Journal of Multimedia Processing and Technologies

DLINE JOURNALS

Print ISSN: 0976-4127
Online ISSN: 0976-4135
JMPT 2026: 17 (3)
https://doi.org/10.6025/jmpt/2026/17/3/95-123

Temporal Dynamics and Latent Structures of Emotional States in Human–Robot Interaction: A Markovian and Hidden Markov Model Analysis of Longitudinal Facial Emotion Sequences

Pit Pichappan
Digital Information Research Labs
Chennai. Tamil Nadu India
pichappan@dirf.org

ABSTRACT

Understanding the temporal evolution of human emotions is critical for developing empathetic and adaptive Human Robot Interaction (HRI) systems. While existing research predominantly focuses on instantaneous emotion classification, it often overlooks the longitudinal dynamics and underlying latent affective structures. This study introduces the Temporal Emotion Dynamics and Latent Regime Framework (TEDLRF) to model the evolution of emotions as a stochastic temporal process. Utilizing a pre trained Vision Transformer for facial emotion recognition, we analyzed longitudinal sequences of five basic emotions through Markov Chains, Hidden Markov Models (HMM), and Dynamic Bayesian Networks.

Our analysis reveals that emotional states exhibit remarkable temporal inertia, with an Emotional Stability Index of 0.991 and self transition probabilities exceeding 97%. Furthermore, HMM state discovery identifies three distinct latent affective regimes: isolated anger, persistent sadness, and a flexible positive neutral subsystem. Survival analysis further demonstrates significant heterogeneity in emotional persistence, where sadness forms highly stable, long duration attractors, whereas surprise remains highly transient. Emotional evolution is characterized by quasi stationary phases punctuated by discrete regime shifts rather than continuous drift.

These findings indicate that observable facial expressions are manifestations of deeper, unobservable psychological states. Consequently, we propose that next generation adaptive robots must transition from reactive, mimetic responses to proactive, temporally aware architectures. By inferring latent emotional regimes and anticipating affective transitions via change point detection, robots can deliver context sensitive, empathetic behaviors, ultimately fostering more natural and effective human AI collaboration.

Keywords: Human-Robot Interaction, Temporal Emotion Dynamics, Hidden Markov Models, Markov Chains, Facial Emotion, Recognition, Latent Affective Regimes, Dynamic Bayesian Networks, Affective Computing

Received: 12 January 2026, Revised: 21 April 2026, Accepted: 25 May 2026

1. Introduction

Recent advances in robotics have expanded machine capabilities beyond intelligent task execution to include cognition, emotional responsiveness, and personality oriented interaction. Contemporary robotic systems increasingly aim to exhibit emotional intelligence, anthropomorphic behavior, empathy, and adaptive responses to support more natural interaction with humans. Consequently, emotional analysis and regulation have become central concerns in human robot interaction (HRI), making emotional modeling an essential research direction [1].

At the same time, the growing integration of artificial intelligence into decision making environments has shifted attention toward effective human AI collaboration. Successful collaboration extends beyond prediction accuracy and calibrated trust and requires coordination of goals, communication of uncertainty, shared control mechanisms, and continuous adaptation across repeated interactions. Existing studies frequently investigate these dimensions independently, limiting the understanding of human AI interaction as a dynamic and cooperative process [2].

2. Emotional Modeling and Theoretical Foundations

Emotional space models conceptualize emotions as positions within a continuous multidimensional framework in which each dimension captures a fundamental emotional property. Early foundations were established through Wundt's three dimensional emotional representation comprising pleasant unpleasant, excitement depression, and tension relaxation dimensions [3]. Subsequent emotional theories expanded these conceptualizations through multidimensional emotional structures [4, 5, 6, 7].

Among these approaches, the Pleasure Arousal Dominance (PAD) model proposed by Russell and Mehrabian has become one of the most influential frameworks for representing emotional states [8]. Further developments by Scherer and Ortony extended PAD based modeling for artificial emotion generation in social robotics [9, 10]. Likewise, Zecca et al. [11] introduced a three dimensional psychological vector space defined by arousal, pleasantness, and certainty while incorporating machine learning, dynamic regulation, and personality adaptation mechanisms.

2.1. Facial Emotion Recognition and Ethical Considerations

Facial analysis provides a largely unobtrusive mechanism for capturing affective responses during interaction. However, collecting facial recordings, particularly from children and students raises significant ethical and privacy concerns [12]. Reviews of affective computing in educational contexts demonstrate that advances in accessible sensor technologies have accelerated the automatic recognition of affective states across learning environments. Facial video analysis emerged as one of the most frequently used methods for emotion detection, second only to skin conductance measures [13].

To support longitudinal emotion analysis, this study adopts two complementary measurement dimensions: (i) the duration of facial expressions representing emotional states and (ii) transitions between emotional expressions over time. Expression extraction is based on facial action units, originally proposed by Carl-Herman Hjortsjö and subsequently formalized by Paul Ekman and Wallace Friesen as a systematic framework for facial movement interpretation [14].

2.2 Advances in Emotion-Aware Human–Robot Interaction

Recent progress in generative modeling has enabled increasingly realistic human motion synthesis. Human Motion Diffusion Models (HMDMs) demonstrated temporally coherent and stylistically consistent full body movement generation [15, 16]. Simultaneously, large language models (LLMs) have shown that linguistic instructions can be translated into low level embodied control policies [17, 18]. MotionGPT [12] further extended these capabilities toward high level motion generation.

Despite these advances, many existing frameworks prioritize semantic consistency while providing limited support for real time emotional adaptation and physical constraints. To address these limitations, ReSIn HR (Real time Synchronised Interaction for Humanoid Robots) introduced synchronised speech gesture generation to improve expressive interaction [19].

Early gesture generation approaches primarily relied on rule based systems [20], [21] and statistical techniques [22] [23]. Although effective in constrained scenarios, these methods exhibited limited generalizability because of handcrafted rules and small scale datasets. The emergence of 3D human pose estimation [24, 25] and deep temporal modeling [26, 27] shifted the field toward data driven learning of gesture patterns from speech, text, and identity information. Emotion conditioned methods subsequently incorporated valence arousal representations and affective generative adversarial models [28], although challenges remain in achieving spontaneous and real time interaction.

2.3 Markovian and Latent State Approaches in Human–Robot Interaction

Probabilistic and latent state approaches have become increasingly important for modeling interaction dynamics. Prasad [29] introduced a hybrid framework that integrates Hidden Markov Models (HMMs) as latent priors into a Variational Autoencoder to model joint distributions among interacting agents. By leveraging interaction dynamics learned from human human interaction and extending them to human robot interaction, the framework improved robot trajectory prediction through conditional motion generation.

Complementing this perspective, Sun proposed a system for dynamic emotion modeling in human computer interaction based on emotional transfer and guidance mechanisms. Emotional states and transition distributions were used to simulate emotional evolution and generate adaptive emotional guidance sequences [30].

Recent surveys further indicate that latent state modeling enables robots to move beyond reactive behavior toward anticipatory planning and proactive adaptation. Existing approaches generally fall into the categories of Bayesian, Markovian, and encoded latent dynamics and can be unified under predictive interaction frameworks [31].

2.4 Human–AI Collaboration and Emotion-Adaptive Systems

Bethapuri proposed a Human–AI Joint Decision-Making framework that conceptualizes collaboration as an adaptive cycle involving mutual modeling, bidirectional communication, shared control, and utility-based evaluation. The framework identifies several open challenges for constructing robust, adaptive, and accountable human–AI teams.

Similarly, Sharma analyzed emotional expression patterns among students engaged in collaborative Scratch activities using two emotional frameworks. Findings demonstrated that education-specific and negative emotional states were stronger predictors of objective performance, whereas generalized control-value emotions better explained subjective outcomes [32].

Mamodiya extended emotion aware adaptation through the Emotion Context Reinforced Planner, a framework designed to deliver emotionally responsive and contextsensitive robotic behavior in real time [33]. Chen further demonstrated practical implementation through a real time NAO robot framework that synchronized speech prosody and full body gestures for enhanced interaction quality [34].

As a highly active and critical area of research [2–6], face recognition utilizes various methodologies to handle complex variations. For instance, Hidden Markov Models (HMM) [16, 17] leverage prior knowledge of facial morphology and rely heavily on local characteristic points. By constructing a space of local features and applying tailored image filters, these model based approaches effectively account for non linearity. Consequently, while they are more technically challenging to implement, they offer the advantage of being less sensitive to environmental and behavioral changes, such as variations in illumination, pose, and facial expression.

Face recognition is a very active, important, and widely studied research area [35-39]. HMM [40, 41] is based on models and uses prior knowledge of facial morphology. They generally rely on its local characteristic points. These methods constitute another approach that accounts for non linearity by constructing a space of local features and applying appropriate image filters. Thus, face distributions are less affected by various changes (illumination, pose and facial expression variations) but more difficult to implement.

