Conversations with Identity Performing Robots: Considerations for Algorithms and Interfaces

Anonymous Author(s)

ABSTRACT

Recent advances in HRI have begun to unravel the longstanding, implicit assumption that a given robot has a single, coherent, and unchanging body and identity. In doing so, these recent advances call into question conversational robots' traditional means of representing and reasoning about who is speaking and who is being spoken to. This provides new opportunities and complications for algorithms and interfaces for conversational human-robot interaction. In this paper, we thus discuss the flexibility of robot identity performance and what considerations must then be made when it comes to the algorithms and interface used to direct robots' conversational capabilities.

KEYWORDS

Human-Robot Interaction, Robot Groups, Identity Performance

1 INTRODUCTION

In many Human-Robot Interaction (HRI) domains, robots need to be able to communicate with humans either autonomously using dialogue models, or in a teleoperated fashion using robot control interfaces. In both of these approaches, it is often important to represent who is speaking to whom. These representations are typically straightforward in dyadic interactions because there typically is one speaker and one listener. However, recent research (e.g. [6, 9]) is showing that apparently even dyadic interactions are not always this simple, and that robots may not have a single coherent static body and identity. For example, if multiple identities inhabit a single body, or identities hop from body to body, it may be necessary to specify which identity is speaking or being spoken to. Even in cases where robots *appear* to have a one-to-one relationship between body and identity, a single robot architecture may be used to control multiple robot bodies, meaning that when speech is generated, it may be necessary to specify which body (or bodies) are used to generate speech. These possibilities complicate algorithm design, as the knowledge representations and the algorithms that operate on them will need to account for the intended bodies and identities involved in communication. Moreover, they complicate interface design, as interfaces for dialogue teleoperation must similarly account for these nuances. In this work, we thus consider how both algorithms and interfaces for language understanding and generation may need to be adapted under flexible body-identity configurations in human-robot interactions.

2 BACKGROUND

Robots can perform identity (a particular persona) through the presentation of a set of design cues such as name, speech, and behavior [3]. For instance, Bejarano et al. [1] discusses design cues that can be used by robot groups to alter their overall identity presentation. Some that are particularly relevant to conversations include how robots refer to themselves/others when they speak and the kinds of robot names and voices used. Different combinations of such design cues can indicate different configurations of mind-body-identity, otherwise referred to as *identity performance strategies*.

Different identity performance strategies (also referred to as social presence options) can indicate different relationships and associations among the robot minds (cognitive architectures), bodies (physical constructs), and identities (performed personas) present [4, 6]. For instance, a *One-for-one* strategy, that is often used in dyadic interactions, presents each robot body as having a single static identity/mind such that there appears to be a 1-1 association between bodies and identities. On the other hand, a *One-for-all* strategy in which multiple robot bodies share a single identity/mind (e.g. coordinated swarm robots) and a *Re-embodiment* strategy in which an identity may move from one body to another (see [6]), may appear to be a many-1 association between bodies and identity. Robots could also present a *Co-embodiment* strategy in which a single robot body may have multiple identities/minds (possibly to provide multiple users with personalized experiences) to signal a 1-many association between body and identities.

These different identity performance strategies can lead to different human perceptions (e.g. regarding human comfort with robots, expertise of robots, perceived trust in robots) and can change how people understand and interact with robots (e.g. what people understand to be the relationships among robots and whether or not they prefer to interact with particular robots) [1, 4, 6, 9, 10]. For instance, in prior work, Bejarano et al. [1] demonstrated that changes in the design cues used to indicate different identity performance strategies can affect user mental models in different ways. For instance, designing robot groups with shared qualities (e.g. using the same name and voice) and collective behaviors (e.g. speaking at the same time or using similar language) can lead to people perceiving a robot group as more entitative and being of a single mind. Meanwhile, designing robot groups to have unique qualities (e.g. having separate individual names and voices per robot body) among the group can lead to people perceiving a robot group as less entitative and each robot body as having its own unique mind.

