UNIWERSYTET JANA DŁUGOSZA W CZĘSTOCHOWIE
Edukacyjna Analiza Transakcyjna 2023, nr 12



http://dx.doi.org/10.16926/eat.2023.12.06

Jacek SZEDEL
https://orcid.org/0000-0002-1925-7186
Silesian University of Technology
e-mail: Jacek.Szedel@polsl.pl

Bożena WIECZOREK
https://orcid.org/0000-0002-4506-2841
Silesian University of Technology
e-mail: Bozena.Wieczorek@Polsl.pl

# Estimating Ego States: The Machine Learning Perspective

**How to cite [jak cytować]:** Szedel, J., Wieczorek, B. (2023). Estimating Ego States: The Machine Learning Perspective. *Edukacyjna Analiza Transakcyjna*, *12*, 103-119.

## Abstract

The interface between Artificial Intelligence (AI), machine learning (ML) and psychology is an intensively explored research area. Specifically, Transactional Analysis (TA), with its structured and precise language, presents a promising area for applying ML techniques, unveiling new potential research avenues. This article explores the intersection of artificial intelligence, machine learning, and psychology, focusing on developing a method and the software environment for estimating ego states using MS Kinect™ sensor data. The research investigates the application of TA theory, emphasizing capturing the behavioural ego indicators. While Kinect skeletal data is considered, the gestures and postures are the primary input. The Authors present an innovative approach to annotate and visualize Kinect data using video streams for further autonomous ego state estimation. Within this study, they collected a dataset of 15 students from The Silesian University of Technology. The data was acquired through the use of both a video camera and a Kinect sensor. The nine distinct labels were employed for data annotation. They reflect Parent, Adult, and Child ego states across different temporal dimensions encompassing the past, present, and future. The study includes preliminary results demonstrating the outcomes of this approach's visualization technique and their interpretation. The final part of the article discusses the potential of applying the presented method in applications for the education field.

**Keywords:** ego state estimation, ego states in temporal dimensions, artificial intelligence and machine learning, MS Kinect™.

## Introduction

Integrating artificial intelligence (AI) and machine learning (ML) with psychology establishes an important interdisciplinary domain, delving into the intricate dynamics between humans and technology. This multifaceted field investigates the capacity of computer algorithms to replicate, simulate, and elevate the human-like cognitive functions of information systems. Inside this realm, artificial intelligence, as a branch of computer science, directs its efforts toward crafting intelligent agents proficient in learning, reasoning, and problem-solving. Simultaneously, machine learning empowers systems to evolve and elevate their functionalities, fostering adaptability and autonomy in their operations.

One of the rapidly advancing research domains in applying machine learning algorithms to psychology is sentiment analysis, a field predominantly concerned with recognizing emotional states. Images, video sequences, and texts serve as inputs for these solutions. Much of the research has focused on discerning the eight Eckman emotions (Eckman, 1984), but recent works have investigated a more nuanced examination of more fine-grained emotion datasets. For example, the 'Go Emotions' project by Google Research (Alon, Jeongwoo, 2021) introduced a dataset that distinguishes 12 positive emotions, 11 negative ones, 4 ambiguous, and one neutral. While most works concentrate on recognizing emotions, some authors also address other tasks like verifying the sincerity of it. One example can be assessing a smile's genuineness (Nurzyńska, Smolka, 2017). This work examines a dataset of video sequences, including spontaneous and posed smiles, of the UvA-NEMO database (Dibeklioğlu et al., 2012).

A particularly intricate challenge within the convergence of computer science and psychology revolves around capturing and emulating personality traits. This complex field encompasses several compelling problems, such as the formulation and examination of questionnaires, mobile and ubiquitous capture, recommender systems, autobiographical analysis, text analysis, and personality profiling derived from the activities of computer game users. These issues are explored and discussed in the book "Personality Capture and Emulation" by William Sims Bainbridge (Bainbridge, 2013), providing comprehensive insights into the intricate processes involved in understanding and replicating human personality within the digital realm.

From the computer scientist's perspective, a promising direction for developing psychology-related applications of artificial intelligence and machine learning is the theory of Transactional Analysis (TA). As elaborated in the following chapter, several papers have investigated this area so far. It is important to highlight that TA provides concise and enumerative structures, such as ego states, types of transactions, game scenarios, live positions, life script drivers, and many more. Given that machine learning algorithms excel at identifying elements within finite sets of objects, this presents an opportunity to leverage TA's well-defined constructs for enhancing the understanding and modelling of human psychological processes just by integrating TA principles into AI and ML frameworks. The mentioned concept considered within the vast field of knowledge of TA is no straightforward task. However, starting from the basics, like analyzing personality structure through estimating ego states, can lead to achieving valuable results with a broad spectrum of applications.

The goal of this work is to develop a method and software for labelling a data set consisting of video recordings and skeletal data sequences obtained from a Kinect sensor. The solution devised aims to autonomously assess ego states by leveraging behavioural ego indicators, emphasizing the nuances of body language and using machine learning techniques. In the current version, skeletal data are analyzed. However, the program is designed to be expanded with functions for capturing data related to other modalities (e.g., facial expressions). The later part of the article presents selected works related to the discussed topic, examines the significance of body language in transactional analysis, refers to the implemented method of marking samples, introduces the proposed approach with an overview of the research methodology and developed software, and presents initial preliminary results. The final chapter contains conclusions and explores potential applications of the developed solution in education.

