Modelling Adaptive Presentations in Human-Robot Interaction using Behaviour Trees

Nils Axelsson
Division of Speech, Music and Hearing
KTH Royal Institute of Technology
Stockholm, Sweden
nilsaxe@kth.se

Gabriel Skantze
Division of Speech, Music and Hearing
KTH Royal Institute of Technology
Stockholm, Sweden
skantze@kth.se

Abstract

In dialogue, speakers continuously adapt their speech to accommodate the listener, based on the feedback they receive. In this paper, we explore the modelling of such behaviours in the context of a robot presenting a painting. A Behaviour Tree is used to organise the behaviour on different levels, and allow the robot to adapt its behaviour in real-time; the tree organises engagement, joint attention, turn-taking, feedback and incremental speech processing. An initial implementation of the model is presented, and the system is evaluated in a user study, where the adaptive robot presenter is compared to a non-adaptive version. The adaptive version is found to be more engaging by the users, although no effects are found on the retention of the presented material.

1 Introduction

Speakers in dialogue cannot just assume that their speech is received by the addressee and understood as intended. They have to continuously monitor the addressee to verify that the information is attended to, perceived, understood and accepted (Clark, 1996). By keeping close track of verbal and non-verbal feedback from the addressee, speakers can alter their presentation in order to accommodate the listener.

In this paper, we explore how this process can be modelled in spoken human-robot interaction. As a test-bed, we have designed a scenario where a robot is presenting visual information (such as a poster or a piece of art) to a human, as seen in Figure 1. This setting allows us to explore how the presentation can be adapted to the audience's level of attention, understanding and engagement.

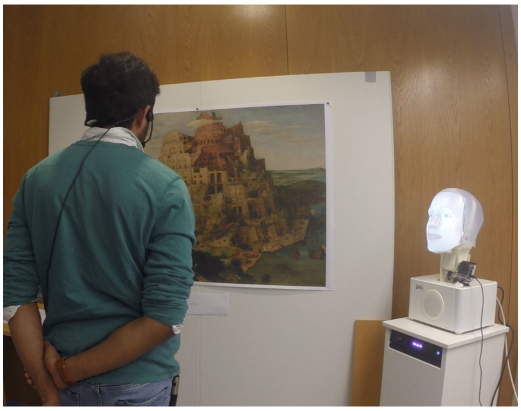


Figure 1: The scenario chosen as a test-bed for the model: a robot presenting a painting to a human.

Modelling adaptive presentation in a human-robot interaction scenario is non-trivial, as the robot needs to pick up feedback from different modalities, and continuously adapt its behaviour to accommodate the listener. It is also not obvious that such a system would be better in terms of teaching the presented material and user experience, compared to a fixed, non-adaptive presentation (such as audio-guides used in museums), as the robot is unlikely to exhibit the same level of adaptation as a human. This paper has two main contributions, which address these concerns. First, we explore the use of Behaviour Trees (Colledanchise and Ögren, 2018) for modelling the adaptive behaviour. Behaviour Trees, a specific formalism for decomposing a plan into a tree structure, have been applied extensively to video games and robotics (Hasegawa et al., 2017; Hu et al., 2015), and systems that break down an interaction or a dialogue to a tree are not new (Smith and Hipp, 1994; Boye, 2007; Bohus and Rudnicky, 2009). However, we are not aware of any previous attempts at applying specifically Behaviour Trees to real-time modelling of spoken interaction. Second, we present an experiment where we compare the adaptive robot presenter to a version where the presentation is statically executed, i.e., where the user's reactions are not taken into account.