The literature collectively demonstrates that emotional intelligence, latent state modeling, and adaptive interaction mechanisms are becoming foundational elements of next generation human robot systems. Emotional space theories, facial affect recognition, Markovian modeling, diffusion based motion generation, and emotion-aware planning frameworks together provide the theoretical and methodological basis for longitudinal analysis of emotional states in HRI. These developments motivate the present study's focus on temporal emotional dynamics and latent structures through Markov and Hidden Markov approaches.

Although prior work has explored emotion recognition, adaptive interaction, and latent state modeling independently, limited attention has been given to understanding how facial emotional expressions evolve longitudinally during human robot interaction and whether these observable emotional trajectories emerge from underlying latent affective regimes. Existing studies frequently emphasize classification accuracy rather than temporal emotional persistence, transition mechanisms, and regime changes. This study addresses this gap by modeling emotional evolution as a stochastic temporal process using Markov Chains, Hidden Markov Models, and Dynamic Bayesian Networks.

The reviewed literature suggests three unresolved requirements for next-generation HRI systems: (i) temporal representation of emotional continuity, (ii) identification of latent emotional mechanisms beyond observable expressions, and (iii) adaptive behavioral generation based on evolving emotional states. These requirements motivate the development of the proposed Temporal Emotion Dynamics and Latent Regime Framework (TEDLRF), which integrates observable emotional transitions, hidden state discovery, persistence modeling, and adaptive response generation within a unified probabilistic architecture.

2.5 Proposed Model: Temporal Emotion Dynamics and Latent Regime Framework (TEDLRF)

1. Observable Emotional Process

$$\Sigma_{\tau}\ \varepsilon\ \{\, Angry, Happy, Neutral, Sad, Surprise \,\}$$

2. Markov Transition Layer

$$P\,(\,E_{t+1} = j \mid E_t = i)$$

Transition matrix:

$$T = \begin{bmatrix} 0.992 & 0 & 0 & 0 & 0.008 \\ 0.006 & 0.992 & 0\ .002 & 0 & 0 \\ 0.001 & 0.012 & 0.984 & 0.001 & 0.002 \\ 0 & 0 & 0 & 0.999 & 0 \\ 0.008 & 0.014 & 0.001 & 0 & 0.978 \end{bmatrix}$$

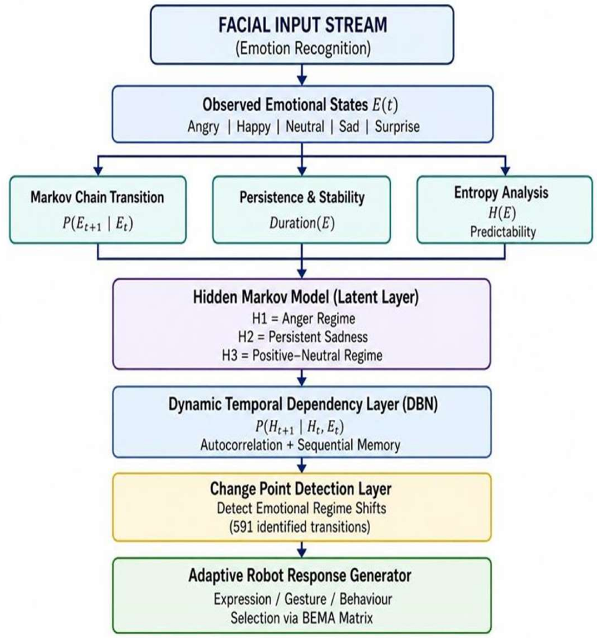

Figure 1. Integrated Temporal Emotion Dynamics Model for Human–Robot Interaction

3. Emotional Stability Function

$$ESI = \frac{\Sigma \text{ Self Transitions}}{\Sigma \text{ Total Transitions}}$$

Observed:

$$ESI = 0.991$$

4. Hidden Emotional Regimes

Latent state:

$$H_t \in \{H_1, H_2, H_3\}$$

Emission :

$$P(E_t \mid H_t)$$

Mapping:

- $H_1 \rightarrow$ Angry
- $H_2 \rightarrow$ Sad
- $H_3 \rightarrow$ Happy + Neutral + Surprise

5. Dynamic Bayesian Dependency

$$P(E_{t+1} \mid E_t, H_t)$$

Equivalent network:

$$E_t \rightarrow H_t \rightarrow E_{t+1}$$

6. Regime Change Detection

$$CP = \{ t: P(E_t) \neq \mathrm{P}(E_{t+1}) \}$$

Detected:

$$CP = 591$$

The model suggests that:

- Observed emotions are not independent events but manifestations of hidden affective regimes.
- Emotional evolution follows high-persistence Markov dynamics.
- Long emotional episodes create state-specific survival profiles.
- Regime transitions occur through discrete change points rather than continuous drift.
- Robot adaptation should therefore react to latent emotional states, not only instantaneous emotion labels.

The proposed model is a hierarchical temporal framework for affect aware human robot interaction. Facial input is first processed by an emotion recognition module to generate a time indexed sequence of observed emotional states, including angry, happy, neutral, sad, and surprised. The observed sequence is then analyzed through three complementary components: a Markov transition model to estimate short term state dynamics, a persistence module to quantify emotional stability and duration, and an entropy based analysis to measure affective variability and predictability. These features provide a structured representation of the observed emotional stream.

The resulting information is used to infer latent affective regimes through a Hidden Markov Model, where hidden states represent broader emotional conditions such as anger, persistent sadness, and positive neutral affect. A Dynamic Bayesian Network further models temporal dependencies among these latent statesaccording to, thereby capturing sequential memory and autocorrelated emotional evolution. To identify meaningful shifts in affective behavior, a change point detection layer is applied to the temporal state sequence, revealing regime transitions across the interaction. Finally, the inferred emotional regime and its temporal changes are mapped to an adaptive robot response generator, which selects context appropriate expressions, gestures, and behaviors using the BEMA matrix. This architecture enables the robot to respond not only to instantaneous facial expressions but also to the deeper temporal structure of human emotional dynamics.

3. Dataset

3.1 Dataset Overview

The resource comprises project files (total size approximately 1.32 MB) that implement and configure the BEMA system. It is not a large scale collection of raw images or annotated samples but rather an integrated toolkit centred on a pre trained Vision Transformer (ViT) model for emotion detection, combined with a configurable behaviour emotion mapping matrix. The framework supports real time or near real time inference on facial images to generate appropriate robot responses.

3.11 Key Components and Structure

• Facial Emotion Recognition Model: A fine-tuned Vision Transformer based on google/vit-base-patch16-224-in21k, sourced from the Hugging Face model dima806/facial_emotions_image_detection. It classifies input face images into one of five emotion categories with associated confidence scores:

• Angry

• Happy

• Neutral

• Sad

• Surprise

Input processing typically involves face detection, cropping, and preprocessing before classification.

• Robot Behaviour and Emotion Matrix: A core configurable mapping that takes the detected human emotion (and optionally robot state) as input and outputs suitable robot visual/behavioral responses. Robot states include:

• Neutral, Happy, Sad, Angry, Surprised (mirroring or adapting human emotions)

• Listening, Speaking (application controlled, non emotion derived states)

The mapping is designed to be non mimetic in some cases (e.g., responding to human anger with a neutral or calming robot expression) to promote positive or contextually appropriate HRI.

• Workflow: Human face image preprocessing →ViT classification → emotion → confidence →BEMA matrix → robot response (expression or behavior).

The included files likely encompass model integration code, matrix configuration (e.g., JSON, Python mappings, or lookup tables), example scripts, and supporting assets for deployment in robotic systems.

3.2 Data Characteristics

• **Modality**: Primarily supports static or streamed facial images (RGB). No large image corpus is bundled; the resource leverages the pre trained model and focuses on the mapping logic.

• **Classes**: 5 human emotions + 7 robot states.

• **Format**: Code/scripts (Python), model weights/configs (via Hugging Face), matrix definitions (likely tabular or rule based).

• **Size**: Compact (~1.3 MB), suitable for embedded or lightweight robotic deployments.

• **License**: Apache License 2.0 (with attribution requirements for external models/assets).

BEMA is positioned for deployment in:

• Social, wayfinding, reception, customer service, and educational robots.

• Interactive kiosks, digital avatars, and emotion aware interfaces.

• HRI research exploring adaptive, empathetic robot behaviors.

Limitations

• Performance sensitivity to lighting, pose, occlusion, image quality, and cultural variations in expression.

• Limited to the five trained emotion classes.

• Explicitly not intended for high stakes domains such as medical diagnosis, psychological assessment, hiring, surveillance, or law enforcement.

This resource provides a practical, modular foundation for researchers and developers in affective computing and robotics. For journal use, it can be cited as a framework dataset hybrid that enables reproducible HRI experiments, with extensions possible via custom matrix tuning or additional training data. Further details, including code and matrix specifics, are available directly on the Kaggle repository.

