Cite as: Suna Bensch, Jiangeng Sun, Juan Pedro Bandera Rubio, Adrián Romero-Garcés and Thomas Hellström. Personalised multi-modal communication for HRI, WARN workshop at the 32nd IEEE International Conference on Robot and Human Interactive Communication, RO-MAN, 2023.

Personalised multi-modal communication for HRI

Suna Bensch[1], Jiangeng Sun[1], Juan Pedro Bandera Rubio[2], Adrián Romero-Garcés[2] and Thomas Hellström[1]

Abstract— **One important aspect when designing understandable robots is how robots should communicate with a human user to be understood in the best way. In elder care applications this is particularly important, and also difficult since many older adults suffer from various kinds of impairments. In this paper we present a solution where communication modality and communication parameters are adapted to fit both a user profile and an environment model comprising information about light and sound conditions that may affect communication. The Rasa dialogue manager is complemented with necessary functionality, and the operation is verified with a Pepper robot interacting with several personas with impaired vision, hearing, and cognition. Several relevant ethical questions are identified and briefly discussed, as a contribution to the WARN workshop.**

I. INTRODUCTION

Understandability, a.k.a. *Intelligibility* is attracting increasing attention in human-robot interaction research. An understandable robot is designed and acts to communicate important information to interacting humans. Such information may concern the robot's past, current, and future actions, but also its limitations, intentions, plans, and beliefs [5]. Understandability affects user experience, trust, safety, and efficiency, and is particularly important in elder care applications, where humans may be fragile, and non-understandable robot actions may negatively affect the humans' integrity [3]. Incorporating understandability in elder care robotics applications is not only important, but also challenging in several respects. Older adults often have reduced hearing, vision, mobility, and cognition - some with impairments that vary largely between individuals. One approach to deal with this challenge is to personalize communication, and also to adapt communication to varying environmental conditions such as ambient light and sound. In this paper we propose a design of a dialogue management system for robots with such functionality. The design has a general applicability, but in this paper we specifically target a robot for cognitive exercises with older adults. We demonstrate how the dialogue system adapts to users with different user models, and to varying environmental conditions. As part of the WARN workshop at RO-MAN 2023, we also provide a discussion on ethical questions related to the proposed solutions.

*This work was supported by the Wenner-Gren foundations, the TAIGA Centre for Transdisciplinary AI, and by the projects TED2021-131739B-C21 and PDC2022-133597-C42, funded by the Gobierno de España and FEDER funds.

[1] Suna Bensch, Jiangeng Sun, and Thomas Hellström are with Department of Computing Science, University of Umeå, Umeå, Sweden suna@cs.umu.se jisu0012@student.umu.se thomash@cs.umu.se

[2] Juan Pedro Bandera Rubio and Adrián Romero-Garcés are with the Department of Electronic Technology, University of Málaga, 29071, Málaga, Spain jpbandera@uma.es argarces@uma.es

II. BACKGROUND

In [5], the following questions are identified as particularly important for an understandable robot to answer in order to communicate in an understandable fashion:

$Q1$: When to communicate?
$Q2$: What to communicate?
$Q3$: How to communicate?

While these questions have been studied in linguistics [11], a fundamental analysis within HRI is largely missing. In the presented work we consider the following examples of how a robot may answer these questions in an adaptive fashion to promote understandability:

1) Ensuring the user's attention before communicating ($Q1$), for example, by suspending communication if the user is not perceived as sufficiently attentive.
2) Adapting the amount of spoken information to the user's estimated cognitive state and to the environment ($Q2$). For example, by adapting the level of a cognitive exercise to the user's ability, or by being brief if the user has cognitive impairments, or if it is very noisy.
3) Adapting modes of communication (e.g. speech vs screen) and associated parameters to the user and environment ($Q3$). For example, by increasing the volume in moderately noisy environments, and using a screen display in very noisy environments, in particular if the human is hearing impaired.

III. METHOD

To address the questions $Q1 - Q3$ listed above we developed a dialogue management system building on the Rasa dialogue manager [2]. Rasa accepts input from a speech recognition module and produces text messages to be output to the user. We propose additional functionality that considers a user model and an environment model as described below. The resulting dialogue system was implemented on a laptop computer interfaced to a Pepper robot[1] using the NAOqi API[2]. To improve speech recognition accuracy, Google's *Speech-to-Text* function was used to convert sound files to text strings. In the remainder of this section, the models and algorithms developed for the additional functionality is described. It should be noted that the models are highly application dependent and are here kept simple to demonstrate the general ideas of adaptation and personalization.

[1] The Pepper robot from Aldebaran.
[2] NAOqi API.

