Exploring Gesture-Based Tangible Interactions with a Lighting AI Agent

Milica Pavlovic[1,2(✉)], Sara Colombo[1], Yihyun Lim[1], and Federico Casalegno[1]

[1] Design Lab, Comparative Media Studies and Writing, Massachusetts Institute of Technology, 20 Ames Street, Cambridge, MA 02142, USA
{milicap, scolombo, yihyun, casalegno}@mit.edu
[2] Interaction and Experience Design Research Lab, Politecnico di Milano, Via Durando 38/a, 20158 Milan, Italy

Abstract. The paper explores a gestural and visual language to interact with an Artificial Intelligent agent controlling connected lighting systems. Six interaction modalities (four gestural and two visual) were designed and tested with users in order to collect feedback on their intuitiveness, comfort and engagement level. A comparison between traditional voice-based interaction modalities with AI and the proposed gesture-based language was performed. Preliminary results are discussed, including the importance of cognitive metaphors in gesture-based interaction, the relation between intuitiveness, innovation, and engagement, and the advantages provided by gesture-based interactions in terms of privacy, subtleness, and pleasantness, versus the limited options and the need to learn a codified language. Insights will help designers in the development of seamless interactions with AI agents for ambient intelligent systems.

Keywords: Ambient UX · Tangible interactions · AI agent · Hybrid materials

1 Introduction

Ambient Intelligence (AmI) [1] system technologies are sensitive, responsive, adaptive, transparent, ubiquitous, and context-aware environments. These systems can be embedded with artificial intelligent (AI) agents, which perform front-end communication with the user. We are observing the rise of different kinds of AI agents designed to communicate with humans through different languages [2].

Finding new languages to communicate with AI agents beyond voice interaction, leaning towards multimodal interactions that engage diverse senses, is a wide and prominent research area. We faced this topic in a project for a lighting AI agent. The study presented in this paper is part of a concluded project aimed at designing a 10-year future vision of a lighting AI agent, Phil. We crafted specific interaction modalities as the communication language between Phil and the user, based on touch inputs rather than the audio channel (used in current well-known personal AI assistants).

Touch-based interactions are perceived by users as natural and seamless [3]. Multi-touch interactions have been analyzed for diverse applications, also in terms of user performance and ergonomics, and design strategies have been proposed in this field [4].

Design strategies for gestural inputs also bring up considerations on interacting with diverse media and surfaces, like, for example, on-body interactions [5].

The aim of this explorative work is to investigate the potential of gesture-based interfaces as an alternative to voice interactions in the design of AI agents. We tested our proposed communication language with a group of potential users. We discuss users' responses and provide preliminary insights that can guide further research on the design of new types of communications with AI agents.

1.1 Phil: The Lighting AI Agent

Phil is an artificial intelligent agent controlling ambient lighting systems. By connecting with light sources, it communicates with users and assists them in diverse outdoor and indoor environments, both private and public, supporting activities at home, at work, in leisure, and on the move. Phil is able to understand the context in which the user is, by accessing their biometrics, calendar, as well as by understanding the social dynamics taking place. It is capable of learning user's needs and daily routines, their preferences and interests. After learning, Phil provides the user with desired ambiance lighting, lighting social triggers, wayfinding, and notifications.

We designed a tangible language for the interaction with Phil based on gestural inputs and visual outputs. In our vision, such interactions will be allowed by the development of smart materials enabling projected capacitive touch sensitive surfaces [6]. Such materials can be embedded into everyday surfaces such as furniture, walls, as well as clothing, thanks to electroactive fabrics [7] enabling seamless interactions.