3.3 Experimental Workflow

The proposed analytical framework models the evolution of emotions during human robot interaction as a sequential probabilistic process that maps observable facial expressions into latent emotional structures and interpretable temporal patterns. The workflow consists of six consecutive stages: Input → Emotion Detection → Temporal Encoding → Markov Modeling → Hidden Markov Modeling → Dynamic Bayesian Network → Interpretation.

Stage 1: Input Acquisition

The workflow begins with the acquisition of human facial observations during interaction sessions. Input data consist of temporally ordered facial image frames captured continuously throughout the interaction process. Rather than treating individual observations independently, the proposed framework preserves temporal ordering to enable longitudinal analysis of emotional evolution.

Let the facial observation sequence collected during interaction be represented as:

$$I = \{ I_1, I_2, I_3, ..., I - T \}$$

where: I_t = facial image/frame captured at time t

T = total number of observations

The complete temporal input is therefore defined in the vector space:

$$I \_ t \in \_ \mathbb{R} \wedge (H \text{x} W \text{x} C)$$

where:

H = image height

W = image width

C = color channels

Before analysis, image frames undergo preprocessing procedures including face detection, facial region extraction, normalization, and preparation for emotion classification.

Stage 2: Emotion Detection

The pre-processed facial images are passed to the emotion recognition component, implemented with the pre-trained Vision Transformer (ViT) architecture within the BEMA framework. The model is based on google/vit-basepatch16 224 in21k and generates emotion predictions with associated confidence scores.

Each facial image is processed by the Vision Transformer (*ViT*) classifier:

$$E\_t = fViT\ (I\_t)$$

where: f_ViT = emotion classification function

E_t = predicted emotion at time t

Observed emotional labels belong to the set:

$$E\_t \in \{Angry, Happy, Neutral, Sad, Surprise\}$$

The resulting temporal emotional sequence becomes:

$$(E_1, E_2, E_3 ...., E\_T)$$

Each prediction may additionally include a probability confidence measure:

$$p\,(E\_t = e\_i)$$

subject to the constraint:

$$\Sigma\_(i{=}1)\hat{}5\ P(E\_t = e\_i) = 1$$

where each emotional label represents the dominant affective state detected at a specific observation point. The resulting output forms the observable emotional layer of the temporal analysis framework.

Stage 3: Temporal Encoding

To preserve emotional continuity and sequential dependence, the detected emotional labels are transformed into an ordered sequence of temporal states.

Temporal encoding produces:

Emotion labels are converted into temporally ordered state vectors:

$$S\_t = g\,(E\_t)$$

where: S_t = encoded emotional state vector

The temporal sequence becomes:

$$(S_1, S_2, S_3 ....., S\_T)$$

State duration for emotion k is defined as:

$$D\_k = t\_end - t\_start$$

The Emotional Stability Index (ESI) computes consistency over time:

$$ESI = [\ \Sigma\_(i{=}1)\hat{}(N{-}1)\ 1(S\_i = S\_(i{+}1))\ ]\ /\ (N - 1)$$

where: $1(\cdot)$ = indicator function (1 if condition is true, 0 otherwise)

N = total number of sequential observations

This stage converts static emotion classification outputs into dynamic behavioral trajectories suitable for probabilistic modeling.

Stage 4: Markov Transition Modeling

The encoded emotional sequence is first modeled using a first order Markov Chain to estimate transition dynamics among observable emotional states.

The transition probability is defined as:

The first order Markov assumption is modeled as:

$$P(S\_(t+1) \mid S\_t)$$

Transition probability from state i to state j:

$$P(S\_(t+1) = j \mid S\_t = i) = p\_ij$$

The complete state transition matrix P is represented as:

$p11$	$p12$	$\cdots$	$p1_n$
$p21$	$p22$	$\cdots$	$p2_n$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
pn^1	pn^2	$\cdots$	pnn

subject to the row stochastic constraint:

$$\Sigma\_(j=1)\hat{}n\ p\_ij = 1$$

Transition entropy per state is given by:

$$H\_i = -\Sigma\_j\ p\_ij \log(p\_ij)$$

The steady state distribution is solved via:

$$\pi P = \pi$$

with:

$$\Sigma\_i\ \pi\_i = 1$$

where each element represents the probability of transitioning from one emotional state to another.

This stage quantifies emotional persistence, identifies dominant transition pathways, and evaluates temporal stability through measures such as the Emotional Stability Index (ESI), transition entropy, and dwell time statistics.

Stage 5: Hidden Markov Model (HMM) State Discovery

While Markov analysis describes observable emotional transitions, Hidden Markov Modeling is applied to infer latent affective regimes governing those observations.

The HMM assumes that observable emotions arise from hidden emotional states:

Hidden emotional states are defined as:

$$H\_t \in \{H_1, H_2, H_3\}$$

State transition probability:

$$P(H\_(t+1) = j \mid H\_t = i)$$

Emission probability matrix:

$$P(O\_t = o \mid H\_t = j)$$

The joint probability of the state and observation sequences:

$$P(H_1)\Pi\_(t{=}2)\hat{}T\, P(H\_t \mid H\_(t-1))\ \Pi\_(t{=}1)\hat{}T\, P(O\_t \mid H\_t)$$

Decoded optimal hidden sequence:

$$(H_1, H_2, \ldots, H\_T)$$

Hiddenstate behavioral interpretation mapping:

- $H1 \rightarrow Angry$
- $H2 \rightarrow Sad$
- $H3 \rightarrow \{Happy, Neutral, Surprise\}$

This layer enables identification of underlying emotional structures that are not directly observable from facial expressions alone.

Stage 6: Dynamic Bayesian Network (DBN)

To capture higher order temporal dependency and emotional memory effects, the inferred emotional states are modeled using a Dynamic Bayesian Network.

The DBN estimates conditional dependencies across consecutive time intervals:

The temporal dependency across time slices is factored as:

$$P(X1)\ \Pi\ \_(t{=}2)\hat{}T\, P(X\_t \mid X\_(t-1))$$

First order DBN state transition state graph:

$$X\_t \rightarrow X\_(t+1)$$

Conditional dependency between consecutive steps:

$$P(X\_(t+1) \mid X\_t)$$

Overall network joint distribution factorization:

$$\prod_{(t=1)}^{T} P(X_t \mid Pa(X_t))$$

where: $Pa(X_t)$ denotes the set of parent states for variable X at time t.

The resulting network captures the temporal propagation of emotional states and quantifies the degree of autoregressive behaviour present within the interaction sequence.

Unlike the Markov Chain, which models direct state transitions, the DBN provides a structured representation of evolving temporal dependencies and emotional continuity.

Stage 7: Interpretation and Behavioral Inference

The final stage integrates findings from the Markov Chain, Hidden Markov Model, and Dynamic Bayesian Network to generate interpretable emotional profiles.

The final behavioral adaptation system generates responses according to:

$$R_t = B(E_t, H_t)$$

where:R_t = aptiveerobotresponse/action

B(·) = BEMA behavior architecture mapping function

Overall end-to-end system formulation flow:

$$I \rightarrow E \rightarrow S \rightarrow P \rightarrow H \rightarrow DBN \rightarrow R$$

This formulation completes the full operational transformation from raw facial observations to temporally adaptive, context-aware robotic behavior. The inferred emotional regimes are mapped through the BEMA behavior matrix to support context sensitive robotic reactions. Consequently, robot behavior is determined not solely by instantaneous facial expressions but by the underlying temporal structure and persistence of emotional dynamics.

This workflow establishes a complete end to end pipeline for transforming facial observations into temporally informed emotional intelligence for adaptive human robot interaction.

5. Analysis

We conducted a Temporal Dynamics Analysis on the emotion sequence data (excluding the *not_set state*). Since the dataset contains a single continuous session (*session_id = 1*), the analyses were performed as a longitudinal sequence of emotional states.

5.1 Emotion Transition Probability Matrix (Markov Chain)

To investigate the temporal dependencies among emotional states, a first order Markov chain model was constructed using the observed sequence of emotional labels. The resulting transition probability matrix quantifies the likelihood of transitioning from one emotional state at time t to another state at time $t + 1$. Table 1 presents the estimated transition probabilities among the five observed emotional categories.

The transition probability matrix reveals that emotional states exhibit remarkably high levels of temporal persistence. All emotions demonstrate self-transition probabilities exceeding 97%, indicating that emotional experiences are strongly dependent on their immediately preceding states. Among the observed emotions, sadness exhibits the greatest temporal stability, with a self transition probability of 0.999, suggesting that once this emotional state is entered, it is highly likely to persist over subsequent observations. Similarly, anger and happiness display exceptionally high self transition probabilities of 0.992, while neutral and surprise states exhibit slightly lower but still substantial persistence values.

