

Citation for published version:

Wortham, RH & Rogers, V 2017, 'The Muttering Robot: Improving Robot Transparency Through Vocalisation of Reactive Plan Execution', Paper presented at 26th IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man) Workshop on Agent Transparency for Human-Autonomy Teaming Effectiveness, Lisbon, Portugal, 1/09/17 - 1/09/17.

Publication date:
2017

Document Version
Early version, also known as pre-print

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The Muttering Robot: Improving Robot Transparency Through Vocalisation of Reactive Plan Execution

Robert H. Wortham¹ and Vivienne E. Rogers²

Abstract—Transparency is an important design consideration for all intelligent autonomous systems. Previous work shows that a real-time visual display of a robot’s decision making produces significantly better understanding of that robot’s intelligence. We investigate vocalisation of behaviour selection as a possible alternative solution for situations where a visual display of decision making is either impractical or impossible. In initial experiments we find that vocalisation is associated with a significant improvement in understanding of the robot, comparable with the results obtained using a real-time display. We also find that vocalisation has no significant effect on participants emotional response, though it may slightly increase positive feelings about the robot. We discuss the relative merits of visual and vocalised transparency mechanisms, and suggest possible applications.

I. INTRODUCTION

The relationship between transparency, trust and utility is complex [1] but nevertheless it is clear that transparency is an important design consideration for all intelligent autonomous systems. Transparency has been shown to improve the quality of teamwork [2] in robot-human collaborative scenarios. It is also a key factor when humans attribute credit and blame in these collaborative scenarios [3]. Increased robot transparency is associated with reduced assignment of credit or blame to the robot, and increased assignment to humans. This increased focus on and facilitation of human agency in collaborative robot-human tasks is a desirable outcome, because it allows automation to empower and enhance its human users. Previous work shows that a real-time visual display of a robot’s decision making produces significantly better understanding of that robot’s intelligence [5], [6]. In this paper we describe a possible alternative solution for situations where a visual display of decision making is either impractical or impossible. We use the Instinct reactive planner [4] to control a small mobile robot, monitor the hierarchical action selection process, and use a novel algorithm to convert the output from the monitor into vocalised (spoken) sentences. Humans have evolved to produce and comprehend language [7]. We are able to perform several tasks simultaneously involving language and sharing mental resources between different cognitive systems [8]. This proposes using language as a likely candidate to enhance robot transparency. The vocalisation of the robot is, however, not an implicit designed behaviour of the robot reactive plan, but rather a separate monitoring channel expressed vocally. The result is a robot that ‘mutters’,

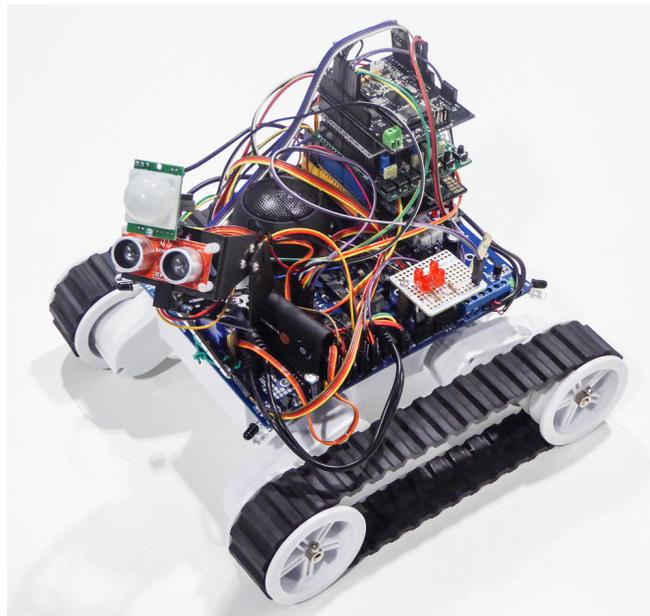


Fig. 1. The R5 Robot [4]. The loudspeaker and blue-tooth audio adapter are visible. The Four red LEDs are powered from the audio output, and serve as a visual indication of the vocalisation.

