Citation: Tejwani, Ravi, Katz, Boris and Breazeal, Cynthia. 2021. "Migratable AI: Investigating Users' Affect on Identity and Information Migration of a Conversational AI Agent." 13086.

As Published: 10.1007/978-3-030-90525-5_22

Publisher: Springer International Publishing

Persistent URL: https://hdl.handle.net/1721.1/146058

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Massachusetts Institute of Technology

Migratable AI : Investigating users' affect on Identity and Information migration of a Conversational AI agent

Ravi Tejwani[1], Boris Katz[1] and Cynthia Breazeal[1]

Massachusetts Institute of Technology, Cambridge, USA
{tejwanir,boris,breazeal}@mit.edu

Abstract. Conversational AI agents are becoming ubiquitous and provide assistance to us in our everyday activities. In recent years, researchers have explored the migration of these agents across different embodiments in order to maintain the continuity of the task and improve user experience. In this paper, we investigate user's affective responses in different configurations of the migration parameters. We present a 2x2 between-subjects study in a task-based scenario using information migration and identity migration as parameters. We outline the affect processing pipeline from the video footage collected during the study and report user's responses in each condition. Our results show that users reported highest joy and were most surprised when both the information and identity was migrated; and reported most anger when the information was migrated without the identity of their agent.

Keywords: conversational ai · affective computing · agent migration

1 Introduction

We are surrounded by conversational AI agents, such as Alexa [1], Jibo [11] or Google Home [7], as they assist us in our daily activities like providing weather and news updates, ordering meal and ride shares, setting room temperature etc. These agents build model of our personal preferences and interests as we interact and develop relationship with them. We also interact with the robotic agents in public setting, such as Pepper [21], Kuri [20], or Moxi [19] at hospitals, restaurants, and grocery stores, where we share our preferences with them. Since these agents exist in different form factors or embodiments and setting, they do not always share information amongst each other. However, the migration of information or identity across embodiments could lead to changes in users' perception [26] and affective states. For example, after your interaction with the home agent, Alexa, you might be surprised to see if the restaurant robot greets you with your name and knows about your food order when you enter the restaurant.

Agent migration is a concept which allows an agent to disembody from its current form and migrate to different embodiments while maintaining the relationship with the user. Prior work has explored the concept of agent migration

through various different architectures[9] [22] [2] [5]. They explored the migration in the form of a synthetic character or a visual entity than compared to a conversational AI agent. Further, several user studies on agent migration [18] [13] [8] explored users impression on the agent such as validating if the users perceived that it was the same agent in another embodiment, or if the users understood the concept of agent migration. For instance, Syrdal et. al. performed series of group discussions with a school class children, aged 3 to 6, on evaluating children's impressions on the understanding of the concept of migration[25]. However, the affective behavior analysis of the users in the context of migration of AI agents have not been studied before. User's affective behavior and autonoumous reactions provide a deeper understanding of user's reaction towards the system in comparison to the subjective reports given by the users [17] [6] [24] which would be beneficial in designing effective migratable AI agents.

In our previous work [26], we proposed a Migratable AI system which allows a conversational AI agent to migrate across different physical embodiments. We measured the user's perception on trust, competence, likability and social presence using information migration and identity migration as parameters. In this paper, we build upon our previous work by analyzing the affective behavior of the users during the migration of the conversational AI agent. We ran a 2x2 between-subjects study in a task-based scenario using information migration and identity migration as parameters to investigate the affective responses of the users. We outline the affect processing pipeline from the video footage collected during our experimental study. The pipeline comprised of two stages: affect detection and affect interpretation. The findings from this paper, can be used for the further development of effective migratable systems.

2   Related Work

2.1   Identity migration

Prior work has explored the concept of agent migration through various different identity migration architectures [2] [22] [5]. The agent migration was first explored by Imai et al. [9] by demonstrating a tour guide application where a personal agent could migrate from mobile device to a physical robot. Later, the research by Martin et al. [14] explored that the identity is not just "Who am I?" but "Who am I in the eyes of others?" where they proposed the visual identity cues in their experiment that characters share a common feature - such as a hat or glasses, common colour scheme, common set of markings, or characters are of the same class of objects. Further, [18] [8] explored the research questions such as "Do participants feel that they are interacting with the same agent across different embodiments?"; "What are the most important aspects of an agent to communicate identity retention?"; or "Do users perceive the agents in different embodiments as the same entity?"