Proceedings of the SIGDial 2019 Conference, pages 345–352
Stockholm, Sweden, 11-13 September 2019. ©2019 Association for Computational Linguistics

2 Background

The scenario of a robot presenting information to an audience (one or several people), has been explored in earlier work (Jensen et al., 2005; Szafir and Mutlu, 2012; Ohya et al., 2006). However, these works have not focused on how the presentation can be adapted based on verbal and non-verbal feedback. Poster presentations between humans have been studied in order to analyse the gaze and backchannel behaviours of participants and presenters (Kawahara, 2012). Hashimoto et al. (2011) and Verner et al. (2016) have shown that more interactive robot teachers lead to better results in learning. Yousuf et al. (2012) and Eichner et al. (2007) show that users prefer presenting agents that adapt their grounding behaviour to their audience.

2.1 Grounding and Adaptation

According to Clark (1996) and Allwood et al. (1992), any coordinated action can be described as an action ladder, with each level requiring the cooperation of speaker and addressee. If the speaker A is presenting to the addressee B, then the levels of the action ladder, bottom-to-top, are **attention** (B must be paying attention to A's presentation), **hearing** (B must hear the words said by A), **understanding** (B must understand the meaning behind the words said by A) and **acceptance** (B must accept, and optionally be interested in, the concept proposed by A's presentation).

The addressee can give positive and negative evidence of each level (feedback), to signal completeness to the speaker. If negative evidence is signalled for a level, all levels above it have failed by extension. If positive evidence is signalled for a level, all levels below it have succeeded by extension. Feedback signals like these can then be used by the speaker to adapt the presentation – by explaining some information in more depth or by making the presentation more interesting – and thereby accommodate the listener. This process is referred to as *Grounding* by Clark (1996). It is not possible to give positive evidence in response to every piece of a conversation, but the important thing is to receive enough evidence to meet the *grounding criterion*, the requirements for evidence needed depending on how important the speakers deem the content of the presentation to be.

2.2 Behaviour Trees

A Behaviour Tree, or BT, is a tree structure that models a plan, initially proposed by Mateas and Stern (2002). Behaviour Trees have been used in video games (Isla, 2005, 2008; Hasegawa et al., 2017) and to model robot behaviours (Hu et al., 2015; Colledanchise et al., 2016). There is previous work applying BTs to virtual agents (Sun et al., 2012; Fujita et al., 2003), but to our knowledge, so far they have not been used to model conversational agents or social behaviour.

The leaves of the tree are the tasks that are executed. All non-leaves are control flow nodes. Execution flows from the root down the tree, starting when some external process *ticks* the root to start execution. Each node in the tree returns one of three values to its parent; `SUCCESS` or `FAILURE` if the task has finished with either result, or `RUNNING` if it has not finished.

The two most common control flow nodes are *Sequence* and *Selector* nodes. *Sequence* nodes run their children in order from left to right until a `FAILURE` or `RUNNING` is encountered, at which point the sequence returns that value. If all child nodes succeed, the sequence returns `SUCCESS`. *Selector* nodes run their children from left to right until a `SUCCESS` or `RUNNING` is encountered, returning that value, or `FAILURE` if all children fail (Colledanchise and Ögren, 2018).

3 Modelling the presentation

In this paper, we propose a Behaviour Tree to model the complex task of poster presentation while taking grounding and adaptation into account. The tree breaks down this complex task into smaller, *independent* tasks. As Section 4 describes, our initial implementations of these individual tasks are greatly simplified, as many of them are indeed challenging research problems in their own right. However, the decomposition into the behaviour tree allows us to start with simpler initial implementations of the individual tasks (some of which can be controlled through *Wizard of Oz*), and then gradually replace them with more complex models (e.g., through machine learning), without changing the structure of the tree, or the implementation of other tasks.

