Toward Modeling Commensal Interactions in Human Dyads

Radoslaw Niewiadomski
University of Genoa
Department of Informatics, Bioengineering, Robotics and Systems Engineering
Genoa, Italy
radoslaw.niewiadomski@unige.it

Cigdem Beyan
University of Verona
Department of Computer Science
Verona, Italy
cigdem.beyan@univr.it

Abstract

We postulate the need for the creation of computational methods to model interactions specific to commensal settings. They would be used to analyze and quantify interactions during shared meals, and to design new devices for commensality. To illustrate the concept, we present algorithms for measuring: 1) food intake ratio and synchronization, and 2) smile ratio and synchronization in pairs of eaters. They process images of two commensals captured simultaneously to extract information specific to their nonverbal behaviors and subsequently apply the Event Synchronization algorithm to compute their degree of synchronization. Next, we test the proposed methods on videos of 12 dyads sharing meals. Our findings suggest that the self-reported strength of the relationship is positively correlated with the degree of food intake synchronization and inversely correlated with the quantity of smiles. We conclude by discussing potential applications for developing artificial companions to support solo eaters.

CCS Concepts

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

Keywords

Commensality, social interactions, interpersonal synchronization, food intake, smile

ACM Reference Format:
Radoslaw Niewiadomski and Cigdem Beyan. 2025. Toward Modeling Commensal Interactions in Human Dyads. In *Designing Interactive Systems Conference (DIS '25 Companion), July 05–09, 2025, Funchal, Portugal.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3715668.3736334

1 Introduction

Understanding and modeling social interactions in dyads and groups is a crucial area of research spanning multiple disciplines, including psychology, sociology, and artificial intelligence. Such interactions involve complex dynamics shaped by verbal and nonverbal communication, individual roles, and shared social contexts, requiring the analysis of interpersonal relationships and coordination patterns. Advances in Social Signal Processing have enabled computational approaches to capture and interpret nonverbal cues such as body posture, gaze, and vocal characteristics, thereby offering valuable insights into the dynamics of social interactions and their analysis [57]. Modeling group interactions in terms of both social exchanges and physical actions has multiple applications in Human-Computer and Human-Robot Interaction [10]. Apart from explaining human social behavior, these models are essential when designing new forms of interaction (e.g., between humans and artificial agents, or in VR-based environments) and new interactive devices (e.g., smart tables [69]). Previous research has primarily focused on structured meetings and free-standing mingling events to investigate various social phenomena [10]. Meetings typically involve dyads or small groups of 3–4 participants seated around a table to perform a predefined task. In contrast, mingling events feature dynamic social interactions in which individuals spontaneously form and dissolve dyads and groups [10]. Such settings have enabled researchers to analyze key aspects of social dynamics, including engagement [7, 46], group cohesion [53, 68], leadership [39, 42, 60], social roles [1, 66], group performance [27, 48], mimicry [43], entrainment [25, 62], and so forth. These studies typically analyze sequences of behaviors exhibited by each interaction partner, including gaze movements, the start and end of utterances, interruptions, hand gestures, and the regularities (e.g., repetitions and patterns) in the behavior of others.

Despite the significant amount of work mentioned above, a considerable gap remains in understanding the dynamics of social interactions in commensal settings. The new trends in HCI [41, 54] postulate using technology to enable, enhance or facilitate commensality experience. This encompasses tele-dining [12], designing new form of interactions such as interactive smart tables [6, 17], interactive games [2], virtual [26, 32, 55, 59] and augmented reality [61] characters, as well as robotic dining companions [19, 21, 33]. Motivated by the aforementioned studies, in this paper, we advocate for the design of new computational approaches for modeling interactions specific to commensality.

The act of eating and sharing food [44] is one of the most frequent social experiences and is as old as humanity. Commensal events are characterized by rich social interactions, where the act of consuming food provides an opportunity to exchange ideas, strengthen social bonds (e.g., friendships), or finalize business matters. However, interactions at the table are highly specific, as participants continuously divide their attention between food and drink preparation, consumption, and social engagement. A person may focus on their plate while listening to a conversation, though the nonverbal cues they exhibit in this context differ from those in other social settings. For example, if a diner does not respond to a question immediately posed by their interaction partner, it does not necessarily mean they are not listening—they may simply be chewing. Similarly, if they do not maintain eye contact while speaking, it

ACM ISBN 979-8-4007-1486-3/25/07
https://doi.org/10.1145/3715668.3736334

could be because they are cutting their food. Instead, in other contexts, e.g., brainstorming meetings [20], these behaviors might be perceived as rude or disruptive to the interaction. Such examples illustrate that existing models may not be well-suited for commensal settings, highlighting the need for dedicated studies. Consequently, herein we focus on investigating food intake and smile exchange in commensal settings between partners with varying levels of acquaintance.

