POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

# Emotion-Driven Architecture for Human-Robot Interaction

Master Thesis in
Automation and Control Engineering - Ingegneria dell'Automazione

Author: **Eleonora Righi**

Student ID: 10696797
Advisor: Prof. Andrea Bonarini
Academic Year: 2024-25

# Abstract

This thesis presents the design and implementation of an affective architecture for a non-anthropomorphic robot capable of interacting with a human actor in an improvisational setting. The objective of the work is to enable autonomous, coherent, and expressive behavior based on internal emotional dynamics rather than predefined scripts, thereby enhancing the naturalness and engagement of human–robot interaction. The proposed system was developed on the *BlackWing* platform and models eight emotions based on Plutchik's framework. Emotional states evolve over time according to external stimuli and internal factors, including personality and mood: personality parameters influence emotional thresholds and temporal persistence, while mood introduces a longer-term affective bias. Sensory inputs from multiple modalities, such as vision, proximity, touch, and sound, are interpreted as interaction events that update the internal emotional state through dedicated equations

When emotional thresholds are exceeded, the dominant state is expressed through coordinated movements and sound, exploiting the expressive potential of the robot's morphology. The proposed architecture integrates perception, emotional processing, and behavioral control within an embedded system. A user study was conducted to evaluate the perceptual effectiveness of the expressions, providing preliminary evidence of their discernibility in human–robot interaction.

**Keywords:**Autonomous Robot, Personality, Emotional Modeling, Human-Robot Interaction, Expressive Robotics, Improvisation

# Abstract in lingua italiana

Questa tesi si pone l'obiettivo di progettare e implementare un'architettura affettiva per un robot non antropomorfo in grado di interagire con un attore umano in un contesto di improvvisazione teatrale. L'elemento centrale del lavoro è abilitare un comportamento autonomo, coerente ed espressivo basato su dinamiche emotive interne, anziché su sequenze predefinite, al fine di migliorare la naturalezza e il coinvolgimento nell'interazione uomo–robot.

Il sistema proposto è stato sviluppato sulla piattaforma *BlackWing* e modella otto emozioni basate sul modello di Plutchik. Gli stati emotivi evolvono nel tempo in funzione degli stimoli esterni e di fattori interni, tra cui personalità e umore: i parametri di personalità influenzano le soglie emotive e la persistenza temporale, mentre l'umore introduce un bias affettivo a lungo termine. Gli input sensoriali provenienti da diverse modalità, quali visione, prossimità, tatto e suono, vengono interpretati come eventi di interazione che aggiornano lo stato emotivo interno tramite specifiche equazioni.

Quando i valori emotivi superano determinate soglie, lo stato dominante viene espresso attraverso movimenti coordinati e segnali sonori, sfruttando il potenziale espressivo della morfologia del robot. L'architettura integra percezione, elaborazione emotiva e controllo comportamentale all'interno di un sistema centralizzato. È stato infine condotto uno studio con utenti per valutare l'efficacia percettiva delle espressioni, fornendo evidenze preliminari della loro riconoscibilità nell'interazione uomo–robot.

**Parole chiave:** Robot Autonomo, Personalità, Modello emozionale, Interazione Uomo-Robot, Robotica Espressiva, Improvvisazione

# Contents

# Introduction

In recent years, the social robotics field has evolved into developing systems capable of interacting with humans in a natural and interesting way. These robots are designed to operate in social and creative environments, where the ability to express emotions, adapt behaviour dynamically and react meaningfully to human actions is essential to achieve believable interactions.

Human communication relies heavily on non-verbal cues such as posture, movement, timing, and expressive dynamics. For non-anthropomorphic robots, which cannot rely on detailed facial expressions or complex body articulation, conveying internal states through movement and behavioral modulation represents a significant challenge. At the same time, these limitations open opportunities to explore alternative forms of expressivity inspired by fields such as animation, theater, and movement studies.

Improvisational theater provides a particularly relevant framework for studying adaptive interaction. Unlike scripted performances, improvisation is based on continuous action–reaction dynamics, where participants must respond coherently and in real time to the evolving situation. Applying these principles to robotics requires the design of systems that can autonomously interpret stimuli, maintain an internal state, and generate context-appropriate responses.

The aim of this thesis is to design and implement an emotionally-driven robotic architecture capable of interacting with a human actor in an unscripted theatrical setting. The goal is to achieve behavior that is not only reactive, but also coherent over time and shaped by internal factors such as personality and mood, thereby enhancing the perceived realism of the interaction.

## 0.1. Description of the work

To address these objectives, an affective control architecture was developed and implemented on the non-anthropomorphic robot *BlackWing*. The proposed system models eight emotions based on Plutchik's framework, each represented as a dynamic internal variable whose evolution depends on external interactions and internal parameters. The work includes the design of the emotional update logic, the integration of hardware and software components, and the development of expressive motion strategies adapted to the robot's physical structure. The system enables the robot to react differently to similar situations depending on its internal configuration, supporting more varied and believable improvisational behavior.

Finally, a user study was conducted to evaluate whether the robot's emotional expressions were correctly perceived by human observers, providing an initial assessment of the effectiveness of the proposed approach.

## 0.2. Thesis Structure

This thesis is organized as follows.

- **Chapter 1: Research Context and Foundations** presents the research context and theoretical foundations, including improvisational theater, emotion models, personality frameworks, and related work on emotional robotics.
- **Chapter 2: Affective Architecture and Emotion-based Logic** describes the proposed affective architecture, detailing the emotional logic, personality and mood modeling, and interaction mechanisms.
- **Chapter 3: Hardware Description** provides an overview of the robot hardware and its expressive capabilities.
- **Chapter 4: Software Description** illustrates the software architecture and the implementation of perception, emotional processing, and expression modules.
- **Chapter 5: Survey-Based Evaluation of Emotional Perception** presents the design and results of the user study conducted to evaluate emotional perception.
- **Chapter 6: Conclusions** summarizes the main contributions of the work and discusses limitations and possible future developments.

# 1 | Research Context and Foundations

## 1.1. Theater and improvisation

Theater has long been one of the highest expressions of human creativity. Unlike scripted performances, improvisational theater thrives on spontaneity, collaboration, and the actors' immediate responses to one another. This dynamic form of performance brings to life moments that are truly unique and unrepeatable. Understanding how improvisation has developed over time offers insight into both its artistic value and its enduring role in theatrical practice.

### 1.1.1. History of Improvisational Theater

Improvisation has ancient origins. In ancient Greece, during Dionysian festivals, comedic and satirical performances often included also a certain degree of improvisation, as actors were allowed to adapt the performance based on the audience reactions. Around the same period, other forms of improvisational theater developed in various forms in Asia and Africa as well [1]. *Fabulae Atellanae* or *Oscan Games* (named after the town of origin, Atella, an Oscan town in Campania) emerged around 300 BC. These performances consisted of masked improvised farces designed to entertain the audience through short sketches, sometimes presented after tragedies or during public festivals. A defining feature of *Fabulae Atellanae* was the use of stock characters, each easily recognizable by its own distinctive mask, while the plot and dialogue were improvised in each performance. This form of entertainment remained popular for several centuries until Emperor Tiberius ultimately banned it due to its perceived licentiousness [40][18].

Between the 16th and 18th centuries in Italy, actors belonging to the *Commedia dell'Arte* tradition performed without a script, relying instead on a "canovacci" that provided basic information such as setting, time and a general plot. Starting from these fixed elements, the actors were then free to improvise dialogues and actions as they pleased, resulting in performances that were always new [1][8]. This tradition gradually declined following the revolution initiated by Carlo Goldoni and other contemporary playwrights, who transformed Italian theater from the improvisational *Commedia dell'Arte* to fully scripted comedies with a strong focus on social realism [6].

Around the mid-20th century in USA, theatrical improvisation was rediscovered and systematized by Viola Spolin, and later by her son Paul Sills, as a pedagogical tool to teach social skills, encourage creativity and develop "spontaneity" in young children through structured *Theater Games* [36]. In later years, one of the most influential approaches that incorporates improvisational principles was the Meisner Technique, developed by Sanford Meisner in 1941. This acting technique, still popular today, relies on structured improvisational exercises and repetition to train actors to listen attentively to one another on stage and to produce spontaneous, truthful, and instinctive reactions [26]. Between 1970 and 1980, Keith Johnstone further developed improvisation as a form of public performance. His work led to the creation of *Theatresports*, a competitive format in which two teams of actors improvise scenes in response to prompts and challenges. The strong and immediate reactions of the audience, often comparable to those observed during sporting events, inspired the name of this theatrical form [38].

Today, improvisational theater is practiced worldwide and has gained widespread popularity, partly due to its presence in television programs and live performance formats. Nevertheless, examples of fully improvised theater still exist, particularly in comedic and *avant-garde* contexts [1]. Beyond public performances, improvisation is still commonly used as a training tool for actors and as a method for developing characters, narratives, and scripts.

A defining feature of improvisational theater is, of course, the complete absence of a script: scenes develop entirely through the choices made by the actors on stage and the interactions between them. These interactions are based on a continuous process of action-reaction, which constitutes the foundation of improvisational theater. The choice of reaction is what makes the performance succeed or fail. Anytime an actor makes a proposal, the other actor on stage is expected to accept it and build upon it, following a principle that is commonly referred to as *"Yes, and"* (not doing so would be detrimental to the collaborative nature of improvisation itself) in a way that is coherent with the story developed so far and the character personality.