2.2   Information migration

Information migration architectures were explored in [22] [16] as generic memory models and persistent memory models. Aylett et al. proposed CMION in [2] [22], an open source architecture comprised of three layers (Mind, Mind-Body, and Body), that served as a framework for the bidirectional mapping of information to different levels of abstraction (i.e., from raw sensory data to symbolic data and vice versa). These models were created to focus on the following three different aspects:

1. **Scope** - Short term memory (STM) was modeled computationally to maintain a companion's current focus and Long term memory (LTM) was used for the artificial companions that interact with human users over a long period of time
2. **Efficiency** - how to optimize the storage and recall of memory contents; forgetting through the processes of generalization and memory restructuring.
3. **Adaptability** - how to use different conversational strategies for information or memory that the robot remembers during the interaction with human (no-memory, partial memory, complete memory).

2.3   User Perception of Agent Migration

User studies have explored users' perception of agents that can migrate across forms in [13] [18] [8] by studying the higher level users' impression on the agent such as validating if the users felt that it was the same agent in another embodiment, or if the users understood the concept of agent migration. Further, [12] [5] explored the users perception on the long term interaction derived from the companion's interaction history both with the environment and the user.

In this paper, we go beyond the users perception on agent's identity or the subjective reports and investigate user's affective state during the migration of an conversational AI agent using information migration and identity migration as parameters.

Table 1: Participant Demographics

Condition	Female	Male	Other	Age(Std. Dev.)
(INF+,ID+)	8	10	0	24.4(5.06)
(INF+,ID-)	9	9	0	24.6(6.09)
(INF-,ID+)	7	10	1	28.2(10.2)
(INF-,ID-)	8	9	1	22.6(3.61)

3   Method

3.1   Participants

We recruited 72 participants from Cambridge area using email advertisements. Participants were between 18 and 54 years old with mean age M=24.2, SD=5.09. Participants were randomly assigned and counterbalanced by gender across the four conditions (n=18 per condition) as described in Table 1. The study was approved by our Institutional Review Board, and participants signed an informed consent form prior to the study.

3.2   Study Protocol

We ran a 2×2 between-subjects study with **Information migration $\times$ Identity migration**. The 4 conditions used in the study are described in Figure 1.
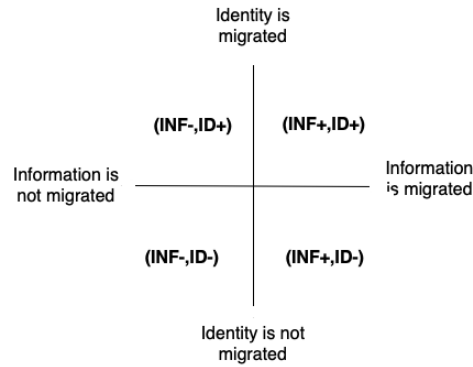

Fig. 1: Study conditions

We used the Migratable AI system [26] in the study. The system allowed the conversational AI agent to migrate across different embodiments by preserving its identity(identity migration) and/or remembering the information context (information migration). Each participant began the study in our lab's study room, modeled as their "home", with the home agent (Alexa) [1]. The home agent delivered the participant's schedule for the day which included a job interview. Throughout the conversation, the home agent learned about the participant such as his/her name and feelings.

Fig. 2: Left to right: Home agent, Home agent migrating to receptionist robot, Home agent migrating to waiting room assistant.

The mobile robot (Kuri) [20] was located in a hallway of the lab which played the role of front desk receptionist robot at the interview location. The receptionist robot, changed its appearance to look and sound like home agent (when identity was migrated) or continued to look and sound like Kuri with a different voice profile (when identity was not migrated). The receptionist robot detected their face, recognized the participant by name, and acknowledged the reason for their visit (when information was migrated) or prompted the participant for their name and reason for their visit (when information was not migrated). During the conversation, the receptionist robot either validated the participant's feelings (when information was migrated) or asked how they were feeling for their interview (when information was not migrated). The receptionist robot also learned the participant's drink preferences (coffee, water or tea) and escorted the participant to the interview waiting area.