## Related work

This chapter provides a brief survey of works that employ the knowledge of transactional analysis for solving problems at the interface of psychology and artificial intelligence. The presented overview focuses on papers that address three essential tasks: ego state estimation, enhancing the affective capabilities of intelligent agents, and simulating crowds or social networks. All selected works aim to capture, model, and emulate human personality at the TA personality structure level. It is worth noting that some elements of Transactional Analysis also appear in the context of other problems related to information technologies, like applying TA-based interviews in cognitive systems (Vartanov et al., 2023), managing interpersonal communication (Zhao et al., 2023), or improving software usability (Yamaoka, Yoshida, 2007). However, we decided to consider only threads closely associated with artificial intelligence and machine learning as they are vital for this work. The remainder of this chapter defines the three issues mentioned earlier and discusses example solutions.

Ego state estimation is a typical machine learning classification problem in which a system assigns the input data a predefined category or label. The number of labels for ego state recognition depends on the model and the level of personality structure analysis. For example, in the structural model, the labels should correspond to the three ego states (P, A, C) and the functional model to the five (NP, CP, A, FC, AC). As with all classification tasks, ego state estimation involves preliminary stages such as data capturing, preprocessing, labelling, and model learning. The input data can come from texts and speech recognition, audio and video recordings, streams, and motion sensors like Microsoft Kinect ™. The classification implements the supervised learning model. Hence, the labels have to be assigned by human annotators (labellers). Minamikawa and Yokoyama (2011, March 2011) proposed a technique for estimating the egograms of Japanese bloggers as an alternate method for those using the Five-Factor Model (FFM). In this approach, the input data for the classification was determined by a statistical analysis of the blog text regarding the presence of specific words and textual emoticons, both called "feature words". The researched dataset included texts from authentic personal weblogs of 551 individuals. The subjects also answered the Todai-shiki Egogram examination ver. 2, TEG2 (TEG Research Group, 2000). The Multinomial Naïve Bayes classifier – a machine learning algorithm usually used for natural language processing tasks, generated predictions of ego state labels based on the test data. The method's efficacy was assessed by comparing prediction outcomes with those derived from the TEG2 questionnaire. The accuracy results for three and five ego states ranged from 12% to 20%.

Intelligent agents are autonomous software programs embedded with artificial intelligence to achieve specific goals. They operate independently, making decisions based on their environment. Widely used in robotics, customer service chatbots, and data analysis, these agents are characterized by autonomy, social ability, responsiveness, and proactiveness. As AI evolves, its complexity and applications in automating and improving efficiency across various sectors increase. As in ego state estimation, TA elements have already been employed in this field. Fujita et al. (2009) described an agent capturing the personality of Japanese contemporary writer Miyazawa Kenji. The authors created an agent capable of interacting with users through the mental replication of the person the writer represents. The human user and system interaction was achieved using universal Ekman emotions (Eckman, 1984). The cited work is an example of the application of so-called historical analysis as a part of the functional analysis subdomain of TA, as described by Stewart and Jones (2016, pp. 62-63). The developed intelligent agent allowed the system users to interact with it, capturing their speech features and facial expressions.

Another field where the personality structure theory of TA was applied is simulating crowds and social networks. It is a complex field combining computer science, sociology, and psychology. It involves creating virtual environments where individual agents representing people interact according to specific rules and behaviours. The aim is to mimic the dynamics of real-world social interactions and crowd behaviours. These simulations are crucial for understanding group dynamics, like how opinions spread through a network or how crowds respond to emergencies. This technology has wide-ranging applications, from urban planning and event management, studying the spread of information or diseases through a population, to creating reliable crowd animations. Rigs and Egbert (2012) introduced a social crowd simulation algorithm designed to improve how crowd behaviour is simulated. At its core, this algorithm is tailored to address the evolving social needs of individual agents, enabling them to dynamically engage in or exit social encounters, thus reflecting real-world social fluidity. Based on the principles of transactional analysis from psychology, the algorithm emulates the emotional engagement of agents in their conversation, which allows for emulating the different characteristics of the duration of the agent's mutual interaction.

## Body language in transactional analysis

According to Alexander Lowen (Lowen, 1971, p. 7), "The living organism expresses itself in movement more clearly than in words. […] In pose, in posture, in attitude and in every gesture the organism speaks a language which antedates and transcends its verbal expression." Body posture requires very strong control of movements. It is possible to comprehend this area with consciousness, but only to a limited extent.

Body language plays a significant role in transactional analysis. Following the official website dedicated to Eric Berne, "one must look at how the words are being delivered (accents on particular words, changes in tone, volume, etc.) as the non-verbal signs accompanying those words (body language, facial expressions, etc.). Transactional Analysts will pay attention to these cues when analyzing a transaction and identifying which ego states are involved" (Berne, 2022).

Recognizing different ego states may be assisted by the table 1.

