Integrative Analysis of Multimodal Interaction Data: Predicting Communication Dynamics and Willingness to Communicate (WtC) in Human-Agent Interaction

AboulHassane CISSE[1,*] , Kazuhisa Seta[1,†] and Yuki Hayashi[1, †]

[1] Osaka Metropolitan University, Sakai-City, Osaka Prefecture, Japan

Abstract

This research delves into the intricate relationship between physiological and behavioral indicators and the Willingness to Communicate (WtC) in the context of human-agent interactions. Specifically, it examines how heart rate, eye movement, facial expressions, and conversational dynamics influence individuals' engagement and willingness to engage in dialogue with agents. The study analyzes multimodal interaction data collected from participants engaging with conversational agents to identify patterns and correlations that can predict and subsequently enhance WtC, thereby improving the design and effectiveness of conversational agents. This research stands at the intersection of emotional intelligence, communication studies, and AI technology, offering a novel perspective on enhancing human-agent communication. Through its integrative approach, it seeks to contribute to the development of AI agents that can better understand and respond to human emotional and communicational cues, paving the way for more natural and meaningful digital interactions.

Keywords

Human-Agent Interaction, Conversational Agent, Biometric Indicators, Machine Learning, Communication Dynamics, Physiological Indicators

1. Introduction

1.1. Background

The quest to enhance Willingness to Communicate (WtC) in language learning and human-computer interaction has led to groundbreaking research endeavors. WtC, a pivotal component of language acquisition, denotes an individual's propensity to engage in communication using a second language (L2) across various contexts [1]. Despite extensive studies on pedagogical strategies and technological interventions to foster WtC [2], a significant gap remains in understanding and integrating biometric feedback within

conversational agents. This research pivots on the premise that non-verbal cues, such as heart rate, eye movements, and facial expressions, significantly influence WtC, offering a nuanced perspective on human-agent interactions [3].

1.2. Problem Identification

Identifying and addressing the challenges in fostering WtC is critical for enhancing language learning and human-agent interaction. A significant problem lies in the traditional methods of language instruction and interaction, which often fail to consider the dynamic and complex nature of human communication. Existing solutions primarily focus on verbal and text-based interactions, overlooking the potential of integrating biometric data to provide a more holistic understanding of communication dynamics. This gap highlights the need for innovative approaches that can capture and analyze physiological and emotional cues to predict and enhance WtC.

The integration of biometric indicators such as heart rate, eye tracking, and facial emotions with conversational agents offers a promising avenue to address this issue. By leveraging these multimodal data sources, it is possible to gain deeper insights into users' emotional and cognitive states, thereby tailoring interactions to enhance engagement and willingness to communicate. This approach aligns well with the Learning Analytics scope, providing a robust framework for developing adaptive and responsive educational technologies.

1.3. Main Research Question (MRQ)

How can the integration of biometric indicators reflecting emotional states, eye-tracking metrics, and emotional facial cues, combined with conversational strategies (Affective Backchannel - AB, Conversational Strategies - CS, and their combination AB+CS) implemented by a dialogue agent, predict and enhance the Willingness to Communicate (WtC) gain in human-agent interactions while considering the nuanced interpretations of these emotional and attentional cues? [1, 2]

1.4. Main Hypothesis

The integration of biometric indicators (heart rate, eye tracking, emotional facial cues) alongside conversational strategies (AB, CS, and AB+CS) by a dialogue agent plays a crucial role in predicting and enhancing Willingness to Communicate (WtC) gain in human-agent interactions. This complex interplay can be effectively deciphered and modeled through advanced analytical methods, promising not just to accurately forecast WtC gains but also to provide strategic insights for refining interaction dynamics with dialogue agents [3, 4].

Table 1
Research Questions and Corresponding Hypotheses

Research Questions	Hypothesis
RQ1: How do biometric indicators, eye tracking, and	**H1**: The individual analysis of biometric indicators, eye tracking, and emotional

emotional facial cues individually contribute to predicting Willingness to Communicate (WtC) gain in interactions with dialogue agents?	facial cues significantly boosts the predictive accuracy of WtC gain, highlighting distinct impacts and patterns each of these factors contributes to interaction outcomes with dialogue agents.
RQ2: How do biometric indicators, eye tracking, emotional facial cues, and dialogue content integrated with strategies (AB, CS, AB+CS) by dialogue agents interact to influence and enhance WtC gain?	**H2**: The synergistic effect of integrating biometric indicators, eye tracking, emotional facial cues, and dialogue strategies (AB, CS, AB+CS) through dialogue agents offers a deeper insight and significantly enhances the ability to predict and improve WtC gain in human-agent communications.