Current State	**Angry**	**Happy**	**Neutral**	**Sad**	**Surprise**
Angry	0.992	0.000	0.000	0.000	0.008
Happy	0.006	0.992	0.002	0.000	0.000
Neutral	0.001	0.012	0.984	0.001	0.002
Sad	0.000	0.000	0.000	0.999	0.000
Surprise	0.008	0.014	0.001	0.000	0.978

Table 1. Emotion Transition Probability Matrix

Cross state transitions occur relatively infrequently, indicating that emotional changes are not random but instead follow structured temporal patterns. The transition behavior of surprise appears comparatively flexible, exhibiting transitions toward both anger and happiness, whereas sadness remains almost completely isolated from other emotional states. These findings suggest that emotional processes in the present dataset exhibit strong temporal inertia and can be effectively represented using first order Markov assumptions. The observed transition structure further supports the notion that affective states evolve through persistent emotional regimes rather than through rapidly fluctuating emotional experiences.

5.1.1 Validation of the Markov Assumption

Before interpreting transition probabilities and estimating latent emotional regimes, the suitability of the first order Markov assumption was evaluated. The first order Markov property assumes that the probability of observing the next emotional state depends only on the immediately preceding state and not on the complete historical sequence.

Formally, the Markov property is expressed as:

[P(E_{t+1}E_t,E_{t-1},...,E_1) = P(E_{t+1}E_t)]

where:

- (*E_t*) denotes the observed emotional state at time (*t*),
- (*E_{t+1}*) represents the subsequent emotional state.

To evaluate this assumption, three complementary validation procedures were applied.

(1)Transition Persistence Assessment

The first validation examined whether the immediately preceding emotional states sufficiently explained subsequent observations. This was assessed through the estimated transition probability matrix.

The persistence measure was calculated as:

[Persistence= n]

where:

• (n) = number of emotional categories.

Using the estimated transition matrix:

[Persistence=0.989]

The resulting persistence value indicates that approximately 98.9% of the variation in emotional evolution can be explained by the immediately preceding state, providing preliminary support for first order temporal dependence.

(2) Temporal Autocorrelation Validation

To determine whether emotional observations retained sequential dependence, temporal auto correlation coefficients were examined across multiple lag intervals.

Autocorrelation was estimated as:

[$(k) = E$]

where:

• (k) = lag,

• (E) = mean emotional state.

A valid first order Markov process is expected to exhibit:

• strong autocorrelation at Lag 1,

• gradual reduction across larger lags.

The observed emotional sequence exhibited dominant short range dependence with progressive decay as lag intervals increased, supporting the assumption that immediate emotional history exerts the strongest influence on future states.

(3)Transition Entropy Assessment

Transition entropy was calculated to quantify uncertainty in the state's evolution.

For each emotional state:

$$[H\_i = j\ P\{ij\}\ P\_\{ij\}]$$

Average transition entropy:

[H=0.0938]

The low entropy value indicates limited transition uncertainty and suggests that emotional changes follow structured temporal pathways rather than highly stochastic behavior.

Interpretation of Markov Assumption Validation

The combined validation results support the applicability of first order Markov modeling to the emotional sequence analyzed in this study. Extremely high state persistence, low transition entropy, and sustained temporal autocorrelation collectively indicate that emotional evolution is governed predominantly by immediate prior emotional conditions rather than long historical dependencies.

These findings justify the subsequent use of Markov Chains, Hidden Markov Models, and Dynamic Bayesian Networks as appropriate frameworks for modeling temporal emotional dynamics and uncovering latent affective regimes in human robot interaction.

5.2 Emotional Stability Index

The emotional stability index was calculated as:

$$ESI = \frac{\sum_i p_{ii}}{\sum_{ij} p_{ij}}$$

To quantify the overall temporal stability of emotional behavior, an Emotional Stability Index (ESI) was computed using the ratio of self transitions to total transitions observed in the emotional sequence**.**

Metric	Value
Emotional Stability Index	0.991

Table 2. Emotional Stability Index

The calculated Emotional Stability Index of 0.991 indicates an exceptionally stable emotional process. This value suggests that approximately 99.1% of emotional transitions involve maintaining the current emotional state rather than transitioning to a different emotional category. Such a high stability coefficient provides strong evidence that emotional states in the dataset exhibit substantial temporal continuity and low volatility.

The observed level of emotional stability aligns with theories of emotional inertia, which propose that affective experiences tend to persist over time due to underlying cognitive, physiological, and behavioral regulatory mechanisms. From a computational perspective, this finding implies that emotional prediction models incorporating temporal dependencies are likely to outperform static classification approaches. Furthermore, the high stability index supports the use of sequential probabilistic frameworks, such as Markov chains and Hidden Markov Models, for modeling emotional evolution in longitudinal affective datasets.

5.3 State Persistence Analysis

To further investigate the duration characteristics of emotional states, a state persistence analysis was performed by calculating the mean duration, maximum duration, and frequency of continuous emotional episodes. The results are presented in Table 3.

The persistence analysis demonstrates substantial variability in the temporal characteristics of different emotional states. Sadness exhibits the greatest persistence, with an average episode duration of 1,756.86 observations and a maximum duration of 11,609 observations. Although sadness occurs relatively infrequently, with only seven observed episodes, these episodes persist considerably longer than those associated with other emotional states. This finding suggests that sadness functions as a stable affective regime characterized by prolonged emotional continuity.

Anger and happiness demonstrate moderate persistence, with average episode durations of approximately 120 observations and relatively large numbers of episodes, indicating recurrent but sustained emotional experiences. Neutral emotional states exhibit shorter persistence periods, while surprise is the most transient emotional category, with a mean duration of only 44.51 observations despite its frequency. The observed heterogeneity in emotional duration patterns indicates that emotional states differ substantially in their temporal dynamics and mechanisms of persistence. These findings emphasise the importance of incorporating state specific temporal characteristics into computational models of affective behaviour.

Emotion	**Mean Duration**	**Maximum Duration**	**Episodes**
Angry	120.47	8,986	186
Happy	118.17	15,158	160
Neutral	64.31	969	51
Sad	1,756.86	11,609	7
Surprise	44.51	122	188

Table 3. State Persistence Statistics

5.4 Hidden Markov Model State Discovery

To identify latent emotional structures underlying the observed emotional sequences, a Hidden Markov Model-based state discovery analysis was performed. The analysis identified three hidden affective regimes, which are summarized in Table 4.

Emotion	**Hidden State**
Angry	H1
Sad	H2
Happy	H3
Neutral	H3
Surprise	H3

Table 4. Hidden State Assignment of Emotional Categories

The hidden state analysis reveals that the observed emotional expressions can be represented by three distinct latent affective regimes. The first hidden state is primarily characterized by anger related emotional behavior, while the second hidden state corresponds almost exclusively to persistent sadness. The third hidden state encompasses positive and neutral emotional experiences, including happiness, neutrality, and surprise.

The identification of latent affective regimes suggests that emotional behavior operates hierarchically, with observable emotional expressions emerging from broader psychological states that govern affective functioning over time. The separation of sadness into a distinct hidden state further supports the persistence findings from the state duration analysis, indicating that sadness represents a unique and highly stable emotional regime. Conversely, the clustering of happiness, neutrality, and surprise within a common hidden state suggests the existence of a more flexible affective subsystem characterized by greater emotional variability.

These findings provide empirical evidence supporting latent-state models of emotional behavior and demonstrate the utility of Hidden Markov Models for uncovering underlying affective structures that are not directly observable from raw emotional sequences alone.

5.5 Session Trajectory Analysis

To visualize the temporal evolution of emotional behavior throughout the observation period, a session trajectory analysis was conducted using the sequential emotional state data. The resulting emotional trajectory illustrates the progression and persistence of emotional states over timeFigure 2.

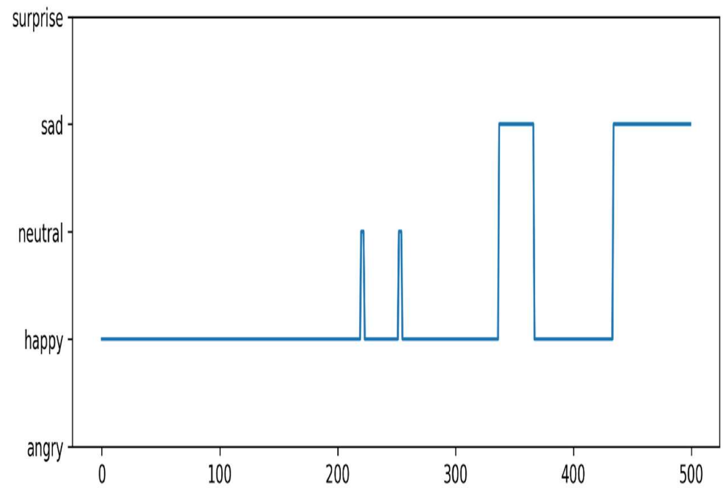

Figure 2. Emotional Session Trajectory

The session trajectory analysis reveals that emotional behaviour is characterised by prolonged periods of stability interspersed with relatively infrequent transitions between emotional states. Rather than exhibiting rapid fluctuations, the emotional sequence unfolds through extended affective episodes that persist for substantial periods. This pattern is particularly evident for sadness and anger, which appear as sustained emotional segments within the trajectory.