2 Related Work

Human behavior understanding through social signals has been extensively examined in a recent survey [10], which categorizes research into three main areas: (1) the detection of social traits such as leadership, dominance, and personality traits, (2) social role recognition and the identification of social relations, and (3) interaction dynamics analysis to assess group cohesion, empathy, rapport, and similar aspects. Several approaches identified leaders and their styles by analyzing speaking, head, and body activity, as well as factors like variations in prosody, the number of speaking turns, and body motion patterns [8, 9, 51]. Another widely studied area has been the automatic recognition of personality traits using cues such as speaking activity, visual focus of attention, prosody, and body movements. In this line, while most studies have concentrated on meetings with 3–4 participants [63, 65], relatively few methods have been tested in settings, including people watching movies together [36], free-standing conversations [14, 15], and surveillance videos [64]. Studies on automated social role detection demonstrated the effectiveness of turn-taking patterns, prosodic features, and lexical information in role recognition [13, 58, 65] while other effective cues are found to be facial expressions, head pose, and body movement e.g., [18, 52, 66]. On the other hand, the automated detection of social relations has largely relied on computer vision techniques, utilizing large datasets from movies, and YouTube videos [30, 31]. Another well-studied area is interaction dynamics in dyads and groups, encompassing aspects such as conversational context e.g., [35], engagement e.g., [56], group cohesion e.g., [53, 67], empathy e.g.,[39, 62], and rapport e.g., [27, 38]. A key factor in these dynamics is vocal entrainment, where individuals synchronize their words or speaking styles. Furthermore, nonverbal behavior analysis has been widely used to assess group performance [37], interaction quality [22], and satisfaction levels [23]. Computational studies have focused on automatically distinguishing different types of group conversations, including brainstorming vs. decision-making [20], formal vs. informal [34], focused vs. unfocused [5], and scenario-based vs. non-scenario interactions [35]. Regarding commensality, Ondras et al. [47] analyze interaction dynamics in a group of human eaters. Their Bite Timing Prediction model predicts socially appropriate bite timing in a 6-second window, using gaze, speech, and skeleton data from all three participants. This model, although designed for assistive robotics, is one of the first to analyze the interaction dynamics of a commensal triad.

3 Quantifying Commensal Interactions

Modeling commensality involves a dyad or a group sharing meals, which traditionally, in many cultures, takes place at a table. At the level of nonverbal communication, this interaction is characterized by a continuous exchange of gaze, smiles, and other social signals that contribute to phenomena such as mimicry, synchronization, or entrainment. At a higher level, these signals and actions may provide information about the quality of the interaction, user satisfaction, and their well-being. Modeling social interaction is realized at two levels: the first (i.e., low-level) consists of modeling single actions (e.g., food intake) and social signals (e.g., gaze direction); the second (i.e., high-level) consists of modeling their meaning on a longer time scales, such as the strength of the relation between the participants, and their enjoyment of the interaction. In our approach, we focus on the latter while addressing low-level modeling, we use the existing approaches frequently used in the literature. Among the cues specific to commensality, the most important are probably related to food and drink intake. During commensal events, individuals alternate between consuming food and engaging socially. Engaging interactions likely involve long moments of desynchronization, where one person speaks while the other listens and/or consumes food. In contrast, less interaction probably can be associated with faster food intake. Thus, food intake (e.g., its speed) and the degree of food intake synchronization can serve as indicators of the quality of interaction, as well as provide insights into the type of relationship between the eaters. More standard measurements include smile quantity and synchronization. The smile is one of the most important social signals, frequently studied by HCI researchers, e.g., [43, 45]. It can convey various meanings, such as an expression of enjoyment and satisfaction, a backchannel signal, and politeness. All these various meanings of the smile may be plausible in a commensal scenario. The frequency and interpersonal smile synchronization may indicate a rewarding and enjoyable interaction.

3.1 Dataset

To illustrate the concept, we use audio-visual recordings of colocated pairs sharing meals. It comprises 12 sessions recorded in the same room, featuring pairs with varying levels of acquaintance. Participants generally consumed similar food, such as pasta or rice. Additionally, water and napkins were provided. The videos were recorded using two cameras. Each recording includes the synchronized view of two participants facing each other. Videos were recorded at a resolution of 1920×1080 with a frame rate of 25 frames per second. In total, 234 minutes were recorded, with the shortest session lasting 8 minutes and the longest 39 minutes. The participants filled out a set of questionnaires before and after dining. A combination of standard and in-house-designed questionnaires was used. Before the meal, each person completed a questionnaire measuring their attitude toward commensality in general (in total 10 questions with a Likert scale from 1 to 5), the frequency of eating together in the last 6 months (in total 5 questions with a Likert scale from 1 to 5), the use of technology during eating and two standard questionnaires measuring their relationship with their partner: the Inclusion of Other in the Self (IOS) Scale [3] and the Quality of Relationships Inventory (QRI) [49]. Additionally, on the day of the experiment, the participants filled out another in-house designed questionnaire (*PRE_Q*) to address their emotions and attitude toward the person they were scheduled to meet that