At the interview waiting area, the participant interacted with the waiting room assistant (Smart TV) which conversed with the participant until the arrival of the interviewer. It changed its appearance to look and sound like home agent (when the identity was migrated) or continued to look and sound like itself (when the identity was not migrated). While the participant waited, it offered the participant their preferred drink (which it remembered in the condition when the information was migrated) or offered the participant a drink while waiting (when the information was not migrated). It also acknowledged the participant's feelings (when the information was migrated) and wished them good luck before the interviewer arrived. The interviewer role was enacted by the experimenter.

For identity migration - the design decisions were informed from the past literature on what helps users perceive an identity of an agent [2] [14] [4]. In the identity migration conditions, the same visual characteristics (Figure 2, panda-esque circular appearance) and voice (Joanna TTS) was used across all embodiments to convey identity continuity.

For information migration - the information parameters such as the person's name, feelings about the interview, drink preference and reason for visit were learned by each agent during the conversation. If the system was configured to

migrate information across embodiments, this information was shared amongst the agents to maintain the continuity of the interaction else the agent had to prompt the user for the information. The number of conversational turns (four in this user study) touch basing the personal and non-personal information between the agent and the participant were kept consistent across all the conditions. This might necessitate the agent to repeat certain questions to the users when the information was not migrated but it was to ensure that we do not create a bias in the study and keep the conversational turns consistent.
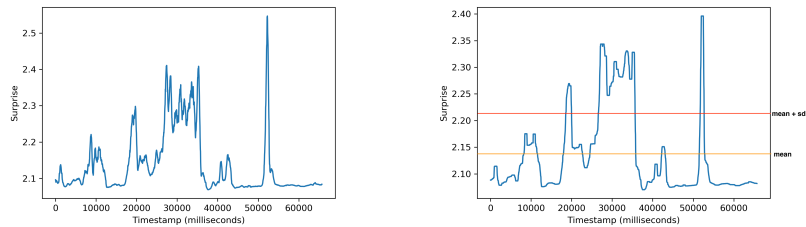
(a) Raw affect data of a participant

(b) Affect data after applying median filter and threshold

Fig. 3: Affect interpretation for surprise of a sample participant

3.3   Data Collection

A front facing USB camera were connected to a Raspberry Pi and attached to each embodiment to record the interaction. The Raspberry Pi used face detection to send a wake up signal to the robot/embodiment. It began recording the video when the participant's face was detected in front of the embodiment and stopped the recording when the interaction with the user ended. The video recordings of all the 72 participants (18 in each condition) were processed using Affdex [15] for affect analysis.

4   Affect processing pipeline

Affect analysis has been performed in the past using several statistical heuristics such as mean value of the pertaining window [10], if at any given point in

Table 2: Affect features by each migration condition

Features	(INF+,ID+)	(INF+,ID-)	(INF-,ID+)	(INF-,ID-)
joy	0.274 ± 0.03	0.188 ± 0.01	0.256 ± 0.02	0.239 ± 0.02
anger	0.179 ± 0.01	0.216 ± 0.02	0.151 ± 0.01	0.157 ± 0.02
surprise	0.181 ± 0.01	0.144 ± 0.02	0.139 ± 0.02	0.157 ± 0.02
smile	0.254 ± 0.03	0.274 ± 0.04	0.306 ± 0.03	0.187 ± 0.02
brow_raise	0.226 ± 0.02	0.188 ± 0.01	0.261 ± 0.02	0.231 ± 0.02
brow_furrow	0.201 ± 0.02	0.201 ± 0.01	0.252 ± 0.03	0.269 ± 0.04
nose_wrinkle	0.258 ± 0.03	0.222 ± 0.02	0.283 ± 0.03	0.298 ± 0.04
upper_lip_raise	0.230 ± 0.02	0.219 ± 0.02	0.219 ± 0.01	0.215 ± 0.02
mouth_open	0.226 ± 0.01	0.205 ± 0.02	0.257 ± 0.03	0.229 ± 0.02
eye_closure	0.214 ± 0.01	0.25 ± 0.02	0.253 ± 0.02	0.289 ± 0.03
cheek_raise	0.204 ± 0.04	0.162 ± 0.02	0.315 ± 0.03	0.278 ± 0.04

