Socially-acceptable Extended Reality Models and Systems

D4.1 Specification of nonverbal communication software interface

30 March 2023

Funded by
the European Union

DELIVERABLE INFORMATION	
Deliverable leader	SUPSI
Document type	R
Document code	D4.1
Document name	Specification of nonverbal communication software interface
Work Package / Task	WP4 / T4.1 Specification of nonverbal communication software interface
Delivery Date (DoA)	30 March 2023
Actual Delivery Date	30 March 2023
Reviewers	D. Petrak (TuDA) T. Tran (TuDA) L. Vigano (KCL) L. Sabattini (UNIMORE)

DELIVERABLE HISTORY			
Date	**Version**	**Author**	**Summary of main changes**
26 Feb. 23	0.1	A. Paolillo (SUPSI) A. Giusti (SUPSI)	Drafting the ToC and general concepts.
9 Mar. 23	1.0	A. Paolillo (SUPSI) A. Giusti (SUPSI)	Filling the sections, adding content and details; minor changes in the titles of the sections; collecting contributions from the other partners.
20 Mar. 23	1.1	A. Paolillo (SUPSI) A. Giusti (SUPSI)	Collecting contributions from other partners and reviewers; writing conclusions; finalizing the draft of the deliverable.

DISSEMINATION LEVEL

PU	Public	

SERMAS partners

Disclaimer

This project has received funding from the Horizon Europe programme under the Grant Agreement 101070351.

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Public Executive Summary

This deliverable

- Presents the overall general design and specification of a software interface between the nonverbal communication and the dialogue system
- Focuses on the perspective of the nonverbal communication block
- Describes the management of two streams of information: *towards* and *from* the dialogue system,
- Describes the stream of information exchanged by the human user and the XR agent, in terms of perception and actuation.

Table of contents

List of Figures

List of Tables

1. Introduction

This deliverable presents the overall general design and specification of a software interface to make the nonverbal communication interact with the dialog management developed in WP 5. In a nutshell, from the perspective of the nonverbal communication block, this software interface must manage two streams of information: the first is *towards* the dialogue system and contains data about nonverbal interaction observed in users; the latter is *from* the dialogue system and contains instructions for nonverbal interaction to be executed by the XR agent.
The software interface must manage the stream of information exchanged by the (human) user and the XR agent, in terms of perception and actuation.

