https://doi.org/10.1145/3652988.3673963

A Virtual Agent as a Commensal Companion

Maurizio Mancini
Sapienza University of Rome
Rome, Italy
m.mancini@di.uniroma1.it

Radoslaw Niewiadomski
DIBRIS, University of Genoa
Genoa, Italy
radoslaw.niewiadomski@dibris.unige.it

Gabriele De Lucia
Sapienza University of Rome
Rome, Italy
delucia.1811284@studenti.uniroma1.it

Francesco Maria Longobardi
Sapienza University of Rome
Rome, Italy
longobardi.1941987@studenti.uniroma1.it

ABSTRACT

Previous work introduced the concept of artificial commensal companions, i.e., embodied agents capable of interacting with humans during meals. They are supposed to bring the benefits of eating together in settings where a human would be forced to eat alone (e.g., elderly, hospitalized patients, self-isolation, etc.). This paper presents an experiment with a virtual agent and a human eating together. We invited volunteers to bring a small meal and let them chat briefly with the agent, simulating eating behaviors during the conversation. After the experience, participants filled out a questionnaire, providing quantitative and qualitative feedback. While results are encouraging (i.e., participants showed interest in eating with an agent), further work is still needed to provide more convincing results.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*.

KEYWORDS

commensality, companion, activity recognition, virtual agent, interaction design

ACM Reference Format:
Maurizio Mancini, Radoslaw Niewiadomski, Gabriele De Lucia, and Francesco Maria Longobardi. 2024. A Virtual Agent as a Commensal Companion. In *ACM International Conference on Intelligent Virtual Agents (IVA '24), September 16–19, 2024, Glasgow, United Kingdom.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3652988.3673963

1 INTRODUCTION

Commensality, the act of eating together, is one of the most frequent human multimodal interactive experiences. It is a social activity, as the participants share food while chatting, discussing, and, often, building or strengthening relationships. The nonverbal behavior of the commensal partners is very rich, while they share attention between food consumption and engagement in conversation with others. Whether it's a business lunch, a romantic dinner, or a meal in a noisy student canteen, gatherings around the table represent a captivating reservoir of multimodal interactions enriched with cultural subtleties and social norms.

There is an increasing interest in technology enhancing, facilitating, or enabling the commensal experience [17, 20, 23]. Recently, Niewiadomski and colleagues introduced the concept of the Artificial Commensal Companion (ACC), an autonomous and "socially intelligent agent designed to interact verbally and nonverbally with humans during mealtime" [16]. Such companions should be able to recognize human activities (e.g., whether the human partner is eating, drinking, and so on), as well as objects on the table (such as plates, types of foods), and be able to perform engaging interaction with them. They are expected to provide human users some of the benefits of traditional human-human commensality. Indeed, eating alone is considered one of the most important factors of unhappiness in developed societies [22]. It is believed, however, that to obtain such a positive impact on human well-being, the role of ACCs should not be reduced to just a dietary coach (e.g., [8]) or physical (e.g., feeding) assistant ([19]), but ACCs need to become social partners aware of cultural and social norms at the table [16]. Some attempts have been proposed for ACCs, mainly using social robots [7, 10, 12, 16]. At the same time, more rare are virtual systems with two notable examples: the virtual co-eating system [24] and the virtual eating companion [13]. The first features a character displayed on a mesh fabric, embedded with a facial expression recognition module, able to make simple conversations. The second is an active listener that facilitates the generation of new ideas, with its eating behavior modeled based on quantitative analysis of human dining behavior.

In the first part of the paper, we present our first realization of such an artificial commensal companion. It is based on a virtual character developed in Unreal that can display verbal and nonverbal behavior. It includes a component responsible for tracking and detecting human activities by analyzing facial expressions. The application is capable of real-time interaction, thanks to a simple webcam that detects the human user and can be displayed on any screen, such as on a tablet or PC. The second part presents an experimental protocol for validating our companion. Five naive participants interacted with our companion while eating lunch. We developed a questionnaire to evaluate the participants' experience and collect suggestions on improving our application.

IVA '24, September 16–19, 2024, Glasgow, United Kingdom

ACM ISBN 979-8-4007-0625-7/24/09
https://doi.org/10.1145/3652988.3673963

2 ARCHITECTURE

Our prototype of ACC is loosely inspired by the SAIBA and SEMAINE architectures [1, 26]. Our embodied agent cannot understand human speech, and its behavior is generated by considering the results of the analysis of the user's facial expressions. The architecture is presented in Figure 1.