## 1.2. Emotions

In the next section, an attempt was made to define the word emotion for this project.

### 1.2.1. What is an Emotion?

The term "emotion" is relatively recent in the English language and its precise meaning has long been subject of debate, as it has been used to refer to a wide range of phenomena and lacks a universally accepted definition [20]. In the 17th century, when the word first appeared in Britain as a translation from the French *émotion*, it primarily referred to a physical disturbance or a movement of the body rather than to a psychological state. Previously emotional experiences were commonly described as "passion" or "affections". The word "emotion" entered academic literature between the 1830s and the 1850s [12].

A key figure in the establishment of this word was Thomas Brown, who has also been described as the "inventor of the emotions" [9]. While recognizing that a universal definition of emotion could not be easily found, he proposed that

> "if any definition of them be possible, they may be defined to be vivid feelings, arising immediately from the consideration of objects, perceived, or remembered, or imagined, or from other prior emotions" [11].

This definition implies that emotions are linked to the perception of an object in the mind, rather than the only presence of the object itself [12].

Shortly thereafter another key figure emerged: Charles Bell described briefly an emotion as "a movement of the mind" that could manifest through "outward signs", thus emphasizing the physical component to the definition of emotions and distancing himself from Brown's conception of emotion as purely mental states [12][5].

Based on these studies, Darwin later published *The Expression of the Emotions in Man and Animals*, explicitly proposing a correlation between emotions and their expression through facial mimic and physical behavior across different ages and different species. Darwin's work will become a fundamental reference for emotion research in the following century [12].

However the definition of the word "emotion" remains incomplete and controversial to these days.

### 1.2.2. Main emotions

In the late 19th century, McCosh, one of the first psychologists that studied emotions, identified more than one hundred feeling states that could be ascribed to this category, subsequent studies in the following decades managed to reduce this broad range to a more manageable set of emotions [25][12].

In the 1990s, Paul Ekman and his colleagues identified 6 "basic" emotions (joy, sorrow, anger, fear, surprise, and disgust). Each one of them could be expressed at varying levels of intensity, in particular also through specific facial expressions [13].

Around the same period, Robert Plutchik proposed a model consisting of 8 primary emotions (adding anticipation and trust to Ekman's 6 basic emotions). Plutchik organized these emotions in a circular structure resembling a color wheel, in which similar emotions are positioned adjacent to one another, while opposing emotions are placed at 180° from each other. Secondary emotions can also be obtained by combining adjacent primary emotions [31].

In order to explicitly represent variations of intensity for these first 8 emotions, Plutchik also introduced a three-dimensional model shaped more or less like a cone [30](see Figure 1.1).

An evolution of this approach is the Geneva Emotion Wheel (GEW), developed by researchers at the University of Geneva [17][35]. This model is based on empirical studies in which participants were asked to place some emotions inside a circular space defined by two axes: one for Negative/Positive Valence and the other for Low/High Power. While the GEW visually resembles Plutchik's model, it provides a more nuanced and multidimensional representation of emotional experience, presented in Figure 1.2.

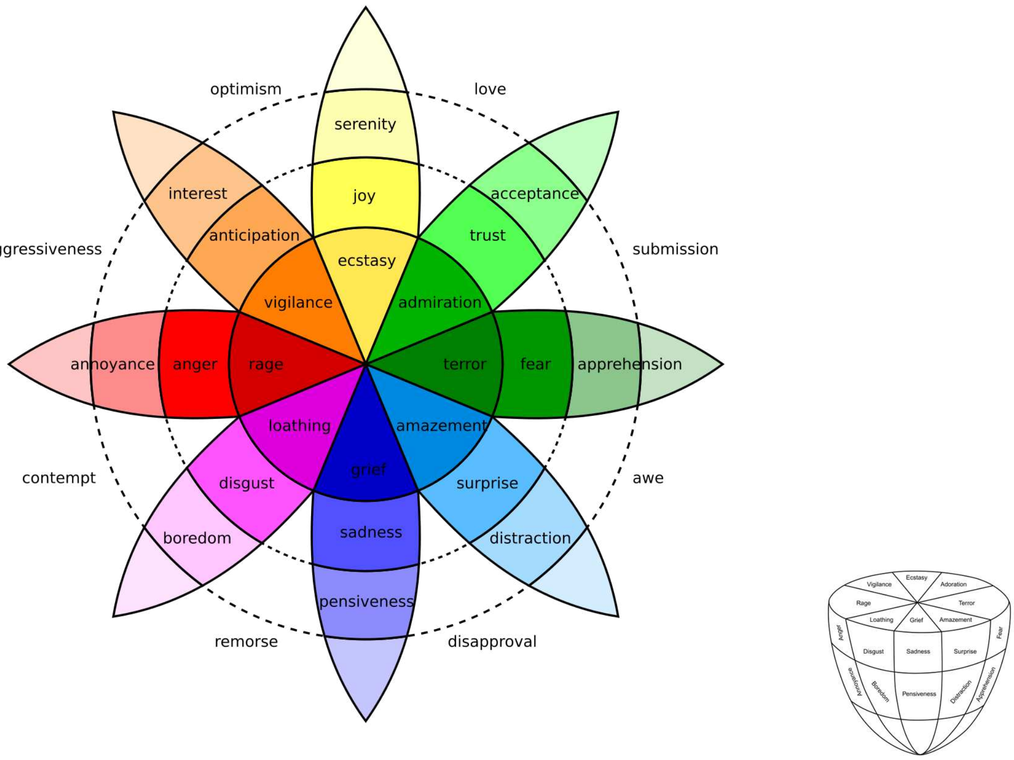
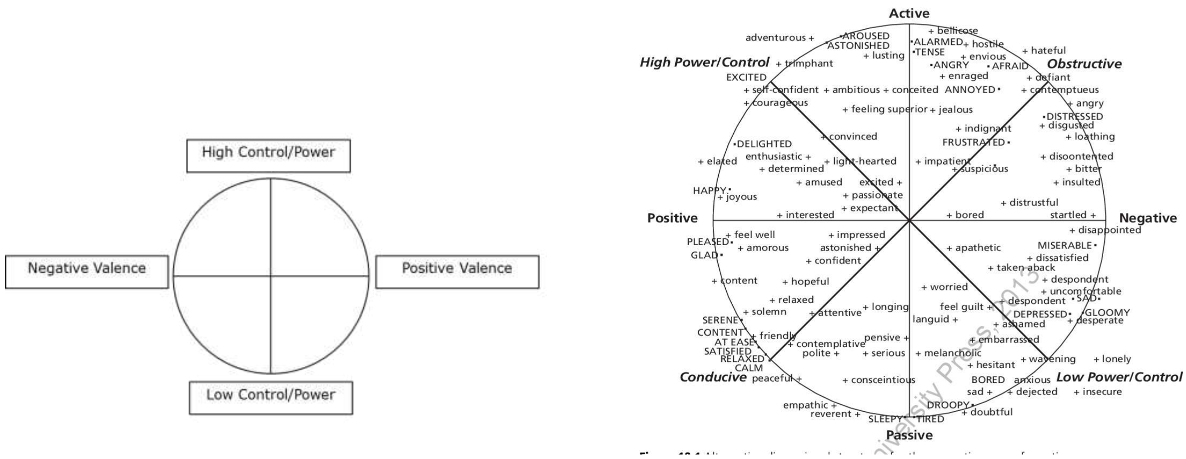



Figure 1.1: Plutchik's model showing 8 principal emotions, each with varying intensity from highest in the center, to lowest along the petals or, in the second picture, intensity is directly linked to the dimension of the section on the cone.[31]




(a) The two main axes are intensity and valence. (b) Each emotion is identified by a single point.

Figure 1.2: The Geneva Emotion Wheel model [17].

For the purposes of this project, Plutchik's model was adopted, since it offers a more complete framework than Ekman's classification while remaining less complex than the

Figure 1.3: Examples of neutral masks. They do not present any facial characteristic that could suggest an emotion[37]

Geneva wheel. Nevertheless, the latter was taken in consideration to better capture and express subtle variations in emotion meanings.

Additionally, Ekman's observation that "a threshold may need to be crossed to bring about an expressive signal, and that threshold may vary across individuals" [13]deeply inspired the interpretation of emotional expression within this study.

### 1.2.3. Expressing emotions

Humans express emotions through subtle movements of the facial muscles, as well as through full-body movements that are readily interpretable by other human beings, independently of age or cultural background. These expressive cues play a fundamental role in interpersonal communication and emotional understanding.

When it comes to robots, however, reproducing the full range of human facial and bodily movements is often technically unfeasible due to mechanical constraints. Nevertheless, it remains possible to convey emotional states in ways that are recognizable and meaningful to human observers. This idea was already explored by Jacques Lecoq, a French actor, mime, and theater pedagogue born in 1921, who deliberately employed the neutral mask (See Figure 1.3) in his theatrical practice. The neutral mask eliminates facial features and micro-movements in order to remove facial expressivity altogether, thereby emphasizing the expressive potential of the body alone. Lecoq's work demonstrates that posture, movement, and gesture can effectively communicate emotions and inner states even in the absence of facial expression, sometimes with equal clarity and intensity [22].

One of the most influential sources of inspiration for non-verbal emotional expression in artificial agents comes from the field of graphic animation. A widely recognized framework for conveying emotion through movement originates from Disney's early animation practices, which were later formalized as the fundamental *principles of traditional animation* and remain extremely influential today. These principles include:

1. **Squash and Stretch**: defining the rigidity and perceived mass of an object by altering its shape during motion. Although this principle is not directly applicable to rigid robotic structures, a sense of elasticity can still be suggested through controlled acceleration and deceleration, such as the characteristic "bounce" observed in the animated lamp *Luxo Jr.*, produced by Pixar in 1986 [21].