Table 1
*Recognizing different ego states*

| | **Controlling Parent** | **Nurturing Parent** | **Adult** | **Natural Child** | **Adapted Child** | **Little professor** |
|---|---|---|---|---|---|---|
| Words | Bad, should, ought, don't | Good, nice, well done | How, why, who, yes, no | Fun, want, mine | Can't, wish, please, thank you | I've got an idea |
| Gestures Postures | Pointing finger, pounding table, shaking head | Open arms | Straight posture, level eye contact | Energetic, loose-imbed | Slumped, dejected, nail-biting | Batting eyelashes |
| Tone of voice | Sneering, condescending | Loving, encouraging, concerned | Calm, clear, even, confident | Loud, free | Whining, sulking, defiant | Teasing |
| Facial expression | Scowl, hostile, disapproving | Smiling | Thoughtful, alert eyes | Joyful, twinkling eyes | Fearful, pouting | Wide-eyed, 'innocent' |

Source: (Enock, 2006, Leigh-Hunt, 2016).

An important facet of understanding the meaning of the human way of self-expression within transactional analysis is its functional part, including the behavioural analysis component. It defines what is known as behavioural indicators of the ego state. Through functional analysis and these behavioural markers, which encompass aspects such as speech content, body posture, voice tone, gestures, and facial expressions, it's possible to infer the structural state in which an individual is currently operating (Pankowska 2010, p. 41).

## Ego state timeline model

Proposed by Zbigniew Wieczorek (Wieczorek, 2017), an alternative to the personality adaptation model uses the classic language of transactional analysis. It can be employed to make a preliminary diagnosis and to determine the direction of work with the client. Assuming that a healthy functioning person is in touch with the past, projects the future, and has easy access to all of the ego states, a procedure to assess and recognize different states of the self during a communication timeline is developed. This approach can be illustrated by the diagram in Figure 1, which shows the smooth boundaries between states of the self in the past, present, and future.

| **The ego state in the past** | **The ego state at present** | **The ego state in the future** |
|---|---|---|
| P | P | P |
| A | A | A |
| C | C | C |

*Figure 1*
Diagram with boxes corresponding to the ego states in the past, present and future perspective

Source: (Wieczorek, 2017).

The Author of this model suggests working with the client by identifying which state of the self dominates at the beginning of the conversation, which states come up later, or which states the client cannot leave. Using the diagram in Figure 1, the therapist creates quick notes during each meeting, analyses changes, and explores possible areas that need strengthening.

In pursuit of the research's primary objective – estimating ego states, a prerequisite task was preparing a dataset for training supervised machine learning algorithms. The methodology advocated by Wieczorek was used to prepare the software for annotating the dataset incorporating Kinect skeletal data and the audio and video recordings.

### The proposed approach

### Method

As mentioned earlier, this research aims to develop a software environment that will allow the users to annotate a dataset that we can use to train machine learning algorithms, whatever they may be, to estimate ego states effectively based on Kinect data. In this study, we focused on Kinect skeletal data; however, future extensions may consider another input source of the motion sensors. The vital element of the applied method is the labelling scheme. In our approach, the starting point is the video recordings in which the labeller can recognize so-called behavioural indicators of ego states as described by Stewart and Joines (2016, pp. 58-60), by Pankowska (2010, p. 41) or Harsh (2021). In this study, we used the previously described nine-element matrix of Parent, Adult, and Child regarding the temporal perspective for the labelling. The primary source of information for the labeller is the recorded movie itself, but in particular, the recorded person's statements content. The data being labelled is the body movement data recorded in parallel with the video using the MS Kinect™ sensor. Kinect skeletal data is a representative data source for that scheme. It removes any unimportant details of the analyzed video sequence and extracts only crucial information about what is called, in Kinect terms, the bones and the joints, and it creates a data stream that is convenient for learning ML algorithms. Therefore, we decided to develop software that allows us to record the Kinect data accompanied by video recordings, supplying the functionality to conjunct those two data sources for further labelling according to observed ego state manifestations. As already mentioned, in this work, the content of the participant's statements was the primary information to explore ego states in the past, present, and future perspectives.

The recordings involved 15 students from the Silesian University of Technology. Participants were tasked with discussing a currently active project they were involved in, focusing on ongoing endeavours rather than past projects. This selection criterion was employed based on the understanding that discussions about present projects inherently encompass past, present, and future elements. At the beginning of the recording session, students provided a concise introduction to the project topic. Subsequently, they were prompted with questions exploring various dimensions of the ego state matrix, spanning past, present, and future (as shown in Figure 1). In instances where responses were deemed too laconic, supplementary questions were posed to elicit more detailed reflections. These additional inquiries aimed to prompt participants to articulate specific opinions, feelings, or judgments while considering the varying ego states and temporal dimensions. The questions posed during the recording sessions are outlined in Table 2. The table provides additional information regarding the temporal perspectives each question addresses.