2. Communication in Social Human-Robot Interaction

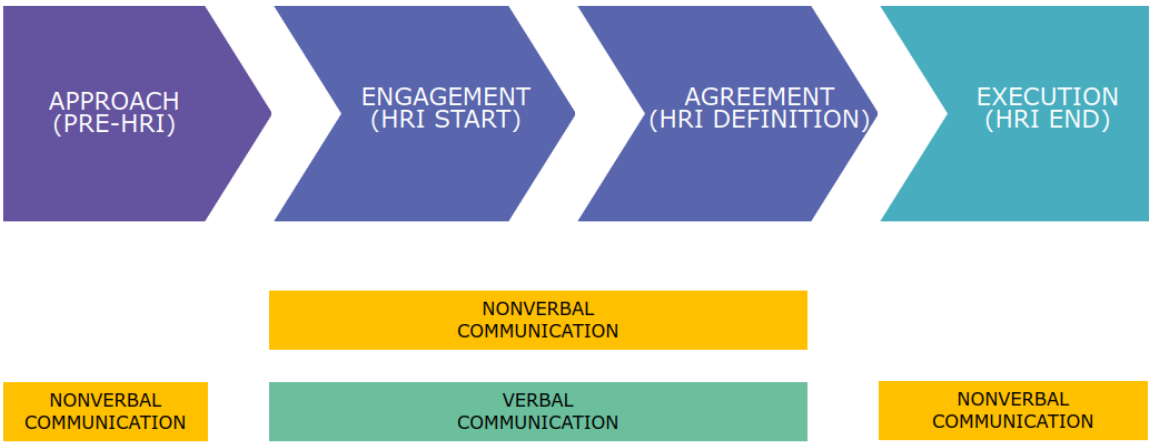

Figure 1. The Social Human-Robot Interaction pipeline

As a visual reference, consider the block diagram presented in Fig. 1. The Social Human-Robot Interaction pipeline can be seen as a cascade of 4 phases: Approach, Engagement, Agreement, and Execution. In the Approach phase, the actors of the interaction (the human user and the XR agent) only share the scene; there is potential of interaction, but it could also not take place. In the Engagement phase, the actors come to contact and actually start the interaction, followed by the Agreement phase, where both the user and the agent need to agree on the task to be executed. Finally, in the Execution phase, the task is actually realized, and the interaction comes to an end.

Note that in the real implementation of a social human-robot interaction pipeline, some or all the phases could be repeated as an iterative process.

During each of these phases, and at the switch between one phase and the other, it is important to establish communication modalities in order to make an effective and successful human-robot interaction. In some phases, verbal or nonverbal communication could be dominant, on other phases, the two modalities could coexist and cooperate.

For example, in the Approach phase, where the two interacting actors just share the scene and nothing is defined yet, it is crucial to understand the possible intention of interacting of the user. This information can be inferred by the XR agent by observing the body language of the user and other motion cues. During

the Engagement and the Agreement phase, it is expectable that the actor will share information through dialog.

From the perspective of nonverbal communication, it is important to design and specify and interface with the dialog system, which would allow an effective and successful interaction experience. For example, if the intention to interact is detected by the nonverbal communication block, this piece of information can be shared with the verbal communication block in order to start a dialog. As another example, during the playing of the dialog, speech can be accompanied by gestures and other nonverbal modalities to make the expression of the messages clearer.

In this deliverable, we focus on the aspects that concern nonverbal communication, and its interface with the dialogue system. In the following sections we will describe the nonverbal communication modalities. For the aspects related to verbal communication, please refer to deliverable D5.1 "Datasets suitable for training, validation, and testing of modules, along with infrastructure to support data collection/processing", which is developed in the framework of the activities in WP5 "Dialogue Management and NLP".

Table 1- An example of nonverbal communication modalities in a classic social human-robot interaction task where the user asks for direction information

	Explicit	**Implicit**
From user to agent	The user asks for information explicitly attracting the attention of the agent, e.g., waving their hands, which is perceived by the agent's sensors	The user implicitly manifests the need of information showing urgency or anxiety, which is perceived by the agent's sensors
From agent to user	The agent uses its actuators to explicitly indicate a direction to the user, e.g., showing an arrow on the screen or pointing with the arm	The agent uses its actuators to implicitly indicate a direction, e.g., by orienting its body or gaze toward the destination

2.1. Nonverbal communication modalities

Nonverbal communication can be divided into *implicit* and *explicit* and refers to the flows from the agent to the user and vice versa. Table 1 presents, for the sake of clearness of the presentation of these modalities, some examples of nonverbal communication, which are then described in detail in the following subsections. However, note that the list of examples presented here is not exhaustive nor implies that such modalities will be actually implemented. It instead serves as a

base of development and aims to clarify the ideas behind the implementation of the software interface that we need to define between nonverbal communication and the dialog system. This list can be modified, updated and changed according to the use-cases need, and the project evolution.

2.1.1. Explicit nonverbal communication

Explicit nonverbal communication regards those modalities that can be clearly perceived (when we talk about the flow of information happening from the user to the agent) or actuated (from the agent to the user). Concrete examples of explicit nonverbal communication are those realized through iconic or pointing gestures. In the following, we present some examples of interaction based on the explicit nonverbal communication that we plan to have in the project.