A. User model

User models have for a long time been used in robotics, and may refer physical, cognitive, or social aspects of the interaction between human and robot [10]. The purpose of the models may be to help the robot to understand the human's behavior and dialogue, and to adapt the robot's behavior and dialogue to the human's abilities, experiences, and knowledge. User models have, for example, been demonstrated for robot supported physical rehabilitation therapies [8], cognitive therapies [4], and coordination of collaborative tasks [7].

In the presented work, each user model S comprises a set of parameters s_i that specify how a specific human H is expected to perceive and react to a robot R's communication:

s_1 : Hearing impairment (0,1,2)
s_2 : Seeing impairment (0,1,2)
s_3 : Cognitive impairment (0,1,2)
s_4 : Level of attention (0,1)
s_5 : Cognitive exercise level (1,2,3)

The parameters $s_1 - s_4$ model how and to what extent H manages to perceive the robot. The parameters $s_1 - s_3$ are coded as 0, 1, or 2 corresponding to normal, mildly impaired, and severely impaired. $s_1 - s_3$ typically remain stable during an interaction session and change slowly e.g. due to natural degradation or progressive improvement through rehabilitation, or rapidly e.g. due to an accident or illness. The choice of parameters in the user model was done in collaboration with staff at an assisted living facility in Málaga, Spain.

The parameter s_4 typically changes during the course of an interaction session, e.g. due to tiredness after a long session of physiotherapy, or external distraction such as other people entering the room. The value 0 corresponds to low attention and 1 to sufficient attention for meaningful interaction.

The parameter s_5 describes the appropriate difficulty level for the user when involved in exercises with the robot. In a real application, this parameter is typically updated based on the user's progress during therapy.

For the evaluation presented in this paper, the parameters $s_1 - s_3$ were assumed to be provided by the caregivers. The parameter s_4 was simulated by a human operator pressing a key on the computer keyboard. In a real setting, s_4 should be estimated by the robot's perception system. Parameter s_5 was preset to a fixed value.

A complete system would handle several user models, and include a method to identify the user such that the appropriate model can be applied. This functionality is not dealt with further in this paper.

B. Environment model

The environment model X comprises information relevant for the interaction between R and H. We use the following parameters, which are assumed to be dynamically estimated by R during the interaction:

x_1 : Ambient noise (0,1,2)
x_2 : Ambient light (0,1,2)

x_1 and x_2 are coded as follows: 0: not disturbing, 1: a bit disturbing, 2: very disturbing. For the evaluations presented in this paper, these values were simulated by a human operator pressing a key on the computer keyboard. In a real setting, both x_1 and x_2 may vary both between and during interaction sessions. For example, the ambient noise may vary depending on the number of people present in the room, and the ambient light level may vary depending on how close to a window the robot is positioned, and the level of sunlight outside.

C. Design of the dialogue management system

To answer question $Q1$, the system uses s_4, the perceived level of attention (simulated during our tests). If the user is non-attentive ($s_4 = 0$), the system tells the user to be more attentive before proceeding with the dialogue.

To answer questions $Q2$ and $Q3$, the dialogue system dynamically assigns values to a vector C comprising parameters c_i that the robot uses to control communication:

c_1 : Modality ("*Speech*","*Screen*", "*None*")
c_2 : Volume of speech (1,2,3)
c_3 : Font size (1,2,3)
c_4 : Amount of information (1,2,3)

The assignments of $c_1 - c_3$ are described in Algorithm 1. c_1 determines if utterances from the dialogue manager

Algorithm 1 Assignment of $c_1 - c_3$

$c_2 \leftarrow 1 + s_1 + x_1$ ▷ Set required volume
$c_3 \leftarrow 1 + s_2 + x_2$ ▷ Set required font size
if $c_2 \leq 3$ **then** ▷ Is volume within limits?
  $c_1 \leftarrow$ "$Speech$" ▷ Speech is the default modality.
else
  if $c_3 \leq 3$ **then** ▷ Is font size within limits?
    $c_1 \leftarrow$ "$Screen$"
  else ▷ Conditions are too hard for communication.
    $c_1 \leftarrow$ "$None$"
  end if
end if

are spoken by the robot, displayed on its tablet, or not communicated at all (when the environmental conditions in combination with the user's impairment is too demanding for communication to be meaningful). c_2 determines the volume of speech and c_3 determines the font size for utterances displayed on the tablet. Hence, the assignment of $c_1 - c_3$ provides an answer to question $Q3$.

In the setup file for the Rasa dialogue manager, each robot utterance is coded in three versions: (1) Short, (2) Medium, and (3) Long. The current value of c_4 determines which version will be communicated to the user. The assignment of c_4 depends on the current ambient noise level and the user's cognitive impairment, as described in Algorithm 2. The assignment of c_4 provides an answer to question $Q2$.