Table 2
*The questions asked during recordings and the time perspectives they cover*

| Questions | Time perspectives |
|---|---|
| How does your current project align with your interests?<br>Is this project a good fun for you? | present |
| What specific knowledge or skills did you need to acquire before initiating the project? | past |
| Could you share challenges you faced initially, current obstacles, and those yet to be resolved? | past, present, future |
| Has anything particularly unpleasant happened while you worked on the project? If so, for what reason, judge the cause. | past |
| Reflecting on your emotions and feelings toward the project, how have they evolved?<br>Is there anything related to the project that you are currently concerned about? | past, present, future |
| Assess the effectiveness of the tools you employed throughout the project.<br>Did any software component you were using malfunction? Was it annoying?<br>Will you be using the same tools in your following projects? | past, future |
| What potential avenues do you envision for the future progression of the project? | future |
| Would you again select the same topic for your project, given the opportunity?<br>Are you okay with the exact topic of your thesis? | past, present |
| Do you anticipate the skills developed during this project will be valuable in your future endeavours? | future |

Source: own research.

## An overview of the designed system

The software designed as a part of the reported research embraces functionalities such as capturing audio and video streams synchronized with MS Kinect™ skeletal data, annotating recordings with a relevant label set, and ultimately visualizing the labelled samples. Following the labelling phase, the Support Vector Machine (SVM), a machine learning algorithm (Cortes, Vapnik, 1995), as an example ML solution, can be trained using chosen samples, and a cross-validation test can be conducted to evaluate the effectiveness of the learning process. The diagram in Figure 2 illustrates the overall structure of the created software environment. The Kinect sensor and a web camera capture the input data in this environment. The system includes two external components: the Kinect Software Development Kit (SDK) and the open-source video and audio processor – Ffmpeg. Facilitating the integration, the Kinect wrapper, a software library, augments and refines the capabilities provided by Kinect SDK. The essential part of the system is the main module that involves the functionalities mentioned earlier. It stores recorded and labelled samples in the system database.

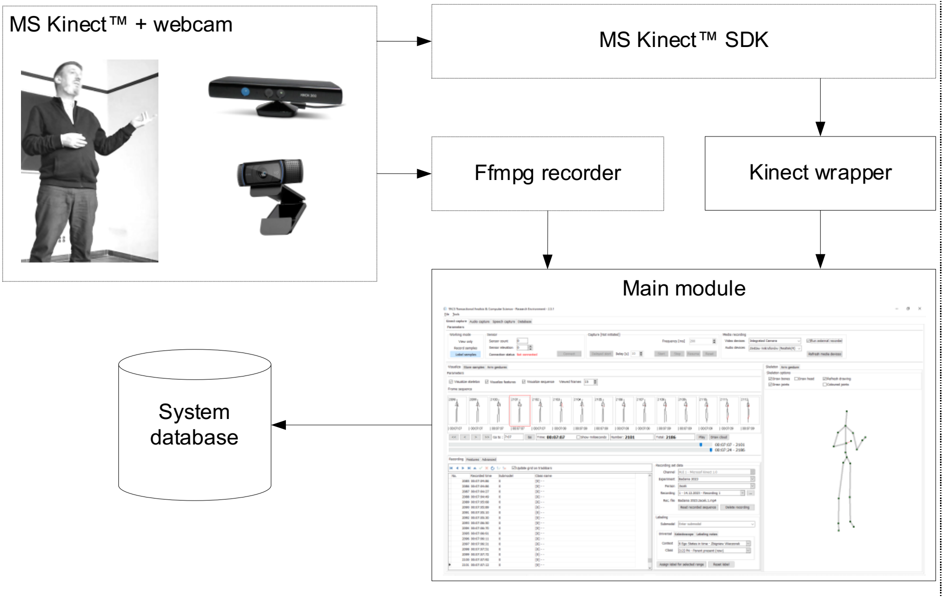


*Figure 2*
The overall structure of the designed software environment (solid lines indicate components developed by the Authors)

Source: own research.

The following figure (Fig. 3) depicts the structural organization of the developed database. The database aspect of system design is crucial, while the correct database design not only facilitates current operations but also lays the foundation for seamlessly incorporating new functionalities in the future. In the current phase of system development, the primary focus is on capturing skeletal data. However, the system design considers the future extension, involving other modalities like skeletal data in a seated position, facial expression, tone of voice, and others. As a result, the initial database object termed the "Capture Channel" is defined as the first element of the database structure. It represents the targeted modality. Subsequently, the "Experiment" becomes the next object in focus. It aggregates "Persons" and "Recordings". Each recording is affiliated with "Video files", and Kinect captured skeletal frames, denoted as "Samples". Following the recording, the ensuing activity involves labelling. Hence, the successive attribute assigned to a sample is the "Class", signifying a specific label. In this study, the utilized labels are limited to the nine categories Wieczorek (2017) advocated, referring to Parent, Adult, and Child ego states across past, present, and future temporal dimensions. The method we developed is based on the TA personality structural model. However, to accommodate future expansions, such as the inclusion of more detailed views like the five ego states of the functional model or even more specific division of the Parent and Child ego states into their positive and negative aspects (as discussed by Pankowska 2010, pp. 41-44), our database includes an additional object named "Classification context" which aggregates "Classes". Introducing the "Classification context" will enable the future classification of samples in versatile ways, accommodating varying levels of analysis. This flexibility presents an option for future research, allowing for dynamic exploration and interpretation of data based on different frameworks.