continually vocalising the initiation and progress through the reactive plan hierarchy. The immediate and obvious difficulty with this approach is that the robot executes multiple reactive plan cycles per second, each traversing many plan elements in the plan hierarchy. It is thus impossible to report vocally on the initiation and progress of each plan element in real-time. Our algorithm first generates pre-defined candidate sentences to be uttered, and then uses a number of novel parametrised approaches to select from these candidates. This algorithm produces understandable vocalised sentences that usefully convey the decision processes or ‘thinking’ taking place within the robot in real time. We deploy this algorithm within a small mobile robot, shown in Figure 1, and show that observers’ models of the robot improve significantly when also exposed to the muttering.

II. THE MUTTERING MECHANISM

Those interested in the theory of reactive planning and the detailed operation of the Instinct Planner and R5 robot should read [4] before they read this explanation of the muttering mechanism. However, this description should suffice unsupported for those mainly interested in the results of this experiment, and who only wish to have an understanding of the mechanism used to achieve the muttering behaviour of

¹Robert H. Wortham, Department of Computer Science, University of Bath, Bath, BA2 7AY, UK. r.h.wortham@bath.ac.uk

²Vivienne E. Rogers, Department of Applied Linguistics, Swansea University, Swansea, SA2 8PP, UK. v.e.rogers@swansea.ac.uk

the robot. The complete source code for the Instinct Planner is available on an open source basis¹, as is the code for the R5 Robot, including the muttering mechanism described in this section².

The robot behaviour (or action) selection is performed by the Instinct Planner [4]. The planner combines sensory information gathered by the robot, with a pre-defined set of *Drives*, each Drive designed to achieve a specific goal or objective. Each Drive is expressed as a hierarchically composed plan of *Competences*, *Action Patterns* and ultimately *Actions*. These actions invoke the behaviour primitives of the robot, such as ‘stop’, ‘turn left’, ‘scan for human’, ‘flash headlight’ and so on. The planner is considered to be *Reactive* due to its ability to switch task focus from one Drive to another, or within a Drive from one branch of the plan to another, based on changing sensory input, its execution path being re-evaluated each plan cycle [9].

The planner produces a transparency feed for each execution cycle of the plan, that is to say, for each cycle the planner produces a stream of data corresponding to the traversal of the plan hierarchy leading to an action being selected. This stream contains the *Plan Element Identifier* (ID) of each plan element, and the status of the plan element. As the planner traverses down the hierarchy it reports plan element IDs together with the status *Executed* (E). As the planner completes the processing of each plan element travelling back up the hierarchy, it again reports the plan element ID, but this time with the outcome of the execution. The outcome is one of four options: *Success*, *In Progress*, *Failed*, *Error*. Success indicates that the plan element has completed successfully. In Progress indicates either that an underlying physical behaviour of the robot is still in the process of execution, or that a more complex element such as an Action Pattern or Competence is part way through its various steps, but as yet not completed. Failed is a common outcome of a reactive plan, arising from the dynamic and unpredictable world in which the robot operates. Error is only returned when an internal programming error occurs, such as a fault in the plan design, or a bug in the software.

In previous work [5], [6] we have used this transparency feed to drive a dynamic visual display showing the plan execution in real-time, as a means to make the operation of the robot more transparent. However, there are limitations to this approach, discussed more fully in section V, and so here we are interested in generating a stream of audible output to convey at least some of the information in the transparency feed in real-time to those observing and interacting with the robot. From the explanation above, readers will already realise that the transparency data feed consists of many tens of plan element notifications per second. The data rate of the transparency feed in notifications per second R_t is given by

$$R_t = 2R_p D_p \quad (1)$$

Where R_p is the rate of plan execution and D_p is the depth

of the plan hierarchy currently being executed. For the R5 robot operating with a plan cycle rate of eight cycles per second, a plan with a hierarchy depth of seven generates 112 notifications per second. It is not possible to generate meaningful speech output at this rate. Therefore we must be selective. The mechanism adopted here uses three stages of selectivity, described in the following three subsections.