The abstract BT is shown in Figure 2. Whereas most traditional dialogue systems process the interaction utterance-by-utterance, the BT allows the system to process the interaction incremen-

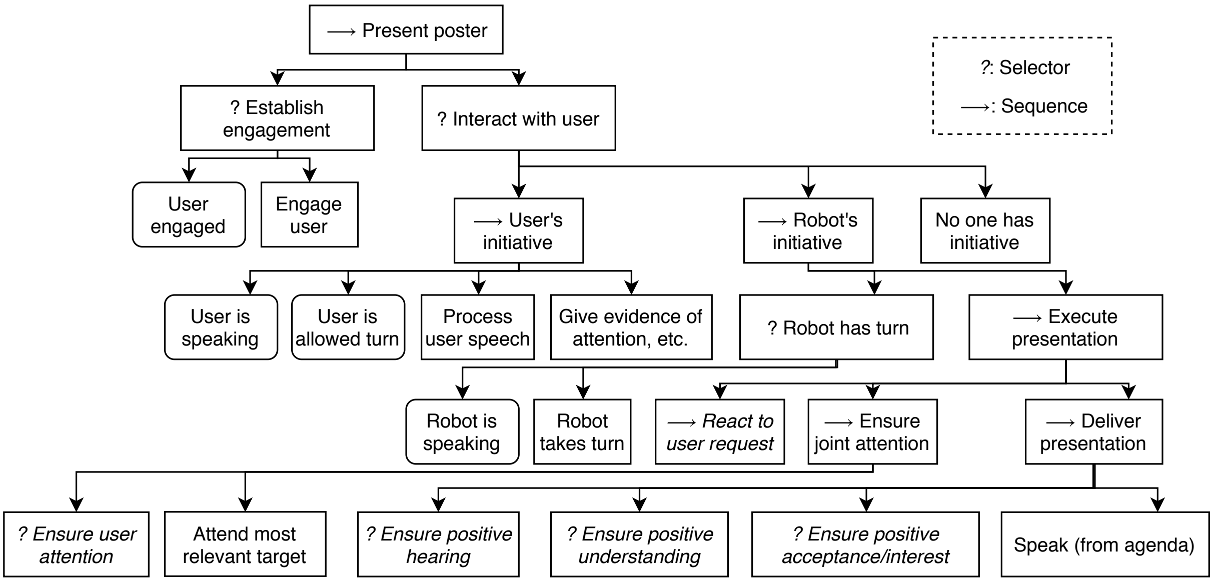


Figure 2: The Behaviour Tree developed as part of this project. Note that the children of any selector or sequence with an italic title are not shown to save room.

tally, in real time (in the vein of Schlangen and Skantze, 2009). Thus, the tree is designed to be executed on the time scale of 10 times per second. The root represents the entire task of presenting a poster. The tree contains both a sub-tree for finding and recruiting participants and presenting to them, and thus will never return `SUCCESS`; the presentation is either going on (`RUNNING`) or impossible (`FAILURE`). The deeper levels of the tree are discussed, top-to-bottom and left-to-right, below.

Dynamic information is not kept in the static tree; instead, it depends on external modules to keep track the joint action ladder (a *knowledge manager*), and where the agent is in its presentation (an *agenda*). These components are not discussed here, as they are less general than the tree.

The system needs to find a user to whom to present, which happens in the **Establish engagement** sub-tree at the top of the tree. After this tree has succeeded at inviting or engaging a user into the presentation, which can be a more or less complicated task (Bohus and Horvitz, 2009, 2014), the system presents its presentation through its **interact with user** sub-tree.

This sub-tree handles turn-taking by offering the turn to the addressee if appropriate, which can be done in multiple ways (Meena et al., 2014; Ström and Seneff, 2000). As the tree runs at its rate of 10 Hz, the user's utterance is processed incrementally, and the system can deploy backchannels and gaze cues in response (Morency et al., 2008).

If the user does not have the turn, the robot either has or takes the turn through its **Robot's initiative** sub-tree, and executes the presentation. Firstly, joint attention is ensured or grabbed (see Yu et al. (2015)) if lost, this can be sensed in multiple ways (Ba and Odobez, 2009; Sheikhi, 2014; Szafir and Mutlu, 2012).