2. **Timing**: the spacing of actions to convey the weight, size, and personality of a character. Timing is a critical factor, as poorly calibrated motion can cause the observer to lose interest or perceive the movement as unnatural.

3. **Anticipation**: the preparatory motion preceding an action. Anticipation signals that an action is about to occur, focuses the viewer's attention, and enhances realism by reflecting the biomechanical principles of human movement, in which muscles must extend before contracting, and vice versa.

4. **Staging**: the clear presentation of an idea or action so that it is immediately understandable to the audience.

5. **Follow-Through and Overlapping Action**: the continuation of motion after the primary action has ended, reflecting the gradual dissipation of energy. Abrupt and rigid terminations of movement are generally perceived as unnatural.

6. **Straight-Ahead Action and Pose-to-Pose Action**: two contrasting methods for constructing movement, the former emphasizing spontaneity and the latter emphasizing structure and planning.

7. **Slow In and Slow Out**: the gradual acceleration and deceleration of motion, achieved by varying the spacing of intermediate positions to create smoother and more natural movement.

8. **Arcs**: the tendency of natural movement to follow curved rather than linear trajectories.

9. **Exaggeration**: the deliberate amplification of essential aspects of an action to enhance clarity and emotional impact, while avoiding excessive distortion that could undermine realism.

10. **Secondary Action**: additional movements that result from or support a primary action, contributing to greater richness and realism.
11. **Appeal**: the creation of movements or poses that are visually engaging and pleasing to the observer, for example by avoiding rigid symmetry.

Several of these principles can be translated into robotic motion design, where they can contribute to more expressive, believable, and emotionally engaging movements [21].

An alternative and complementary approach to understanding expressive movement is provided by the work of Hungarian choreographer and movement theorist Rudolf Laban. Together with his student Irmgard Bartenieff, Laban developed the Laban Movement Analysis (LMA), a framework for describing, interpreting, and documenting human movement [16].

LMA identifies four interrelated components that contribute to the qualitative aspects of movement: Body, Effort, Shape, and Space, within a broader context that also accounts for phrasing and thematic structure [24][10].

- **Body**: identifies which parts of the body are involved in the movement and examines how different body segments initiate and coordinate action.
- **Effort**: focuses on the dynamic qualities of movement, such as strength, control, and timing. It comprises four sub-components: space, weight, time, and flow.
- **Shape**: describes how the body changes form over time and how it relates to both itself and the surrounding environment.
- **Space**: refers to the spatial characteristics of movement, including volume, area, and distance, all of which evolve dynamically over time.

In particular, the dimensions of Shape and Space have been shown to be highly relevant for the generation of expressive movement in robots, as demonstrated by several studies, including those by Sara J. Burton [7] and by Nakata, Mori, and Sato [27], which will be examined in more detail in the concluding section of this chapter.

Another field that bears great resemblance with the robotic one, sharing the same issues regarding emotion expression is glove puppet animation. Most of the time, puppets' only moving part is their mouth, so while some of the traditional principles of animation can be followed, many other can't be applied due to the fact that the puppet is a physical object, subject to the limitations of physics. So this genre developed its own expressive language, solely relying on mouth and whole body movement [41].

## 1.3. Personality

The American Psychological Association (APA) Dictionary [4] defines personality as "the enduring configuration of characteristics and behavior that comprises an individual's unique adjustment to life, including major traits, interests, drives, values, self-concept, abilities, and emotional patterns". Personality can be shaped by multiple factors, including genetic predispositions, education, cultural background, and life experiences [2]. Numerous theories have been proposed to explain how personality is structured and how it develops over time; however, they all agree that personality plays a fundamental role in influencing individual behavior in everyday life.

### 1.3.1. Some theories of personality

One of the most widely accepted models in contemporary personality psychology is the Five-Factor Model (FFM), also known as the Big Five. This model is based on the assumption that personality can be described using five broad trait dimensions. These domains are:

- Factor I: Extraversion
- Factor II: Agreeableness
- Factor III: Conscientiousness
- Factor IV: Neuroticism (or Emotional Stability)
- Factor V: Openness to Experience (or Intellect)

Each of these broad dimensions encompasses a wide range of specific traits. For example, Extraversion includes traits such as talkativeness and assertiveness as opposed to silence and passivity, while Agreeableness contrasts kindness and warmth with hostility and selfishness. Conscientiousness includes reliability and organization as opposed to carelessness and negligence. Neuroticism refers to traits such as moodiness, nervousness, and emotional instability, whereas Openness to Experience contrasts creativity and curiosity with conventionality and resistance to novelty [19].

Despite various criticisms raised over the years, the Five-Factor Model remains the most empirically supported and widely used framework for the study of personality.

An alternative approach to personality theory was proposed by Carl Gustav Jung. Jung's classification identifies eight psychological types derived from the combination of four fundamental psychological functions. These functions, influenced by earlier philosophical traditions, operate in a compensatory manner: when conscious functioning becomes overly one-sided, the unconscious tends to activate the opposite or less-developed function in order to restore balance [33].

- **Thinking versus Feeling** Thinking and Feeling are considered rational functions, as they relate to the evaluation and interpretation of information and to decision-making processes.
- **Sensation versus Intuition** Sensation and Intuition are defined as irrational functions, in the sense that they are perceptive modes concerned with data gathering rather than judgment.

These functions are combined with two attitudinal types, derived from the works of Nietzsche and the psychologist William James, which describe habitual orientations toward the world:

- Introversion
- Extraversion

By combining psychological functions and attitudinal types, Jung identified eight personality types [33]:

- **Extraverted Thinking**: principled, objective, rational.
- **Introverted Thinking**: idea-oriented, independent, often reserved.
- **Extraverted Feeling**: adaptive, socially attuned.
- **Introverted Feeling**: reserved, value-driven, empathetic.
- **Extraverted Sensation**: realistic, concrete, socially engaging.
- **Introverted Sensation**: calm, restrained, internally focused.
- **Extraverted Intuition**: enterprising, outgoing, novelty-seeking.
- **Introverted Intuition**: imaginative, visionary, introspective.

### 1.3.2. Myers-Briggs studies

The Myers-Briggs Type Indicator (MBTI) is one of the most widely known tools for personality assessment. It was developed during World War II by Katharine Cook Briggs and her daughter Isabel Briggs Myers and was first published in 1962. Drawing on Carl Jung's theory of psychological types, Myers and Briggs proposed a model that classifies individuals into sixteen personality types, each represented by a four-letter acronym reflecting an individual's preferences across four dimensions [23] (Figure 1.4).

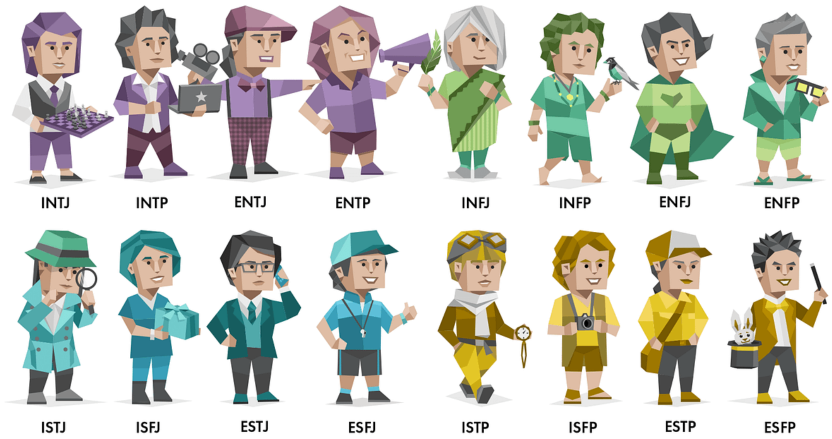



Figure 1.4: All 16 possible personality types according to the MBTI model [23]

1. **Extraversion (E) / Introversion (I)** This dimension describes how individuals orient their energy, either toward the external world or toward their internal thoughts and reflections.
2. **Sensing (S) / Intuition (N)** This dimension concerns how individuals perceive information. Sensing types focus on concrete facts and present realities, whereas intuitive types are more oriented toward patterns, possibilities, and future implications.
3. **Thinking (T) / Feeling (F)** This dimension relates to decision-making processes. Thinking types tend to prioritize logic and objective criteria, while feeling types place greater emphasis on values and emotional consequences.
4. **Judging (J) / Perceiving (P)** This dimension reflects an individual's approach to structure and planning. Judging types prefer organization and predictability, whereas perceiving types tend to be more flexible and adaptable.

Myers and Briggs also developed a questionnaire designed to help both psychologists and individuals gain insight into personality-related preferences. The MBTI questionnaire is now widely accessible online. A fifth dimension, Assertive (A) versus Turbulent (T), was introduced in later adaptations of the model; however, it will not be considered for the purposes of this study.

Despite its popularity, the MBTI has been widely criticized in scientific literature for its limited reliability and validity. Studies have shown that a significant proportion of individuals receive different results when retaking the test after a short period of time, and factor-analytic research has failed to provide strong empirical support for the proposed type structure. Furthermore, correlations between some of the MBTI scales challenge the theoretical independence of the dimensions. For these reasons, the MBTI has often been described as lacking a solid empirical foundation, with the U.S. Army Research Institute suggesting that its types may reflect stereotypical categorizations rather than stable psychological constructs [29].