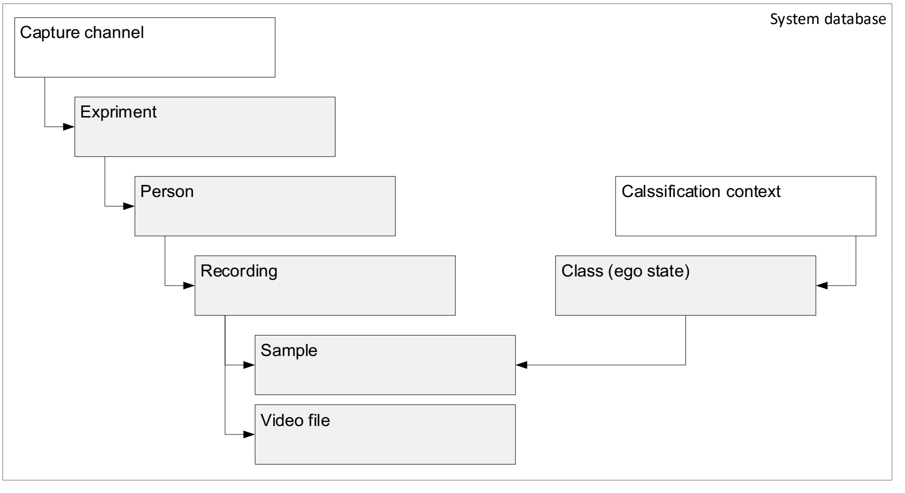


*Figure 3*
The organization of the system database

Source: own research.

## Data recording and labelling scenarios

The developed system incorporates two essential functionalities supporting the planned research methodology: sample recording and sample labelling, with an additional option to preview the captured data stream without immediate storage. In the recording mode, users select the relevant experiment and individual and then add a new recording to the database. The recording process necessitates the sensor to be in a "connected" state, and the subject must be within the frame (approximately 2 meters from the sensor). After initiation, the Ffmpg recorder launches, and the main module frame viewer displays captured frames. Figure 4 depicts the main module layout, highlighting key user interface elements such as the frame stream viewer, current frame viewer, recorded frames grid, recording selector, and label selector. The video recording program (Ffmpg) opens in a separate window, with frames synchronized with the recorded stream.

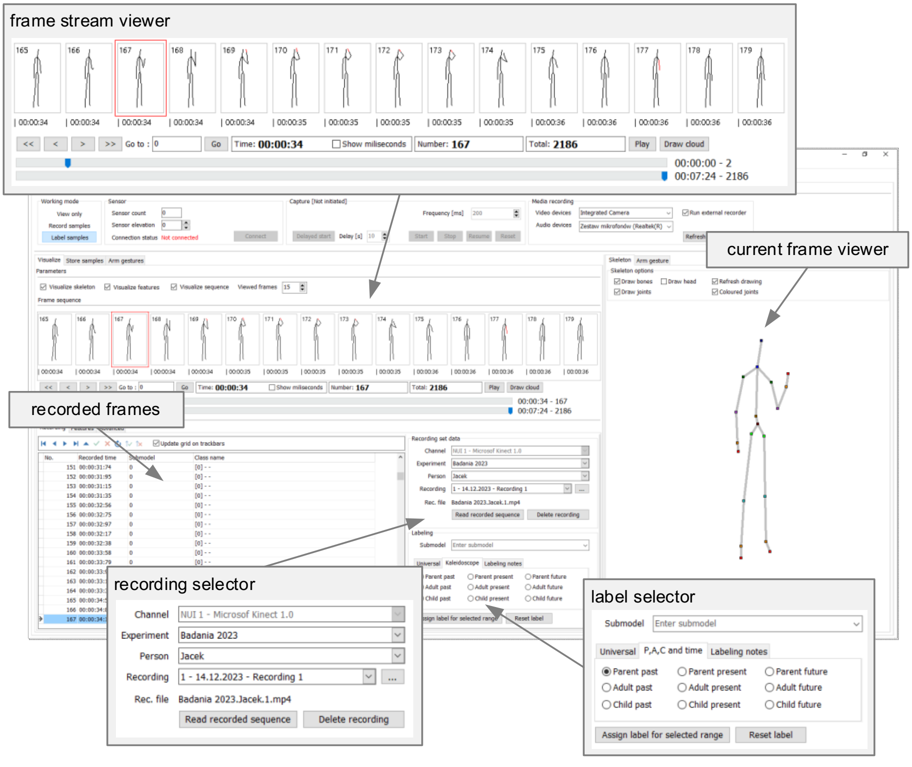


*Figure 4*
The illustration showing the layout of the main module and the key user interface elements

Source: own research.

The second scenario provided by the system involves the labelling of the recorded samples. In this scenario, the users initiate the labelling mode and proceed to select the experiment, individual, and recording. Subsequently, the frame sequence must be loaded into the visualization tool, and the playback of the video recording is then started. The labeller is tasked with observing individual Kinect frames and the video recording, actively seeking moments that exhibit the external signs of the ego states – in this study, the labeller focuses mainly on the content of statements made by participants. As emphasized in our conclusions, it is worth noting that the majority of recorded frames lack significance for the labelling process. Throughout our experiments, we set the frame-grabbing frequency at 200 ms, a parameter we identified as optimal. Consequently, only several essential frames create sequences that surfaced, deserving emphasis and appropriate labelling.