The User Behavior Detection module recognizes main commensal activities by analyzing the video stream from the webcam in front of the user. A simple Behavior Planner module then uses the detected labels to plan the agent's behavior. Finally, the Behavior Realizer displays the animation through a three-dimensional full-body virtual woman character (see Figure 2).

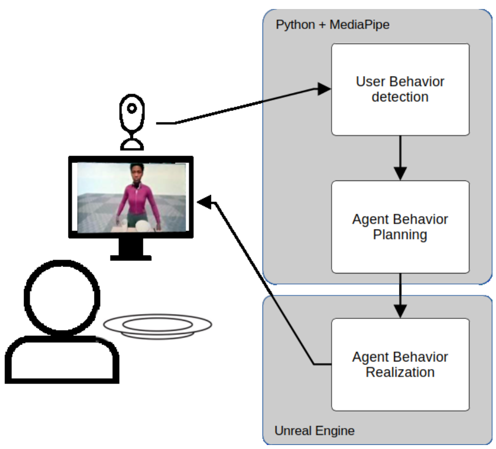

Figure 1: Overview of the Commensal Companion architecture: the user's activities (i.e., no activity, speaking, chewing, other) are detected; the agent's response behavior (i.e., speaking, eating, smiling) is planned, based on a set of pre-defined rules; the agent's behavior is realized and displayed.

2.1 User Behavior Detection

Some relevant works on activity recognition in commensality were recently proposed. Single activities, such as food intake or chewing, are detected with wearable or specifically designed devices, such as a smart fork [3, 6, 11]. [4, 21] detects when persons are eating using video data, and [9] proposes a bite and chews counter based on a pre-trained AlexNet network. Finally, [18] proposes an offline model that recognizes actions related to food consumption and social signals. Also, our model exclusively uses video data captured by a standard video webcam. Intentionally, we do not collect the audio data for privacy reasons, and recent studies show that speaker detection is possible from video data only [2]. Currently, our module distinguishes nearly in real-time between four classes: *no activity*, *speaking*, *chewing*, and a garbage class of behavior *other*.

We leverage MediaPipe [14], which allows us to extract facial features almost in real time. The input vector comprises 52 features extracted on 0.5 seconds-long segments (i.e., 15 video frames). Next, a fully connected 3-layer Multi-Layer Perceptron is employed with RELU, 50 epochs, and a batch size 64. The model was trained offline on the only freely available dataset for commensality [5] using training/testing split 80%/20%, achieving results up to 0.76 F-score (four-class problem with the fourth class being a garbage class).

2.2 Agent Behavior Realization

We rely on the Unreal Engine[25] for the agent behavior realization and, more specifically, the MetaHuman[15] plugin for rigging and skinning. At the moment, the agent shows behavior to initiate the conversation, displays some feedback to the speaking human user, displays smiles, and simulates food consumption. Regarding the latter, some previous prototypes consider such ability, e.g., [7, 12, 13].

More specifically, the following animations were created:

- *speaking 0*: the agent utters "hi, how are you doing?";
- *speaking 1*: the agent utters "yeah!";
- *speaking 2*: the agent utters "why?";
- *speaking 3*: the agent utters "and how?";
- *speaking 4*: the agent utters "really? tell me more about that!"
- *speaking 5*: the agent utters "let's talk about something positive happening to you at the moment!";
- *smiling*: the agent smiles for a few seconds;
- *eating*: the agent takes a spoon of food and chews it for a few seconds.

Figure 2: Control rig manipulation in Unreal Engine for creating commensal behavior animations with MetaHuman.

2.3 Agent Behavior Planning

The Behavior Planner uses a rules-based system to schedule actions for the Behavior Realizer. The agent's behavior depends on the detected human behaviors. For instance, the agent intakes food whenever it detects the human is eating, and it initiates conversation when it detects that the human is not speaking for a longer period. The commands are sent through UDP packets so the User Behavior Detection and Agent Behavior Planning modules can run on a machine while the Agent Behavior Realizer is on a separate machine. That will lower the computer load, as analyzing the user and generating the agent animation in real time are GPU-intensive tasks. The behavior planning rules are:

- if the interaction just started, after 5 seconds, send the *speaking 0* command;
- if the user was talking, then they stopped talking for 0.5 seconds, send the *speaking N* command, with $1 \leq N \leq 4$;

- if the user is not talking for 3 seconds, send the *speaking 5* command;
- if the user is chewing, send the *eating* command;
- if the user is smiling, send the *smiling* command.