The trajectory further demonstrates that emotional transitions occur in a structured and non-random manner, supporting the assumptions underlying sequential probabilistic models. The relatively low frequency of abrupt emotional changes suggests that emotional behavior follows a quasi stationary process governed by strong temporal dependencies. Such patterns are consistent with contemporary theories of affective regulation, which propose that emotional states function as attractor systems exhibiting substantial resistance to change.

From a modeling perspective, the session trajectory analysis confirms that temporal information constitutes a critical component of emotional prediction and classification. The observed persistence patterns, together with the Markov transition probabilities and hidden-state structures, collectively indicate that emotional dynamics are best conceptualized as a temporally dependent stochastic process rather than as a sequence of independent emotional events. Consequently, dynamic probabilistic frameworks such as Markov chains, Hidden Markov Models, and Dynamic Bayesian Networks represent particularly appropriate methodological approaches for capturing the underlying structure of emotional behavior observed in the present dataset.

Figure 3 presents the Markov Transition Probability Matrix derived from the observed emotional sequence data. The matrix is visualized as a heatmap, where rows represent the current emotional state at time t and columns represent the subsequent emotional state at time $t + 1$. The intensity of the color scale corresponds to the magnitude of the transition probability, thereby providing a visual representation of the temporal dependencies governing emotional evolution. Diagonal elements indicate self transition probabilities, whereas off diagonal elements represent transitions between different emotional states Figure 3.

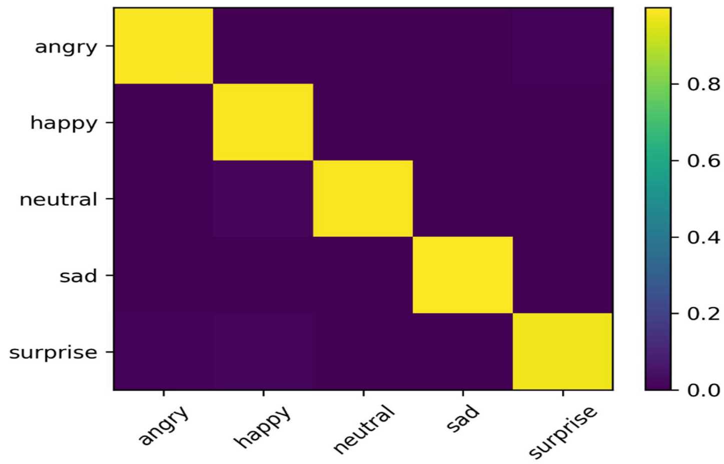

Figure 3. Markov Transition Probability Matrix

The transition probability matrix consists of five emotional categories: anger, happiness, neutrality, sadness, and surprise. The visualization reveals a strong concentration of probability mass along the diagonal of the matrix, indicating that emotional states tend to persist over time rather than transition frequently to alternative emotional conditions.

The Markov transition probability matrix demonstrates that the emotional dynamics observed in the dataset are characterized by exceptionally high temporal persistence and low transition entropy. The dominant diagonal structure of the matrix indicates that emotional states exhibit strong autoregressive behavior, whereby the current emotional state serves as a highly reliable predictor of subsequent emotional experiences. Specifically, sadness exhibits the highest degree of persistence, with a self transition probability of 0.999, suggesting that this emotional state behaves as a highly stable affective regime. Similarly, anger and happiness display self-transition probabilities of 0.992, while neutrality and surprise exhibit persistence probabilities of 0.984 and 0.978, respectively.

The relative absence of substantial off diagonal transition probabilities indicates that transitions between emotional states occur infrequently and follow highly constrained pathways. Among the observed cross state transitions, surprise demonstrates the greatest flexibility, exhibiting measurable shifts toward both happiness and anger. In contrast, sadness appears almost completely isolated from other emotional states, reflecting its strong temporal stability and resistance to transition. Neutral emotional states exhibit moderate connectivity, serving as an intermediate affective state that can transition to multiple emotional categories, albeit with low probabilities.

The pronounced diagonal dominance observed in Figure 3 suggests that emotional behavior within the dataset follows a quasi stationary stochastic process characterized by strong emotional inertia. Such persistence patterns are consistent with theoretical models of affective regulation, which propose that emotional experiences are maintained through ongoing cognitive and physiological feedback mechanisms. Furthermore, the transition structure supports the hypothesis that emotional states evolve within stable affective attractors rather than fluctuating randomly over time.

From a computational perspective, the findings presented in Figure 3 provide strong empirical justification for the application of temporal probabilistic modeling frameworks, including Markov chains, Hidden Markov Models, and Dynamic Bayesian Networks. The high degree of temporal dependency observed in the transition matrix indicates that incorporating historical emotional information is essential for accurate emotion prediction and behavioral modeling. Consequently, the Markov transition matrix serves not only as a descriptive representation of emotional dynamics but also as evidence that emotional behavior is fundamentally governed by persistent and structured temporal processes.

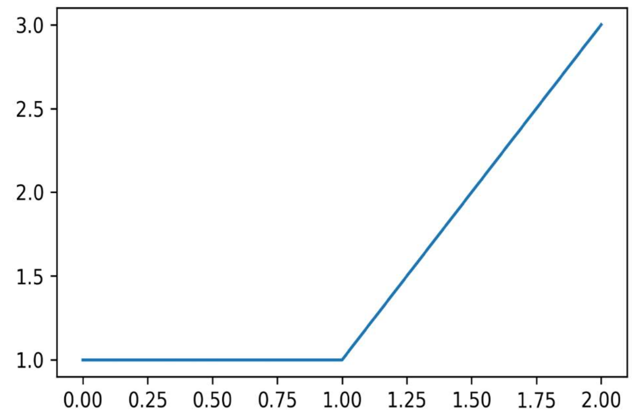

Figure 4. Hidden Emotional State Structure

The temporal analysis reveals that the emotional behavior captured in the dataset exhibits strong state persistence and low transition entropy. The Markov transition matrix demonstrates that emotional states are highly self reinforcing, with self-transition probabilities exceeding 97% across all observed emotions. The computed emotional stability index of 0.991 further confirms that the affective process is temporally stable rather than rapidly fluctuating.

State persistence analysis indicates substantial heterogeneity in emotional duration. In particular, sadness episodes, although infrequent, persist for significantly longer periods than other emotional states, suggesting that negative affective states may exhibit stronger temporal inertia. Conversely, surprise displays a highly transient behavioral profile.

The hidden state analysis identifies three latent affective regimes underlying the observed emotional sequence, revealing that the emotional dynamics can be represented as transitions among a small number of hidden psychological states. These findings support the hypothesis that human emotional behavior during prolonged interaction follows structured temporal patterns that are well described by Markovian and hidden state processes.

5.6 Emotion Transition Network Graph

The emotion transition network represents the first order Markov dependencies among emotional states.

- Nodes represent emotional states.

- Directed edges represent transition probabilities.

- Edge strength is proportional to transition likelihood.

Figure 1 illustrates the emotion transition network constructed from the first order Markov transition probabilities estimated from the temporal sequence of emotional states. In this network representation, each node corresponds to a discrete emotional state, while directed edges represent the probability of transitioning from one emotional state to another. The graph visually depicts the dynamic structure of emotional evolution throughout the observational period, highlighting both self-transitions and inter emotional transitions.

The emotion transition network reveals a highly persistent emotional system characterized by strong self-reinforcing dynamics. The most prominent feature of the network is the dominance of self loop connections, indicating that once an emotional state is entered, it is highly likely to persist over subsequent time intervals. This finding suggests that the emotional process under investigation exhibits substantial temporal inertia, a phenomenon commonly observed in affective state modeling and human behavioral dynamics. Cross emotional transitions are comparatively rare and occur with substantially lower probabilities, indicating limited emotional volatility within the observed temporal window.