## Preliminary results

The primary output of this research is the software environment presented in the previous chapter. Besides the described functionalities, the software also allows users to present the recording and labelling results by visualizing recorded frames drawn superimposed on each other, which we refer to as the frame cloud. The first simple visualization option presents a frame cloud of one individual without distinguishing the assigned classes (labels). This straightforward method captures well the unique characteristics of a person's expression, reflected by the dynamics of their gestures and movements. The insights gained through this visualization technique allow a rough assessment of factors such as students' freedom of expression and emotional involvement during communication. The frame clouds obtained for three individuals are shown in Figure 5. The first participant (a) showed signs of stage fright, which they confirmed in a short interview immediately after the recording. In contrast, another participant (c) demonstrated a very uninhibited way of speaking by making vigorous movements. The frame cloud in the middle, designated as (b), represents a student who responded to the questions with a composed demeanour and emphasized vital points throughout their speech by incorporating hand gestures, adding a dynamic element to their calm and measured delivery.

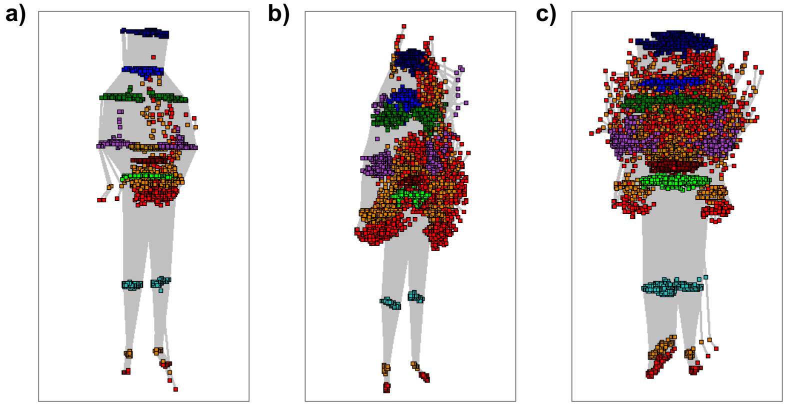


*Figure 5*
Three cloud frames showcase varying levels of expressiveness: a) the participant who showed signs of stage fright and nervousness, b) the second participant who delivered a measured and controlled response, and c) the third participant who showed dynamic movements, illustrating an exceedingly expressive and animated manner of speaking; joints are assigned colours, the red colour indicates hands and legs

Source: own research.

The developed program also implements a more advanced visualization in a view that presents clouds of frames organized according to labels associated with the Parent, Adult, and Child ego states and the temporal dimensions. This view is represented as a matrix of 12 panels, each displaying individual frame clouds. There are also additional panels presenting a sum of frames in rows and columns. The first row represents a sum of frames assigned to the past, present, and future, respectively – the first column panels aggregate frames within the specific ego states. Depending on the user's selection, the top-left panel shows all recorded frames or all with assigned labels.

An illustration of this visualization, generated for a sample individual, is presented in Figure 6

. This visualization enables drawing some conclusions. For instance, it is notable that no frames from the Parent state were obtained, which is a fact attributable to the context of the conversation: it was conducted in an adult student and academic teacher setting. A relatively high number of frames associated with the adult state was particularly noticeable, which is also related to the context of the recording. When the participant spoke about the past, they leaned to the left side and extended their left arm. There are also frames relevant to the Child's ego state. They are linked to hand movements at shoulder height and towards the head. As we expected, the visualization not only explains the dynamics of the conversation but also provides insights into the non-verbal cues associated with different ego states and temporal references.

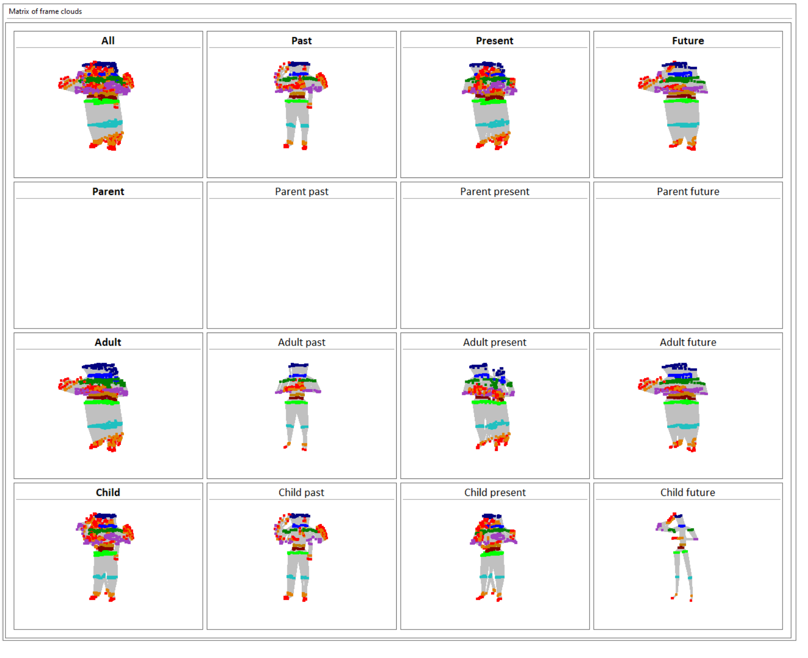


*Figure 6*
The visualization of the cloud frames corresponding to the individual ego states and time dimensions recorded for a single individual

Source: own research.