A. Transparency Execution Stack

First we detect when there are changes in the execution pattern occurring between two consecutive plan cycles. Real world actions typically take much longer than a single plan cycle to execute, and so frequently the same route is traversed through the plan hierarchy many times, with the leaf node Action repeatedly having a status of In Progress. In order to detect these changes we implement a stack arrangement. Starting from the initial Drive notification, we store reports of element executions in a stack structure. Once the leaf node is reached, we traverse the stack in the opposite direction completing the element status for each plan element. On subsequent plan cycles we check whether the same element IDs are being executed at each position in the stack. If there is a change we mark it and clear out all records at lower levels in the stack. This mechanism allows us to gather information about whether this is the first time a plan element has been executed in this context, and also whether the execution status has changed from previous executions.

B. The Speak Rules Engine

Based on information obtained from the Transparency Execution Stack, we now make decisions about whether to generate a candidate sentence about the event, based on a set of *Speak Rules*. The robot holds a matrix of Speak Rule values for each plan element type. The default values for the Action element type are shown in Table I. Similar tables are stored for each of the other plan element types. The Timeout defines

TABLE I
SPEAK RULES FOR THE ACTION PLAN ELEMENT TYPE.

Event	Timeout	RptMyself	RptTimeout	AlwaysSpeak
Executed	800	FALSE	3333	TRUE
Success	500	FALSE	3333	FALSE
In Progress	0	FALSE	0	FALSE
Failed	5000	FALSE	3333	TRUE
Error	5000	FALSE	3333	TRUE

how long the generated sentence will be stored awaiting presentation to the speech output system. After this time-out the sentence will be discarded. RptMyself is a boolean flag to specify whether the sentence should be repeated, should it match the last thing uttered by the robot. RptTimeout determines the time after which the utterance would not be considered to be a repeat. The time-out values are specified in milliseconds. Finally AlwaysSpeak is a boolean that will force the sentence to be spoken next, irrespective of whether other sentences are queued, see subsection II-C below. Considering the settings in Table I, we see that when an Action is In

¹<http://www.robwortham.com/instinct-planner/>

²<http://www.robwortham.com/r5-robot/>

Progress, no candidate sentence will be generated. However, when an element is first Executed, Fails or an Error occurs, a higher priority is given to the candidate sentence that is generated. The R5 robot includes a command line interface accessible via its wifi link. This interface includes commands to change each of these parameters individually, and to save them to a permanent storage area within the robot. Tuning these parameters is at present a matter of iterative human design.

The actual candidate sentences that are produced by the Speak Rules Engine are pre-defined for each plan Element type and Event type combination. For example, new executions of Competence elements create sentences of the form ‘Attempting {plan-element-name}’. The plan element names are stored within the robot alongside the plan itself, and the plan element IDs from the transparency feed are used to locate the correct plan element name relating to the sentence to be constructed. These element names are processed using ‘camel case’ rules to generate speakable names for the plan elements. Camel case is a convention where each new word starts with a capital letter, for example ActivateHumanDetector or ForwardAvoidingObstacle. The processing can also deal with numbers such as Sleep10Seconds. These speakable plan element names are inserted into the pre-defined sentences to create sentences of the form ‘Attempting Forward Avoiding Obstacle’ and ‘Doing Sleep 10 Seconds’.

C. The Vocaliser

Despite the filtering achieved by the Speak Rules Engine, many more candidate sentences are still generated than can be spoken. Each is presented to the Vocaliser, along with the various timeout parameters and boolean flags. The Vocaliser uses a double buffered approach to store sentences to be spoken. Once the buffers are full, further candidate sentences are discarded until the sentences are either spoken or replaced, according to the Speak Rule parameters. The actual vocalisation is performed by a low cost text to speech synthesiser module and a small onboard loudspeaker. The audio is also available via a blue-tooth transmitter, in order that it can be accessed remotely.