Combining Algorithm 1 and 2, we can formulate a function F that maps a user model S and an environment model X to a vector C that is used to control the robot's generation

Algorithm 2 Assignment of c_4

if $x_1 = 2$ **then** ▷ Very noisy
    $c_4 \leftarrow 1$ ▷ Short utterance
else
    if $s_4 = 1$ **then** ▷ User attentive
        $c_4 \leftarrow 3 - s_3$ ▷ Consider cog. impairment
    else ▷ User not attentive
        $c_4 \leftarrow 1$ ▷ Short utterance
    end if
end if

of communicative actions:

$$C = F(S, X). \tag{1}$$

IV. Evaluation

The mechanisms described above were implemented for a simple application where the Pepper robot conducted a simple cognitive exercise together with an interacting human. For evaluation, interaction with four simulated personas, with varying kinds and levels of impairments, change of environmental conditions, were recorded and evaluated. A short video highlighting some interesting parts of the interaction or the complete video recordings are available from the first author on request.

Figure 1 shows an excerpt of an interaction between the persona *Jack* and the Pepper robot, which adapts communication modality, font size, and volume to changes in the environment and to Jack's user profile, including the dynamically perceived attention level. The persona Jack has mild hearing and cognitive impairment and no seeing impairment. The interaction exemplified in Figure 1 begins with the robot uttering *Hi, are you ready for cognitive exercises?*. The ambient noise and light are not disturbing, Jack is attentive (looking at the screen), and the modality is set to *Speech* with medium volume (see Line 0). In Line 3, Jack becomes unattentive and does not answer the robot's question. The attention parameter switches from 1 (attentive) to 0 (non-attentive). The robot waits and then calls for Jack's attention in Line 4, uttering *Sorry, I can't hear you.* In Line 5, Jack is attentive again and the parameter value changes accordingly. In Line 7, the robot asks Jack to multiply two numbers, and in Line 8, the ambient noise increases and the parameter value changes accordingly from 0 to 1, where 1 stands for 'a bit disturbing'. This results in an increase of the volume from Medium to High (see Line 8).

Jack and the robot continue the dialogue and in Line 11 the ambient noise becomes severely disturbing ($x_1 = 2$) which causes the robot to switch communication modality to *Screen*. The robot continues the dialogue via the screen. In Line 15, there is a severe light disturbance, and parameter x_2 changes from 0 to 2, which results in an increase of the font size from Small to Big on the screen for better readability.

V. ETHICAL CONCERNS

While the intention with the work presented in this paper is to improve understandability and communication between robots and users, we are aware of several relevant ethical questions related to the project. Some of these are:

- How should privacy and data security be handled in an appropriate way? In the application addressed in this paper, the system will store sensitive personal data regarding progress in cognitive training, and about levels of impairment. A related matter is that, as mentioned above, a complete application would typically involve automated identification of the users. These matters clearly have privacy and data security implications that have to be considered appropriately.
- How should integrity be handled in an appropriate way? The choice of modality and parameters for communication affect not only how the user, but also bystanders, hear and see what the robot communicates. If this involves sensitive personal information, the optimal choice becomes an ethical trade-off between understandability and personal integrity. A second aspect related to integrity is that personalized communication may make the user feel "classified" as impaired in a way that is embarrassing, stereotyping, or simply incorrect.
- How are bystanders affected by the adaptive communication? The optimal communication settings for the user may not be optimal for bystanders. A more developed solution should, for example, not increase the volume for a hearing impaired user without taking into account how bystanders may be negatively affected. Hence, the chosen means of communication is an ethical trade-off between understandability and the bystanders' experience.

Obviously, these and other relevant ethical questions would have to be properly addressed before a system such as the one proposed is put into real use.

VI. Discussion

A proposed solution for adaption of communication to both the interacting human and the environment was presented. The evaluation with personas demonstrated and verified the basic operation. It should be noted that while the parameters in our simplified user and environment models are discrete, the equations to compute the communication parameters $c_1 - c_4$ can be easily modified to allow for several discrete steps, or continuous parameters.

Future work will include experiments with real users, to calibrate scales for the variables in the models, perceived environmental factors, and communication parameters.

Another important part of future work is to develop robust perception methods for estimation of the parameters in the environment model X and of the level of attention (s_4). Perception of x_1 and x_2 could be done by sensing and averaging ambient sound and light levels respectively. Perception of the level of attention is much more complex but solutions could be inspired by methods developed for the automotive industry to detect whether drivers are fully attentive [12].

