahmer, E. Child-robot interactions for second language tutoring to preschool children. *Front. Hum. Neurosci.* **2017**, *11*, 73. [CrossRef] [PubMed]
5. Esteban, P.G.; Baxter, P.; Belpaeme, T.; Billing, E.; Cai, H.; Cao, H.L.; Coeckelbergh, M.; Costescu, C.; David, D.; De Beir, A.; et al. How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn J. Behav. Robot.* **2017**, *8*, 18–38. [CrossRef]
6. Belpaeme, T.; Baxter, P.E.; Read, R.; Wood, R.; Cuayáhuitl, H.; Kiefer, B.; Racioppa, S.; Kruijff-Korbayová, I.; Athanasopoulos, G.; Enescu, V.; et al. Multimodal child-robot interaction: Building social bonds. *J. Hum. Robot. Interact.* **2012**, *1*, 33–53. [CrossRef]
7. Cao, H.L.; Esteban, P.G.; Bartlett, M.; Baxter, P.; Belpaeme, T.; Billing, E.; Cai, H.; Coeckelbergh, M.; Costescu, C.; David, D.; et al. Robot-enhanced therapy: Development and validation of a supervised autonomous robotic system for autism spectrum disorders therapy. *IEEE Robot. Autom. Mag.* **2019**, *26*, 49–58. [CrossRef]
8. Loza-Matovelle, D.; Verdugo, A.; Zalama, E.; Gómez-García-Bermejo, J. An Architecture for the Integration of Robots and Sensors for the Care of the Elderly in an Ambient Assisted Living Environment. *Robotics* **2019**, *8*, 76. [CrossRef]

9.　Palacín, J.; Clotet, E.; Martínez, D.; Martínez, D.; Moreno, J. Extending the Application of an Assistant Personal Robot as a Walk-Helper Tool. *Robotics* **2019**, *8*, 27. [CrossRef]

10.　Burgard, W.; Cremers, A.B.; Fox, D.; Hähnel, D.; Lakemeyer, G.; Schulz, D.; Steiner, W.; Thrun, S. Experiences with an interactive museum tour-guide robot. *Artif. Intell.* **1999**, *114*, 3–55. [CrossRef]

11.　Yamazaki, A.; Yamazaki, K.; Ohyama, T.; Kobayashi, Y.; Kuno, Y. A techno-sociological solution for designing a museum guide robot: Regarding choosing an appropriate visitor. In Proceedings of the 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Boston, MA, USA, 5–8 March 2012; pp. 309–316.

12.　Kidd, C.D.; Breazeal, C. Effect of a robot on user perceptions. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), Sendai, Japan, 28 September–2 October 2004; Volume 4, pp. 3559–3564.

13.　Xu, J. Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction. Ph.D. Thesis, Delft University of Technology, TU Delft, The Netherlands, 2015.

14.　Mower, E.; Feil-Seifer, D.J.; Mataric, M.J.; Narayanan, S. Investigating implicit cues for user state estimation in human-robot interaction using physiological measurements. In Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2007), Jeju, Korea, 26–29 August 2007; pp. 1125–1130.

15.　Pitsch, K.; Kuzuoka, H.; Suzuki, Y.; Sussenbach, L.; Luff, P.; Heath, C. "The first five seconds": Contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI. In Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2009), Toyama, Japan, 27 September–2 October 2009; pp. 985–991.

16.　Ahmad, M.I.; Mubin, O.; Orlando, J. Adaptive social robot for sustaining social engagement during long-term children–robot interaction. *Int. J. Hum. Comput. Interact.* **2017**, *33*, 943–962. [CrossRef]

17.　Coninx, A.; Baxter, P.; Oleari, E.; Bellini, S.; Bierman, B.; Henkemans, O.B.; Cañamero, L.; Cosi, P.; Enescu, V.; Espinoza, R.R.; et al. Towards long-term social child-robot interaction: using multi-activity switching to engage young users. *J. Hum. Robot. Interact.* **2016**, *5*, 32–67. [CrossRef]

18.　Komatsubara, T.; Shiomi, M.; Kanda, T.; Ishiguro, H.; Hagita, N. Can a social robot help children's understanding of science in classrooms? In Proceedings of the Second International Conference on Human-Agent Interaction, Tsukuba, Japan, 29–31 October 2014; pp. 83–90.