If the system has the user's attention, it ensures hearing, understanding, and acceptance, in order, according to the respective grounding criteria. As these sub-trees have had their chance to change the presentation agenda to address negative evidence of hearing, understanding and acceptance (see (Vaufreydaz et al., 2016; Aly and Tapus, 2015; Sidner et al., 2006; Skantze et al., 2014) for examples on how to measure these), the system then **speaks from the agenda**, driving the presentation forward. Only if the tree reaches this leaf without any previous leaf returning `RUNNING` does the system speak, resulting in incremental, adaptive speech synthesis in the vein of Skantze and Hjalmarsson (2010); Buschmeier et al. (2012); Kopp et al. (2014).

4 Implementation

We developed an initial implementation of a system containing the Behaviour Tree model proposed in Section 3 as an extension to the *IrisTK* dialogue framework (Skantze and Al Moubayed, 2012). The *Furhat* robot head (shown in Figure 1) served as the robot platform (Al Moubayed et al., 2012).

The *agenda* of the implemented system tracked entire lines of the presentation's script. To adapt the presentation, evidence of understanding was thus tracked on a line-by-line basis, and the system could explain a line for which understanding had not been shown, by finding other lines that explained the misunderstood line.

The system modelled attention by treating users as attentive if they were looking at the system or the poster, using their head pose (estimated via *Kinect*) as a proxy of gaze direction. Upon inattention, the system would restart its current utterance, similar to the stop-and-restart method employed by Yousuf et al. (2012). A *Wizard of Oz* setup was used to tag positive and negative evidence of hearing, attention and acceptance.

5 Experiment

To evaluate the system and tree, we set up an experiment where the system described in Section 4 had two modes: in the **adaptive mode**, the system fully used its adaptive behaviour. In the **non-adaptive mode**, the system always assumed positive feedback on all four levels of the joint action ladder. The non-adaptive system also never yielded the turn to the user. The non-adaptive mode presented the same surface-level five-minute presentation every time, so a five-minute time limit was also set for the adaptive mode, which would end its presentation after that time. The agent's gaze behaviour was the same in both modes, shifting between the participant's head and the poster.

We used a within-subject experimental design, where each subject interacted with the two versions of the system. Two posters with 16th-century paintings were created: Gentile Bellini's *Miracle of the Cross fallen into the channel of Saint Lawrence* (*Croce*, for short), and *Great Tower of Babel*, by Pieter Bruegel the Elder. The orders of the two paintings and modes were both counterbalanced between subjects.

30 subjects participated in the experiment, 16 male and 14 female. A majority of participants were undergraduate university students. Participants were not told about the differences between the adaptive and non-adaptive modes, other than that only the adaptive mode could answer questions. Participants were otherwise encouraged to give active feedback to the agent regardless of condition (even though the non-adaptive version would actually ignore this feedback).

Conditions were evaluated immediately following the end of the respective presentation. Firstly, in order to evaluate retention of the information presented, participants were given an electronic form where they answered questions about the presentation and painting. Secondly, they were asked to fill in adapted versions of the *Godspeed* questionnaire by Bartneck et al. (2009), and the *Networked Minds social presence* questionnaire by Biocca and Harms (2011). Participants were rewarded with a cinema ticket.

6 Results

The results of 2 participants had to be excluded due to technical problems during the experiment, yielding 28 data points (16 male, 12 female), of which 14 indicated that they had previous experience with a social robot, two indicated that they had seen the *Croce* painting before, and eight indicated they had seen the *Babel* painting before.

The Wilcoxon paired signed-rank test (Wilcoxon, 1945) was used to compare the answers given in the *Social Presence* and *Godspeed* forms. The questions were grouped by categories in each test, and the answers to them were averaged. This compensated for the large number of questions.

Five out of ten categories (*anthropomorphism* ($p = .0342, \delta = 0.4 \pm 0.4$), *animacy* ($p = .00770, \delta = 0.63 \pm 0.46$), *perceived safety* ($p = .0128, \delta = 0.58 \pm 0.42$), *perceived emotional contagion* ($p = .000999, \delta = 0.47 \pm 0.22$), *perceived behavioural interdependence* ($p = 2.77 * 10^{-5}, \delta = 0.96 \pm 0.29$)) show statistically significant differences between the adaptive and non-adaptive modes, with the adaptive scoring higher. One additional category, *likeability* of the robot, shows a statistically significant difference ($p = .0148, \delta = 0.70 \pm 0.60$) between the first and the second presentation given to participants, the first scoring higher. No statistically significant differences were found between the two paintings.