2.1.1.1. From user to agent

Examples of explicit nonverbal communication happening from user to agent are:

- The user waves at the agent to check if the agent reacts;
- The user while interacting with the agent indicates a location with their arm;
- The user uses iconic gestures to explicitly inform the agent about something, e.g., answer yes or no to a question.

2.1.1.2. From agent to user

Examples of explicit nonverbal communication happening from agent to user are:

- The agent indicates a direction by pointing with its arm;
- The agent indicates a direction showing an arrow on a screen;
- The agent communicates an emotional feeling by showing a face on a screen;
- The agent clearly answers a specific question of the user using nonverbal modalities, e.g., by showing proper graphics on a display.

2.1.2. Implicit nonverbal communication

Implicit nonverbal communication regards those modalities that can be inferred (when we talk about the flow of information happening from the user to the agent)

or shown (from the agent to the user) through gestures such as positioning in space, facial expressions and other emotional or attentional cues.

2.1.2.1. From user to agent

Examples of implicit nonverbal communication happening from user to agent are:

- The user approaches the agent while looking at it (user is probably going to interact);
- The user is in proximity of the agent but looks and moves elsewhere (user is not interested in interaction);
- The user looks confused while the interaction is in progress;
- The user is in the environment monitored by the agent and is looking around, seemingly confused (the user probably needs help, but did not notice the agent).

2.1.2.2. From agent to user

Examples of implicit nonverbal communication happening from agent to user are:

- The agent moves fast or slow, to show urgency or quietness;
- The agent moves away if the user shows discomfort;
- The agent gets closer if the user gets distracted;
- The agent shakes the body or the arm to attract the user's attention.

2.2. Security and Privacy issues

For each of these nonverbal communications, a number of security and privacy issues will need to be considered to protect the system and the data exchanged. For instance, the following questions will need to be answered, by formally modeling the communication and formally analyzing its properties. How will the agent authenticate the user (and, possibly, how will the user authenticate the agent)? How will authorization and access control be enforced? Which of the exchanged data is private and will need to be protected? And so on. To carry out such analyses, we will need to specify the communication and its concrete steps (e.g., via message sequence charts), specify if and how encryption will be used, and specify which sensible data needs to be protected and how. To that end, we will need to extend existing approaches for the formal and automated analysis of so-called security ceremonies (in which human agents exchange data with the

other, software agents in a system) to the case of nonverbal communications such as those considered for the SERMAS agent. This deliverable and the following ones in WP 4 will provide the basis for the security and privacy validation that will be carried out in task “T 4.5: Validation”, which will start at project month 18 and end at month 30.

3. Integration of nonverbal and verbal modalities

In this section, we focus on the integration between nonverbal and verbal modalities, from the perspective of the nonverbal communication block.

With reference to the block diagram in Fig. 1, we essentially have two kinds of integration between nonverbal and verbal communication modalities. The first, relatively simpler, is about switching between the two modalities and it is based on triggering signals. The latter, more sophisticated, is about the handling of both modalities at the same time, and it is based on synchronization signals. These signals are detailed in the following subsections.

Combining different channels of communication means that the agent can produce nonverbal behaviors given verbal information. For example, it could be co-speech gesture generation.

3.1. Triggering signals

This is the simplest kind of integration we can think about, actually happening at the cross between the Approach and Engaging phase, and between the Agreement and Execution phase. A triggering signal can be used to activate specific motion such as waving hands (e.g., when the intention to interact of a user has been detected and an interaction can start).

3.2. Synchronization signals

This is a more complicated integration since it deals with the coexistence of both the verbal and nonverbal communication modalities during the same phases of the interaction, e.g., during the Engagement and Agreement phase.