Interaction Modalities. In this paper, we describe and assess six interaction modalities with Phil: four user's gestural inputs, and two visual communications to the user. The interaction language is based on a circular shape, which represents Phil. The user can play with it in different ways, in order to input specific commands. Following are the four gestural interactions and two visual communications (Fig. 1):

1. *Calling Phil*, i.e. activating him by drawing a circle with one finger. After drawing a circle, a visual feedback would show that Phil is active in that particular ambient.
2. *Dismissing a notification* by touching a surface with one finger and swiping the visual circle. Notifications consist of personal messages such as "remember to take your medicine", conveyed through ambient light changes. In this case, the person just wants to confirm the reception of a message, but does not require any further action from Phil.
3. Confirming the reception of a notification and *telling Phil to execute a connected activity*, e.g. saving a suggested event to the calendar. In order to do this, the user engages two fingers on the surface. While one is stable on the visual dot, the other finger starts to draw a larger circle. When the large circle closes by making a full round, it indicates that Phil has finalized the required activity.
4. *Turning Phil off*, i.e. deactivating him in a particular ambiance. This interaction is performed by using three hands and swiping the visual circle towards the users' body. The visual would gradually decrease in intensity and finally disappear from the surface.

5. *Phil showing its presence* to the user by two circles that change sizes and rotate. The same movement is used to indicate that Phil is working in the background, this time by making the two circles spinning faster.
6. *Phil providing spatial directions* for users, when guiding them in the environment. In this case, a number of circles move dynamically towards the intended direction.
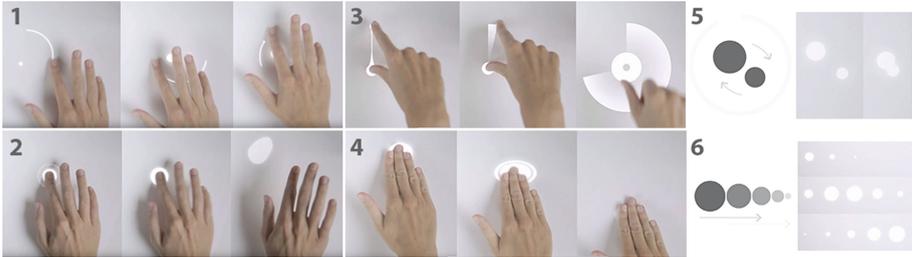

Fig. 1. Six interaction modalities, four based on hand gesture input and two on visual feedback.

2   User Tests

In order to evaluate the proposed tangible interactions, we conducted a user study involving 12 participants (7 females), in a setting within our research lab. Participants' age ranges between 19 and 35, and they all work or study in design and/or engineering fields. They have a high level of tech-savviness (average of 6 on a 7-point Likert scale) and 75% were familiar with AI agents (e.g. Google Home, Alexa, Siri).

Participants were asked to sign a consent and were provided with an overview of the design vision and the elements of the Phil system. As part of the introduction, two video use cases were shown, in order to give users examples of everyday contexts for the interaction with Phil. In one video, a 20-year-old student was guided by Phil in social interactions with her peers, within a co-working environment. The second video showed a young mother being supported and motivated by Phil in her daily workout activities.

After the introduction, users were shown six tutorial videos (see Fig. 1) describing the proposed interaction modalities. After each video, they were asked to repeat each gesture and to assess its intuitiveness, easiness to perform, and level of fun/engagement on a 5-point Likert scale. Additional qualitative questions were asked on each response. For each interaction, we also invited participants to propose alternatives and to indicate if and where they would perform the same interaction on the body (i.e. through smart clothing). Finally, participants were asked if they would prefer interacting with AI agents through voice or gestures.

3   Results

Intuitiveness. Most participants interpreted the set of interactions as intuitive (Table 1), particularly when they could associate them with *cognitive metaphors* or *well-known*

interaction modalities with existing gestural systems (e.g. gestures on the Mac touchpad). For instance, gesture 4 was metaphorically associated to the movement performed on the light switch to turn the light off. Similarly, a participant stated that gesture 4 was intuitive because the downward movement is usually associated to the idea of "reducing" or "turning off". Relying on cognitive metaphors helped participants to interpret the gestures correctly. Users assessed as neutral or little intuitive interactions that lacked any reference to existing systems or that were associated to metaphors not consistent with the gesture meaning (e.g. gesture 1 was associated to controlling the volume by one user, who gave a score of 3). Even when assessed less than 5, all interactions but gesture 3 were described as easy to learn even if not immediately understandable.