Figure 1: The time-flow of the data collection.

day, which includes questions, e.g, "My feelings/attitude towards the person X before this meeting are ..." and "I am happy to meet the person X right now" (in total 7 questions with a Likert scale from 1 to 7). Finally, another set of questions (*POST_Q*) was posed after the experiment (25 questions), composed of e.g., "In general, I would rate this experience..." or "My feelings/attitude towards the person X after this meeting are ...", and so on. Questionnaires are available at https://zenodo.org/records/15399785. All stages of the data collection are explained in Fig. 1.

3.2 Food Intake Ratio and Synchronization

Given a video of two eaters, A and B, we measure the distance between i) the center of the mouth area and ii) the wrist of the dominant hand of each eater, A and B, independently to estimate their food intake. The coordinates of the wrist and face are extracted using MediaPipe's Pose and FaceMesh models [28]. The mouth center is computed by averaging the x and y coordinates of two landmarks: the center of the upper and lower lip. Next, the Euclidean distance between the mouth center and the wrist is calculated, and the resulting value is normalized by computing its ratio relative to the face dimensions in the same image frame. The same operations are repeated for the second eater, resulting in two vectors d_A and d_B of distances, with lengths corresponding to the number of frames in the original video.

In the following step, we detect significant events in d_A and d_B that correspond to food intake. More specifically, we search for moments when the normalized distance remains below the empirically chosen threshold of 1.25 for more than 0.35 but less than 3.5 seconds. According to our observations, this time interval typically corresponds to one food intake. Actions shorter than this may correspond to gestures performed close to the face, while longer durations may occur, e.g., when the person is not eating but rather resting their head on their hand. Next, we create two binary time series of events ts_A and ts_B with the same length as the original distance vectors, such that the detected beginnings of food intake events are marked with ones, while all remaining positions are filled with zeros. Finally, we compute the degree of synchronization for each pair using the Event Synchronization technique (EV) [50] originally proposed to analyze brain signals by measuring synchronization between events occurring in two binary time series. More recently, this technique was successfully applied to analyze, e.g., intra-personal synchronization [29]. The computed values express the degree of synchronization of food intake between commensal partners. This approach is illustrated in Fig. 2, which presents four extracts of dyads with varying levels of acquaintance. It can be seen that, in the first row, the events (in green and orange) are more synchronized. In contrast, in the

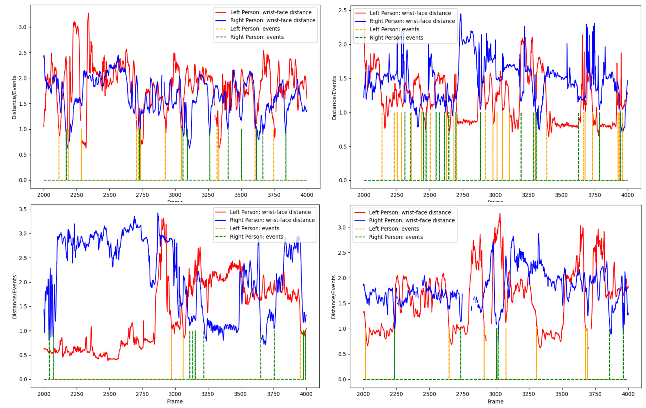

Figure 2: Four extracts: the first row shows the wrist-to-hand distances and synchronization events for two couples that reported a strong relationship, while the second row shows corresponding data for individuals who reported a low acquaintance.

second row, there are long periods without food intake, by one or both eaters. Next, we compare the degree of synchronization with the responses given by participants for the IOS Scale [3], and 14 questions of QRI [49]. To obtain the scores that describe a dyad, we sum the responses of both eaters on: 1) the IOS Scale and 2) the IRQ, and compute their correlations with EV across all videos for 4 buffers: τ = 50, 100, 200, and 400 frames (i.e., from 2 to 16 seconds). The resulting correlations fall within the intervals of 0.66−0.69 (IRQ) and 0.53−0.66 (IOS). While these results are promising, they should be interpreted with caution due to the small number of participants considered. We also examined the correlations for the frequency of food intakes, calculated as the sum of food intakes by both participants, obtaining 0.56 and 0.53 for *IOS* and *IRQ*, respectively.