## Conclusions

This article described the methodology and the software for applying artificial intelligence and machine learning algorithms for automatically estimating ego states based on data obtained from the Kinect sensor. The presented method used video recordings as a starting point for Kinect data labelling. Behavioural indicators of ego states were sought in the recordings, including their temporal dimensions: past, present and future. The Author's work results in the developed software environment that offers functions for recording the input data, labelling it, and visualizing the results. This software was designed to be expandable in the future, including modalities other than skeletal data and testing various data classification algorithms.

The analysis of the preliminary results has yielded several noteworthy observations. Firstly, based on the collected samples, we observed a significant variation in the expression among individual participants, manifested as differences in gestural dynamics. Secondly, we noticed the presence of individual indicators of ego states, validating observations documented in the literature where general and exclusive behavioural indicators were postulated (Stewart, Johnes, 2016, pp. 58-60). Occasionally, we observed signs of stage fright in some of the research participants. We tried to create a casual atmosphere during the recordings. Still, the conversations sometimes followed a somewhat formalized way. A valuable aspect of our developed method is the visualization it offers in the form of a frame cloud, particularly as a matrix representing ego states and temporal dimensions. This visual representation enabled us to analyse ego states' activation among individual participants. The analysis concluded that activating parent states is difficult in the setting in which university students talk to their teachers. We recognize the importance of minimizing any potential influence from the person conducting the study on the participants. However, due to the constraints of this particular experiment, we could not arrange the recordings in a way that eliminates such influence. In the future, a procedure will be developed to reduce this interaction.

There are undoubtedly open challenges and problems to be resolved, as well as exciting opportunities for the future evolution of our developed method. In this article, we have highlighted the first avenue of development, which involves expanding the software's capabilities to capture additional information, such as facial expressions or gesture analysis in a seated position. Another possibility lies in streamlining the labelling process through automatic text description; however, despite the excellence of current speech-to-text algorithms, exploring more advanced voice recording tools would be imperative in such a solution. While analysing the applied method, we stated that creating a recording environment capable of activating the balanced ego state set (including the Parent) is pivotal in developing our approach, and it should be a crucial future consideration of the Authors.

The developed method finds its primary application in supporting participants within the educational sphere, encompassing both educators and students. One compelling application lies in teacher training, where recordings with the designed software would serve as a valuable tool for instructors to refine their expressive techniques. This approach shall strike a balance incorporating the Adult ego state's characteristics in expression while infusing a sense of passion and engagement reminiscent of the Child. Another avenue for the presented utility is in the creation of multimedia educational content, such as podcasts where the presenters feature. In this scenario, the goal is to analyze ego states conducive to effective teaching and learning processes. Moving towards statement analysis – the speech capture channel in the software analysing ego states, there's the potential to employ a similar method for examining texts within online communications or asynchronous text exchanges between students and educators. Beyond this, there's an intriguing prospect of delving into automatic transaction analysis, another facet of transactional analysis, where the solutions delineated in this paper could, as we believe, be beneficial.

## References

Alon, D., Jeongwoo K. (2021). *GoEmotions: A Dataset for Fine-Grained Emotion Classification.* Software Engineers. Google Research. https://blog.research.google/2021/10/goemotions-dataset-for-fine-grained.html [Accessed on 12.2023]

Bainbridge, W.S. (2013). *Personality capture and emulation*. Springer Science & Business Media; http://dx.doi.org/10.1007/978-1-4471-5604-8.

Berne's family members (2022). *Transactional analysis.* https://ericberne.com/transactional-analysis/ [Accessed on 12.2023].

Cortes, C., Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*, 273–297; http://dx.doi.org/10.1007/BF00994018.

Dibeklioğlu, H., Salah, A.A., Gevers, T. (2012). Are you really smiling at me? spontaneous versus posed enjoyment smiles. In: *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision*, *Florence, Italy, October 7–13, 2012, Proceedings, Part III 12* (pp. 525–538). Springer Berlin Heidelberg; http://dx.doi.org/10.1007/978-3-642-33712-3_38.

Ekman, P. (1984). Expression and the nature of emotion. *Approaches to Emotion*, *3*(19), 344.

Enock, K. (2006), Leigh-Hunt, N. (2016). Behaviour change in individuals and organisations. Retrieved from http://www.healthknowledge.org.uk/public-health-textbook/organisation-management/5a-understanding-itd/transactional-analysis [Accessed on 12.2023].

Fujita, H., Sawai, N., Hakura, J., Kurematsu, M. (2009). An action decision model for emotions based on transactional analysis. In: *Recent Advances in Artificial Intelligence, Knowledge Engineering and Data Bases. Proceedings of the 8th WSEAS international conference on Artificial intelligence, knowledge engineering and data bases* (*AIKED '09*). *Cambridge, UK February 21–23, 2009* (pp. 79–88). WSEAS Press.