Table 1. Average scores for each of the six interaction modalities according to their Intuitiveness (I), Comfort (C), and Engagement (E), based on the 5-point Likert scale.

Interaction modalities	1	2	3	4	5	6
I	3.92	4.25	3.25	4.17	4.25	4.65
C	4.67	4.67	4.00	4.75	/	/
E	4.00	4.25	3.67	3.92	4.5	5.00

Comfort. Participants were asked to repeat the gestures and to assess how easy they were to perform. Gesture 3 was assessed as the less comfortable one (with an average score of 4, compared to the others assessed 4.67 and higher). A negative comment regarded the need to use two fingers asymmetrically, which made the participant "impatient". Another participant complained about the need to twist the wrist, which made the gesture unnatural and difficult to perform.

Engagement. The two visual communications were evaluated as very fun, pleasant and engaging, with 100% of participants strongly agreeing on communication 6 being fun/engaging. Gestures received lower scores, with gesture 3 being the less engaging one. Surprisingly, the more intuitive the interaction, the more engaging it was assessed (e.g. in interactions 6, 5, and 2). Similarly, less intuitive interactions were also assessed as the less engaging ones (i.e. gesture 1 and 3). Users U3 U5 and U10, who assessed gesture 3 as little intuitive (with a score of 3), also gave low scores on engagement (3), even if comfort received the highest score (5) in all three cases.

These results may suggest that engagement is more influenced by intuitiveness than comfort when it comes to a new interactions language. Indeed, higher levels of engagement and pleasure might depend on reduced cognitive load instead of the easiness to perform a movement, at least when approaching a new set of gestures for the first time. This outcome is also confirmed by the fact that, during the interviews, gestures assessed as fun were connected to the use of well-known gestures: "it's like saying 'nope, don't like you' on a dating app" (U2 commenting gesture 2, which was rated 5 in engagement).

Alternative Interactions. When asked to suggest alternative interactions to the ones experienced, participants heavily relied on the use of cognitive metaphors. A user proposed a "double tap almost like knocking" to call Phil. As an alternative to gesture

3, one participant suggested wrapping the hand "as to tell I'm closing it", or "I'm saving this information". Some participants proposed gestures more similar to human interactions. For instance, one respondent suggested a "soft touch holding for some seconds to say 'goodnight'" as an alternative to gesture 4. Another stated "I would like to use the whole hand as to closing or opening something, to make the intelligence appear or disappear."

Bodily Interactions. Participants were asked to indicate if they would use on-body interactions in alternative to the proposed on-surface ones. Certain areas kept repeating as the most natural or convenient ones (see Fig. 2), such as palm (voted 26 times), upper wrist (voted 21 times), thigh (voted 18 times), and forearm (voted 13 times). For gestural interactions, the most convenient three areas were palm, thigh and outer wrist, while for visual communications, the three highest ranked areas were palm, outer wrist and forearm. These three areas are indeed in the periphery of the visual field, but still visible. Seven participants observed that they would prefer having a private gestural communication with Phil and not having other people in the surrounding seeing it. The majority of them chose to interact with the palm for this particular reason.
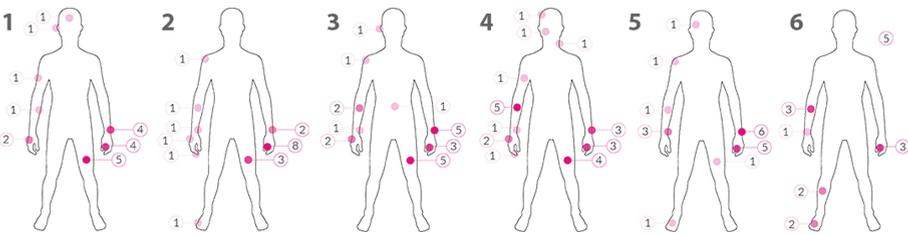

Fig. 2. On-body zones indicated by the study participants for the six interaction modalities; numbers in circles indicate how many times the same zone repeated.