ID+ or ID- represents identity is migrated or not migrated.
INF+ or INF- represents information is migrated or not migrated.

the window the value of the metric exceeds a given threshold [27], or if the mean value of the metric over pertaining window exceeds a given threshold [3]. We implemented the pipeline for affect detection and interpretation using smoothing and a threshold technique. Most of our pipeline overlaps with the approach detailed by Spaulding and Breazeal [24] and D'Mello, Kappas, and Gratch [6].

The data pipeline is as follows: Raw data, $RD = rd_0$, rd_1, ..., rd_n, is used for an interaction time window W, where $|RD| >> W$ and rd_x is x participant's raw data. The raw data, RD, is further processed by an affect detector which produces feature vectors of metrics, $M = m_0$, m_1, ..., m_n (e.g., the degree to which 'joy' 'smile', 'brow raise' etc. are expressed), for each data point. The affect interpreter further analyzes these metric vectors for the time window and produces a feature label, l, for that window.

4.1 Affect Detection

The facial expressions of the participants were evaluated from the video captured by the front facing USB camera mounted on each of the embodiments at 30fps. The camera would get activated at the detection of the participant's face and record the video for the time frame of the interaction between the agent and the participant at each embodiment (Figure 2). We processed the frames from each embodiment, using the Affdex [15] as the affect detector, which detected features such as: *'joy', 'anger', 'surprise', 'smile', 'brow_raise', 'brow_furrow', 'nose_wrinkle', 'upper_lip_raise', 'mouth_open', 'eye_closure', 'cheek_raise'*.
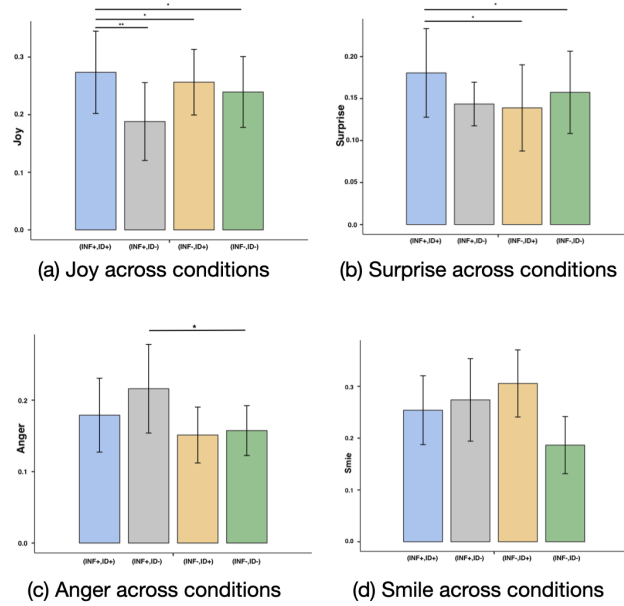
Fig. 4: Box-plot for the normalized affective measures across conditions. The boundary of the box closest to zero indicates the 25th percentile, the line within the box marks the mean and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the box indicate the 10th and 90th percentiles. * means p<.05, ** means p<.01

4.2   Affect Interpretation

The affect data for the interaction duration between the agent and participant on an embodiment, is collected and converted to a feature vector. We interpret each feature as a binary indicator variable whose value is determined by smoothing and a threshold technique. The raw affect data, M, is initially passed through a median-filter smoothing. Further, if the maximum value of the median-smoothed detected peaks exceeds the threshold, then the feature value is interpreted with an indicator value of 1. For the given time window, the interpreted affect feature vector is comprised of set of observed feature indicators. The threshold value for the feature is set at the mean value of the feature across the entire time window plus a standard deviation (Figure 3). Finally, each of the affect feature score is further normalized for the data analysis.