Harsh, B. (2021). *Diagnosis of ego states. Institute for Counselling Transactional analysis*, Pune Mumbai Chapter. https://ictapunemumbai.org/diagnosis-of-ego-states [Accessed on 12.2023].

Lowen, A. (1971). *The language of the body*. New York: Collier Books.

Minamikawa, A., Yokoyama, H. (2011). Personality estimation based on weblog text classification. In: *Modern Approaches in Applied Intelligence: 24th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2011, Syracuse, NY, USA, June 28–July 1, 2011, Proceedings, Part II 24* (pp. 89–97). Springer Berlin Heidelberg; http://dx.doi.org/10.1007/978-3-642-21827-9_10.

Minamikawa, A., Yokoyama, H. (2011, March). Blog tells what kind of personality you have: egogram estimation from Japanese weblog. In: *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 217–220); http://dx.doi.org/10.1145/1958824.1958856.

Nurzyńska, K., Smolka, B. (2017). Spontaneous Smile Detection with Application of Landmark Points Supported by Visual Indications. In: *CS & IT Conference Proceedings* (Vol. 7, No. 1). CS & IT Conference Proceedings. http://dx.doi.org/10.5121/csit.2017.70118.

Pankowska, D. (2010). *Nauczyciel w perspektywie analizy transakcyjnej*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Ricks, B.C., Egbert, P.K. (2012). More realistic, flexible, and expressive social crowds using transactional analysis. *The Visual Computer*, *28*, 889–898.

Stewart, I., Joines, V. (2016). *Analiza transakcyjna dzisiaj*. Rebis. Poznań.

TEG Research Group – the University Tokyo Medical School. (2000). *TEG: Tokyo University Egogram* (*New Ver.*). Kaneko Shobou.

Vartanov, A.V., Izbasarova, S.A., Neroznikova, Y.M., Artamonov, I.M., Artamonova, Y.N., Vartanov, I.I. (2023). The effect of psychological mirroring in telecommunicative dialogue. *Cognitive Systems Research*, *80*, 110–117; http://dx.doi.org/10.1016/j.cogsys.2023.02.008.

Wieczorek, Z. (2017). Język zmiany w analizie transakcyjnej. *Edukacyjna Analiza Transakcyjna*, *6*, 145–156.

Yamaoka, A.T., Yoshida, M. (2007). Basic consideration for various interfaces from viewpoint of service. In: *Universal Access in Human-Computer Interaction. Applications and Services: 4th International Conference on Universal Access in Human-Computer Interaction, UAHCI 2007 Held as Part of HCI International 2007 Beijing, China, July 22-27, 2007 Proceedings, Part III 4* (pp. 990–996). Springer Berlin Heidelberg; http://dx.doi.org/10.1007/978-3-540-73283-9_107.

Zhao, X., Tao, M., Chen, Y., Yin, B., Markopoulos, P. (2023). Tailoring Persuasive Health Messages to the Predominant Ego State of Patients. In: *Lecture Notes in Computer Science* (*including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), 13832 LNCS, pp. 234–247; http://dx.doi.org/10.1007/978-3-031-30933-5_15.

## Estymacja stanów ego: perspektywa uczenia maszynowego

### Streszczenie

Styk dziedzin uczenia maszynowego (UM) i psychologii jest obszarem intensywnych badań od wielu lat. W tym zakresie Analiza Transakcyjna (AT), ze swoim strukturalnym i precyzyjnym językiem, stanowi obiecujący obszar dla zastosowań technik UM, wytyczając nowe potencjalne kierunki badawcze. Prezentowany artykuł odnosi się do połączenia metod sztucznej inteligencji, uczenia maszynowego i psychologii, skupiając się na opracowaniu metody i środowiska oprogramowania pozwalającego na estymację stanów Ja przy użyciu danych z sensora MS Kinect™. Referowane badania koncentrują się na rejestrowaniu behawioralnych wskaźników stanów Ja. Danymi wejściowymi są nagrania audio i wideo oraz tzw. dane szkieletowe z sensora MS Kinect™. Autorzy prezentują metodę etykietowania i wizualizowania danych z sensora Kinect. W ramach badania zebrano zestaw danych obejmujący nagrania 15 studentów z Politechniki Śląskiej. Do etykietowania danych wykorzystano dziewięć odrębnych etykiet, odzwierciedlających stany Ja Rodzic, Ja Dorosły i Ja Dziecko w różnych wymiarach czasowych obejmujących przeszłość, teraźniejszość i przyszłość. Podstawowym rozpatrywanym elementem stanowiącym źródło informacji wykorzystane podczas etykietowania była treść wypowiedzi uczestników. Zasadniczym rezultatem pracy jest oprogramowanie stanowiące środowisko badawcze, którego architektura pozwala na badanie różnych modalności a także różnych algorytmów klasyfikujących. W pracy omówiono wstępne wyniki obejmujące opracowane techniki wizualizacji oraz interpretację wynikających z nich obserwacji. Ostatnia część artykułu omawia możliwości zastosowania przedstawionej metody w obszarze edukacji.

**Słowa kluczowe:** estymacja stanów Ja, stany ja w ujęciu czasu, sztuczna inteligencja, uczenie maszynowe, MS Kinect™.