Furthermore, Williams et al. [10] indicate that the number of bodies and identities involved in a user's mental model may dictate where and how users place trust. This is important to be aware of because, depending on how people understand a particular identity performance strategy, the number of bodies and identities perceived to be present in interaction may change. In other words, identity performance can complicate understandings about the number of "robots" present in a given conversation as robot bodies *and* identities may be considered separately as a "robot". This then complicates who/what people think they can interact with, how people have conversations with robots using different identity performance strategies, and how people control dialogue for robots.

As such, HRI researchers and robot designers need to carefully consider the flexibility of robot identity in algorithms and interfaces for conversational human-robot interactions.

3 CONSIDERATIONS FOR FUTURE ALGORITHMS AND INTERFACE

In this section, we discuss the particular ways that algorithms and interfaces for conversational human-robot interactions might account for the flexible relationship between robot body and identity. In particular, we discuss how existing systems could be modified or extended to handle the scenarios and strategies mentioned.

3.1 Algorithms

We'll use DIARC (Distributed Integrated Affect Reflection and Cognition) [7] as a case study to see how accounting for the flexibility of robot identity might manifest in different components of a robot architecture. In DIARC, natural language understanding and generation is performed through a set of core components [8]: *Automatic Speech Recognition (ASR)*, *Parsing*, *Reference Resolution (RR)*, *Referring Expression Generation (REG)*, *Pragmatics*, *Dialogue Manager*, and *Belief*. The information from these components is then used by a *Goal Manager* component to generate an action by an agent such as a particular speech output. To account for a possibly flexible relationship between robot body and identity, DIARC's core natural language components could be modified or extended to consider the following three design goals:

(1) Components should account for robot identity and body *separately*.
(2) The architecture should track the bodies and identities being used, and the association between these bodies and identities.
(3) The architecture should track, and appropriately use, the design cues used to differentially signal each identity.

In the following subsections we will discuss how each of the core language components of DIARC would need to be modified to account for these three design goals. Because many of these components will require some understanding of the relationship between body and identity, we first propose that DIARC should be extended to include an ***Identity Management*** component that maintains information about (1) the set of robot bodies that are (or can be) part of the multi-robot system; (2) the set of identities that can embody those robots; (3) the current mapping from bodies to identities; (4) the identity performance strategy being used; (5) the set of identity performance cues that would be used to convey each identity.

3.1.1 Speech Recognition, Parsing, and Pragmatic Understanding. In DIARC, the *ASR* component recognizes speech input and converts it to a text representation that can be shared with other components. The *Parsing* component then takes that text representation and turns it into a representation of the utterance's surface-level meaning, comprised of a set of logical predicates. The Understanding module of the Pragmatics Component then takes these surface-level meanings and attempts to infer the intended meaning behind the utterance.

The output typically produced by the Pragmatics component typically takes the form of an *unbound utterance with supplemental semantics* $\langle U(s, h, m), p\rangle$, where U is the utterance type, s is the speaker, h is the hearer, m is a logical predicate representing the intended meaning of the utterance, and p is a set of supplemental semantics that can be later used to identify the correct variable bindings for variables appearing in m.

For these three components to fulfill our stated design goals, the components must ensure that the body and identity of the speaker and hearer are accounted for in these representations. While the body and identity of the human speaker will be the same, we argue that they should be recorded separately so that later on these same representations can be used to represent robot utterances. This yields an utterance form $U(\langle s_b, s_i\rangle, \langle h_b, h_i\rangle, m)$, where $\langle x_b, x_i\rangle$ indicate a body,identity pair.

In order to use this knowledge representation, the speech recognition system will have needed to determine which body and identity are speaking and which body and identity are being addressed. While speaker body and identity can thus be inferred based on speaker identification, robot body and identity are more complicated. If the interactant is physically gazing at or facing a particular body, it may be inferred to be the intended hearer-body. From this, the intended identity can be inferred based on the body-identity mapping maintained by the proposed Identity Management component. Alternatively, an intended identity may be inferred based on dialogue cues from the dialogue manager, and from this, the intended body can be inferred. Finally, we note that it is possible that speakers may intend a group of identities or group of bodies as the intended hearer. However, we leave this consideration for future work.