3 EVALUATION

The evaluations of ACC prototypes are rare [7, 16, 24], and consequently, user expectations and opinions regarding this technology are poorly understood. The main aim of this preliminary evaluation is to collect suggestions for possible improvements and to guide the development of the next version through an iterative interaction design process.

3.1 Procedure

A convenience sample of five participants was invited into a lab on a specific day at lunchtime. They were instructed to bring a meal they liked to eat for lunch, like a pizza slice or a homemade salad. They were not provided any details about the experiment, i.e., they were unaware that they would eat with a virtual agent. When each participant arrived at the lab, they were provided two forms to sign. The first was a written consent to participate in the experiment, and the second was regarding GDPR. No personal data (e.g., demographics) was collected during the test.

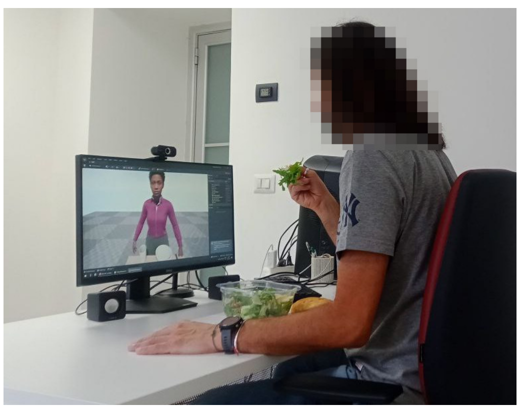

Figure 3: An example of a participant interacting with the commensal companion while eating.

Once the participants were ready to eat their meal the application was launched, and the experimenter left the room for 3 minutes, leaving the participant alone (see Figure 3). The agent is displayed on a screen in front of the participant, and on the top is a webcam used to track human behavior. Afterward, the experimenter re-entered the room and provided the ad hoc questionnaire to evaluate the user experience. The questionnaire is a combination of open (Q1-Q4, Q6-Q11), binary (Q5, Q18), and 5-point Likert scale questions (Q12-Q17). It can be accessed at DOI 10.17605/OSF.IO/FJ65Z.

3.2 Results

In the pilot study, we collected answers from five participants. The answers indicate that the participants favor the general idea of an artificial commensal companion; however, they promptly enumerated the shortcomings of the current prototype. This can be deduced from the answers to Q16 and Q17. Four out of five participants (avg $Q17 = 3.6$) would like to repeat the experience with a different ACC. At the same time, the answers to Q16 show no interest in repeating interaction with this specific system (avg $Q16 = 1.6$). In a similar vein, most participants would prefer to have the company of ACC than eat alone (Q5). The participants liked the general idea, visual quality, animation, and gestures (Q1, Q2, Q4). They described the experience as positive, weird, unusual, or fun (Q1). They mainly criticized verbal communication, i.e., the agent was repetitive and interrupting the human, and its verbal messages were incoherent with the human's verbal content (Q1, Q3, Q4). The participant felt comfortable with the agent (avg $Q12 = 3.4$), which was perceived as neither irritating (avg $Q15 = 2.6$) nor engaging (avg $Q14 = 1$). Finally, when asked how much they enjoyed eating with the ACC (Q13), the answers varied greatly, with an average of 2.4.

Due to the small number of participants, these results cannot be considered conclusive. More interesting are the comments about expectations and possible improvements of the agent. The participants indicate a need for more rich interaction and better synchronization with human interaction partners (e.g., distinguishing better moments when the human eats and speaks). Indeed, the negative remarks about agents' tendency to interrupt humans are probably caused by the insufficient generalization capabilities of the User Behavior Detection module. This appears to be the main problem, alongside the repetitiveness of behaviors. Interestingly, one of the participants declared they would prefer the agent not to ask questions but only answer their questions. Suggested applications include the elderly and individuals in solitude, socio-healthcare settings, and fast-food locals. The proposed risks primarily revolve around privacy concerns, tendencies toward self-isolation, and the abandonment of social habits.