III. EXPERIMENTAL METHODS

An experiment was conducted over three days in December 2016, at the At-Bristol Science Learning Centre, Bristol, UK. This context was chosen because of available subjects in a controlled setting. The robot operated on a large blue table as a special interactive exhibit within the main exhibition area, see Figure 2. Visitors, both adults and children, were invited to stand and observe the robot in operation for several minutes whilst the robot moved around the pen and interacted with the researchers. Subjects were expected to watch the robot for at least three minutes before being handed a paper questionnaire to gather both participant demographics and information about the participants’ perceptions of the robot. During each day, the robot operated for periods in each of two modes; silent (Group 1 results), or with muttering enabled (Group 2 results). The R5 robot carries an on-board speaker

to produce the ‘muttering’, see Figure 1. Typically this is sufficiently loud to be heard in most indoor environments. However, as the At-Bristol environment was particularly noisy with children playing, participants were encouraged to wear headphones to better hear the audio output.

A. Post-Treatment Questionnaire

Table II summarises the questions asked after the participant had observed the robot in operation. In order to facilitate cross-study comparison, the questions match those presented in previous studies that investigate the use of real-time visual displays to provide transparency [6]. These

TABLE II
POST-TREATMENT QUESTIONS

Question	Response	Category
Is robot thinking?	Y/N	Intel
Is robot intelligent?	1-5	Intel
Feeling about robot?	Multi choice	Emo
Understand objective?	Y/N	MM
Describe robot task?	Free text	MM
Why does robot stop?	Free text	MM
Why do lights flash?	Free text	MM
What is person doing?	Free text	MM
Happy to be person?	Y/N	Emo
Want robot in home?	Y/N	Emo

questions are designed to measure various factors: the measure of intelligence perceived by the participants (Intel), the emotional response (if any) to the robot (Emo), and—most importantly—the accuracy of the participants’ mental model of the robot (MM). For analysis, the four free text responses were rated for accuracy with the robot’s actual Drives & behaviours and given a score per question of 0 (inaccurate or no response), 1 (partially accurate) or 2 (accurate). The marking was carried out by a single researcher for consistency, without access to knowledge of which group the subject was in. No special vocabulary was expected. The questions used in the questionnaire are deliberately very general, so as not to steer the subject. Similarly, the marking scheme used is deliberately coarse grained because we are looking for a significant effect at the general level of understanding, not for a nuanced improvement in the subject’s model.

B. Affect Questions

The questionnaire includes a question concerning how participants feel about the robot, specifically they are asked to complete a multiple choice section headed ‘How do you feel about the robot? Please choose one option from each row.’ with options ranging from ‘Not at all’ through to ‘Very’ [10]. A standard two dimensional model of affect is used, with dimensions of Valence and Arousal. The specific feelings interrogated are detailed in Table III together with their assignment to an assumed underlying level of Valence W_{vf} and Arousal W_{af} . These weights are based on values specified for these specific words by Bradley et al [11] scaled within



Fig. 2. The arrangement of the Muttering Robot experiment at At-Bristol. The obstacles are made from giant lego bricks. Participants are wearing headphones fed from a standard headphone amplifier, which in turn is connected to a bluetooth receiver. This receives the audio output from the robot's bluetooth transmitter. This enables participants to hear the robot clearly with high levels of background noise.

TABLE III

AFFECT QUESTIONS AND ASSIGNMENT TO VALENCE AND AROUSAL. PARTICIPANTS WERE ASKED 'HOW DO YOU FEEL ABOUT THE ROBOT? PLEASE CHOOSE ONE OPTION FROM EACH ROW.' OPTIONS WERE NOT AT ALL (0), A LITTLE (1), SOMEWHAT (2), QUITE A LOT (3), VERY (4).

Feeling f	Valence W_{vf}	Arousal W_{af}
Happy	+1.00	+0.51
Sad	-1.00	-0.46
Scared	-0.65	+0.65
Angry	-0.62	+0.79
Curious	+0.35	+0.24
Excited	+0.78	+1.00
Bored	-0.59	-1.00
Anxious	-0.03	+0.69
No Feeling	0	0