4. Software interface

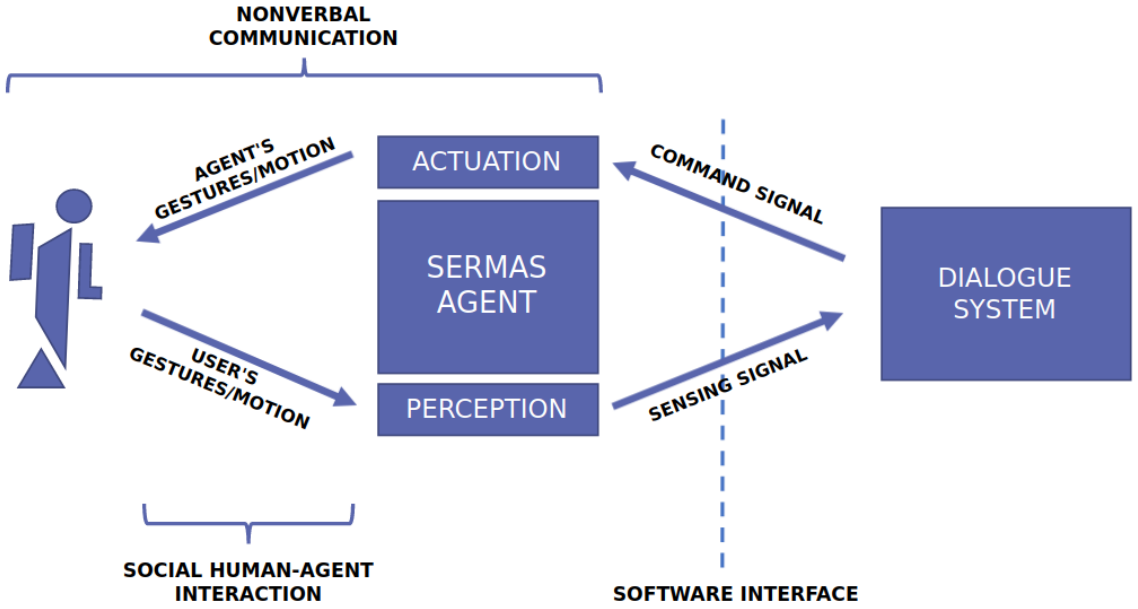

Figure 2. Schematic of the interface between the nonverbal communication block and the dialog system.

In this section, we tackle the concepts introduced in Sections 2 and 3 in concrete terms, giving an overview of how they are going to be implemented. In terms of implementation, from the perspective of the nonverbal communication block, we refer to two main flows of information (see Fig. 2). The first is happening between the agent and the user and it is bidirectional:

- from the user to the agent (this information is captured by the perception of the agent),
- from the agent to the user (which is expressed by the actuation of the agent).

The handling (perception and actuation) of this kind of signal, happening during the social human-agent interaction, is the objective of nonverbal communication. The second flow of information is the one happening between nonverbal communication and the dialog system. At the bridge between these two modules, there is the software interface between nonverbal communication and the dialogue system, which consists of a high-level interface that oversees the handling of

- *Sensing* signals, which are the interpretation of the robot perception and comes from the nonverbal communication to provide the dialogue system with further information
- *Command* signals, which come from the dialogue system to provide the nonverbal module with command to be executed by the agent.

In general, the sensing signal is denoted with the letter S and has this shape

S = <user_id, it, at_1, at_2, …, at_N, timestamp>

where:

- *user_id* refers to an identification number for each user detected in the interacting area around the agent;
- *it* refers to the interaction state of the detected user and it is a Boolean variable set to true if the user is currently interacting with the agent, or false otherwise;
- *at_i,* with *i = 1,…,N*, is a list of attributes describing the state of the user. For example, a possible attribute can be the "level of attention" of the user, the probability that the user is confused, and so on. These pieces of information can be used to trigger the end of the interaction if, e.g., the level of attention is too low, or be used to re-start a piece of dialogue during the interaction if the probability of confusion increases. Another attribute of interest is the "the intention to interact" of the user, which is used to predict the willingness of the user to approach the agent and elaborate proactive strategies of the agent;
- *timestamp*: time associated to the sensing information.