4 CONCLUSIONS

In this paper, we have introduced an autonomous virtual agent acting as an Artificial Commensal Companion. The preliminary evaluations show several interesting insights into this system that will be considered in future versions of our system.

The obvious limitation of the current version consists of the number and variety of behaviors the agent can display. The participants' comments indicate shortcomings of the detection module in a real-life setting. The agent requires richer and more multimodal behavior, and the evaluation should be repeated with more participants, including longer and multiple interaction sessions with the same participant. Future steps beyond addressing the shortcomings above include testing different visual models (e.g., gender, sitting agent) and exploring various embodiments, including social robots. Nonetheless, the general comments regarding Artificial Commensal Companions are positive and encourage further research.

ACKNOWLEDGMENTS

This work is supported by the PRIN 2022 project COCOA (D53D23008850001), funded by the Next Generation EU (NGEU) Programme, National Recovery and Resilience Plan (PNRR), Italian Ministry of University and Research (MUR).

REFERENCES

[1] E. Bevacqua, S. Pammi, S. J. Hyniewska, M. Schröder, and C. Pelachaud. 2010. Multimodal Backchannels for Embodied Conversational Agents. In *Intelligent Virtual Agents*, J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 194–200.

[2] Cigdem Beyan, Muhammad Shahid, and Vittorio Murino. 2021. RealVAD: A Real-World Dataset and A Method for Voice Activity Detection by Body Motion Analysis. *IEEE Transactions on Multimedia* 23 (2021), 2071–2085. https://doi.org/10.1109/TMM.2020.3007350

[3] Yin Bi, Wenyao Xu, Nan Guan, Yangjie Wei, and Wang Yi. 2014. Pervasive Eating Habits Monitoring and Recognition Through a Wearable Acoustic Sensor. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare* (Oldenburg, Germany) *(PervasiveHealth '14)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, 174–177. https://doi.org/10.4108/icst.pervasivehealth.2014.255423

[4] Steven Cadavid and Mohamed Abdel-Mottaleb. 2010. Exploiting Visual Quasi-periodicity for Automated Chewing Event Detection Using Active Appearance Models and Support Vector Machines. In *2010 20th International Conference on Pattern Recognition*. 1714–1717.

[5] Eleonora Ceccaldi, Radoslaw Niewiadomski, Maurizio Mancini, and Gualtiero Volpe. 2022. What's on your plate? Collecting multimodal data to understand commensal behavior. *Frontiers in Psychology* 13 (2022). https://doi.org/10.3389/fpsyg.2022.911000

[6] J. M. Fontana, M. Farooq, and E. Sazonov. 2014. Automatic Ingestion Monitor: A Novel Wearable Device for Monitoring of Ingestive Behavior. *IEEE Transactions on Biomedical Engineering* 61, 6 (June 2014), 1772–1779.

[7] Ayaka Fujii, Kanae Kochigami, Shingo Kitagawa, Kei Okada, and Masayuki Inaba. 2020. Development and Evaluation of Mixed Reality Co-eating System: Sharing the Behavior of Eating Food with a Robot Could Improve Our Dining Experience. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 357–362. https://doi.org/10.1109/RO-MAN47096.2020.9223518

[8] Paula M Gardiner, Kelly D McCue, Lily M Negash, Teresa Cheng, Laura F White, Leanne Yinusa-Nyahkoon, Brian W Jack, and Timothy W Bickmore. 2017. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient education and counseling* 100, 9 (2017), 1720–1729.

[9] Delwar Hossain, Tonmoy Ghosh, and Edward Sazonov. 2020. Automatic count of bites and chews from videos of eating episodes. *IEEE Access* 8 (2020), 101934–101945.

[10] Yugyeong Jung, Gyuwon Jung, Sooyeon Jeong, Chaewon Kim, Woontack Woo, Hwajung Hong, and Uichin Lee. 2023. "Enjoy, but Moderately!": Designing a Social Companion Robot for Social Engagement and Behavior Moderation in Solitary Drinking Context. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 237 (oct 2023), 24 pages. https://doi.org/10.1145/3610028

[11] Azusa Kadomura, Cheng-Yuan Li, Yen-Chang Chen, Koji Tsukada, Itiro Siio, and Hao-hua Chu. 2013. Sensing fork: eating behavior detection utensil and mobile persuasive game. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1551–1556.