19.　Jimenez, F.; Yoshikawa, T.; Furuhashi, T.; Kanoh, M. An emotional expression model for educational-support robots. *J. Artif. Intell. Soft Comput. Res.* **2015**, *5*, 51–57. [CrossRef]

20.　Moshkina, L.; Trickett, S.; Trafton, J.G. Social engagement in public places: A tale of one robot. In Proceedings of the 2014 ACM/IEEE international conference on Human-Robot Interaction, Bielefeld, Germany, 3–6 March 2014; pp. 382–389.

21.　Ivaldi, S.; Lefort, S.; Peters, J.; Chetouani, M.; Provasi, J.; Zibetti, E. Towards engagement models that consider individual factors in HRI: On the relation of extroversion and negative attitude towards robots to gaze and speech during a human–robot assembly task. *Int. J. Soc. Robot.* **2017**, *9*, 63–86. [CrossRef]

22.　Kuno, Y.; Sadazuka, K.; Kawashima, M.; Yamazaki, K.; Yamazaki, A.; Kuzuoka, H. Museum guide robot based on sociological interaction analysis. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, April 28–3 May 2007; pp. 1191–1194.

23.　Sidner, C.L.; Lee, C.; Kidd, C.D.; Lesh, N.; Rich, C. Explorations in engagement for humans and robots. *Artif. Intell.* **2005**, *166*, 140–164. [CrossRef]

24.　Yamazaki, A.; Yamazaki, K.; Burdelski, M.; Kuno, Y.; Fukushima, M. Coordination of verbal and non-verbal actions in human-robot interaction at museums and exhibitions. *J. Pragmat.* **2010**, *42*, 2398–2414. [CrossRef]

25.　Corrigan, L.J.; Basedow, C.; Küster, D.; Kappas, A.; Peters, C.; Castellano, G. Perception matters! Engagement in task orientated social robotics. In Proceedings of the 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Kobe, Japan, 31 August–4 September 2015; pp. 375–380.

26.　Sidner, C.L.; Kidd, C.D.; Lee, C.; Lesh, N. Where to look: A study of human-robot engagement. In Proceedings of the 9th International Conference on Intelligent User Interfaces, Funchal, Portugal, 13–16 January 2004; pp. 78–84.

27.　Bohus, D.; Horvitz, E. Managing human-robot engagement with forecasts and... um... hesitations. In Proceedings of the 16th International Conference on Multimodal Interaction, stanbul, Turkey, 12–16 November 2014; pp. 2–9.

28. Chan, J.; Nejat, G. Promoting engagement in cognitively stimulating activities using an intelligent socially assistive robot. In Proceedings of the 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Montreal, ON, Canada, 6–9 July 2010; pp. 533–538.

29. Mubin, O.; Stevens, C.J.; Shahid, S.; Al Mahmud, A.; Dong, J.J. A review of the applicability of robots in education. *J. Technol. Educ. Learn.* **2013**, *1*, 13. [CrossRef]

30. Leite, I.; Castellano, G.; Pereira, A.; Martinho, C.; Paiva, A. Empathic robots for long-term interaction. *Int. J. Soc. Robot.* **2014**, *6*, 329–341. [CrossRef]

31. Pandey, A.; Gelin, R. A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind. *IEEE Robot. Autom. Mag.* **2018**, *25*, 40–48. [CrossRef]

32. Lewis, J.R. Sample sizes for usability studies: Additional considerations. *Hum. Factors* **1994**, *36*, 368–378. [CrossRef] [PubMed]

33. Nielsen, J. Usability 101: Introduction to Usability. Available online: https://www.nngroup.com/articles/usability-101-introduction-to-usability/ (accessed 8 November 2019).

34. Virzi, R.A. Refining the test phase of usability evaluation: How many subjects is enough? *Hum. Factors* **1992**, *34*, 457–468. [CrossRef]

35. Cooney, G.; Gilbert, D.T.; Wilson, T.D. The novelty penalty: Why do people like talking about new experiences but hearing about old ones? *Psychol. Sci.* **2017**, *28*, 380–394. [CrossRef] [PubMed]

36. Ortony, A.; Clore, G.L.; Collins, A. *The Cognitive Structure of Emotions*; Cambridge University Press: Cambridge, UK, 1990.