3.3 Smiling Ratio and Synchronization

The main visible indicator of a smile is Action Unit 12 (AU12) in the FACS notation [16], which refers to the activation of the zygomaticus major muscle. While different smile types may include other Action Units, the AU12 is present in all of them. We use OpenFace [4] to extract the AU12 activity for each frame, and each eater. We normalize extracted values and filter them, obtaining two time series s_A and s_B of distances, with lengths corresponding to the number of frames in the original video. Next, we extract significant events from s_A and s_B, which occur when AU12 activity remains above the empirically chosen threshold, which is lasting more than 0.6 seconds (to exclude short muscle activation related to chewing or speaking). We create two binary time series of events ts_A and ts_B with the same length s_A and s_B: detected beginnings of smiles are marked with ones, while all remaining positions are filled with zeros. Finally, we compute the degree of synchronization on s_A, s_B using EV [50]. In the final step, we analyze the correlations. A slight inverse correlation was observed: the strength of the relationship appears to correspond to reduced smile synchronization (the strongest correlation with *IOS* was −0.57, τ = 200). We also examined the correlations for the frequency of smiles (i.e., the sum of detected smile events of both participants), obtaining −0.67 and −0.66 for the IOS and IRQ, respectively.

3.4 Discussion

The results for 12 videos are presented in Fig. 3. The average acquaintance measured with the IOS is 3.4 on a 7-point scale, with a relatively large standard deviation (SD = 2.24); 5 people reported the maximum score, 9 reported the minimum. This shows that there is substantial variability in the reported acquaintance across studied pairs. It appears that interactions between dyads reporting a strong relationship showed more synchronized food intake, but fewer smiles and lower smile synchronization. This may seem surprising at first glance, but two points should be considered. First, even individuals who did not know each other well voluntarily participated in the data collection (the results might differ significantly if random people were forced to eat together). Therefore, it cannot be excluded, participants (especially if they did not know each other well) were adapting their behavior, including smiling and quantity of conversation, to create favorable impressions (so-called impression management). When meeting new people, individuals may engage more to establish rapport and appear likable. In contrast, with familiar companions, the need for such impression management diminishes. Moreover, various types of smiles exist and are used in different social situations [11, 24, 45]. In future work, we need to differentiate between different types of smiles. This highlights the need for multifactorial models and large-scale data collection to better understand interaction dynamics.

All participants reported being satisfied with their experience (Q1 in *POST_Q*), with an average score of 6.1 on a 7-point scale (SD = 0.82). Interestingly, at the same time, some participants reported feeling uncomfortable (Q9 in *POST_Q*), with 5 participants giving average or higher ratings (average score of feeling unconformable was 1.56 on a 5-point scale, SD = 0.84), and not relaxed (Q10 in *POST_Q*), with 6 participants providing average or lower ratings (average score of feeling relaxed was 4.13 on a 5-point scale, SD = 1.04), during the interaction. Furthermore, 9 persons reported an increase in positive attitude towards their interaction partner after eating (measured as the difference before and after the session, that is: Q3 in *POST_Q* - Q2 in *PRE_Q*), while 6 participants reported a decrease in positive attitudes (in total, an average increase of 0.2 on a 7-point scale was observed). Interestingly, there was a much larger consensus regarding the imagined attitude of the interaction partners towards the participants themselves. Twelve participants believed that their attitude towards them had improved, while only four thought it had worsened (measured as the difference before and after the session, that is: Q19 *POST_Q* - Q6 in *PRE_Q*). These factors (e.g., feeling uncomfortable) could influence the behaviors of some of the participants and should be taken into account in future models, when more data will be collected. However, at this stage of research, too few participants reported any negative outcomes, while they were generally satisfied, to perform such analyses.

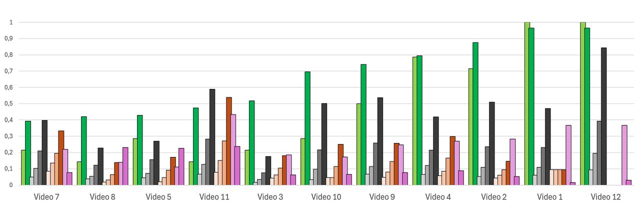

Figure 3: The synchronization results with respect to the questionnaire responses: shades of green correspond to the answers given in the questionnaires, different shades of gray represent the degree of food intake synchronization, and the various shades of brown show smile synchronization values.