[12] R. A. Khot, E. S. Arza, H. Kurra, and Y. Wang. 2019. FoBo: Towards Designing a Robotic Companion for Solo Dining. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. ACM, New York, NY, USA, LBW1617:1–LBW1617:6.

[13] R. Liu and T. Inoue. 2014. Application of an Anthropomorphic Dining Agent to Idea Generation. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 607–612.

[14] MediaPipe. [n. d.]. https://developers.google.com/mediapipe/solutions. Accessed: 2024-06-22.

[15] MetaHuman. [n. d.]. https://www.unrealengine.com/marketplace/en-US/product/metahuman-plugin. Accessed: 2024-06-22.

[16] Radoslaw Niewiadomski, Merijn Bruijnes, Gijs Huisman, Conor Patrick Gallagher, and Maurizio Mancini. 2022. Social robots as eating companions. *Frontiers in Computer Science* 4 (2022). https://doi.org/10.3389/fcomp.2022.909844

[17] R. Niewiadomski, E. Ceccaldi, G. Huisman, G. Volpe, and M. Mancini. 2019. Computational Commensality: from theories to computational models for social food preparation and consumption in HCI. *Frontiers in Robotics and AI* 6 (2019), 1–19.

[18] Radoslaw Niewiadomski, Gabriele De Lucia, Gabriele Grazzi, and Maurizio Mancini. 2022. Towards Commensal Activities Recognition. In *Proceedings of the 2022 International Conference on Multimodal Interaction* (Bengaluru, India) *(ICMI '22)*. Association for Computing Machinery, New York, NY, USA, 549–557. https://doi.org/10.1145/3536221.3556566

[19] Daehyung Park, Yuuna Hoshi, Harshal P. Mahajan, Ho Keun Kim, Zackory Erickson, Wendy A. Rogers, and Charles C. Kemp. 2020. Active robot-assisted feeding with a general-purpose mobile manipulator: Design, evaluation, and lessons learned. *Robotics and Autonomous Systems* 124 (2020), 103344. https://doi.org/10.1016/j.robot.2019.103344

[20] Maína Ribeiro Pereira-Castro, Adriano Gomes Pinto, Tamila Raposo Caixeta, Renata Alves Monteiro, Ximena Pamela Díaz Bermúdez, and Ana Valéria Machado Mendonça. 2022. Digital Forms of Commensality in the 21st Century: A Scoping Review. *International Journal of Environmental Research and Public Health* 19, 24 (2022). https://doi.org/10.3390/ijerph192416734

[21] Philipp V. Rouast and Marc T. P. Adam. 2020. Learning Deep Representations for Video-Based Intake Gesture Detection. *IEEE Journal of Biomedical and Health Informatics* 24, 6 (2020), 1727–1737.

[22] Sainsbury. 2018. The Sainsbury's Living Well Index. https://www.about.sainsburys.co.uk/~/media/Files/S/Sainsburys/living-well-index/sainsburys-living-well-index-may-2018.pdf. Accessed January 26, 2022.

[23] C. Spence, M. Mancini, and G. Huisman. 2019. Digital commensality: Eating and drinking in the company of technology. *Frontiers in psychology* 10 (2019), 2252.

[24] M. Takahashi, H. Tanaka, H. Yamana, and T. Nakajima. 2017. Virtual Co-Eating: Making Solitary Eating Experience More Enjoyable. In *Entertainment Computing – ICEC 2017*, N. Munekata, I. Kunita, and J. Hoshino (Eds.). Springer International Publishing, Cham, 460–464.

[25] Unreal. [n. d.]. https://www.unrealengine.com. Accessed: 2024-06-22.

[26] Hannes Vilhjálmsson, Nathan Cantelmo, Justine Cassell, Nicolas E. Chafai, Michael Kipp, Stefan Kopp, Maurizio Mancini, Stacy Marsella, Andrew N Marshall, Catherine Pelachaud, et al. 2007. The behavior markup language: Recent developments and challenges. In *Intelligent Virtual Agents: 7th International Conference, IVA 2007 Paris, France, September 17-19, 2007 Proceedings 7*. Springer, 99–111.