3.1.2 Reference Resolution. These utterance forms would then be passed, as part of an Unbound Utterance with Supplemental Semantics to the Reference Resolution (*RR*) component. Currently, the *RR* component uses the supplemental semantics to resolve references made to objects, people, locations, and so forth. In a multiple-identity context, it may be the case that different robot identities are presumed to know different things about the world. If this is the case, the *RR* component would need to determine which architectural consultants (e.g., which *Belief* component) maintain the knowledge for each of the managed identities. Alternatively, a simplifying assumption could be made, such that each identity is presumed to know the same things. However, this could lead to uncanny valley effects [11].

3.1.3 Dialogue Management. Once reference resolution is performed, Bound Utterances are passed to the *Dialogue Component*. The Dialogue Component is responsible for managing the flow of dialogue. Critically, the Dialogue Component uses a set of dyadic interaction templates: a Question from S to H, for example, might be responded to with a Statement from H to S. This becomes more complicated in a multi-identity context. A first solution might be to assume dyadic interactions. For example, a question from $\langle s_b, s_i\rangle$ to $\langle h_b, h_i\rangle$ might be a Statement from $\langle h_b, h_i\rangle$ to $\langle s_b, s_i\rangle$. A slightly more complex solution might involve more intentionally selecting *which* body (based, e.g., on consistency and proximity) and *which* identity (based, e.g., on consistency and expertise) to use to respond

to a question. This would require interfacing with the Identity Management component and consulting the dialogue history. Finally, as mentioned above, it is possible that future work might need to account for multiple robots being addressed, or multiple robots speaking at once. If these phenomena were to be accounted for, the Dialogue Manager might need to use different interaction templates depending on the identity performance strategy activated in the Identity Management component, and these interaction templates might need to either involve the generation of multiple utterances from multiple bodies; or might need to involve the understanding and generation of utterances that encode multiple intended hearers or multiple intended speakers.

3.1.4 Pragmatic Generation. Once a robot has decided the utterance it wants to communicate, the Utterance form is sent to the Generation module of the Pragmatics component, which takes utterance forms and transforms them back into surface-level representations, with the help of a set of pragmatic rules. In a multi-identity context, different identities might use these pragmatic rules to different extents, reflecting, e.g., different levels of politeness in their personalities. In such a case, the pragmatic generation component would need to coordinate with the Identity Management component to encode a set of pragmatic rules for each identity, or a set of weightings over those rules for each identity.

3.1.5 Referring Expression Generation. Once a surface level meaning representation is created, it is sent to Referring Expression Generation to flesh entity references (e.g., $object_1$) out into full referring expressions (e.g., *the red box*). As with reference resolution, this may differ based on identity and the presumed knowledge of each of those identities. As such, to fulfill our stated design goals, the robot would once again need to maintain knowledge of which consultant to go to for information about a target referent, based on the intended speaker. In addition, the Referring Expression Generation component would need to coordinate with the Identity Management component in order to decide how to refer to the speaker, or other robot bodies and identities involved in the architecture, as the choice of self- and other-identifying language may depend on identity performance strategy [1].

3.1.6 Speech Synthesis. Finally, the Speech Synthesis component outputs text through a robot speaker or other modality. To fulfill our stated design goals, the information from the utterance representations as to which body and identity to use for speech must be passed through all the way to speech synthesis. The Speech Synthesis component must then use this information to decide which body to use to communicate (alternatively, this information could be used earlier to decide *which* Speech Synthesis component should be sent a message to communicate). Finally, the Speech Synthesis component must coordinate with the Identity Management component to determine what prosodic cues to use to appropriately communicate the intended identity.

3.2 Interfaces

While in the previous section we considered how the architectural components of a robot architecture would need to be modified to facilitate sensitivity to multiple and distributed identities during autonomous communication, in this section we expand our consideration to how this sensitivity can be achieved when robot dialogue is teleoperated rather than autonomous.