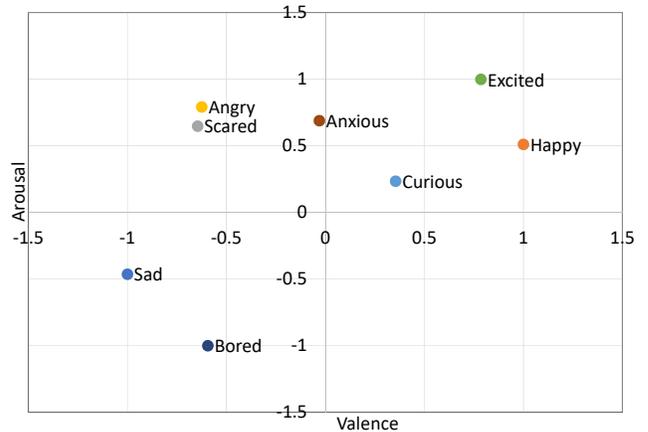


Fig. 3. This diagram shows the Valence and Arousal weightings for the feelings specified in the questionnaire (see Table III) plotted as a simple two dimensional model of Affect.

the range -1 to +1 on both axes.

$$V_F(p) = \frac{1}{|F|} \sum_F W_{vf} V_{pf} \quad (2)$$

$$A_F(p) = \frac{1}{|F|} \sum_F W_{af} A_{pf} \quad (3)$$

These Valence and Affect weightings are also shown graphically in Figure 3. The Valence value $V_F(p)$ and Arousal value $A_F(p)$ for each participant p are therefore calculated by multiplying the scores for each feeling word V_{sf} and A_{sf} by the weightings W_{vf} and W_{af} respectively, and then summing

for each feeling f in the set of feelings F , as shown in Equations 2 and 3.

IV. RESULTS

For each group of participants, the demographics are shown in Table IV. Given the public engagement nature of the experimental location, it was not possible to accurately match each group for age, gender, education and experience with computers and robots. However, mean age and gender are both

fairly well matched. The mix of graduates to non-graduates is also well matched. Group Two contains proportionately more participants identifying themselves as having prior experience of working with robots.

TABLE IV
DEMOGRAPHICS OF PARTICIPANT GROUPS ($N = 68$)

Demographic	Group One (silent)	Group Two (sound)
Total Participants	32	36
Mean Age (yrs)	44.1	47.36
Gender Male	14	17
Gender Female	18	19
STEM Degree	6	11
Other Degree	18	16
Ever worked with a robot?	2	10
Do you use computers?	23	30
Are you a Programmer?	7	10

A. Main Findings

The primary results are shown in Table V. Most importantly, in Group Two (observe robot whilst listening to muttering) there is a marked improvement in the accuracy of participants' reports about the robot's function and capability. This confirms a significant correlation between the accuracy of the participants' mental models of the robot, and the provision of the additional transparency data provided by the muttering ($N=68$, unpaired t test, $p=0.0057$, $t(66)=2.86$). This compares favourably with the results obtained using the ABOD3 real-time display [6].

TABLE V
MAIN RESULTS. BOLD FACE INDICATES RESULTS SIGNIFICANT TO AT LEAST $p = .05$.

Result	Group One (silent)	Group Two (sound)	Effect Size Cohen's D
Is thinking (0/1)	0.50 (sd=0.51)	0.44 (sd=0.50)	.085
Intelligence (1-5)	2.56 (sd=1.32)	2.67 (sd=1.10)	.086
Objective (0/1)	0.78 (sd=0.42)	0.81 (sd=0.40)	-.184
Accuracy (0-8)	1.94 (sd=1.39)	3.19 (sd=2.11)	.696

In both groups, participants almost equally report that they understand the objective of the robot, showing no difference across the groups ($N=68$, unpaired t test, $p=0.81$, $t(66)=0.24$). Note the high level of reported understanding compared with the much lower report accuracy. This indicates that significant numbers of participants in both groups perceive that they have a good model of the robot, when in reality they do not. Finally, there is no significant difference in participants perceived intelligence of the robot, or their reports that the robot is 'thinking'.

B. Affect - Self Report of Feelings

The results obtained from the affect questions detailed in Table III did not yield significant differences between Groups

One and Two, however the findings shown in Table VI do bear some analysis. Firstly, no severe or adverse changes in

TABLE VI
RESULTS BASED ON REPORTED FEELINGS.