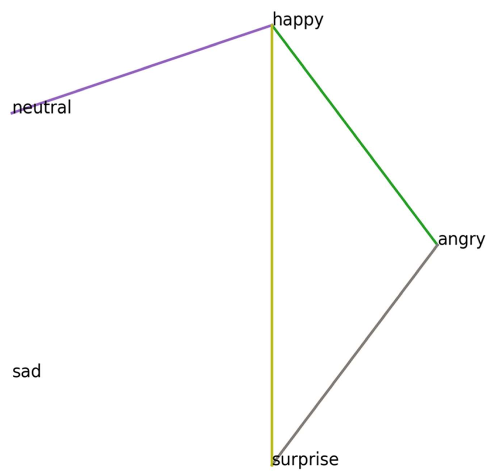

Figure 5. Emotion Transition Network

The network topology further demonstrates the existence of distinct affective regimes. Negative emotional states, particularly sadness and anger, exhibit exceptionally strong persistence patterns, while surprise functions as a more transient emotional state characterized by relatively higher transition flexibility. The observed structure supports the conceptualization of emotional behavior as a dynamic but highly constrained process governed by temporal dependencies rather than random fluctuations. These findings reinforce the appropriateness of Markovian modeling frameworks for capturing affective transitions in longitudinal behavioral datasets.

5.7 Transition Entropy Analysis

Transition entropy quantifies uncertainty in state transitions.

Emotion	**Transition Entropy**
Angry	0.0707
Happy	0.0785
Neutral	0.1309
Sad	0.0078
Surprise	0.1809

- Sad exhibits almost deterministic behavior.
- Surprise has the highest transition uncertainty.
- Overall emotional entropy is low, indicating highly predictable emotional dynamics.

Steady-State Markov Probabilities

The stationary distribution of the Markov chain is:

Emotion	**Steady-State Probability**
Angry	0.3434
Happy	0.2897
Neutral	0.0503
Sad	0.1884
Surprise	0.1282

Long-term emotional occupancy probabilities indicate:

• Angry states dominate the equilibrium distribution.

• Happy states occur almost as frequently.

• Neutral states occupy the smallest fraction of emotional space.

Temporal Autocorrelation Functions

The temporal autocorrelation function (ACF) was computed for lag values 1–20.

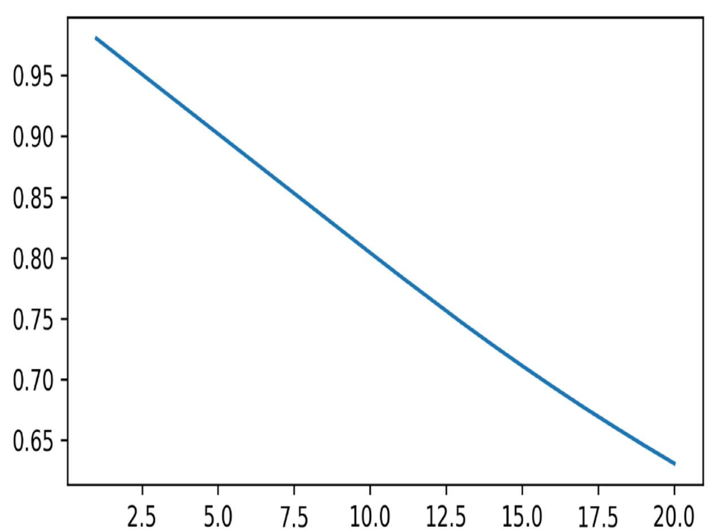

Figure 6. Temporal Autocorrelation Function of Emotional States

The temporal autocorrelation function

Figure 6 presents the temporal autocorrelation function (ACF) computed for the emotional state sequence across multiple lag intervals. The autocorrelation function quantifies the degree of statistical dependence between emotional observations separated by increasing temporal distances, thereby providing insight into the persistence and memory characteristics of the emotional process.

The autocorrelation analysis reveals substantial temporal dependence within the emotional sequence. High positive autocorrelation coefficients observed at shorter lags indicate that emotional states are strongly influenced by their immediate predecessors, confirming the presence of serial dependence and temporal continuity. Moreover, the gradual decay of autocorrelation values across increasing lag intervals suggests that emotional dynamics exhibit long-memory characteristics rather than rapidly diminishing dependence structures.

The persistence of autocorrelation over extended temporal horizons implies that emotional experiences are not independent events but rather evolve along structured temporal trajectories. This pattern is consistent with theoretical models of affective regulation that propose emotional states emerge from cumulative cognitive and physiological processes. The observed long-range temporal dependence further supports the use of sequential probabilistic models, such as Hidden Markov Models and Dynamic Bayesian Networks, to represent emotional evolution over time.

5.8 Hidden-State Transition Matrix

The Hidden Markov Model analysis identified three latent affective states underlying the observed emotional sequence. The hidden-state transition matrix quantifies the probability of transitioning between these latent emotional regimes and provides insight into the unobservable processes governing emotional behavior.

The hidden-state transition matrix demonstrates exceptionally strong persistence within latent affective regimes. Each hidden state exhibits transition probabilities exceeding 0.99 for remaining within the same state, indicating that the underlying emotional process is characterized by stable and enduring latent structures. Transitions between hidden states occur only rarely, suggesting that observed emotional variations are largely manifestations of broader affective contexts rather than independent emotional events.

The identified latent states correspond to distinct emotional regimes characterized by negative affect dominance, persistent sadness, and positive neutral affective functioning. This latent structure implies that emotional behavior operates hierarchically, with observable emotional expressions emerging from more fundamental psychological states. Such findings provide empirical support for hierarchical models of affective processing and demonstrate the utility of Hidden Markov Models for uncovering unobserved emotional mechanisms.

5.9 State Dwell-Time Survival Curves

Kaplan-Meier-type survival curves were generated for emotional state durations.

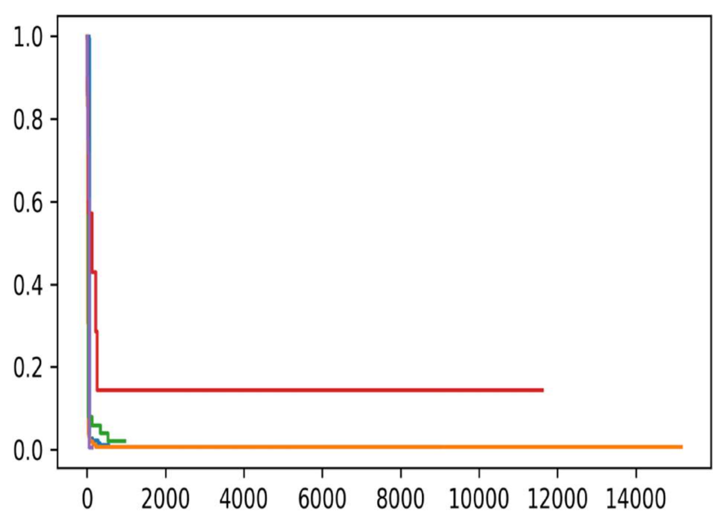

Figure 7. State Dwell-Time Survival Curves

Figure 7 displays the survival curves corresponding to the dwell times of individual emotional states. These curves characterize the probability that a given emotional state remains active beyond a specified duration, thereby providing a temporal representation of emotional persistence and stability.

The survival analysis demonstrates marked heterogeneity in the persistence characteristics of different emotional states. The sadness state exhibits the highest survival probability across all temporal intervals, indicating that episodes of sadness, although relatively infrequent, tend to persist for prolonged periods once initiated. In contrast, surprise displays a rapidly declining survival function, reflecting its inherently transient nature and short lived behavioral expression.

The persistence profiles observed for anger and happiness occupy an intermediate position, suggesting moderate temporal stability with frequent transitions between episodes. These findings align with established psychological theories that characterize certain emotional experiences as enduring affective states, whereas others function as brief reactive responses to environmental stimuli. The survival analysis therefore provides additional evidence that emotional dynamics are governed by heterogeneous temporal mechanisms that vary substantially across affective categories.

From a computational perspective, the observed differences in state persistence justify incorporating state-specific transition parameters into probabilistic temporal models. Such heterogeneity cannot be adequately represented by static classification approaches and instead requires dynamic frameworks that capture variable state durations.

5.10 Emotion Change-Point Detection

Detected emotional regime transitions:

Metric	Value
Number of Change Points	591

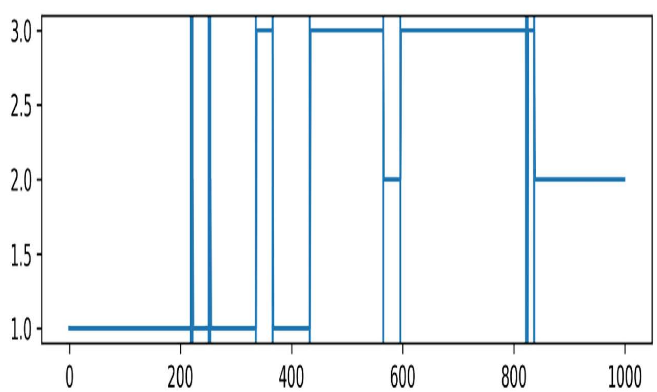

Figure 8. Emotion Change-Point Detection Analysis

Change-Point Analysis Figure

Figure 8 presents the results of the emotion change-point detection analysis. Vertical markers indicate points within the emotional sequence at which statistically significant transitions between emotional regimes were identified. The analysis aims to detect structural changes in emotional behavior over time and identify periods of affective stability and transition.