For the analysis of the retention questionnaire, one additional subject had to be excluded due to technical problems. Eleven questions per poster were graded on a scale from zero to eleven based on correctness, normalising to only count questions that were possible to answer based on the presentation the user received. The answers in the *Babel* questionnaire ($M = 6.938$, $Mdn = 7.542$, $SD = 1.989$) were found to have a statistically significantly ($p = .04235$) different distribution than those in the *Croce* questionnaire ($M = 6.270$, $Mdn = 6.758$, $SD = 1.771$), but no statistically significant differences were found when comparing the adaptive mode and the non-adaptive mode ($p = .449$), or the first and second presentation participants received ($p = .990$).

7 Discussion

The results from the Social presence and Godspeed questionnaires showed that the adaptive version was perceived to have a higher Animacy, Anthropomorphism, Safety, Emotional contagion, and Behavioural interdependence. These are all aspects that relate to higher interactivity, and are all associated with positive values, which indicates that an interactive presenter that takes the user's attention and understanding into account is indeed perceived to be more engaging. When asking the subjects about the difference between the two versions after the experiment, they typically had a hard time identifying the exact difference in terms of interactivity. This is interesting, as it indicates that they were not aware of the specific reason for why they preferred the adaptive version. The gaze behaviour of the robot, which followed users around even in the otherwise non-adaptive mode, may have led to the perception that the system was paying attention to the user even in this mode.

There was a somewhat unexpected difference between the first and second presentation, where the former had a somewhat higher Likeability of the robot, regardless of painting and mode. One potential explanation for this is that users were aware of the format of the evaluation the second time, and might have been more stressed about it.

However, no statistically significant differences were found in the user's retention of the two presentations. There was a large variation in how much the individual subjects remembered from each presentation. Certain participants remembered almost nothing of either presentation. Others were able to quote the robot on every question in both the adaptive and non-adaptive modes. This introduces noise and makes the comparison hard to perform, given the relatively small number of participants.

7.1 Future work

Although the agent developed in our initial implementation does adapt its presentation based on feedback from the user, this adaptation was mostly done on a semantic level (i.e., updating its agenda). In future studies, we will explore how the system could also adapt factors like turn length, speech rate, the frequency with which the agent would require evidence of understanding, and what the system would consider as evidence of understanding.

Classifying negative and positive evidence based on multi-modal signals is indeed a very challenging task, as these cues could be very subtle (e.g., facial expressions of boredom or interest). In this experiment, this classification was done by a human *Wizard of Oz*. The data collected through this experiment could potentially be used to train specific models for this, as they have already been partially annotated by the Wizard.

A natural extension of the model is to also allow several users to take part in the presentation. This would give rise to new challenges when it comes to determining who should be considered to be engaged in the presentation, and how to adapt the presentation, since the different users in the audience might show evidence of understanding to various degrees. Also, if a new user appears in the middle of the presentation, it is not clear how to proceed with the agenda.

8 Conclusions

This paper presents a first step towards a system that uses Behaviour Trees to create an adaptive presentation agent. Initial results show that users find a system that attempts to adapt its presentation to their reception of the presentation more positive along several dimensions. Our initial implementation of the proposed Behaviour Tree model is a promising first step towards a complex adaptive behaviour model for conversational interaction, where the complex task of making an adaptive presentation has been decomposed into smaller tasks, which can gradually be replaced by more and more sophisticated models.

Acknowledgements

This work is supported by the SSF (Swedish Foundation for Strategic Research) project COIN.

References

Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive behavioural systems*, pages 114–130. Springer.

Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of semantics*, 9(1):1–26.

Amir Aly and Adriana Tapus. 2015. An online fuzzy-based approach for human emotions detection: An overview on the human cognitive model of understanding and generating multimodal actions. In *Intelligent Assistive Robots*.

Sileye O Ba and Jean-Marc Odobez. 2009. Recognizing visual focus of attention from head pose in natural meetings. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):16–33.

Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81.

F Biocca and C Harms. 2011. Networked minds social presence inventory (scales only version 1.2). *East Lansing: MIND Labs, Michigan State University. Retrieved from http://cogprints. org/6742*.

Dan Bohus and Eric Horvitz. 2009. Learning to predict engagement with a spoken dialog system in open-world settings. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 244–252. Association for Computational Linguistics.

Dan Bohus and Eric Horvitz. 2014. Managing human-robot engagement with forecasts and... um... hesitations. In *Proceedings of the 16th international conference on multimodal interaction*, pages 2–9. ACM.

Dan Bohus and Alexander I. Rudnicky. 2009. The RavenClaw dialog management framework: Architecture and systems. *Computer Speech & Language*, 23(3):332 – 361.

Johan Boye. 2007. Dialogue management for automatic troubleshooting and other problem-solving applications. In *Proc. of 8th SIGdial Workshop on Discourse and Dialogue*, pages 247–255. Citeseer.

Hendrik Buschmeier, Timo Baumann, Benjamin Dosch, Stefan Kopp, and David Schlangen. 2012. Combining incremental language generation and incremental speech synthesis for adaptive information presentation. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 295–303. Association for Computational Linguistics.

Herbert H. Clark. 1996. *Using language*. Cambridge University Press, Cambridge, UK.

M. Colledanchise and P. Ögren. 2018. *Behavior Trees in Robotics and Al: An Introduction*. Chapman & Hall/CRC artificial intelligence and robotics series. Taylor & Francis Limited.

Michele Colledanchise, Alejandro Marzinotto, Dimos V Dimarogonas, and Petter Ögren. 2016. The advantages of using behavior trees in multi-robot systems. In *Proceedings of ISR 2016: 47st International Symposium on Robotics*, pages 1–8. VDE.

Tobias Eichner, Helmut Prendinger, Elisabeth André, and Mitsuru Ishizuka. 2007. Attentive presentation agents. In *International Workshop on Intelligent Virtual Agents*, pages 283–295. Springer.

Masahiro Fujita, Yoshihiro Kuroki, Tatsuzo Ishida, and Toshi T Doi. 2003. Autonomous behavior control architecture of entertainment humanoid robot sdr-4x. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, volume 1, pages 960–967. IEEE.

Isamu Hasegawa, Tomohiro Hasegawa, Kazutaka Kurosaka, Akihiko Kishi, Akira Iwasawa, and Youichiro Miyake. 2017. How to build a fantasy world based on reality: A case study of final fantasy xv: Part ii. In *SIGGRAPH Asia 2017 Courses*, SA '17, pages 7:1–7:149, New York, NY, USA. ACM.

Takuya Hashimoto, Naoki Kato, and Hiroshi Kobayashi. 2011. Development of educational system with the android robot saya and evaluation. *International Journal of Advanced Robotic Systems*, 8(3):28.

Danying Hu, Yuanzheng Gong, Blake Hannaford, and Eric J Seibel. 2015. Semi-autonomous simulated brain tumor ablation with ravenii surgical robot using behavior tree. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3868–3875. IEEE.

Damián Isla. 2005. Handling complexity in the Halo 2 AI. *Proceedings of the Game Developers' Conference*.

Damián Isla. 2008. Building a better battle. In *Game Developers Conference, San Francisco*, volume 32.

Björn Jensen, Nicola Tomatis, Laetitia Mayor, Andrzej Drygajlo, and Roland Siegwart. 2005. Robots meet humans-interaction in public spaces. *IEEE Transactions on Industrial Electronics*, 52(6):1530–1546.