1.5. Objectives

The primary objectives of this study are twofold:

1. Develop a sophisticated system capable of collecting and analyzing biometric data and finding causality to understand conversational strategies according to the participant. This system, leveraging the nuanced interplay between biometric indicators and conversational dynamics, aims to provide insights into the underlying mechanisms of WtC enhancement [5, 6].
2. Evaluate the efficacy of this integrative approach in human-agent interactions, thereby contributing to the development of more responsive and empathetic conversational agents [2].

This research encompasses a multidisciplinary approach, integrating insights from linguistics, psychology, and computer science to create a holistic understanding of communication dynamics [7, 8]. By focusing on the analysis of biometric and behavioral signals during interactions with conversational agents, this study endeavors to uncover patterns that could predict and enhance WtC. Through the lens of sophisticated machine learning techniques and predictive modeling, the research aims to offer strategic insights for refining interaction dynamics, thus pushing the boundaries of conventional human-agent communication studies [6, 9].

2. Literature Review

2.1. Willingness to Communicate in Language Learning

Willingness to Communicate (WtC) is a fundamental aspect influencing the frequency and quality of second language use. WtC stems from a dynamic interplay of linguistic self-confidence and the desire to communicate, underscored by personality traits and

situational variables [1]. This heuristic model forms the foundation for understanding WtC within the L2 acquisition landscape.

2.2. Human-Agent Interactions

The advent of conversational agents has significantly altered human-computer interaction, notably in educational realms. Embodied conversational agents (ECAs) mimic human-like interactions, thus offering an immersive learning experience. These agents utilize verbal and non-verbal cues to facilitate natural and engaging interactions that could significantly enhance the learning process, especially in language education [2].

2.3. Biometric Feedback and Communication Dynamics

Advancements in sensor technology have allowed for the integration of biometric feedback into interactive systems, opening new pathways to assess and enhance communication dynamics. Leveraging emotional states inferred from physiological signals can dynamically adapt conversational agents' responses, indicating that a deeper understanding of the emotional underpinnings of communication can lead to more effective and personalized educational experiences [3].

2.4. Embodied Conversational Agents in Educational Contexts

The use of ECAs in education, particularly for language learning, has garnered considerable interest. ECAs can act as effective tutors, offering personalized feedback and fostering an encouraging learning environment [4]. In language learning, ECAs' potential to simulate conversational contexts and provide immediate, context-relevant feedback presents a promising avenue for enhancing WtC.

A web-services based conversational agent designed to encourage WtC in the EFL context emphasizes the potential of conversational agents to simulate natural conversations and enhance learners' WtC in specific social contexts [5]. Further research explores the addition of communicative and affective strategies to an embodied conversational agent, aiming to increase second language learners' WtC. This research focuses on dialogue management models based on communication strategies (CS) and affective backchannels (AB) to foster natural and WtC-friendly conversations with learners. Findings suggest that incorporating both CS and AB into conversational agents can significantly improve learners' WtC, marking a crucial step towards creating more interactive and supportive language learning environments [6].

3. Methodology

3.1. System Architecture

The system architecture for this research integrates a comprehensive setup designed to analyze human-agent interactions, particularly in a restaurant context where the conversational agent, Peter, acts as a virtual waiter. This system employs various data collection methods to gather biometric and behavioral data, which are then used to

determine causality and inform the development of Peter's conversational strategies. The key components of the system include:

- **Heart Rate Monitoring**: The RookMotion Device, a wearable technology, is used to continuously monitor participants' heart rates, providing insights into their physiological responses during interactions with Peter.
- **Eye Tracking**: The Tobii Nano Pro Device captures participants' gaze patterns, including fixation duration and saccades, to infer levels of attention and engagement during the ordering process.
- **Facial Emotion Recognition**: A built-in camera, integrated with OpenFace software (an open-source software), analyzes facial expressions to identify emotional states such as satisfaction, confusion, or frustration.
- **Peter Conversational Agent**: An advanced agent developed in Unity, designed to simulate realistic dialogues with participants in a restaurant setting.

Figure 1 : The user interface of Peter - Conversational Agent

- **Aguida Dashboard GUI**: This graphical user interface is used to connect and configure the devices, collect, visualize, and manage the collected data, providing feedback and insights.

3.2. Data Analysis

The analysis of the collected biometric and behavioral data employs machine learning techniques to infer user states and adjust the conversational agent's responses accordingly. The methodology includes several key steps and techniques:

- **Preprocessing:** This initial step involves cleaning and normalizing the biometric data (heart rate, eye tracking, facial expressions) to prepare it for analysis.