Nevertheless, although the MBTI does not represent the most scientifically robust personality assessment method, it provides a psychological framework that is intuitive, easily interpretable, and highly communicable. These characteristics make it particularly suitable for contexts such as the one of the present project, where an artificial personality is designed to interact in a realistic manner with human users.

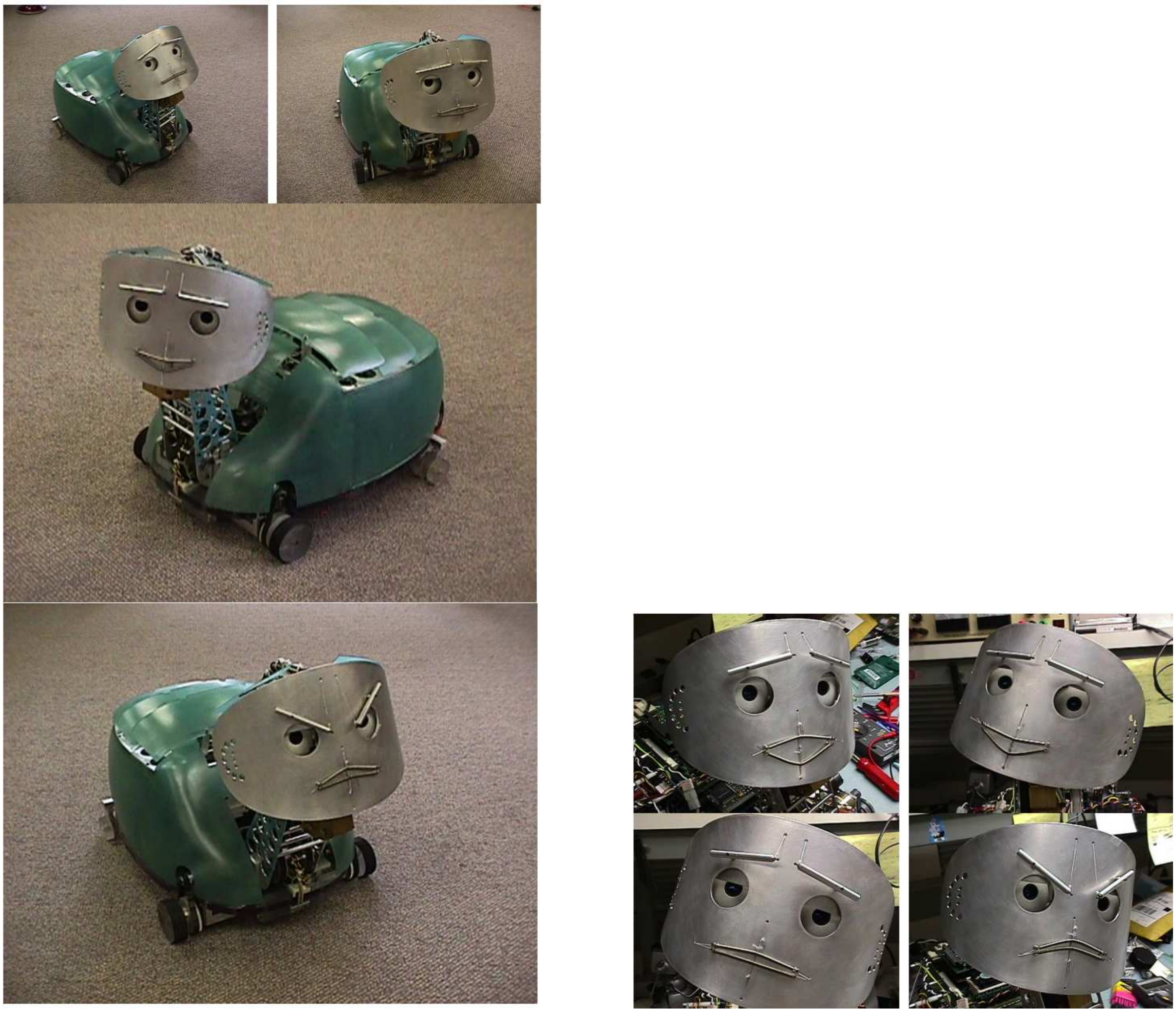
## 1.4. Robot emotions: State of the art

Together with the field of emotional robotics, some artists are making use of especially designed robot as working actors in theater plays. The aim is similar in both cases, but the effect is reached through very different paths.

### 1.4.1. Emotional robots

The following examples illustrate three complementary approaches to robotic emotional expression: facial animation, full-body movement, and theatrical embodiment.

**Sparky** is a tele-operated social robot developed by researchers at Interval Research Corporation with the goal of interacting with humans through the expression of emotions. The primary design objective was to create an engaging and visually appealing robot equipped with an active, integrated body and a face capable of conveying emotional states.



(a) Left to right: afraid and inquisitive. Below: happy and angry [34]

(b) Clockwise from upper left: surprise, happiness, anger, sadness [34]

Figure 1.5: The robot Sparky expressing different emotional states

Sparky is capable of expressing nine emotional states: neutral, happy, sad, angry, surprised, fearful, inquisitive, nervous, and sleepy. These emotions are conveyed through a combination of facial expressions, body posture, and the overall quality of the robot's motion. For instance, when Sparky is in a happy state, it approaches people while smiling and producing cheerful vocalizations. Conversely, when expressing sadness, fear, or nervousness, its movements and sounds are altered accordingly. The robot can also display anger by charging toward a person who is mistreating it (Figure 1.5). Public interaction experiments yielded generally positive results: both children and adults approached the robot spontaneously and tended to treat Sparky similarly to an animal or a small child [34].

**EMYS** is an educational robot developed by researchers at the *Universidade Técnica de*

*Lisboa* with the aim of interacting with children and supporting English language learning. The robot has eleven degrees of freedom, including head rotation and mouth opening and closing, but lacks lower eyelids and independent eye movement. These characteristics make its expressive capabilities resemble those of a mechanical puppet (Figure 1.6).

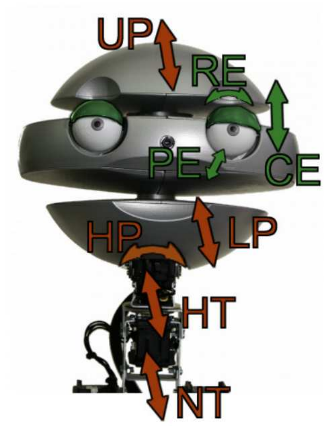
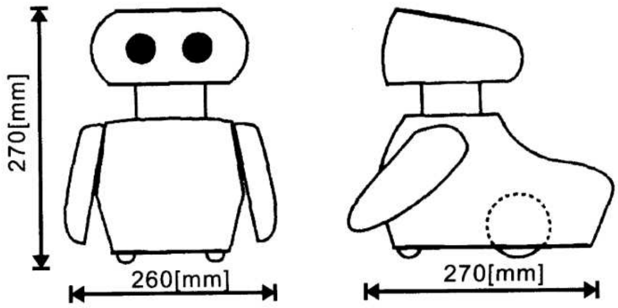


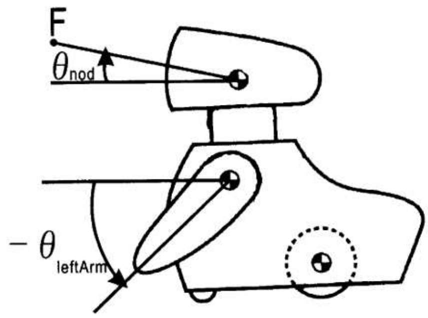

Figure 1.6: EMYS face structure, with all movable components highlighted [41].

To enhance EMYS's emotional expressiveness, researchers applied classical animation principles such as exaggeration, timing, and staging, drawing inspiration from animators like Chuck Jones and Jim Henson [21]. Initially, EMYS was designed to express the six basic emotions identified by Ekman, each at three different intensity levels. Emotional states were mapped onto the robot's facial features by assigning numerical values to each movable component within a predefined range of positive, negative, and neutral values. Each emotion was then represented by a specific configuration of these parameters. Although the resulting expressions appeared more cartoon-like than realistic, user studies showed that most participants were able to correctly recognize the intended emotions [41].

An alternative approach based on Laban Movement Analysis was explored by researchers at the National Institute of Advanced Industrial Science and Technology in Japan. Their experimental **dancing robot** was equipped with one degree of freedom for each arm, one degree of freedom for head nodding, and two wheels that enabled free movement on the floor (Figure 1.7).




(a) Experimental dancing robot [27]




(b) Experimental dancing robot [27]

Figure 1.7: Experimental dancing robot used in emotion expression studies.

In this study [27], particular emphasis was placed on quantifying Laban's concepts of Effort and Shape and translating them into computable parameters that could be evaluated in real time. Feedback collected from public interaction experiments indicated a strong correlation between movements characterized by pronounced Laban features and the perception of basic emotional states, thus verifying the study's initial assumption.

These studies provide important insights for the design of emotionally expressive artificial agents, particularly in contexts involving human–robot interaction and performative behavior.

### 1.4.2. Theatrical robots

Many notable examples of robots in theatrical performances originate from Japan. Among these, the work of playwright and director Oriza Hirata is particularly significant, not merely for featuring robots alongside human actors, but for moving beyond the use of robots as mere spectacle. In Hirata's productions, robots take distinct dramatic roles, and theatrical conventions are employed to explore fundamental questions surrounding human–robot coexistence.