Result	Group One (silent)	Group Two (sound)	Change	Sig.
Valence	0.383 (sd=0.354)	0.437 (sd=0.253)	% 12.4	$p=0.47$
Arousal	0.350 (sd=0.291)	0.346 (sd=0.231)	% -1.27	$p=0.95$

feeling were found as a result of adding muttering to the robot, and this in itself is an important result if muttering is to be considered for practical applications. There was a measurable, though not statistically significant increase in Valence. Valence is a measure of the extent of positive feelings about the robot, whilst Arousal is a measure of strength of feeling. Thus this result gives tentative indication that whilst participants did not have overall stronger feelings about the robot, their feelings were marginally more positive. However, a larger study would be necessary to obtain statistically convincing evidence of this finding.

V. DISCUSSION

This approach of vocalising transparency through muttering may have applications where users are visually impaired, or may need to concentrate their vision elsewhere whilst working with an autonomous system. Applications may include divers working underwater with robots, as commercial divers have good audio systems for communication, but work in environments where visibility may be very poor.

A. Visual Display versus Vocalisation

Where robots are operating with very large hierarchical reactive plans, or where another action selection method is being used, it is hard to decouple the design and operation of a real-time visual display from the plan itself. If the display is to be mounted on the robot this also impacts the design of robot. For hierarchical plans, the visual display needs to either only display the highest level elements of the plan, or must move and scale to move around the plan as the focus of execution changes. For designers this can be undertaken manually, but for end users or observers of a robot manual interaction is impractical, and so some automated pan and zoom mechanisms would be required.

In contrast, a vocalised transparency output has the benefit that it is independent of the physical layout and structure of the plan and it scales indefinitely with the size of the plan. A vocalised transparency feed could likely be configured to work with any structured action selection mechanism, not necessarily a hierarchical structure. Conversely, due to the much lower bandwidth of a vocalised output, much of the fine detail of plan execution is lost. Also, if a plan is suspended in a debugging mode, the vocalised output would cease, but a visual display would continue to display a useful trace of activity to assist with debugging. The Speak Rules described in subsection II-B must also be tuned manually, and may

vary by robot and application, although this has yet to be investigated.

The authors had expected that participants might find the muttering robot to be somewhat irritating. It is therefore very interesting that this was not borne out in the data, in fact if anything there is a marginal improvement in the attitude of participants to the robot. In practical applications we envisage the muttering to be able to be turned on and off by users at will, possibly using a voice activated interface. Perhaps asking the robot to explain itself would turn on the muttering, and telling it to ‘shut up’ would restore silence.

The results provide evidence to support the case that we can add the transparency measure (muttering) without affecting the user experience. The user has substantially the same qualitative experience, measured by the first three results in Table V and by the emotional model measures in Table VI, but in fact has a better internal model to understand the robot. This is very important for social robotics applications as it counters the argument that making the robot transparent might reduce its effectiveness, for example in companionship or human care related applications [1], [12].

Having discussed some advantages and disadvantages of visual and vocalised transparency, it seems they are complementary and reinforcing. Developers might switch back and forth between both mechanisms, or use them in parallel. It is certainly easier to observe a robot whilst listening to it, than to observe a robot and a remote visual display concurrently. End users might have the option to see a visual display on their tablet or laptop, but when this is inconvenient they could enable the muttering, and then eventually turn off all transparency once they have a good working model of the robot, enabling them to understand and predict its behaviour without further recourse to the transparency mechanisms.

VI. CONCLUSIONS AND FURTHER WORK

As in previous studies [6] the results indicate that significant numbers of participants in both groups perceive that they have a good model of the robot, when in reality they do not. This leads us to conclude that reports of understanding by those interacting with robots should be treated with healthy scepticism. However, in this study we show that the vocalised transparency feed produces a marked improvement in the accuracy of participants’ reports about the robot’s function and capability, confirming a significant correlation between the accuracy of the participants’ mental models of the robot, and the provision of the additional transparency data provided by the muttering.