4 Conclusions

We presented a method to compute degrees of food intake synchronization and smile synchronization of eating partners to model commensal interactions. Our analysis shows that the former appears to correlate with the reported strength of the relationship. In the future, we will include other cues based on gaze direction and speech. For instance, we will examine whether, when one person is speaking, the other maintains eye contact or looks at their plate, focusing on food intake. This could be measured as the percentage of time Partner B maintains eye contact while Partner A is speaking, and vice versa. Similarly, we could measure the percentage of time Partner B is chewing while Partner A is speaking, as well as the percentage of time spent i) in silence (i.e., when neither person is speaking), and ii) in mutual eye contact. Most of these cues can be potentially extracted, like how we analyzed food intake. At the same time, we will analyze other factors included in the questionnaires, such as attitude changes during the meal, reciprocal expectations regarding the behavior of the interaction partners, reported relax/stress levels, as well as general preferences regarding the commensality. Our methods need to be tested on larger datasets, preferably recorded in diverse settings (e.g., public places), contexts (e.g., dates, family meetings, business lunches), and with more people. Another interesting research direction is to study potential imbalance in groups. While at the moment we simply sum up the responses of the two persons, they often did not agree even on the reported strength of the relationship. It would be worth to examine whether these reported inequalities are reflected in their behaviors.

We believe that such methods can serve as tools to evaluate the commensal experience in terms of satisfaction from the shared meal, group cohesion, but also the type of commensal event, and more. Beyond that, they can be used for designing new interactive systems for commensality. Several researchers [19, 21] postulate the creation of artificial commensal companions, that is, "active partners during mealtime, able to interact with a human partner and influence their eating experience" [40]. Such agents may bring the benefits of commensality, such as healthier food choices, increased well-being, and more, especially for individuals who are constrained to eat alone. However, existing virtual commensal companions (e.g., [26, 32, 55]) still lack the skills to perform human-like interactions, as often reported in their evaluations. Models of interaction dynamics, such as those presented in this paper, are essential for enabling such virtual companions to cultivate rewarding and enjoyable interactions when people are eating in their company.

Acknowledgments

This work is supported by the PRIN 2022 project COCOA, PRIN 2022T8ZNNM, funded by the European Union - Next Generation EU (NGEU) Programme, CUP D53D23008850001.

References

[1] Xavier Alameda-Pineda, Yan Yan, Elisa Ricci, Oswald Lanz, and Nicu Sebe. 2015. Analyzing Free-Standing Conversational Groups: A Multimodal Approach. In *Proc. of ACMMM.* 5–14.

[2] Peter Arnold. 2017. You Better Eat to Survive! Exploring Edible Interactions in a Virtual Reality Game. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 206–209. doi:10.1145/3027063.3048408

[3] A. Aron, E. N. Aron, M. Tudor, and G. Nelson. 1991. Close relationships as including other in the self. *Journal of Personality and Social Psychology* 60, 2 (1991), 241–253.

[4] T. Baltrusaitis, P. Robinson, and L-P. Morency. 2016. OpenFace: An open source facial behavior analysis toolkit.. In *Proc. of IEEE WACV.* 1–10.

[5] Sophia Bano, Jianguo Zhang, and Stephen J. McKenna. 2017. Finding Time Together: Detection and Classification of Focused Interaction in Egocentric Video. In *IEEE CVPR workshops.* 2322–2330.

[6] Pollie Barden, Rob Comber, David Green, Daniel Jackson, Cassim Ladha, Tom Bartindale, Nick Bryan-Kinns, Tony Stockman, and Patrick Olivier. 2012. Telematic dinner party: designing for togetherness through play and performance. In *Proceedings of the Designing Interactive Systems Conference.* ACM, 38–47.

[7] Roman Bednarik, Shahram Eivazi, and Michal Hradis. 2012. Gaze and conversational engagement in multiparty video conversation: an annotation scheme and classification of high and low levels of engagement. In *Proc. Workshop on eye gaze in intelligent human machine interaction.* 1–6.

[8] Cigdem Beyan, Francesca Capozzi, Cristina Becchio, and Vittorio Murino. 2018. Prediction of the Leadership Style of an Emergent Leader Using Audio and Visual Nonverbal Features. *IEEE Trans. Multimedia* 20, 2 (2018), 441–456.

[9] Cigdem Beyan, Muhammad Shahid, and Vittorio Murino. 2018. Investigation of Small Group Social Interactions Using Deep Visual Activity-Based Nonverbal Features. In *Proc. of ACMMM.* 311–319.

[10] Cigdem Beyan, Alessandro Vinciarelli, and Alessio Del Bue. 2023. Co-located human-human interaction analysis using nonverbal cues: A survey. *Comput. Surveys* 56, 5 (2023), 1–41.

[11] Yevgen Bogodistov and Florian Dost. 2017. Proximity Begins with a Smile, But Which One? Associating Non-duchenne Smiles with Higher Psychological Distance. *Frontiers in Psychology* Volume 8 - 2017 (2017). doi:10.3389/fpsyg.2017.01374

[12] 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), 911000.

[13] Marco Cristani, R. Raghavendra, Alessio Del Bue, and Vittorio Murino. 2013. Human behavior analysis in video surveillance: A Social Signal Processing perspective. *Neurocomputing* 100 (2013), 86 – 97.