37. Ortony, A.; Norman, D.; Revelle, W. Affect and Proto-Affect in Effective Functioning. In *Who Needs Emotions, The Brain Meets the Robot*; Oxford University Press: New York, NY, USA, 2005.

38. Sloman, A. Varieties of meta-cognition in natural and artificial systems. *Metareasoning: Thinking about Thinking.* **2011**, 307–323. [CrossRef]

39. Sloman, A.; Logan, B. Evolvable Architectures for Human-Like Minds. In *Affective Minds*; Elsevier: Amsterdam, The Netherlands, 2000; pp. 169–181.

40. Mutlu, B.; Forlizzi, J.; Hodgins, J. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In Proceedings of the 2006 6th IEEE-RAS International Conference on Humanoid Robots, Genova, Italy, 4–6 December 2006; pp. 518–523.

41. Yoshikawa, Y.; Shinozawa, K.; Ishiguro, H.; Hagita, N.; Miyamoto, T. Responsive Robot Gaze to Interaction Partner. In Proceedings of the Robotics: Science and Systems, Philadelphia, PA, USA, 16–19 August 2006.

42. Lazzeri, N.; Mazzei, D.; Zaraki, A.; De Rossi, D. Towards a believable social robot. In *Biomimetic and Biohybrid Systems*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 393–395.

43. Saldien, J.; Vanderborght, B.; Goris, K.; Van Damme, M.; Lefeber, D. A motion system for social and animated robots. *Int. J. Adv. Robot. Syst.* **2014**, *11*. [CrossRef]

44. Gómez Esteban, P.; Cao, H.L.; De Beir, A.; Van de Perre, G.; Lefeber, D.; Vanderborght, B. A multilayer reactive system for robots interacting with children with autism. In Proceedings of the 5th International Symposium on New Frontiers in Human-Robot Interaction, Sheffield, UK, 5–6 April 2016; pp. 1–4.

45. Richmond, V.P.; Gorham, J.S.; McCroskey, J.C. The relationship between selected immediacy behaviors and cognitive learning. *Ann. Int. Commun. Assoc.* **1987**, *10*, 574–590. [CrossRef]

46. Shim, J.; Arkin, R.C. Other-oriented robot deception: A computational approach for deceptive action generation to benefit the mark. In Proceedings of the 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014), Bali, Indonesia, 5–10 December 2014; pp. 528–535.

47. Szafir, D.; Mutlu, B. Pay attention!: Designing adaptive agents that monitor and improve user engagement. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012; pp. 11–20.

48. Baroni, I.; Nalin, M.; Zelati, M.C.; Oleari, E.; Sanna, A. Designing motivational robot: How robots might motivate children to eat fruits and vegetables. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (2014 RO-MAN), Edinburgh, UK, 25–29 August 2014; pp. 796–801.

49. Chidambaram, V.; Chiang, Y.H.; Mutlu, B. Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, Boston, MA, USA, 5–8 March 2012; pp. 293–300.

50. Brown, L.; Kerwin, R.; Howard, A.M. Applying behavioral strategies for student engagement using a robotic educational agent. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 October 2013; pp. 4360–4365.

51. Crumpton, J.; Bethel, C.L. A survey of using vocal prosody to convey emotion in robot speech. *Int. J. Soc. Robot.* **2016**, *8*, 271–285. [CrossRef]

52. Lim, A.; Okuno, H.G. The mei robot: Towards using motherese to develop multimodal emotional intelligence. *IEEE Trans. Auton. Ment. Dev.* **2014**, *6*, 126–138. [CrossRef]

53. Bennewitz, M.; Faber, F.; Joho, D.; Behnke, S. Intuitive multimodal interaction with communication robot Fritz. In *Humanoid Robots, Human-Like Machines*; Itech: Vienna, Austria, 2007.

54. Gonsior, B.; Sosnowski, S.; Buß, M.; Wollherr, D.; Kühnlenz, K. An emotional adaption approach to increase helpfulness towards a robot. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura, Portugal, 7–12 October 2012; pp. 2429–2436.

55. Bartneck, C.; Kulić, D.; Croft, E.; Zoghbi, S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* **2009**, *1*, 71–81. [CrossRef]