Tatsuya Kawahara. 2012. Multi-modal sensing and analysis of poster conversations toward smart posterboard. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 1–9. Association for Computational Linguistics.

Stefan Kopp, Herwin van Welbergen, Ramin Yaghoubzadeh, and Hendrik Buschmeier. 2014. An architecture for fluid real-time conversational agents: integrating incremental output generation and input processing. *Journal on Multimodal User Interfaces*, 8(1):97–108.

Michael Mateas and Andrew Stern. 2002. A behavior language for story-based believable agents. *IEEE Intelligent Systems*, 17(4):39–47.

Raveesh Meena, Gabriel Skantze, and Joakim Gustafson. 2014. Data-driven models for timing feedback responses in a map task dialogue system. *Computer Speech & Language*, 28(4):903–922.

Louis-Philippe Morency, Iwan de Kok, and Jonathan Gratch. 2008. Predicting listener backchannels: A probabilistic multimodal approach. In *International Workshop on Intelligent Virtual Agents*, pages 176–190. Springer.

Taku Ohya, Tatsuya Hiramatsu, Yong Xu, Yasuyuki Sumi, and Toyoaki Nishida. 2006. Towards robot as an embodied knowledge medium. In *the 5th international workshop of social intelligence design (SID2006)*.

David Schlangen and Gabriel Skantze. 2009. A general, abstract model of incremental dialogue processing. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 710–718. Association for Computational Linguistics.

Samira Sheikhi. 2014. Inferring visual attention and addressee in human robot interaction. Technical report, École Polytechnique Fédérale de Lausanne (EPFL).

C Sidner, C Lee, L-P. Morency, and C ForLines. 2006. The effect of head-nod recognition in human-robot conversation. In *Proceedings of the 1st Annual Conference on HumanRobot Interaction*, pages 290–296. ACM Press.

Gabriel Skantze and Samer Al Moubayed. 2012. IrisTK: a statechart-based toolkit for multi-party face-to-face interaction. In *Proceedings of the 14th ACM international conference on Multimodal interaction*, pages 69–76. ACM.

Gabriel Skantze and Anna Hjalmarsson. 2010. Towards incremental speech generation in dialogue systems. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 1–8. Association for Computational Linguistics.

Gabriel Skantze, Anna Hjalmarsson, and Catharine Oertel. 2014. Turn-taking, feedback and joint attention in situated human-robot interaction. *Speech Communication*, 65:50–66.

Ronnie W Smith and D Richard Hipp. 1994. *Spoken natural language dialog systems: A practical approach*. Oxford University Press on Demand.

Nikko Ström and Stephanie Seneff. 2000. Intelligent barge-in in conversational systems. In *Sixth International Conference on Spoken Language Processing*.

Libo Sun, Alexander Shoulson, Pengfei Huang, Nicole Nelson, Wenhu Qin, Ani Nenkova, and Norman I Badler. 2012. Animating synthetic dyadic conversations with variations based on context and agent attributes. *Computer Animation and Virtual Worlds*, 23(1):17–32.

Daniel Szafir and Bilge Mutlu. 2012. Pay attention!: designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 11–20. ACM.

Dominique Vaufreydaz, Wafa Johal, and Claudine Combe. 2016. Starting engagement detection towards a companion robot using multimodal features. *Robotics and Autonomous Systems*, 75:4 – 16. Assistance and Service Robotics in a Human Environment.

Igor M Verner, Alex Polishuk, and Niv Krayner. 2016. Science class with robothespian: using a robot teacher to make science fun and engage students. *IEEE Robotics & Automation Magazine*, 23(2):74–80.

Frank Wilcoxon. 1945. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83.

Mohammad Abu Yousuf, Yoshinori Kobayashi, Yoshinori Kuno, Akiko Yamazaki, and Keiichi Yamazaki. 2012. Development of a mobile museum guide robot that can configure spatial formation with visitors. In *Intelligent Computing Technology*, pages 423–432, Berlin, Heidelberg. Springer Berlin Heidelberg.