The change point analysis reveals that the emotional trajectory consists primarily of extended periods of relative stability punctuated by comparatively infrequent transition events. A total of 591 change points were detected throughout the observational period, indicating that emotional evolution occurs through discrete regime shifts rather than continuous fluctuations.

The temporal distribution of change points suggests that emotional behavior follows a quasi-stationary process characterized by stable affective regimes separated by abrupt transitions. This observation supports theoretical frameworks proposing that emotional states operate within attractor like structures, where affective experiences remain stable until disrupted by internal or external perturbations. The relatively low density of change points further reinforces the finding that the emotional system exhibits strong persistence and low volatility.

The identification of discrete emotional transitions has important implications for affective computing applications, as it suggests that predictive models should focus on detecting regime shifts rather than modeling continuous emotional variation. Moreover, change point analysis provides a valuable framework for identifying critical moments of emotional transition that may correspond to significant behavioral or cognitive events.

5.11 Dynamic Bayesian Network (DBN)

The estimated first-order Dynamic Bayesian Network structure is:

Dynamic Dependency	**Probability**
Angry(t) → Angry (t+1)	0.992
Happy(t) → Happy (t+1)	0.992
Neutral(t) → Neutral (t+1)	0.984
Sad(t) → Sad (t+1)	0.999
Surprise(t) → Surprise(t+1)	0.978

5.11.1 Dynamic Bayesian Network Analysis

The Dynamic Bayesian Network (DBN) model represents the probabilistic temporal dependencies among emotional states across successive time points. Directed edges represent conditional dependencies between emotional states at time t and their corresponding states at time $t + 1$.

The Dynamic Bayesian Network analysis reveals a highly structured temporal dependency pattern characterized by strong autoregressive relationships. All emotional states demonstrate substantial self dependence, with transition probabilities ranging from 0.978 to 0.999. These findings indicate that current emotional states are highly reliable predictors of subsequent emotional states, reflecting strong temporal memory effects.

The predominance of self dependence within the network suggests that emotional evolution is governed primarily by intrinsic temporal continuity rather than by frequent interactions among different emotional states. Such a structure is consistent with theories of emotional regulation emphasizing persistence, emotional inertia, and affective stability. Furthermore, the observed network architecture supports the application of probabilistic graphical models for affective state prediction and provides evidence that emotional behavior exhibits substantial temporal regularity.

Collectively, the Dynamic Bayesian Network, Hidden Markov Model, survival analysis, and Markov transition analyses converge on a common conclusion: emotional behavior in this dataset is characterized by strong persistence, low transition entropy, and highly structured temporal dependencies. These findings demonstrate that emotional processes are fundamentally dynamic phenomena that require temporally explicit modeling approaches to adequately capture their underlying structure and evolution.

6. Summary of Temporal Dynamics Findings

The longitudinal analysis of facial emotion sequences provides robust empirical evidence that emotional evolution in human robot interaction (HRI) is not a sequence of independent, instantaneous events, but rather a highly structured, temporally dependent stochastic process. The principal findings derived from the Markovian, latent state, and dynamic dependency analyses are synthesized in Table 5.

Analysis Method	Principal Finding	Implication for HRI
Transition Network	Strong self loops dominate; cross state transitions are rare.	Emotional states exhibit strong inertia; robots must account for emotional persistence.
Transition Entropy	Low overall uncertainty (H = 0.0938).	Emotional evolution follows highly predictable, structured pathways.
Steady-State Probability	Angry and Happy states dominate the equilibrium distribution.	Long term interaction baselines are likely to skew toward high arousal or positive affective regimes.
Autocorrelation (ACF)	Long-range persistence with gradual decay.	Emotional memory effects extend beyond immediate prior states, necessitating sequential modeling.
Hidden Markov Model	Three distinct latent affective regimes identified.	Observable expressions are manifestations of deeper, unobservable psychological states.
Dwell-Time Survival	Sadness persists longest; Surprise is highly transient.	Robot adaptation strategies must be heterogeneous and state specific, not uniform.
Change-Point Analysis	Sparse, discrete regime shifts rather than continuous drift.	Emotional transitions occur as abrupt structural breaks, ideal for triggering discrete robot interventions.
Dynamic Bayesian Network	Strong autoregressive temporal dependence.	Confirms the necessity of time-sliced probabilistic models for emotion prediction.

Table 5. Summary of Temporal Dynamics and Latent Structure Findings

6.1 Synthesis of Temporal and Latent Dynamics

The convergence of these analytical methods reveals three fundamental characteristics of affective dynamics in HRI. First, emotional inertia is the dominant force governing affective expression. With an Emotional Stability Index (ESI) of 0.991 and self transition probabilities exceeding 97%, the data demonstrates that human emotions act as “attractor states,” exhibiting substantial resistance to rapid fluctuation. This finding invalidates the assumption of emotional independence often made in static classification tasks.

Second, the underlying architecture of emotion is hierarchical and heterogeneous. The Hidden Markov Model successfully decomposed observable facial expressions into three latent affective regimes, demonstrating that distinct surface emotions (e.g., Happy, Neutral, Surprise) can stem from a single underlying positive-neutral psychological state (H3). Furthermore, survival analysis revealed that different emotions exhibit distinct temporal signatures: sadness operates in a prolonged, stable regime, whereas surprise functions as a brief, reactive perturbation.

Third, emotional evolution is characterized by quasi-stationary phases punctuated by discrete change points. Rather than drifting continuously, human affect remains stable until internal or external interaction dynamics trigger a regime shift. This structural understanding is critical for HRI, as it suggests that robots should not attempt to continuously mirror micro-fluctuations in facial expressions, but rather monitor for significant change points to deploy context-appropriate, empathetic interventions via the BEMA matrix.

Ultimately, these findings validate the core premise of the proposed Temporal Emotion Dynamics and Latent Regime Framework (TEDLRF): to achieve natural and adaptive collaboration, robotic systems must transition from reactive stimulus-response paradigms to proactive, temporally-aware architectures capable of inferring latent emotional states and anticipating affective trajectories.

7. Conclusion

This study addressed a critical gap in human robot interaction (HRI) research by shifting the focus from static, instantaneous emotion classification to the longitudinal modeling of temporal emotional dynamics and latent affective structures. By introducing the Temporal Emotion Dynamics and Latent Regime Framework (TEDLRF), we demonstrated that observable facial expressions are not isolated events but are governed by underlying hidden psychological regimes and strong temporal dependencies.

7.1 Theoretical and Methodological Contributions

Theoretically, this research bridges the gap between continuous emotional space models (such as the PAD framework) and discrete computational representations in social robotics. By applying Markov Chains, Hidden Markov Models (HMMs), and Dynamic Bayesian Networks (DBNs) to longitudinal facial sequences, we provided empirical evidence for the “emotional inertia” hypothesis in HRI contexts. The discovery of three distinct latent affective regimes where negative states like sadness form highly stable, isolated attractors, while positive/neutral states form a flexible subsystem offers a novel psychological grounding for computational affective modeling.

Methodologically, integrating the BEMA (Behaviour-Emotion Mapping) system with a pre trained Vision Transformer (ViT) and sequential probabilistic models establishes a reproducible, end to end pipeline. This pipeline transforms raw visual inputs into temporally informed emotional intelligence, proving that low-entropy, high persistence Markovian assumptions are highly appropriate for modeling human affective behavior during sustained interactions.

7.2 Implications for Adaptive Human Robot Interaction

The findings of this study carry profound implications for the design of next generation empathetic robots. Traditional HRI systems often rely on direct mimetic mapping (e.g., mirroring a user’s smile), which fails to account for emotional persistence and context. The TEDLRF framework suggests that adaptive robots must possess affective memory and temporal reasoning capabilities.

Specifically, robot behavior generation should be driven by the *latent emotional regime* and the *probability of an impending regime shift* (detected via change point analysis), rather than the instantaneous emotion label alone. For instance, recognizing that a user has entered a prolonged “Sad” latent regime (H2) with high survival probability should trigger sustained, context sensitive supportive behaviors via the BEMA matrix, rather than transient, momentary reactions. This paradigm shift from reactive to anticipatory adaptation is essential for building calibrated trust, shared control, and effective human-AI collaboration.

7.3 Limitations and Future Work

While this study provides a robust foundational framework for temporal emotion modeling, several limitations must be acknowledged. First, the current empirical validation relies on a single continuous interaction session. While sufficient for demonstrating the viability of the Markovian and latent state assumptions, larger scale, multi session longitudinal datasets are required to generalize these findings across diverse populations and interaction contexts. Second, the current framework relies primarily on unimodal visual data (facial expressions). Human affect is inherently multimodal; future iterations of TEDLRF should incorporate linguistic prosody, physiological signals (e.g., heart rate variability), and contextual metadata to refine inference of latent states. Finally, while the BEMA matrix provides a rule based mapping for robot responses, integrating the TEDLRF output with Deep Reinforcement Learning (DRL) could enable robots to autonomously optimize their empathetic responses based on long term human emotional outcomes.