Research shows that multiple communication modalities are often preferred over single modality [6], [9] and more

Persona *Jack* with mild hearing impairment (s_1=1), no seeing impairment (s_2 = 0), and mild cognitive impairment (s_3 = 1).								
	Utterances [Changes in environment/attention]	Ambient Noise x_1	Ambient Light x_2	Attention Level s_4	Modality c_1	Font size c_3	Volume c_2	Amount of information c_4
0	*Robot: Hi, are you ready for cognitive exercises?*	0 (not disturbing)	0 (not disturbing)	1 (attentive)	*Speech*	1 (small)	2 (medium)	*medium*
1	*Jack: Yes, I'm ready.*							
2	*Robot: Math or memory game?*							
3	[Jack non-attentive]			0 (not attentive)				
4	*Robot: Sorry, I can't hear you.*							
5	[Jack attentive]			1 (attentive)				
6	*Jack: yeah, sorry, can you repeat?*							
	...							
7	*Robot: Ok, let's go. The result of 4 x 5?*							
8	[Noise disturbance: 0 -> 1]	1 (a bit disturbing)					2 (high)	
9	*Jack: The result is 20!*							
10	*Robot: Correct! Continue?*							
11	[Noise disturbance: 1 -> 2]	2 (severely disturbing)			*Screen*		4 (volume off)	
12	*Jack: Yes!*							
13	*Robot: The result of 3 x 5?*							
14	*Jack: The answer is 15.*							
15	[Light disturbance: 0 -> 2]		2 (severely disturbing)			3 (big)		
16	*Jack: No.*							

Fig. 1. Excerpt of the dialogue between the persona *Jack* and the Pepper robot. The values for the parameters x_1, x_2, s_4, c_1, c_2, c_3, and c_4 are given in the respective columns and hold until the value changes (read from top to bottom). The excerpt exemplifies how the dialogue manager adapts communication modality, font size, and volume to changes in environment and user attention.

engaging [1]. While we in the presented work focus on selecting and adapting single modalities to the environmental status and the human's cognitive state, future work will involve communicative actions combining several modalities, also including gestures.

While evaluating the system it became apparent that an adaptive dialogue system should adapt not only outgoing messages from the system, but also adaptively support different input modalities from the user. For example, if the dialogue system switches from speech to screen output because the environment is too noisy, it cannot reasonably expect to hear the user's voice, but should rather offer a touch screen based input.

Another important line for future work is to consider related ethical questions, such as the ones discussed in Section V.

The developed source code can be downloaded from GitHub. A video showing the robot functionalities, including adaption to changing environmental factors and user attention levels can be accessed here.

References

[1] A. Bliek, S. Bensch, and T. Hellström. How can a robot trigger human backchanneling? In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 96–103, 2020.

[2] T. Bocklisch, J. Faulkner, N. Pawlowski, and A. Nichol. Rasa: Open source language understanding and dialogue management, 2017.

[3] P. Bustos, L. Manso, A. Bandera, J. Bandera, I. García-Varea, and J. Martínez-Gómez. The cortex cognitive robotics architecture: Use cases. *Cognitive Systems Research*, 55:107–123, 2019.

[4] D. Conti, G. Trubia, S. Buono, S. Di Nuovo, and A. Di Nuovo. *Evaluation of a Robot-Assisted Therapy for Children with Autism and Intellectual Disability*, pages 405–415. Springer Verlag, 07 2018.

[5] T. Hellström and S. Bensch. Understandable robots - What, Why, and How. *Paladyn, Journal of Behavioral Robotics*, 9(1):110–123, 2018.

[6] D. O. Johnson and A. Agah. Human robot interaction through semantic integration of multiple modalities, dialog management, and contexts. *International Journal of Social Robotics*, 1(283), 2009.

[7] G. Maeda, G. Neumann, M. Ewerton, R. Lioutikov, O. Kroemer, and J. Peters. Probabilistic movement primitives for coordination of multiple human-robot collaborative tasks. *Autonomous Robots*, 41, 03 2017.

[8] A. Martin, J. C. Pulido, J. C. Gonzalez, A. García-Olaya, and C. Suárez. A framework for user adaptation and profiling for social robotics in rehabilitation. *Sensors*, 20(17), 2020.

[9] K. Pollmann. The modality card deck: Co-creating multi-modal behavioral expressions for social robots with older adults. *Multimodal Technologies and Interaction*, 5(7), 2021.

[10] S. Rossi, F. Ferland, and A. Tapus. User profiling and behavioral adaptation for hri: A survey. *Pattern Recognition Letters*, 99:3–12, 06 2017.

[11] J. Sidnell. *Conversation Analysis: An Introduction*. Wiley-Blackwell, Oxford, 2010.

[12] L. G. Véras, A. K. F. Gomes, G. A. R. Dominguez, and A. T. Oliveira. Drivers' attention detection: a systematic literature review, 2022.