When considering the out-the-box speech control interfaces for language capable robots like Misty[1] and Pepper[2], human operators are often limited to only being able to control a single robot body at a time unless a custom interface is created for multi-robot control (e.g. see [2, 5]). However, even custom dialogue teleoperation interfaces presented in the literature have not accounted for flexibility in the relationship between robot body and identity. Instead, the only methods of teleoperating robot speech in such interfaces consist of predefined buttons and/or an input text box to send custom text-to-speech commands to a particular robot body.

Additionally, dialogue teleoperation interfaces may allow for customization of certain identity-laden design cues, like the pitch of a robot's voice, or the color of LEDs on its body. However, if an operator wanted to change the identity performance of multiple robots on-the-fly or synchronize the dialogue of multiple robot bodies, this would prove impossible or time-consuming in current dialogue teleoperatoin interfaces.

To address these potential difficulties, and account for a possibly flexible relationship between robot body and identity, existing interfaces could be modified or extended to consider the following two design goals:

(1) Teleoperators should be given a way to create *identity profiles*, each with a particular set of design cues that can be quickly applied to robot bodies.
(2) Teleperators should be able to flexibly change which identity is associated with each robot body.

In Figure 1, we demonstrate an interface prototype that accounts for these considerations. To enable faster teleoperation (regardless of whether an interaction involves multiple, distributed, dynamic identities), this interface has the following features:

(1) *Operator defined identities:* The identities "Buddy", "Honey", and "Bumble" are predefined by the operator with a given set of identity cues. In this case, the cues for each identity consist of a specific name and a particular voice that can only be used by that identity.
(2) *Connection to multiple robot bodies:* Dialogue involving multiple robots (labeled robot bodies 1, 2, and 3) can be controlled by the interface.
(3) *Flexible control of body-identity associations:* Teleoperators can change which identities are associated with each body, on-the-fly. Operators can choose multiple bodies for a given identity and can send synchronous commands to each of those bodies. That is, more than one robot body can say the same thing at the same time using either the same identity, or different identities.
(4) *Options for desired speech output:* We included the options of creating predefined buttons and also using an input text box to send custom text-to-speech commands to any of the robot bodies connected. An extension of this could include template speech buttons in which a button may be formatted as "Hello, my name is [robot name]" where