This study indicates the possibility that participants feel more positive about the robot when it is muttering, but with the limited study size these results are not statistically significant, and in comparison with the much stronger effect of the transparency on accuracy of mental model, this emotional effect appears to be much weaker. Indeed, there was almost no difference in the levels of Arousal between the two groups, which in itself is an interesting result as we had expected to see some increase due to the increased stimulation of participants by the vocal output from the robot. Further, larger

studies would therefore be required to explore the extent of the effect of muttering on positive feelings toward the robot.

In this experiment, as with the experiments considering a visual real-time display, we have concentrated on the output of the decision making process. We have therefore not considered making available the sensory model that exists within the robot, nor making transparent the various thresholds that must be crossed to release the various elements of the reactive plan. Perhaps to do so would overload the user with data, but in some applications it may be helpful to gain an insight about how the world is perceived by the robot, as this would aid an understanding of its subsequent decision making processes. It might also be useful to investigate the benefits of a more complex sentence generation algorithm, able to generate varying sentences that might make the vocalisation sound less ‘robotic’. Finally, we have yet to expose participants to the simultaneous use of visual and vocalised transparency. Whether there are further gains to be made by employing both together, or whether they would interfere with one another is at present an open question.

REFERENCES

- [1] R. H. Wortham and A. Theodorou, “Robot transparency, trust and utility,” *Connection Science*, vol. 29, no. 3, pp. 242–248, 2017.
- [2] C. Breazeal, C. Kidd, A. Thomaz, G. Hoffman, and M. Berlin, “Effects of nonverbal communication on efficiency and robustness in human-robot teamwork,” in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Alberta, Canada: Ieee, 2005, pp. 708–713. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1545011>
- [3] T. Kim and P. Hinds, “Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction,” *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pp. 80–85, 2006.
- [4] R. H. Wortham, S. E. Gaudl, and J. J. Bryson, “Instinct : A Biologically Inspired Reactive Planner for Embedded Environments,” in *Proceedings of ICAPS 2016 PlanRob Workshop*, London, UK, 2016. [Online]. Available: <http://icaps16.icaps-conference.org/proceedings/planrob16.pdf>
- [5] R. H. Wortham, A. Theodorou, and J. J. Bryson, “What Does the Robot Think? Transparency as a Fundamental Design Requirement for Intelligent Systems,” in *IJCAI-2016 Ethics for Artificial Intelligence Workshop*, New York, USA, 2016. [Online]. Available: [http://opus.bath.ac.uk/50294/1/WorthamTheodorouBryson\[_\]EFAI16.pdf](http://opus.bath.ac.uk/50294/1/WorthamTheodorouBryson[_]EFAI16.pdf)
- [6] —, “Improving Robot Transparency : Real-Time Visualisation of Robot AI Substantially Improves Understanding in Naive Observers,” in *2017 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Lisbon: {in-press}, 2017.
- [7] R. C. Berwick and N. Chomsky, *Why only us: Language and evolution*. MIT press, 2015.
- [8] G. Kempen, N. Olsthoorn, and S. Sprenger, “Grammatical workspace sharing during language production and language comprehension: evidence from grammatical multitasking,” *Language and Cognitive Processes*, vol. 27, no. 3, pp. 345–380, 2012.
- [9] J. J. Bryson, “Intelligence by design: Principles of modularity and coordination for engineering complex adaptive agents,” Ph.D. dissertation, MIT, Department of EECS, Cambridge, MA, June 2001, AI Technical Report 2001-003.
- [10] Z. Dörnyei and T. Taguchi, *Questionnaires in second language research: Construction, administration, and processing*. Routledge, 2009.
- [11] M. M. Bradley and P. J. Lang, “Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings,” The Center for Research in Psychophysiology, University of Florida, Tech. Rep., 1999. [Online]. Available: <http://www.uvm.edu/pdodds/teaching/courses/2009-08UVM-300/docs/others/everything/bradley1999a.pdf>
- [12] T. J. Prescott, “Robots are not just tools,” *Connection Science*, vol. 29, no. 2, pp. 142–149, 2017. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1080/09540091.2017.1279125>