Hirata's first robot theater performance, *I, Worker* (Figure 1.8), premiered in 2008. Set in the near future, the play portrays the everyday life of a married couple living with two service robots, Takeo and Momoko. Both robots are of the Robovie R3 type and are distinguished primarily by color, with the "female" robot further characterized by the use of an apron.

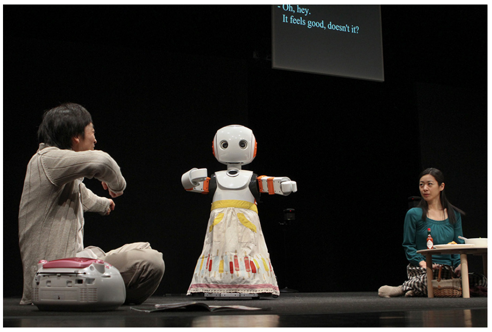



Figure 1.8: *I, Worker* by Oriza Hirata (2008) [32].

In 2014, as part of an international collaboration, Hirata directed the robot theater project *La Métamorphose* (Figure 1.9), featuring the android Repliee S1 in the role of Gregor Samsa. In this adaptation, Kafka's story is relocated to a near-future French town, where Gregor Samsa awakens to discover that he has transformed not into an insect, but into an android.

Figure 1.9: *The Metamorphosis: Android Version* by Oriza Hirata (2014) [39].

Although robots may experience technical malfunctions, they do not fail in the same way as human performers. Since the robots featured in Hirata's productions lack artificial intelligence, they are unable to autonomously respond to unexpected situations and therefore do not possess the capacity for improvisational cooperation or creative agency. As a result, the performative process becomes asymmetrical: the robot executes a predefined sequence of actions, while the human actor adapts, reacts, and compensates in real time. This distinction becomes less pronounced in cases where the robot is remotely controlled in real time [39].

# 2 | Affective Architecture and Emotion-based Logic

## 2.1. Project Logic

The goal of this project was to design a robot capable of interacting with a human actor on a stage, without a script, in a way that is interesting but also realistic. To do so, the robot adapts its reactions based on three main factors:

- its **personality**,
- its **current mood**, and
- the **interactions** received from the human actor.

### 2.1.1. Emotions

In order to model emotional dynamics, each emotion is associated to a *"jar"* whose filling level varies over time. Each emotional jar represents the drive to express a specific emotion, gradually filling (or emptying) itself in response to events that enforce (or dampen) that emotion. Depending on the robot's personality, its current mood, and the specific interaction occurring at a given moment, each emotional jar is filled or emptied at different rates.

The *level* of the content of each jar is modeled as a continuous floating-point value in the interval $[0, 1]$, where 0 corresponds to a completely empty jar (no expressive drive), and 1 represents the maximum level of accumulated expressive pressure.

It is important to clarify that, in this model, the jars do not represent the emotion itself, but rather the internal need or pressure to express that emotion. The filling process therefore reflects the accumulation of expressive tension, rather than the continuous presence of an emotional state.

The emotional architecture considers a total of eight emotions, based on Plutchik's emotion model [30]:

- Anger
- Anticipation
- Disgust
- Fear
- Joy
- Sadness
- Surprise

Among these, five emotions (*Anger*,*Anticipation*,*Joy*, and *Sadness*) are modeled as *cumulative dimensions*, meaning that activation of their expression results from the progressive accumulation of interaction effects over time through the jar mechanism previously described. Two additional emotions (*Disgust* and *Surprise*) do not follow this cumulative mechanism and exhibit a different activation logic. Since their dynamics are event-based and cannot be meaningfully represented through gradual accumulation, they are treated separately and will be discussed later in this chapter. The eighth emotion, *Trust*, is not explicitly modeled in the current system. Its representation would require more complex mechanisms related to long-term interaction and relational dynamics, which are beyond the scope of the current implementation. For this reason, Trust was not included in the model and will not be further considered here.

For each emotion, three thresholds are defined, corresponding to increasing intensity levels inspired by Plutchik's emotion wheel [30]. When the value of an emotional jar exceeds one of these thresholds, the robot expresses the corresponding emotion at the associated intensity level. If multiple emotions cross a threshold simultaneously, only the emotion with the highest value is expressed. After expression, the selected emotion is reset (its jar is emptied), while all other emotional values remain unchanged until the next interaction.

Emotion values are stored in an array structure, where each element contains the floating-point value representing the fullness of the corresponding emotional jar. After each interaction, dedicated update equations modify these values according to the robot's personality traits, current mood, and the type of interaction received. These mechanisms, which regulate the loading dynamics of the jars, will be detailed in the following sections.

### 2.1.2. Interactions with the Human Actor

Six types of interactions were considered in this project. Additional interactions can easily be introduced in future extensions of the system.

- **Hit**: the robot receives a punch or strong shove. Negative interaction.
- **Caress**: the robot receives a gentle stroke. Positive interaction.
- **Like**: the robot perceives something it evaluates as pleasant (in the implemented prototype, the color *orange*). Positive interaction.
- **Dislike**: the robot perceives something it evaluates as unpleasant (here, the color *blue*). Negative interaction.
- **Noise**: the robot detects a loud sound. Negative interaction.
- **Closeness**: the human actor approaches the robot. The interpretation of this action depends on the robot's emotional state. A joyful robot may interpret proximity as friendly, while a fearful robot interprets it as threatening.

Each interaction is therefore assigned five values (positive or negative), one for each of the cumulative emotions, indicating the degree to which the interaction increases or decreases the corresponding expressive drive.

| **Emotions/Interactions** | Hit | Caress | Like | Dislike | Noise | Closeness |
|---|---|---|---|---|---|---|
| Anger | 0.55 | -0.70 | -0.55 | 0.35 | 0.15 | 0.20* |
| Anticipation | -0.2 | 0.05 | 0.85 | 0.73 | 0.55 | 0.40* |
| Fear | 0.75 | -0.65 | -0.2 | 0.03 | 0.50 | 0.45* |
| Joy | -0.77 | 0.40 | 0.40 | -0.50 | -0.20 | 0.25* |
| Sadness | 0.30 | -0.55 | -0.5 | 0.50 | 0 | -0.60* |

Table 2.1: Interaction values for all suitable emotions

As it is possible to infer from Table 2.1, closeness behaves differently from the other interactions, since its influence on the emotions depends entirely on the prevalent emotion in that specific moment; so it will only affect that one, while having zero influence on the others.

### 2.1.3. Mood

Mood is defined as "a disposition to respond emotionally in a particular way that may last for hours, days, or even weeks, perhaps at a low level and without the person knowing what prompted the state. Moods differ from emotions in lacking an object; for example, the emotion of anger can be aroused by an insult, but an angry mood may arise when one does not know what one is angry about or what elicited the anger" [3].

In this project, mood is modeled as a global affective state that modulates the robot's emotional dynamics over time. Unlike emotions, which are associated to discrete triggered events related to specific interactions, mood acts as a slowly varying background condition that influences how the robot reacts to stimuli.

Mood is represented through two components: a valence and an intensity. The valence can be either positive or negative, while the intensity is modeled as a continuous value in the interval $[0, 1]$, representing the strength of the current mood state. Values close to zero indicate a neutral or negligible mood influence, whereas values close to one correspond to a strongly polarized affective condition.

At the beginning of the performance, the initial mood is set by the user. Over time, mood intensity gradually decreases following a linear decay process, simulating the natural fading of affective states. The rate of this decay depends on the robot's personality configuration and on the number of interactions in which it has participated. In this way, some personality profiles tend to maintain mood states for longer periods, while others return more quickly to a neutral condition.

Mood is also influenced by the robot's own emotional expressions. Each time an emotion is expressed, a feedback effect modifies the current mood. Expressions with a positive valence reinforce a positive mood (or attenuate a negative one), whereas expressions with a negative valence reinforce a negative mood (or weaken a positive one). The strength of this feedback depends on personality traits related to affective sensitivity: more emotionally sensitive configurations produce stronger mood changes, while more emotionally stable configurations result in weaker modulation.

If no active mood is present (intensity equal to zero), the expression of an emotion establishes a new mood with the corresponding valence. If the expressed emotion has the same valence as the current mood, the mood intensity increases, approaching its maximum value asymptotically. Conversely, if the expressed emotion has an opposite valence, the mood intensity decreases proportionally to its current strength and may eventually return to neutrality.

Through this mechanism, mood functions both as a slowly decaying affective background and as a self-reinforcing process influenced by emotional expression. This bidirectional interaction between mood and emotions contributes to generating coherent and temporally structured affective behavior, preventing purely reactive or memoryless responses.

### 2.1.4. Personality Features

Personality influences the evolution of emotions over time. To preserve generality, specific personality traits have been mapped to computational effects. Four main rules have been implemented:

- **Extraversion vs. Introversion**: for extraverted personalities, all emotional expression thresholds are multiplied by a scaling factor `IE_scaler`, making emotion expression easier to trigger.

$$Threshold(t) = Threshold(t-1) \cdot IE\_scaler \tag{2.1}$$

- **Feeling vs. Thinking** Mood modifications are much more accentuated in Feeling-oriented personalities, due to their naturally stronger response to emotions. To model this difference, all mood variations produced by emotional activation are multiplied by an amplifying factor `TF_scaler` for Feeling-inclined personalities.

$$MoodEffect(t) = MoodEffect(t-1) \cdot TF\_scaler \tag{2.2}$$

- **Intuition vs. Sensing**: intuitive personalities assign stronger weight to expected future patterns. For example, one caress may increase *Joy* more strongly for an Intuitive personality, which tends to anticipate repetition of a pleasant stimulus. This effect is computed through the term $INCLINATION$ in the update equations, which will be introduced in the next paragraph.
- **Perceiving vs. Judging**: perceiving personalities adapt more quickly and rely less on past experience. Judging personalities, by contrast, weight previous emotional states more strongly. This distinction is modeled through the parameter $\alpha$ in the update equations, which will be explained in the following paragraph.