5   Results

Normality was first checked for the affective measures from the visual inspection of Q-Q plots and Shapiro-Wilk's test. With all p-values < 0.05, the Shapiro-Wilk test rejects the null hypothesis of data normality, hence, we perform the Kruskal-Wallis H test over the data. Furthermore, the Dwass-Steel-Critchlow-Fligner test was used for the pair-wise comparisons. All the statistical analysis were performed using R(version 3.6.1) and Jamovi[23].

We found a significant effect in joy, surprise and anger amongst the users when configuration of the identity and information of the agent was changed during the migration.

Joy: There was a statistically significant difference in joy scores across different conditions, $\chi^2(3) = 7.560$, p = 0.016. The pair wise comparisons showed that mean joy score of 0.274 ± 0.03 for (INF+, ID+) was significantly greater than 0.188 ± 0.01 for (INF+,ID-) with p=0.008, 0.256 ± 0.02 for (INF-,ID+) with p=0.024 and 0.239 ± 0.02 for (INF-,ID-) with p=0.039 (Figure **??**).

Surprise: There was a statistically significant difference in surprise scores across different conditions, $\chi^2(3) = 4.40$, p = 0.033. The pair wise comparisons showed that mean surprise score of 0.181 ± 0.01 for (INF+, ID+) was significantly greater than 0.139 ± 0.02 for (INF-,ID+) with p=0.036 and 0.157 ± 0.02 for (INF-,ID-) with p=0.042 (Figure **??**).

Anger: There was a statistically significant difference in anger scores across different conditions, $\chi^2(3) = 3.157$, p = 0.041. The pair wise comparisons showed that mean anger score of 0.216 ± 0.02 for (INF+,ID-) was significantly greater than 0.157 ± 0.02 for (INF-,ID-) with p=0.032 (Figure **??**).

Other affective measures: The analysis results for the other affective scores were not significantly different across the different conditions: smile ($\chi^2(3)$ = 2.986, p = 0.091), brow_raise ($\chi^2(3)$ = 2.602, p = 0.080), brow_furrow ($\chi^2(3)$ = 1.391 p = 0.842), nose_wrinkle ($\chi^2(3)$ = 1.462, p = 0.924), upper_lip_raise ($\chi^2(3)$ = 2.753, p = 0.178), mouth_open ($\chi^2(3)$ = 1.115, p = 0.273), eye_closure ($\chi^2(3)$ = 1.394, p = 0.307) and cheek_raise ($\chi^2(3)$ = 2.171, p = 0.092).

Table 2 summarizes the mean scores with their standard deviation for all the affective features across all the conditions.

6   Discussion and Conclusions

We presented the results from one of the first systematic investigations of users affective behavior on the migration of the conversational AI agent. We ran a 2x2 between-subjects study in a task-based scenario with 72 participants using information migration and identity migration as parameters to investigate the affective behavior of the users. We outlined an affect processing pipeline from the video footage collected during the study.

We inferred that users expressed most joyfulness and surprise when they saw their agent in a different embodiment and the agent remembered their preferences and context **(both the information and identity of the agent was**

migrated). This was corroborated by the participant's comments during the post-study interview. Participant P21 said *"I think it remembers that I am anxious about the interview. That means it cares about me and makes it different from, for example, a coffee machine."* Another participant, P34, said, *"Being familiar with Alexa, allowed me to trust Receptionist and TV agent"*.

The users were most disappointed and angry when they found out that their information was shared with different agents **(information was migrated but identity was not migrated)**. P51 said *"I did not trust the agents well because they seemed to share all of the information about me, and I did not want to disclose more."*. Also, P39 said *"... especially after the receptionist agent knew what I told Alexa, I no longer trusted Alexa"*. The insights gained from this research could be used for further development of affective migratable systems.