Zhou Yu, Dan Bohus, and Eric Horvitz. 2015. Incremental coordination: Attention-centric speech production in a physically situated conversational agent. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 402–406.

A Appendices

The Godspeed forms included the questions as found at http://www.bartneck.de/2008/03/11/the-godspeed-questionnaire-series/. The Social Presence forms includes the questions as referenced (Biocca and Harms, 2011), but the following questions were removed:

- I often felt as if (my partner) and I were in the same (room) together.
- I think (my partner) often felt as if we were in the same room together.
- I often felt as if we were in different places rather than together in same (room)
- I think (my partner) often felt as if we were in different places rather than together in the same (room).

I was sometimes influenced by the robot's moods.

Strongly disagree	(First session)	□ □ □ □ □ □ □	(First session)	Strongly agree
	(Second session)	□ □ □ □ □ □ □	(Second session)	

Question (Babel)	Question (Croce)	Answer type
Have you interacted with a social robot like the one in this experiment before?		Yes/No
In what context have you interacted with a system like the one used in the experiment?		Text
Had you seen the painting before the presentation?		Text
What was the name of the painting?		Text
Who was the artist who painted the painting?		Text
From roughly what year was the painting?		Number
Briefly describe the contents of the painting, i.e. what you saw, not what the robot told you.		Text
Who were the men on the bottom right of the painting?	Who was the person on the bottom left of the painting?	Text
Who was the woman on the left of the river, at the bottom left?	What was the design of the tower itself based on?	Text
Why did the cross fall into the water?	What does the tower symbolise?	Text
What was special about the cross?	From what country was the artist?	Text
Who was the man who was retrieving the cross from the water?	The painting is an example of a certain technique; what technique?	Text
In what Italian city does the scene take place?	There are many examples of small details in the painting: give some examples.	Text
The artist had relatives who also became artists: who were they?		Text

Table 1: The questions that measured retention.

An example Social Presence question is shown above Table 1. Godspeed questions were presented identically (with the same seven-point scale), but the ends of the scale were instead the two adjectives or adjective phrases connected to the specific Godspeed question.

The full questionnaires can not be presented here because of space issues. Table 1 on the bottom of this page shows the retention-based questions that were part of the electronic questionnaire.

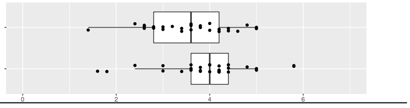
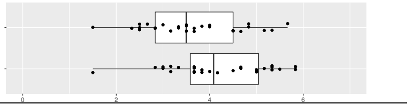
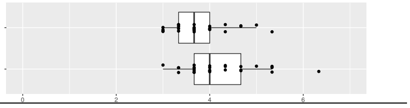
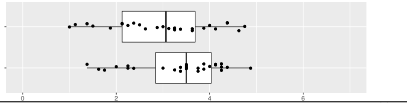
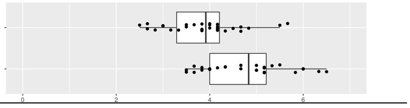
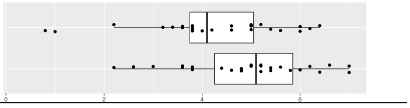
Category (Form)		
Anthropomorphism (Godspeed)		$p = 0.00700$
Non-ada. Adaptive		
Animacy (Godspeed)		$p = 0.00700$
Non-ada. Adaptive		
Perceived safety (Social presence)		$p = 0.0128$
Non-ada. Adaptive		
Perceived emotional contagion (Social presence)		$p = 0.000999$
Non-ada. Adaptive		
Perceived behavioural interdependence (Social presence)		$p = 2.77 * 10^{-5}$
Non-ada. Adaptive		
Likeability (*of robot*) (Godspeed)		$p = 0.0147$
Second First		

Table 2: Visualisation of numbers given in Section 6.