### 2.1.5. Emotion Update Equations

The eight considered emotions can be distributed in two valence categories:

- **Positive**: *Anticipation*, *Joy*, *Surprise*, *Trust*,
- **Negative**: *Anger*, *Disgust*, *Fear*, *Sadness*.

At the start of the simulation, the initialization of emotional states depends on the robot's prevailing mood valence.

**Positive mood** When the robot begins in a positive mood, baseline levels of positive emotions are slightly elevated, even before any external stimuli or events occur.

**Negative mood** Conversely, an initial negative mood produces a slight elevation of negative emotional states.

In both cases, these effects are amplified for robots whose personality profile belongs to the **Feeling (F)** type. Robots with a Feeling-oriented personality trait exhibit heightened emotional sensitivity, so their initial emotional states are more strongly influenced by mood valence.

From an implementation perspective, the array of emotional jars is first initialized with neutral baseline values. A multiplicative factor associated with the Thinking/Feeling trait( called `TF_scaler`) is then computed: this factor is increased for Feeling personalities in order to model their greater emotional responsiveness. Subsequently, small positive or negative offsets are applied to selected emotions according to the initial mood valence. A negative mood slightly increases baseline levels of negative emotions, while a positive mood produces the opposite effect by enhancing positive emotions and attenuating negative ones. This initialization step therefore introduces a coherent emotional bias that reflects both the robot's mood and its personality before any interaction takes place.

For all emotions except *Disgust* and *Surprise*, the value after each interaction is computed using the following equations.

**Matching valence** between mood and emotion:

$$emotion_i[t] = emotion_i[t-1] + \alpha \cdot emotion_i[t-1] + |1-\alpha| \left(INTERACTION_j \cdot mood + Inclination(\beta)\right) \tag{2.3}$$

**Opposite valence** between mood and emotion:

$$emotion_i[t] = emotion_i[t-1] + \alpha \cdot emotion_i[t-1] + |1-\alpha| \left(INTERACTION_j \cdot (1-mood) + Inclination(\beta)\right) \tag{2.4}$$

Where:

- $emotion_i[t]$ is the emotional value of emotion $i$ after the current interaction,
- $emotion_i[t-1]$ is the level of the emotional jar before the interaction, it contains the "emotional history" of the robot.
- $INTERACTION_j \in [-1, 1]$ is the effect of interaction $j$ on emotion $i$ before personality scaling,

- $\beta \in \mathbb{N}$ is the number of consecutive repetitions of that same interaction before the current one,

  **Example:** At the beginning of the simulation, the robot receives a "Hit":
  $\beta = 0$
  Then another one:
  $\beta = 1$
  Then again and again, $N$ more times:
  $\beta = 2$
  ...
  $\beta = N - 1$
  until an interaction that is different from "Hit" occurs, then the *beta* counter starts again:
  $\beta = 0$

- $Inclination(\beta)$ is an additional term derived from the derivative of a function that depends on the second MBTI personality dimension (*Intuition/Sensing*). This function models the increment of $emotion_i$ proportionally to the number of repetitions of a given interaction ($\beta$). Only the derivative with respect to $\beta$ is considered, as it represents the rate at which the emotion value increases. For characters with a stronger Intuition preference, each action contributes not only its immediate effect but also potential future implications. Mathematically, this function is modeled as an exponential growth, which is faster for intuitive personalities and slower for sensing personalities. As an example, Figure 2.1 illustrates the behavior of the $Inclination(\beta)$ function for Sensing and Intuitive personalities (ISTJ and INTJ, respectively). In the Figure:

  - $f_N$: emotional weight for **Intuitive** personalities, showing faster exponential growth and greater sensitivity to repeated actions.

  - $f_S$: emotional weight for **Sensing** personalities, showing slower growth.

  - Dashed lines: tangents to the functions, representing the rate of change captured by the $Inclination(\beta)$ term in the equations.

  Slightly different functions are considered for different emotions, in order to keep the robot's reactions coherent.

- $\alpha \in [0, 1]$ weights the influence that past emotional states maintain on the present and depends on the fourth personality dimension of the MBTI. Since both these

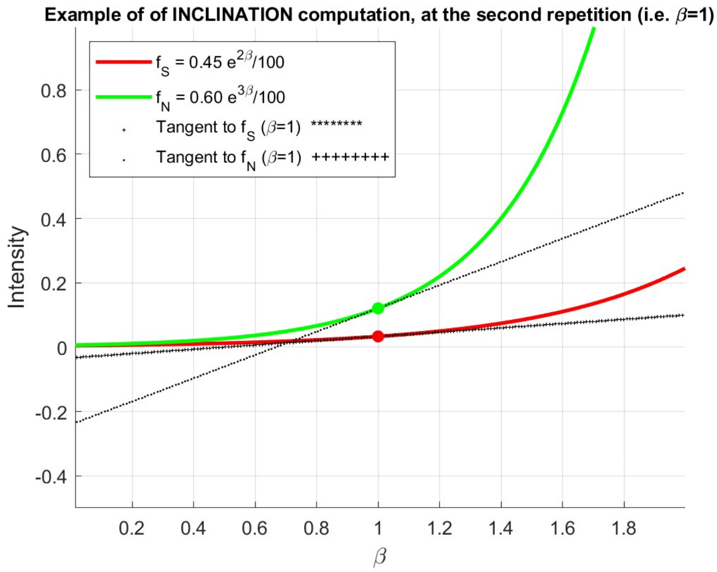



Figure 2.1: Example: Illustration of the $Inclination(\beta)$ functions for INTJ (Intuitive) and ISTJ (Sensing) personalities, at the second repetition of a certain action (i.e $\beta = 1$).

equations can be rewritten like this:

$$emotion_i[t] = EffectsOfPastEvents + \alpha \cdot EffectsOfPastEvents + |1 - \alpha| \cdot PresentEvents$$

A smaller $\alpha$ indicates that present events have a stronger influence on the emotional state, whereas a larger $\alpha$ increases the persistence of past interactions, reducing the relative impact of the current input. This behavior reflects differences between *Perceiving (P)* and *Judging (J)* personalities: P types are more responsive to current events and adapt quickly, while J types tend to retain past emotional states longer, allowing prior experiences to influence their reactions. To implement this effect, the previous emotional value $EffectsOfPastEvents$ is carried over into the current computation as $EffectsOfPastEvents \cdot (1 + \alpha)$. This ensures that the emotional state at time $t$ is influenced not only by the current interaction but also by a weighted memory of past states. In other words, a larger $\alpha$ gives more "emotional inertia", so that past experiences continue to affect future emotional dynamics, while a smaller $\alpha$ emphasizes immediate reactions. Again, slightly different values of $\alpha$ can be assigned to different emotions to achieve more nuanced and realistic interactions.

### 2.1.6. Special Case: Disgust and Surprise

The emotions *Disgust* and *Surprise* do not accumulate over time in a meaningful way.

- **Surprise** decreases immediately upon the presentation of the event, since the event is no longer unexpected. Also it is has priority of expression on all other emotions, reflecting the need for an immediate reaction to novel or unexpected stimuli.
- **Disgust** does not strengthen simply due to repeated exposure.

For this reason, these emotions follow a discrete-level model: each interaction directly assigns a value corresponding to its intensity, overwriting the previous level rather than building upon it.

#### Surprise

The modulation of Surprise is based on three main principles:

- **Personality influence (Judging vs. Perceiving)**: The Judging/Perceiving dimension of the MBTI affects the sensitivity to unexpected events. Judging personalities, which tend to prefer structure and predictability, are modeled as more susceptible to surprise than Perceiving personalities. To account for this effect, a scaling factor, denoted as `JP_scaler`, is introduced. This factor takes higher values for personalities including Judging (ending in J)and lower values for those including the perceiving character (ending in P).
- **Habituation to repeated stimuli**: The intensity of surprise generated by an unexpected external event (e.g., a loud noise) is maximal at its first occurrence and decreases with repeated exposure. This effect is modeled conceptually as:

$$emotion[5] = 0.5 * JP\_scaler - (\beta/10) \tag{2.5}$$

  where $\beta$ represents the number of consecutive occurrences of the same interaction.
- **Context-dependent surprise**: Surprise can also arise from sequences of events, not only from individual ones. In particular, transitions between categories of stimuli — for example, from a positive event to a negative one, or vice versa — are treated as unexpected and elicit stronger Surprise. This conceptual approach generalizes the effect beyond specific preselected actions, capturing the idea that novelty or contrast between consecutive events drives the emotional response.

Formally, this can be expressed as:

$$emotion[5] = BaseValue \cdot JP\_scaler \cdot \delta_{\mathrm{P}\to\mathrm{N}}$$

where $\delta_{\mathrm{P}\to\mathrm{N}} = 1$ if an unexpected sequence of positive-to-negative (or negative-to-positive) actions occurs, and 0 otherwise.