3.1 Advantages of Gestural Interactions

67% of respondents stated they would use a gestural interaction system in public spaces, because it provides more privacy. "Tactile has limitation but it's better for public spaces". "I was suggesting my thigh because no one can notice what I'm doing". Tangible interactions also eliminate the issue of AI listening to private conversations: "I like that it is tangible, a vocal interaction means it always has to listen…".

Voice interactions were described as not suitable for public spaces also because in such environments people do not want to disturb others. As opposite to voice-based AI, Phil was assessed as "less annoying and invasive, also in public contexts". Gestural interactions, being a codified language, were evaluated as a more efficient way to interact with an AI agent than vocal language. "The tangible interface is way more efficient […] the AI doesn't have to interpret what you are saying, so less possibility for misunderstanding, because each input corresponds to a specific output."

3.2 Disadvantages of Gestural Interactions

Compared to voice-based interactions, gestural interfaces were assessed as more effortful: "when I'm home I want to be lazy and this [making a gesture] takes energy". The need of touching surfaces was assessed as potentially negative in certain contexts. A user stated: "there are considerations for dirty surfaces, I wouldn't touch […] the kitchen with dirty hands or outdoors, where billions of people touch the same surface." Gesture-based interactions are simplified forms of communication, and provide less flexibility compared to voice interactions: "With talking you can say so many things… tactile has limitation". However, in some cases simple gestures were considered a preferable option: "If it's an interaction that contains few steps/commands, then I would rather have a conversation; if it's one step, touch is good."

4 Conclusion and Future Work

Gestural interfaces do not provide the same advantages of text and voice communication, because they require users to learn a new codified language. However, it emerged that the use of cognitive metaphors can increase intuitiveness and engagement. Though reference to existing interactions (e.g. the ones performed on the phone) was appreciated, users also pointed out the need for novel interaction modalities when communicating with an AI agent. Users much appreciated simple and easy-to-remember gestures. When more than one finger was involved, symmetry was preferred (as in twisting or pinching). Tangible interactions were considered very suitable for office and public spaces, while voice was considered convenient at home. Privacy was seen as one of the main advantages of gestures, especially thanks to the use of on-body gestures, which give an extra layer of intimacy and make the interaction less noticeable.

We tested a limited number of interaction modalities, therefore the preliminary insights on the use of gesture-based interactions with AI agents cannot be considered comprehensive. Additional gestures and visual communications will be tested, especially based on the alternatives provided by users. More detailed analysis through statistical methods will be performed on the collected data, in order to confirm the correlation between elements such as intuitiveness and engagement. This work sets the basis for further investigations on new communication languages in the interaction with AI agents.

References

1. Cook, D.J., Augusto, J.C., Jakkula, V.R.: Ambient intelligence: technologies, applications, and opportunities. Pervasive Mob. Comput. **5**(4), 277–298 (2009)
2. Boukricha, H., Ipke W.: Modeling empathy for a virtual human: how, when and to what extent? In: The 10th International Conference on Autonomous Agents and Multiagent Systems, vol. 3, pp. 1135–1136. International Foundation for Autonomous Agents and Multiagent Systems (2011)

3. O'hara, K., Harper, R., Mentis, H., Sellen, A., Taylor, A.: On the naturalness of touchless: putting the "interaction" back into NUI. ACM Trans. Comput. Hum. Interact. (TOCHI) **20**(1), 5 (2013)
4. Ingram, A., Wang, X., Ribarsky, W.: Towards the establishment of a framework for intuitive multi-touch interaction design. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 66–73. ACM (2012)
5. Mujibiya, A., Cao, X., Tan, D.S., Morris, D., Patel, S.N., Rekimoto, J.: The sound of touch: on-body touch and gesture sensing based on transdermal ultrasound propagation. In: Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces, pp. 189–198. ACM (2013)
6. Barrett, G., Omote, R.: Projected-capacitive touch technology. Inf. Disp. **26**(3), 16–21 (2010)
7. Syduzzaman, M., Patwary, S.U., Farhana, K., Ahmed, S.: Smart textiles and nano-technology: a general overview. J. Text. Sci. Eng **5**, 1000181 (2015)