References

1. Alexa, A.: Alexa (2014), `https://developer.amazon.com/alexa`
2. Aylett, R., Kriegel, M., Cuba, P., Vala, M., Paiva, A., Koay, K., Arent, K., Márquez Segura, E., Nylander, S., Cramer, H., et al.: Body–hopping: Migrating artificial intelligent agents between embodiments
3. Bernin, A., Müller, L., Ghose, S., von Luck, K., Grecos, C., Wang, Q., Vogt, F.: Towards more robust automatic facial expression recognition in smart environments. In: Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments. pp. 37–44 (2017)
4. Cuba, P.: Agent migration between bodies and platforms (2010)
5. Duffy, B.R., O'Hare, G.M., Martin, A.N., Bradley, J.F., Schon, B.: Agent chameleons: Agent minds and bodies. In: Proceedings 11th IEEE International Workshop on Program Comprehension. pp. 118–125. IEEE (2003)
6. D'Mello, S., Kappas, A., Gratch, J.: The affective computing approach to affect measurement. Emotion Review **10**(2), 174–183 (2018)
7. Google: Google Home (2013), `https://store.google.com/gb/product/google_home`
8. Grigore, E.C., Pereira, A., Yang, J.J., Zhou, I., Wang, D., Scassellati, B.: Comparing ways to trigger migration between a robot and a virtually embodied character. In: International Conference on Social Robotics. pp. 839–849. Springer (2016)
9. Imai, M., Ono, T., Etani, T.: Agent migration: communications between a human and robot. In: IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028). vol. 4, pp. 1044–1048. IEEE (1999)
10. Jeong, S., Breazeal, C.L.: Improving smartphone users' affect and wellbeing with personalized positive psychology interventions. In: Proceedings of the Fourth International Conference on Human Agent Interaction. pp. 131–137 (2016)
11. Jibo: Jibo (2017), `https://www.jibo.com`
12. Kriegel, M., Aylett, R., Cuba, P., Vala, M., Paiva, A.: Robots meet ivas: a mind-body interface for migrating artificial intelligent agents. In: International Workshop on Intelligent Virtual Agents. pp. 282–295. Springer (2011)
13. Kriegel, M., Aylett, R., Koay, K.L., Casse, K., Dautenhahn, K., Cuba, P., Arent, K.: Digital body hopping-migrating artificial companions. Proc. of Digital Futures' 10 (2010)

14. Martin, A., O'hare, G.M., Duffy, B.R., Schön, B., Bradley, J.F.: Maintaining the identity of dynamically embodied agents. In: International Workshop on Intelligent Virtual Agents. pp. 454–465. Springer (2005)
15. McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., Kaliouby, R.e.: Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit. In: Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems. pp. 3723–3726 (2016)
16. Ono, T., Imai, M., Nakatsu, R.: Reading a robot's mind: A model of utterance understanding based on the theory of mind mechanism. Advanced Robotics **14**(4), 311–326 (2000)
17. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence: Analysis of affective physiological state. IEEE transactions on pattern analysis and machine intelligence **23**(10), 1175–1191 (2001)
18. Ready, P.C.: Migration between two embodiments of an artificial pet
19. Robotics, D.: Moxi (2013), `https://diligentrobots.com/moxi`
20. Robotics, M.: Kuri (2018), `https://www.heykuri.com/explore-kuri/`
21. Robotics, S.: Pepper (2015), `https://www.softbankrobotics.com/us/pepper`
22. with Robots, L., Companions, I.: Lirec (2009), `http://lirec.eu`
23. Source, O.: Jamovi (2018), `https://www.jamovi.org/`
24. Spaulding, S., Breazeal, C.: Frustratingly easy personalization for real-time affect interpretation of facial expression. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). pp. 531–537. IEEE (2019)
25. Syrdal, D.S., Koay, K.L., Walters, M.L., Dautenhahn, K.: The boy-robot should bark!-children's impressions of agent migration into diverse embodiments. In: Proceedings: New Frontiers of Human-Robot Interaction, a symposium at AISB (2009)
26. Tejwani, R., Moreno, F., Jeong, S., Park, H.W., Breazeal, C.: Migratable ai: Effect of identity and information migration on users' perception of conversational ai agents. In: 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). pp. 877–884. IEEE
27. Zhang, L., Tjondronegoro, D., Chandran, V., Eggink, J.: Towards robust automatic affective classification of images using facial expressions for practical applications. Multimedia Tools and Applications **75**(8), 4669–4695 (2016)