[14] Dario Dotti, Esam Ghaleb, and Stylianos Asteriadis. 2020. Temporal Triplet Mining for Personality Recognition. In *Proc. of IEEE FG.* 171–178.

[15] Dario Dotti, Mirela Popa, and Stylianos Asteriadis. 2020. Being the Center of Attention: A Person-Context CNN Framework for Personality Recognition. 10, 3 (2020).

[16] Paul Ekman and Wallace V. Friesen. 1978. Facial Action Coding System (FACS).

[17] Hasan Shahid Ferdous, Bernd Ploderer, Hilary Davis, Frank Vetere, Kenton O'Hara, Geremy Farr-Wharton, and Rob Comber. 2016. TableTalk: integrating personal devices and content for commensal experiences at the family dinner table. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM, 132–143.

[18] Gaurav Fotedar, Aditya Gaonkar P., Saikat Chatterjee, and Prasanta Kumar Ghosh. 2016. Automatic recognition of social roles using long-term role transitions in small group interactions. In *Proc. of INTERSPEECH.* 2065–2069.

[19] Ayaka Fujii, Kei Okada, and Masayuki Inaba. 2021. A Basic Study for Acceptance of Robots as Meal Partners: Number of Robots During Mealtime, Frequency of Solitary Eating, and Past Experience with Robots. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN).* 73–80. doi:10.1109/RO-MAN50785.2021.9515451

[20] Dineshbabu Jayagopi, Taemie Kim, Alex Pentland, and Daniel Gatica-Perez. 2012. Privacy-sensitive recognition of group conversational context with sociometers. *Springer Multimedia Systems* 18 (2012), 3–14.

[21] Rohit Ashok Khot, Eshita Sri Arza, Harshitha Kurra, and Yan 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, Article LBW1617, 6 pages. doi:10.1145/3290607.3313069

[22] Juha M. Lahnakoski, Paul A.G. Forbes, Cade McCall, and Leonhard Schilbach. 2020. Unobtrusive tracking of interpersonal orienting and distance predicts the subjective quality of social interactions. *Royal Society Open Science* 7 (2020).

[23] Catherine Lai and Gabriel Murray. 2018. Predicting Group Satisfaction in Meeting Discussions. In *Proc. of the Workshop on Modeling Cognitive Processes from Multimodal Data.*

[24] Mark Leary and Robin Kowalski. 1990. Impression Management: A Literature Review and Two-Component Model. *Psychological Bulletin - PSYCHOL BULL* 107 (01 1990), 34–47. doi:10.1037/0033-2909.107.1.34

[25] Chi-Chun Lee, Athanasios Katsamanis, Brian R Baucom, Panayiotis G Georgiou, and Shrikanth S Narayanan. 2012. Using measures of vocal entrainment to inform outcome-related behaviors in marital conflicts. In *Proc. of APSIPA.* 1–5.

[26] Rui Liu and Tomoo 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* (Seattle, Washington) *(UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 607–612. doi:10.1145/2638728.2641342

[27] Nichola Lubold and Heather Pon-Barry. 2014. Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. In *Proc. ACM Multimodal Learning Analytics Workshop and Grand Challenge.* 5–12.

[28] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. MediaPipe: A Framework for Building Perception Pipelines. arXiv:1906.08172 [cs.DC] https://arxiv.org/abs/1906.08172

[29] Vincenzo Lussu, Radoslaw Niewiadomski, Gualtiero Volpe, and Antonio Camurri. 2020. The role of respiration audio in multimodal analysis of movement qualities. *Journal on Multimodal User Interfaces* 14 (2020), 1–15. doi:10.1007/s12193-019-00302-1

[30] Jinna Lv, Wu Liu, Lili Zhou, Bin Wu, and Huadong Ma. 2018. Multi-stream Fusion Model for Social Relation Recognition from Videos. In *MultiMedia Modeling.* Cham, 355–368.

[31] Jinna Lv and Bin Wu. 2019. Spatio-Temporal Attention Model Based on Multi-view for Social Relation Understanding. In *MultiMedia Modeling.* Cham, 390–401.

[32] Maurizio Mancini, Radoslaw Niewiadomski, Gabriele De Lucia, and Francesco Maria Longobardi. 2024. A Virtual Agent as a Commensal Companion. In *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents* (GLASGOW, United Kingdom) *(IVA '24)*. Association for Computing Machinery, New York, NY, USA, Article 26, 4 pages. doi:10.1145/3652988.3673963

[33] Maurizio Mancini, Radoslaw Niewiadomski, Gijs Huisman, Merijn Bruijnes, and Conor Patrick Gallagher. 2020. Room for One More? - Introducing Artificial Commensal Companions. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts* (Honolulu, HI, USA) *(CHI'20)*. Association for Computing Machinery, New York, NY, USA, 1–8.