- **Feature Extraction:** Key features indicative of participants' engagement levels, emotional states, and interaction patterns are identified from the raw data.
- **Classification and Regression Models:** Supervised learning algorithms, such as Support Vector Machines (SVM), Random Forests, and Neural Networks, are used to classify emotional states and predict engagement levels.
- **Time-Series Analysis:** This technique analyzes the sequential nature of biometric data to understand the dynamics of participants' actions over time.
- **Evaluation Metrics:** The effectiveness of classification models is assessed using metrics such as accuracy, precision, recall, and F1 score. For regression predictions, mean squared error (MSE) is used. Additionally, user experience and willingness to communicate during the ordering process are evaluated through measures such as task completion time, the number of errors, and participant feedback.

4. Evaluation

4.1. Evaluation Methods

To assess the system's impact on users' Willingness to Communicate (WtC) and their overall interaction experience, a mixed-methods approach was adopted, encompassing both qualitative and quantitative evaluation strategies:

4.1.1. Quantitative Analysis

- **Pre- and Post-Interaction Surveys**: Changes in WtC were measured using pre- and post-interaction surveys.
- **Interaction Log Analysis**: Usage patterns and engagement frequency with the conversational agent were analyzed.
- **Performance Metrics**: Task completion times and error rates during interactions with the agent were recorded and analyzed.
- **Statistical Analysis**: Correlations between biometric data (heart rate, eye movement, facial expressions) and levels of engagement and WtC improvement were evaluated.

4.1.2. Qualitative Analysis

- **Content Analysis**: Verbal responses were analyzed for indications of confidence, nervousness, and willingness to communicate.

4.1.3. Usability Testing

- **User Task Analysis**: Observations of how participants interacted with the system helped identify potential points of friction.

4.1.4. Additional Evaluation Metrics

- **Preference Survey**: At the end of the experiment, participants completed a survey to express their system preferences.
- **Observation of Task Duration and Completion**: The duration and completion of tasks were observed to evaluate efficiency and engagement.

4.1.5. Preliminary Results and Feedback

Preliminary results from the interactions with nine participants show positive user engagement:

- **Increased Willingness to Communicate**: Users exhibited a higher willingness to engage in conversations with the agent.
- **System Reliability**: The system effectively integrated biometric data with conversational strategies, though some improvements in response times and data synchronization are noted.

5. Discussion

5.1. Analysis of Findings

The analysis focused on the relationship between the system's input variables and the improvement of users' Willingness to Communicate (WtC). Key findings include:

- **Pre- and Post-Interaction Comparison**:
  - WtC scores showed significant improvement post-interaction, indicating the system's effectiveness. Specifically, the average WtC score increased by 20%, demonstrating the positive impact of the system.
  - Confidence levels also saw a notable rise from an average score of 3.2 to 4.5 on a 5-point scale, indicating increased self-assurance in participants during interactions.
  - For instance, one participant's WtC score improved from 2.00 to 3.00 on a 3-point scale.
- **Biometric Data Correlations**:
  - Higher engagement levels were correlated with specific biometric indicators, such as consistent eye contact and lower heart rate variability. For instance, participants with stable heart rate variability were 30% more likely to maintain eye contact, which is a key indicator of engagement.
  - Facial emotion analysis showed that positive emotions (e.g., happiness) were prevalent in 60% of the interactions, as measured by Action Unit (AU) intensities related to smiling.
  - The average heart rate during interactions varied significantly, with standard deviations ranging from 3.46 to 45.32, showing different stress and engagement levels among participants.
  - Example: The average heart rate decreased from 89.20 to 72.20 during the interaction sessions, indicating reduced nervousness and increased comfort.

- **Conversational Strategies**:
  - Strategies such as Affective Backchannels (AB) and Conversational Strategies (CS) proved effective in enhancing user engagement and communication willingness. The use of AB alone resulted in a 15% increase in WtC scores, while the combination of AB+CS resulted in a 25% increase.
  - Preferred system choices among participants were:
    - AB: Chosen by 22.2% (2 out of 9 participants).
    - CS: Chosen by 55.6% (5 out of 9 participants).
    - AB+CS: Chosen by 22.2% (2 out of 9 participants).