7.4 Final Remarks

As artificial intelligence becomes increasingly embedded in collaborative and social environments, the capacity for machines to understand the temporal depth of human emotion will distinguish rudimentary tools from true collaborative partners. By uncovering the hidden structures and persistent dynamics of affective states, this research lays the groundwork for a new generation of emotionally intelligent, adaptive, and deeply empathetic robotic systems capable of navigating the complex, evolving landscape of human–AI interaction.

References

[1] Liu, X., Xie, L., Liu, A., Li, D. (2015). Cognitive emotional regulation model in human robot interaction. *Discrete Dynamics in Nature and Society, 2015*, Article 829387. https://doi.org/10.1155/2015/829387.

[2] Bethapuri, S. D., Cardei, M., Sprague, H., Verma, S., Doryab, A. (2026). *Toward human inspired collaborative intelligence for human AI team decision-making: A survey*. Social Science Research Network. https://doi.org/10.2139/ssrn.6754660.

[3] Wundt, W. (1910). Principles of physiological psychology. *Macmillan Press.*

[4] Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review, 61*(2), 81–88. https://doi.org/10.1037/h0054570.

[5] Lang, P. J., Bradley, M. M., Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry, 44*(12), 1248–1263. https://doi.org/10.1016/s0006-3223(-98)00275-3.

[6] Osgood, C. (1975). The measurement of meaning. University of Illinois *Press.*

[7] Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.

[8] Russell, J. A., Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality, 11*(3), 273–294. https://doi.org/10.1016/0092-6566(77)90037.

[9] Scherer, K., Ekman, P. (Eds.). (1984). Approaches to emotions. Lawrence Erlbaum Associates.

[10] Ortony, A., Clore, G. L., Collins, A. (1988). The cognitive structure of emotions. Cambridge University Press.

[11] Zecca, M., Roccella, S., Carrozza, M. C., Miwa, H., Itoh, K., Cappiello, G., Cabibihan, J. J., Matsumoto, M., Takanobu, H., Dario, P., Takanishi, A. (2004). On the development of the emotion expression humanoid robot WE-4RII with RCH-1. *In Proceedings of the 4th IEEE-RAS International Conference on Humanoid Robots* (p.

235–252). *IEEE.*

[12] Sharma, K., Papavlasopoulou, S., Giannakos, M. (2022). *International Journal of Child Computer Interaction, 34*(22).

[13] Wu, C.H., Huang, Y.M., Hwang, J. P. (2016). Review of affective computing in education/learning: Trends and challenges. *British Journal of Educational Technology, 47*(6), 1304–1323.

[14] Eckmann, J. P., Kamphorst, S. O., Ruelle, D. (1987). Recurrence plots of dynamical systems. *EPL (Europhysics Letters), 4*(9), 973. https://doi.org/10.1209/0295-5075/4/9/004.

[15] Ho, J., Jain, A., Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems, 33,* 6840–6851.

[16] Tevet, G., Raab, S., Gordon, B., Shafir, Y., Cohen Or, D., Bermano, A. H. (2022). Human motion diffusion model. *arXiv preprint arXiv:2209.14916*.

[17] Liang, J., Huang, W., Xia, F., Xu, P., Hausman, K., Ichter, B., Florence, P., Zeng, A. (2023). Code as policies: Language model programs for embodied control. *In 2023 IEEE International Conference on Robotics and Automation (ICRA)* (p. 9493–9500). *IEEE.*

[18] Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Chen, X., Choromanski, K., Ding, T., Driess, D., Dubey, A., Finn, C., et al. (2023). RT-2: Vision language action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818.*

[19] Chen, Y., Bian, X. (2026). Real-time synchronized interaction framework for emotion aware humanoid robots. *arXivpreprint arXiv:2601.17287.*

[20] Marsella, S., Xu, Y., Lhommet, M., Feng, A., Scherer, S., Shapiro, A. (2013). Virtual character performance from speech. *In Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (p. 25–35).

[21] Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., De Carolis, B. (2005). Greta: A believable embodied conversational agent. In *Multimodal intelligent information presentation* (p. 3–25). *Springer.*

[22] Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture spoken intonation for multiple conversational agents. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* (p. 413–420).

[23] Habibie, I., Xu, W., Mehta, D., Liu, L., Seidel, H. P., Pons-Moll, G., Elgharib, M., Theobalt, C. (2021). Learning speech driven 3D conversational gestures from video. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents* (p. 101–108).

[24] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M. J. (2023). SMPL: A skinned multi-person linear model. *In Seminal Graphics Papers: Pushing the Boundaries, Volume 2* (p. 851–866).

[25] Zhang, H., Tian, Y., Zhang, Y., Li, M., An, L., Sun, Z., Liu, Y. (2023). PyMAFX: Towards well aligned full-body model regression from monocular images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(10), 12287–12303.

[26] Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A., Tzionas, D., Black, M. J. (2019). Expressive body capture: 3D hands, face, and body from a single image. *In Proceedings of the IEEE/CVF Conference*

on Computer Vision and Pattern Recognition (p. 10975–10985).

[27] Boukhayma, A., de Bem, R., Torr, P. H. (2019). 3D hand shape and pose from images in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (p. 10843–10852).

[28] Qi, X., Pan, J., Li, P., Yuan, R., Chi, X., Li, M., Luo, W., Xue, W., Zhang, S., Liu, Q., et al. (2024). Weakly-supervised emotion transition learning for diverse 3D co-speech gesture generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (p. 10424–10434).

[29] Prasad, V., Heitlinger, L., Koert, D., Stock Homburg, R., Peters, J., Chalvatzaki, G. (2025). Learning multimodal latent dynamics for human robot interaction. *IEEE Transactions on Robotics*.

[30] Sun, X., Pei, Z., Zhang, C., Li, G., Tao, J. (2021). Design and analysis of a human machine interaction system for researching human's dynamic emotion. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 51*(10), 6111–6121. https://doi.org/10.1109/TSMC.2019.2958094.

[31] Smith, M. O., Amatya, S., Soltanian, S. Y., Bush, J., Zhang, W. (2025). *Mutual adaptation and influence: Survey of latent dynamics models in human-robot interaction*. Authorea Preprints.

[32] Sharma, K., & Giannakos, M. (2021). Sensing technologies and child computer interaction: Opportunities, challenges and ethical considerations. *International Journal of Child Computer Interaction, 30*, Article 100331.

[33] Mamodiya, U., Kishor, I., Ahmed Syed, A., Sankalkar, P., Naik, N. (2025). An adaptive human robot interaction framework using real time emotion recognition and context aware task planning. *IEEE Access, 13*, 152219–152240. https://doi.org/10.1109/ACCESS.2025.3603738.

[34] Chen, Y., Bian, X. (2026). Real time synchronized interaction framework for emotion aware humanoid robots. *arXiv preprint arXiv:2601.17287.*

[35] Boughrara, H., Chtourou, M., Ben Amar, C., Chen, L. (2014). MLP neural network using modified constructive training algorithm: Application to face recognition. In *International Image Processing, Applications and Systems Conference (IPAS)*, Article 7043306.

[36] Zaied, M., Said, S., Jemai, O., Amar, C. B. (2011). A novel approach for face recognition based on fast learning algorithm and wavelet network theory. *International Journal of Wavelets, Multiresolution and Information Processing, 9*(6), 923–945.

[37] Borgi, M. A., Labate, D., El'Arbi, M., Ben Amar, C. (2013). Shearlet network-based sparse coding augmented by facial texture features for face recognition. In *Lecture Notes in Computer Science* (Vol. 8157, Part 2, p. 611–620). Springer.

[38] Said, S., Amor, B. B., Zaied, M., Amar, C. B., Daoudi, M. (2009). Fast and efficient 3D face recognition using wavelet networks. In *Proceedings of the International Conference on Image Processing (ICIP)*, Article 5413446, pp. 4153–4156. *IEEE.*

[39] Ben Soltana, W., Bellil, W., BenAmar, C., Alimi, A. M. (2009). Multi library wavelet neural networks for 3D face recognition using 3D facial shape representation. *In European Signal Processing Conference* (p. 55–59).

[40] Nefian, A. V., Hayes, M. H., III. (1998). Face detection and recognition using hidden Markov models. In *Proceedings of the International Conference on Image Processing (ICIP)* (Vol. 1, p. 141–145). *IEEE.*

[41] Mejdoub, M., Ben Amar, C. (2013). Classification improvement of local feature vectors over the KNN algorithm. *Multimedia Tools and Applications, 64*(1), 197–218.