[34] Aleksandar Matic, Venet Osmani, and Oscar Mayora-Ibarra. 2014. Mobile Monitoring of Formal and Informal Social Interactions at Workplace. In *Proc. of ACM UbiComp.* 1035–1044.

[35] Chreston Miller and Christa Miller. 2019. Timing is Everything: Identifying Diverse Interaction Dynamics in Scenario and Non-Scenario Meetings. In *International Conference on eScience.* 203–212.

[36] Juan Abdon Miranda Correa, Mojtaba Khomami Abadi, Nicu Sebe, and Ioannis Patras. 2018. AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. *IEEE Trans. on Affective Comput.* (2018).

[37] Go Miura and Shogo Okada. 2019. Task-Independent Multimodal Prediction of Group Performance Based on Product Dimensions. In *Proc. of ACM ICMI.* 264–273.

[38] Philipp Muller, Michael Xuelin Huang, and Andreas Bulling. 2018. Detecting Low Rapport During Natural Interactions in Small Groups from Non-Verbal Behaviour. In *Proc. of Inter. Conf. on Intelligent User Interfaces.* 153–164.

[39] Philipp Matthias Muller and Andreas Bulling. 2019. Emergent Leadership Detection Across Datasets. In *Proc. of ACM ICMI.* 274–278.

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

[41] Radoslaw Niewiadomski, Eleonora Ceccaldi, Gijs Huisman, Gualtiero Volpe, and Maurizio Mancini. 2019. Computational Commensality: From Theories to Computational Models for Social Food Preparation and Consumption in HCI. *Frontiers in Robotics and AI* 6 (2019). doi:10.3389/frobt.2019.00119

[42] Radoslaw Niewiadomski, Lea Chauvigne, Maurizio Mancini, Gualtiero Volpe, and Antonio Camurri. 2024. Nonverbal Leadership in Joint Full-Body Improvisation. *IEEE Transactions on Affective Computing* (2024), 1–13. doi:10.1109/TAFFC.2024.3514933

[43] Radoslaw Niewiadomski, Ken Prepin, Elisabetta Bevacqua, Magalie Ochs, and Catherine Pelachaud. 2010. Towards a smiling ECA: studies on mimicry, timing and types of smiles. In *Proceedings of the 2nd international workshop on Social signal processing* (Firenze, Italy) *(SSPW '10)*. ACM, New York, NY, USA, 65–70.

doi:10.1145/1878116.1878134

[44] Elinor Ochs and Merav Shohet. 2006. The cultural structuring of mealtime socialization. *New directions for child and adolescent development* 2006, 111 (2006), 35–49.

[45] Magalie Ochs, Radoslaw Niewiadomski, Paul Brunet, and Catherine Pelachaud. 2012. Smiling virtual agent in social context. *Cognitive Processing* 13, 2 (2012), 519–532. doi:10.1007/s10339-011-0424-x

[46] Catharine Oertel and Giampiero Salvi. 2013. A Gaze-Based Method for Relating Group Involvement to Individual Engagement in Multimodal Multiparty Dialogue. In *Proc. of ACM ICMI*. 99–106.

[47] Jan Ondras, Abrar Anwar, Tong Wu, Fanjun Bu, Malte Jung, Jorge Jose Ortiz, and Tapomayukh Bhattacharjee. 2022. Human-robot commensality: Bite timing prediction for robot-assisted feeding in groups. In *6th Annual Conference on Robot Learning*.

[48] Ehsan Othman, Frerk Saxen, Dmitri Bershadskyy, Philipp Werner, Ayoub Al-Hamadi, and Joachim Weimann. 2019. Predicting Group Contribution Behaviour in a Public Goods Game from Face-to-Face Communication. *Sensors* 19, 12 (2019), 2786.

[49] Gregory Pierce, Irwin Sarason, and Barbara Sarason. 1991. General and Relationship-Based Perceptions of Social Support: Are Two Constructs Better Than One? *Journal of personality and social psychology* 61 (12 1991), 1028–39. doi:10.1037/0022-3514.61.6.1028

[50] Rodrigo Quian, Thomas Kreuz, and Peter Grassberger. 2002. Event Synchronization: A simple and fast method to measure synchronicity and time delay patterns. *Physical review. E, Statistical, nonlinear, and soft matter physics* 66 (11 2002), 041904. doi:10.1103/PhysRevE.66.041904

[51] Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2012. A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups. *IEEE Trans. Multimedia* 14, 3 (2012), 816–832.

[52] Ashtosh Sapru and Herve Bourlard. 2015. Automatic recognition of emergent social roles in small group interactions. *IEEE Trans. Multimedia* 17, 5 (2015), 746–760.