In the specific case of this project, a sudden *Hit* following a *Caress* is treated as such an unexpected sequence, producing a high Surprise value:

$$emotion[Surprise] = 0.65 \cdot JP\_scaler \cdot \delta_{\mathrm{Caress}\to\mathrm{Hit}}$$

with

$$\delta_{\mathrm{Caress}\to\mathrm{Hit}} = \begin{cases} 1 & \text{if the current action is } \textit{Hit} \text{ and the previous action was } \textit{Caress} \\ 0 & \text{otherwise.} \end{cases}$$

### Disgust

Similarly, for disgust modulation the Extraversion/Introversion dimension is used as a scaling factor. Extraverted personalities are modeled as more likely to express higher levels of disgust, whereas Introverted personalities tend to exhibit more restrained reactions. To reflect this behavior, the values assigned to disgust for specific interactions (or interaction sequences) are calibrated to lay close to the activation thresholds. This allows the same stimulus to trigger a higher intensity level in Extraverted personalities and a lower level in Introverted ones.

Formally, the Disgust value generated by an interaction can be expressed as:

$$emotion[2] = BaseValue \cdot IE\_scaler$$

where $IE\_scaler$ takes higher values for Extraverted personalities and lower values for Introverted ones.

For example, assuming a threshold value of `THRESHOLD_2` $= 0.75$, the interaction sequence in which a *Closeness* action follows a *Dislike* event is modeled like this:

$$emotion[5] = 0.85 \cdot IE\_scaler$$

and produces a high-level response for Extraverted personalities while remaining below the highest threshold for Introverted ones.

Since the base value is intentionally set close to the activation threshold, this sequence produces a high-level response for Extraverted personalities (for which $IE\_scaler > 1$), while remaining below the highest threshold for Introverted ones (for which $IE\_scaler \leq 1$).

## Example 1:

In this example, the robot is configured with an **ENFP** personality and an initial **positive mood** with an intensity of **99%**. Listing 2.1 shows a sample interaction sequence with the human actor. The robot first receives a caress, then is presented with an object associated with a negative preference (a blue pencil case), followed by an object associated with a positive preference (a yellow/orange scarf).

It is worth noting that the robot produces two different responses(first *Interest*, then *Serenity*) to the same positive stimulus ("like"), while maintaining behavioral coherence with the interaction context. This variability is generated by the evolution of the internal emotional state and contributes to making the interaction less predictable and more dynamic.

The example illustrates how personality, mood, and accumulated emotional changes influence the robot's reactions over time, allowing the system to generate adaptive and context-dependent expressive behaviors.

**Listing 2.1:** Output of the program: case 1

```
——INITIALIZATION——
    Anger:0.100
    Anticipation:0.150
    Disgust:0.000
    Fear:0.100
    Joy:0.250
    Sadness:0.000
    Surprise:0.000
    Trust:0.150

//... some variable checks...

——It's SHOWTIME!———
```

```
Input received: 'caress'
    Anger: 0.126
    Anticipation: 0.377
    Disgust: 0.000
    Fear: 0.083
    Joy: 0.516
    Sadness: 0.000
    Surprise: 0.000
    Trust: 0.150

Expressing SERENITY
[ESP] Serenity triggered

Mood pos increased by 0.01, new mood value: 0.92
---
Input received: 'dislike'
    Anger: 0.223
    Anticipation: 0.128
    Disgust: 0.650
    Fear: 0.145
    Joy: 0.000
    Sadness: 0.112
    Surprise: 0.000
    Trust: 0.150

Expressing DISGUST
[ESP] Disgust triggered

Mood n decreased by 0.17, new mood value: 0.58
---
Input received: 'like'
    Anger: 0.000
    Anticipation: 0.538
    Disgust: 0.000
    Fear: 0.000
    Joy: 0.249
    Sadness: 0.000
```

```
    Surprise: 0.000
    Trust: 0.150

Expressing INTEREST
[ESP] Interest triggered!

Mood p increased by 0.04, new mood value: 0.59
---
Input received: 'like'
    Anger: 0.000
    Anticipation: 0.446
    Disgust: 0.000
    Fear: 0.000
    Joy: 0.596
    Sadness: 0.000
    Surprise: 0.000
    Trust: 0.150

Expressing SERENITY
[ESP] Serenity triggered

Mood p increased by 0.04, new mood value: 0.57
---
//...end of simulation
```

### Example 2:

In this second example, the robot is configured with an **INFJ** personality and an initial **negative mood** with an intensity of **99%**. Listing 2.2 shows a sample interaction sequence with the human actor. In this case, it is important to observe that the selected personality exhibits lower reactivity, meaning that some interactions do not immediately trigger an emotional expression. This behavior reflects higher activation thresholds and greater emotional persistence associated with the current internal configuration. Another relevant aspect is that, despite the reduced reactivity, higher-level emotional states can still be reached either through specific combinations of stimuli or through repeated interactions. This demonstrates the cumulative nature of the emotional dynamics, where repeated or consistent inputs progressively influence the internal state until the activation threshold for expression is exceeded.

**Listing 2.2:** Output of the program: case 2

```
——INITIALIZATION——
    Anger:0.250
    Anticipation:0.000
    Disgust:0.100
    Fear:0.100
    Joy:0.000
    Sadness:0.250
    Surprise:0.000
    Trust:0.000

//... some variable checks...

——It's SHOWTIME!———
Input received: 'dislike'
    Anger: 0.568
    Anticipation: 0.000
    Disgust: 0.650
    Fear: 0.189
    Joy: 0.000
    Sadness: 0.688
    Surprise: 0.000
    Trust: 0.000

Expressing PENSIVENESS
[ESP] Pensiveness triggered

Mood n increased by 0.01, new mood value: 0.94
——
Input received: 'closeness'
    Anger: 0.568
    Anticipation: 0.000
    Disgust: 0.900
    Fear: 0.189
```

```
    Joy: 0.000
    Sadness: 0.000
    Surprise: 0.000
    Trust: 0.000

Expressing DISGUST
[ESP] Disgust triggered

Mood n increased by 0.03, new mood value: 0.89
---
Input received: 'dislike'
    Anger: 1.000
    Anticipation: 0.000
    Disgust: 0.700
    Fear: 0.365
    Joy: 0.000
    Sadness: 0.327
    Surprise: 0.000
    Trust: 0.000

Expressing RAGE
[ESP] Rage animation started!

Mood n increased by 0.05, new mood value: 0.92
---
Input received: 'caress'
    Anger: 0.000
    Anticipation: 0.025
    Disgust: 0.000
    Fear: 0.000
    Joy: 0.361
    Sadness: 0.000
    Surprise: 0.000
    Trust: 0.000

//no reaction
---
```

```
Input received: 'caress'
    Anger: 0.000
    Anticipation: 0.094
    Disgust: 0.000
    Fear: 0.000
    Joy: 0.754
    Sadness: 0.000
    Surprise: 0.000
    Trust: 0.000

Expressing JOY
[ESP] Joy triggered

Mood p decreased by 0.13, new mood value: 0.45
---
Input received: 'hit'
    Anger: 0.367
    Anticipation: 0.026
    Disgust: 0.000
    Fear: 0.462
    Joy: 0.000
    Sadness: 0.248
    Surprise: 0.000
    Trust: 0.000

//no reaction
---
Input received: 'hit'
    Anger: 0.822
    Anticipation: 0.061
    Disgust: 0.000
    Fear: 0.996
    Joy: 0.000
    Sadness: 0.604
    Surprise: 0.000
    Trust: 0.000
```

```
Expressing TERROR
[ESP] Terror triggered
---
//...end of simulation
```

# 3 | Hardware Description

## 3.1. General description

The robot employed in this project is named **Blackwing**. Its first prototype was developed in 2021at AIRLab, Politecnico di Milano, and then used also within the framework of the *My Robot Body* project in 2024. The goal of this project was to investigate the relationship between human bodies and non-anthropomorphic robots, with the aim of fostering personal body awareness and enhancing interpersonal communication [14]. At that stage, Blackwing consisted solely of its distinctive black wing mounted on an omnidirectional wheelbase (see Figure 3.1).

Figure 3.1: The robot *Blackwing* at work for the project *My Robot Body* in 2024. Controlled by a participant with sound and motion of the head and wrist [14].

Following the initial experimental phase, the robot underwent several hardware upgrades aimed at increasing its expressive capabilities and autonomy. In particular, a more powerful onboard computer was integrated, along with an additional articulated limb designed to function as a neck. This limb supports a camera that acts as the robot's head, enabling basic visual perception and head-based expressive movements.

The current version of Blackwing thus represents an evolution from a single expressive element to a more articulated robotic body, while still preserving its non-anthropomorphic design philosophy(see Figure 3.2).

Figure 3.2: The robot *Blackwing* today

Blackwing's overall dimensions are approximately 175 cm in height, 40 cm in width, and 60 cm in depth, when the robot is idle, but of course this measurements can change depending on the wing position. It could be divided into 3 main sections:

- **Top Section**: includes the head and the *wing*, a black trapezoidal piece of cloth. It connects to the middle section through a wooden plane.
- **Middle Section**: is mostly hodden under a layer of clothes, hiding the power supply, the GPU and most of the cabling. Right below the three sonars are placed in the front and one in the back. This section is about 65cm high.
- **Bottom Section**: is composed of the wheels and the wheelbase. It is semi-transparent and some of the machinery can be seen through the transparent plastic cover.