Note, once again, that we refer to sensing signal as those pieces of information extrapolated and interpreted from the sensors (and not the raw sensory information).

The attributes represent an important piece of information, which is at the base of the integration of nonverbal communication and the dialogue system. They are used to

- Perceive events, to send the triggering signals introduced in Sec 3.1 to the dialogue system, and to

- Perceive states, to send the synchronization signals introduced in Sec. 3.2 to the dialogue system.

The commands signal is denoted with the letter C, and has the shape

C = <user_id, ac_1, ac_2, …, ac_M, timestamp>,

where:

- the "*user_id*" (optional) is the user to which the actions are directed, and
- *ac_i,* with *i = 1,…,N*, is a list of action that the agent should take, according to the current sensing input or the input coming from the dialogue system;
- *timestamp*: time associated to the command signal.

Similarly, as the attributes of the sensing signals, the actions represent another important piece of information at the base of the integration of nonverbal communication and the dialogue system. They are used to

- Actuate instantaneous actions (e.g., movements), to show that a triggering signal has been captured, and to
- Actuate continuous actions, in order to accompany the dialogue system and synchronize with the nonverbal communication mechanisms.

Such messages can be implemented using different technologies. One convenient solution, widely used and adopted as a standard in the robotics community, would be to use the Robotic Operating System (ROS) and its infrastructure allowing the easy transmission of messages through a publisher-and-subscriber protocol.

Signal S and C can be used to handle the nonverbal communication between the user and the agent. This aspect regards the implementation of perception and actuation of these modalities and is addressed in Sec. 4.1-4. For a list of possible attributes and actions, we refer to Section 2. In this section, instead, we focus on how these attributes can be used to realize a software interface.

4.1. From user to agent (perception)

The perception is used to *fill* signal S with the proper information.

The information coming from the sensor of the agent (in the case of physical agents such as robots) or the virtual environment (in the case of agents in virtual, augmented, or extended reality) is processed to interpret the state of the user.

4.2. From agent to user (actuation)

The actuation is a way to *use* the information contained in the signal C.
More into detail, the detected attributes of the user can be used by the agent to implement reaction strategies or other kinds of motion that enhance the effectiveness of the interaction experience.
For example, if the agent perceives the intention to interact with the user, the agent can use this same piece of information to wave the arm (if any) or blink the lights (if any) to show availability for interacting with the user.

4.3. From dialogue management to nonverbal communication block

The information contained in the dialogue management system can be used to fill actions of the signal C and used by the nonverbal communication module to actuate motions or behavior that accompany the dialogue and makes the interaction experience more advanced and effective.
For example, if the dialogue reaches a particular point where it requires special attention by the user, the verbal communication could fill the attribute “need attention” and be used to blink lights or waive hands.

4.4. From nonverbal communication to dialogue management

The attributes of signal S can be used to both trigger the start or the end of dialogues, or to synchronize the motion and actuation of the agent with pieces of verbal communication.
For example, if the attribute “intention to interact” of the signal S rises, this information can be used to start a dialogue to greet and welcome the user.
The attribute “probability of confusion”, instead, can be used to repeat part of the dialogue, interrupt the dialogue, or ask the user what is not clear.

5. Conclusion

In this deliverable we have specified the nonverbal communication software interface. We describe the messages that should be exchanged between the nonverbal communication block and the dialogue system, to realize a successful socially accepted human-robot interaction pipeline. We envision that the interface actually consists of two kinds of message. The first, which we call sensing signal, is used to collect pieces of information about the user reconstructed by interpreting the agent's sensor data. The second, called command signal, is used to actuate motions or behaviors to communicate something to the user. These signals, shared with the dialogue system, would allow an integration of the nonverbal communication modalities with the dialogue system. In terms of implementation, standard tools borrowed from the robotics communities, such as the ROS infrastructure and its publisher-subscriber communication protocol, can be used to realize the interface.