[53] G. Sharma, S. Ghosh, and A. Dhall. 2019. Automatic Group Level Affect and Cohesion Prediction in Videos. In *Proc. of ACII*. 161–167.

[54] Charles Spence, Maurizio Mancini, and Gijs Huisman. 2019. Digital commensality: Eating and drinking in the company of technology. *Frontiers in psychology* 10 (2019), 460197.

[55] Monami Takahashi, Hiroki Tanaka, Hayato Yamana, and Tatsuo Nakajima. 2017. Virtual Co-Eating: Making Solitary Eating Experience More Enjoyable. In *Entertainment Computing – ICEC 2017*, Nagisa Munekata, Itsuki Kunita, and Junichi Hoshino (Eds.). Springer International Publishing, Cham, 460–464.

[56] Arno Veenstra and Hayley Hung. 2011. Do they like me? Using video cues to predict desires during speed-dates. In *IEEE ICCV Workshops*. 838–845.

[57] Alessandro Vinciarelli, Maja Pantic, Hervé Bourlard, and Alex Pentland. 2008. Social signal processing: state-of-the-art and future perspectives of an emerging domain. In *Proceedings of the 16th ACM International Conference on Multimedia* (Vancouver, British Columbia, Canada) *(MM '08)*. Association for Computing Machinery, New York, NY, USA, 1061–1070. doi:10.1145/1459359.1459573

[58] Alessandro Vinciarelli, Fabio Valente, Sree Harsha Yella, and Ashtosh Sapru. 2011. Understanding Social Signals in Multi-party Conversations: Automatic Recognition of Socio-Emotional Roles in the AMI Meeting Corpus. In *Proc. of IEEE SMC*. 374–379.

[59] Jui-Ying Wang and Tomoo Inoue. 2023. The Similarity of Virtual Meal of a Co-eating Agent Affects Human Participant. In *Collaboration Technologies and Social Computing*, Hideyuki Takada, D. Moritz Marutschke, Claudio Alvarez, Tomoo Inoue, Yugo Hayashi, and Davinia Hernandez-Leo (Eds.). Springer Nature Switzerland, Cham, 115–132. doi:10.1007/978-3-031-42141-9_8

[60] Yanbang Wang, Pan Li, Chongyang Bai, VS Subrahmanian, and Jure Leskovec. 2020. Generic Representation Learning for Dynamic Social Interaction. In *Proc. of ACM SIGKDD Int. Conf. on KDDM MLG Workshop*.

[61] Philip Weber, Kevin Krings, Julia Nießner, Sabrina Brodesser, and Thomas Ludwig. 2021. FoodChattAR: Exploring the Design Space of Edible Virtual Agents for Human-Food Interaction. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 638–650. doi:10.1145/3461778.3461998

[62] Bo Xiao, Zac E Imel, David C Atkins, Panayiotis G Georgiou, and Shrikanth S Narayanan. 2015. Analyzing speech rate entrainment and its relation to therapist empathy in drug addiction counseling. In *Proc. of ISCA*.

[63] Shen Yan, Di Huang, and Mohammad Soleymani. 2020. Mitigating Biases in Multimodal Personality Assessment. In *Proc. of ACM ICMI*. 361–369.

[64] Gloria Zen, Bruno Lepri, Elisa Ricci, and Oswald Lanz. 2010. Space Speaks: Towards Socially and Personality Aware Visual Surveillance. In *Proc. of ACM MPVA*. 37–42.

[65] Lingyu Zhang, Mallory Morgan, and Indrani et al. Bhattacharya. 2019. Improved Visual Focus of Attention Estimation and Prosodic Features for Analyzing Group Interactions. In *Proc. of ACM ICMI*. 385–394.

[66] Lingyu Zhang and Richard J. Radke. 2020. A Multi-Stream Recurrent Neural Network for Social Role Detection in Multiparty Interactions. *IEEE Journal of Selected Topics in Signal Processing* 14, 3 (2020), 554–567.

[67] Yanxia Zhang, Jeffrey Olenick, Chu-Hsiang Chang, Steve W. J. Kozlowski, and Hayley Hung. 2018. The I in Team: Mining Personal Social Interaction Routine with Topic Models from Long-Term Team Data. In *Proc. of IUI*. 421–426.

[68] Yanxia Zhang, Jeffrey Olenick, Chu-Hsiang Chang, Steve W. J. Kozlowski, and Hayley Hung. 2018. TeamSense: Assessing Personal Affect and Group Cohesion in Small Teams through Dyadic Interaction and Behavior Analysis with Wearable Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3 (2018).

[69] Bo Zhou, Jingyuan Cheng, Mathias Sundholm, Attila Reiss, Wuhuang Huang, Oliver Amft, and Paul Lukowicz. 2015. Smart table surface: A novel approach to pervasive dining monitoring. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 155–162.