5.2. Implications for Future Research

The initial results and user feedback provide a foundation for future research, which could explore:

- **Longitudinal Impact**: Extending the study to evaluate the sustained impact on WtC and language proficiency. This would involve tracking participants over several months to assess long-term benefits and retention of language skills.
- **Advanced AI Integration**: Investigating the potential of incorporating more advanced AI features, such as natural language understanding and generation, to further personalize interactions and learning experiences. This could enhance the system's ability to respond to complex and nuanced user inputs.
- **Cross-Cultural Communication**: Exploring the system's effectiveness in diverse cultural contexts, which might influence communication styles and WtC. Understanding cultural differences could lead to more tailored and culturally sensitive conversational strategies.

5.3. Potential Enhancements to the System

Based on initial feedback, several enhancements to the system are considered:

- **Improved Responsiveness**: Enhancing the system's ability to process and respond to biometric data in real-time to create more fluid interactions. This includes optimizing the data processing pipeline to reduce latency.
- **User Interface Customization**: Developing a more customizable UI that can adjust to user preferences and learning styles. This could involve offering different themes, interaction modes, and personalized feedback mechanisms.
- **Data Integration**: Streamlining the integration of conversational data and biometric feedback to provide more coherent and contextually relevant responses from the agent. This could improve the overall interaction quality and user satisfaction.
- **Expansion of Conversational Domains**: Including a wider array of conversational topics and scenarios to cater to different interests and needs of language learners. This could make the system more engaging and relevant to users' real-life communication needs.

The study involved 9 participants, and detailed reports were generated to analyze the effectiveness of the system across various metrics, including confidence, nervousness, and WtC during both the first and second experiments. The results indicated that the

implemented strategies significantly improved the participants' willingness to communicate, with notable enhancements in confidence. However, regarding the reductions in nervousness, more insights are needed to fully understand the impact. The findings suggest that integrating biometric feedback with conversational strategies can effectively enhance language learning experiences.

6. Conclusion

This study explored the integration of biometric feedback with conversational strategies to enhance Willingness to Communicate (WtC) in human-agent interactions. The system, incorporating heart rate monitoring, eye tracking, and facial emotion recognition, demonstrated significant improvements in WtC scores, confidence levels, and engagement during interactions with the conversational agent, Peter. The findings underscore the potential of using biometric indicators to inform and refine conversational strategies, thereby creating more adaptive and responsive educational technologies.

The results of this study align with the initial hypotheses. The individual analysis of biometric indicators, eye tracking, and emotional facial cues significantly boosts the predictive accuracy of WtC gain. This hypothesis was supported by the data, as the distinct impacts of these factors were clearly demonstrated through improved WtC scores and engagement levels. Additionally, the synergistic effect of integrating biometric indicators, eye tracking, emotional facial cues, and dialogue strategies (AB, CS, AB+CS) significantly enhances the ability to predict and improve WtC gain in human-agent communications. The combination of these elements resulted in a deeper understanding and more effective enhancement of WtC, particularly evident in the substantial increases in WtC scores and confidence levels.

Key outcomes highlighted the effectiveness of Affective Backchannels (AB), Conversational Strategies (CS) and their combination (AB+CS) in enhancing user engagement and communication willingness. The strategies proved beneficial in increasing confidence and reducing nervousness to some extent, although further insights are needed to fully understand the impact on nervousness.

Acknowledgements

References

[1] MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. The Modern Language Journal, 82(4), 545-562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

[2] Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). Embodied conversational agents. MIT Press. ISBN: 9780262032780

[3] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine, 18(1), 32-80. https://doi.org/10.1109/79.911197

[4] Ayedoun, E., Hayashi, Y., & Seta, K. (2016). Web-services based conversational agent to encourage willingness to communicate in the EFL context. Information and Systems in Education, 15(1), 15–27. https://doi.org/10.12937/ejsise.15.15

[5] Ayedoun, E., Hayashi, Y., & Seta, K. (2018). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. International Journal of Artificial Intelligence in Education. https://doi.org/10.1007/s40593-018-0171-6

[6] Picard, R. W. (1997). Affective computing. MIT Press. Paperback. ISBN: 9780262661157

[7] Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. Image and Vision Computing, 27(12), 1743-1759. https://doi.org/10.1016/j.imavis.2008.11.007

[8] Baron-Cohen, S., Tager-Flusberg, H., & Lombardo, M. V. (Eds.). (2013). Understanding other minds: Perspectives from developmental social neuroscience (3rd ed.). Oxford University Press. ISBN: 9780199692972

[9] Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. University of Texas at Austin. https://repositories.lib.utexas.edu/handle/2152/31333

[10] Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2016). OpenFace: An open source facial behavior analysis toolkit. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), 1-10. https://doi.org/10.1109/WACV.2016.7477553