[1] https://www.mistyrobotics.com
[2] https://www.aldebaran.com/en/pepper

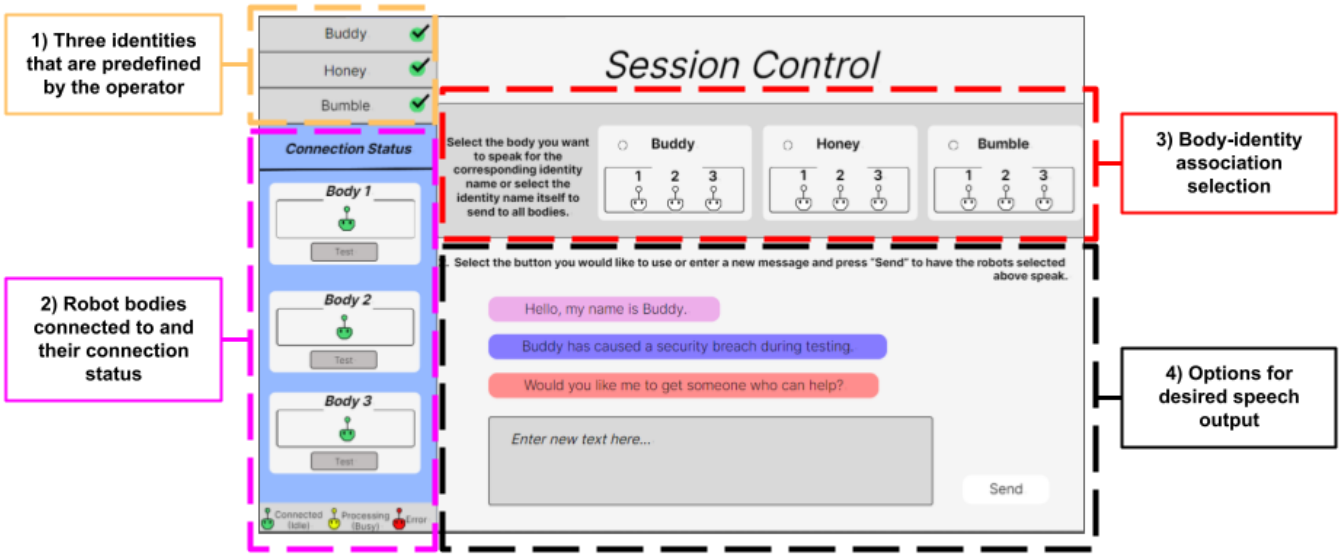

Figure 1: Prototype of a robot speech control interface that accounts for identity.

the information in brackets can be automatically filled by identity cues associated with a selected identity (in this case the name of the identity).

4 CONCLUSION

In this paper, we discuss the flexibility of robot identity performance and the opportunities and complications that flexibility provides for the algorithms and interfaces used to direct robots' conversational capabilities in HRI. Through a case study with the DIARC architecture and a design prototype, we are able to identify five overarching design goals when robot architectures involve multiple, distributed, dynamic, identities. Specifically, we argue that in such cases, it is important to (1) account for speaker identity and body separately, (2) specify the bodies and identities being used, and the association between those bodies and identities, (3) specify what design cues are to be used to signal a particular identity, (4) offer teleoperators a way to create identity profiles with a particular set of design cues that can be quickly applied to robot bodies, and (5) allow teleoperators to flexibly change which identity is associated with which robot body.

REFERENCES

[1] Alexandra Bejarano, Samantha Reig, Priyanka Senapati, and Tom Williams. 2022. You Had Me at Hello: The Impact of Robot Group Presentation Strategies on Mental Model Formation. In *Proceedings of the 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[2] Dylan F Glas, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2008. Simultaneous teleoperation of multiple social robots. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*. 311–318.

[3] Ryan Blake Jackson, Alexandra Bejarano, Katie Winkle, and Tom Williams. 2021. Design, performance, and perception of robot identity. In *Workshop on Robo-Identity: Artificial identity and multi-embodiment at HRI*.

[4] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Re-Embodiment and Co-Embodiment: Exploration of social presence for robots and conversational agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 633–644.

[5] Matthew Marge, Stephen Nogar, Cory J Hayes, Stephanie M Lukin, Jesse Bloecker, Eric Holder, and Clare Voss. 2019. A research platform for multi-robot dialogue with humans. *arXiv preprint arXiv:1910.05624* (2019).

[6] Samantha Reig, Michal Luria, Elsa Forberger, Isabel Won, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2021. Social robots in service contexts: Exploring the rewards and risks of personalization and re-embodiment. In *Designing Interactive Systems Conference 2021*. 1390–1402.

[7] Matthias Scheutz, Gordon Briggs, Rehj Cantrell, Evan Krause, Tom Williams, and Richard Veale. 2013. Novel mechanisms for natural human-robot interactions in the diarc architecture. In *Proceedings of AAAI Workshop on Intelligent Robotic Systems*. Palo Alto, CA, 66.

[8] Matthias Scheutz, Thomas Williams, Evan Krause, Bradley Oosterveld, Vasanth Sarathy, and Tyler Frasca. 2019. An overview of the distributed integrated cognition affect and reflection diarc architecture. *Cognitive architectures* (2019), 165–193.

[9] Ravi Tejwani and Cynthia Breazeal. 2021. Migratable AI: Investigating users' affect on identity and information migration of a conversational AI agent. *Proceedings of the International Conference on Social Robotics*.

[10] Tom Williams, Daniel Ayers, Camille Kaufman, Jon Serrano, and Sayanti Roy. 2021. Deconstructed Trustee Theory: Disentangling Trust in Body and Identity in Multi-Robot Distributed Systems. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 262–271.

[11] Tom Williams, Priscilla Briggs, and Matthias Scheutz. 2015. Covert robot-robot communication: Human perceptions and implications for human-robot interaction. *Journal of Human-Robot Interaction* 4, 2 (2015), 24–49.