## 3.2. Top: Head and Wing

The upper section of the robot is actuated by four servo motors, which are controlled at the lowest control level by an ESP32 board (specifications in tab 3.1 and figure 3.3). The

ESP32 receives high-level commands from the main computer specifying the emotions to be expressed and moves the servos accordingly.

Table 3.1: ESP32-WROOM-32 Key Specifications [15]

| Feature | Specification |
|---|---|
| Microcontroller | Dual-core Tensilica LX6, up to 240 MHz |
| Memory | 520 KB SRAM, 4 MB Flash |
| Wireless Connectivity | Wi-Fi 802.11 b/g/n, Bluetooth v4.2 (BLE) |
| GPIO | 34 pins (digital/analog functions) |
| ADC / DAC | 12-bit ADC, 2 × 8-bit DAC |
| Interfaces | UART, SPI, I²C, PWM |
| Power Supply | 3.0-3.6 V DC |
| Dimensions | ∼18 × 25.5 mm |

Figure 3.3: ESP32-WROOM-32 Pin Layout (Top View) [15]

### 3.2.1. Wing: open/close motion

The wing is supported by two aluminum rods, each approximately 95 cm long and 6 mm in diameter. These rods are connected to a mechanism consisting of a vertical metal plane on which a DM996 servo motor is mounted (for specifications see Tab 3.2). The servo actuates the open/close motion of the wing via two small connecting rods linked to 3D-printed crescent-shaped parts, which in turn hold the metal rods of the wing.

The DM996 servo provides high torque, making it suitable for handling the inertia and weight of the wing structure during dynamic movements.

Table 3.2: DM996 Specifications

| Parameter | DM996 |
|---|---|
| Power Supply | 4.8-6 V |
| Maximum Torque | 11 kg·cm |
| Maximum Speed | 0.18 s/60° |
| Range | 180° |
| PWM | 50 Hz |

Due to mechanical constraints and inertia, the servo operates within the range of [10°,100°], with a default idle position of 40°(see Figure 3.5a).

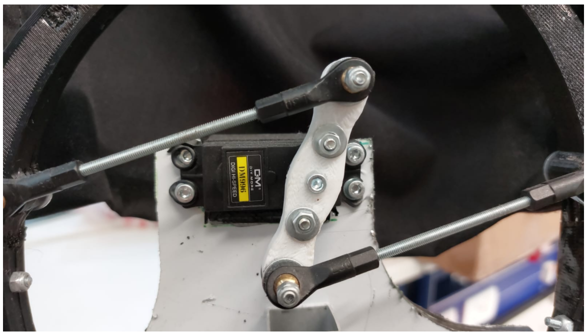


Figure 3.4: Open-close mechanism

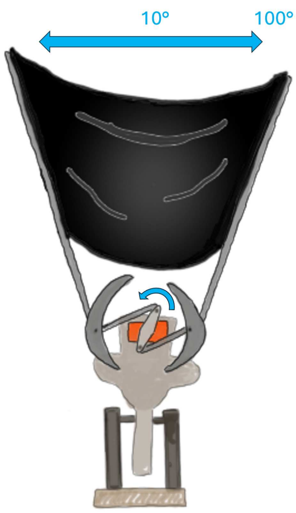
### 3.2.2. Wing: tilt motion

The metal frame of the wing is mounted on an Alucobond (PVC enclosed in thin aluminum sheets) base supported by two vertical columns approximately 20 cm long. A DS3225WG servo motor (see Tab 3.3 for specifications) allows the entire wing to tilt forward and backward by approximately 40°.

Table 3.3: DS3225WG Specifications

| Parameter | DS3225WG |
|---|---|
| Power Supply | 4.8-6.8 V |
| Maximum Torque | 21 kg·cm |
| Maximum Speed | 0.16 s/60° |
| Range | 180° (configurable up to 270°) |
| PWM | 50 Hz |

Although the available angular range is limited by structural constraints and by the wing's length, even small tilting movements result in visually salient expressive gestures. The tilt servo operates within the range of [70°,110°], with a default idle position of 90°(see Figure 3.5b).




(a) Qualitative drawing of the open/close wing mechanism


(b) Qualitative drawing of the tilting mechanism for the wing

### 3.2.3. Head: rotation

The head is mounted on the same wooden base as the wing tilt mechanism. From this base, a 20 cm "neck" connects to a smaller platform that holds a 3001HD servo motor (specifications in Table 3.4). This motor rotates a square-shaped metal structure, on which another motor is mounted horizontally, to enable the head's rotation.

Table 3.4: 3001HD Specifications

| Parameter | 3001HD |
|---|---|
| Power Supply | 4.8-6 V |
| Maximum Torque | 3.2 kg·cm |
| Maximum Speed | 0.14 s/60° |
| Range | 180° |
| PWM | 50 Hz |

(a) (b)

Figure 3.5: Front and side view of the robot's head

The motor's operating range is [5°,175°], with a default idle position of 90°(see Figure 3.6a).

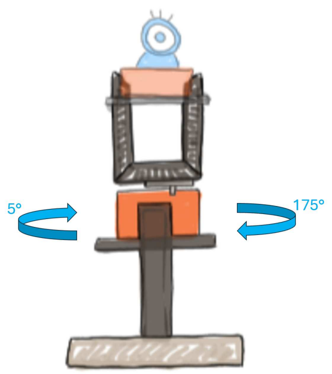
### 3.2.4. Head: tilt motion

On top of the square metal structure, a MG996R servo motor (for specifications see Tab 3.5) is mounted horizontally. This servo rotates itself relative to the metal structure, tilting the head forward and backward. A USB camera is glued to the side of the servo facing upwards.

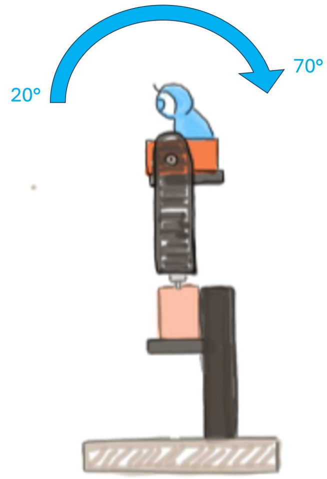
The servo operates within the range of [20°,70°], with a default idle position of 40°(see Figure 3.6b).

Table 3.5: MG996R Specifications

| Parameter | MG996R |
|---|---|
| Power Supply | 4.8-7.2 V |
| Maximum Torque | 11 kg·cm |
| Maximum Speed | 0.17 s/60° |
| Range | 180° |
| PWM | 50 Hz |




(a) Qualitative drawing of the head rotation mechanism




(b) Qualitative drawing of the head tilting mechanism

## 3.3. Bottom: Wheels

The robot is mounted on an omnidirectional wheel base consisting of 3 wheels, each with a diameter of 58 mm. The wheels are made of nylon and rubber, to enhance durability, and allow smooth movement in any direction. The motors driving the wheels are controlled via an Arduino Mega 2650 board (specifications in tab 3.6 and figure 3.6), with power distributed evenly among the three wheels. Maximum speed is approximately 1.4 m/s, and encoder feedback allows precise motion control.

Table 3.6: Arduino Mega 2560 Key Specifications [28]

| Feature | Specification |
|---|---|
| Microcontroller | ATmega2560, 8-bit AVR, 16 MHz |
| Flash Memory | 256 KB (8 KB used by bootloader) |
| SRAM | 8 KB |
| EEPROM | 4 KB |
| Digital I/O Pins | 54 (15 PWM capable) |
| Analog Input Pins | 16 (10-bit ADC) |
| SPI / I$^2$C | SPI, I$^2$C (TWI) |
| Operating Voltage | 5 V |
| Input Voltage (recommended) | 7-12 V |
| Dimensions | 101.52 mm × 53.3 mm |

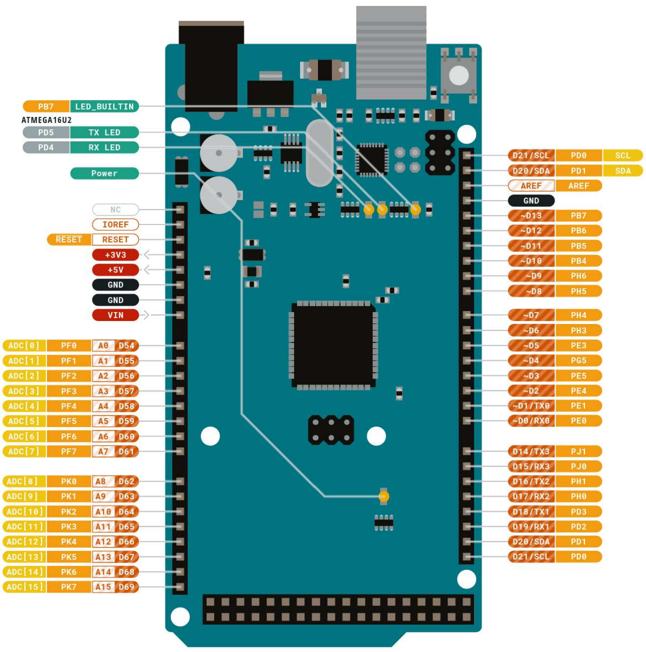



Figure 3.6: Arduino Mega 2560 Pin Layout (Top View) [28]

## 3.4. GPU: the brain

The main algorithms and computational tasks are executed on an NVIDIA Jetson Nano embedded computing board, equipped with 4 GB of 64-bit LPDDR4 RAM, a 128-core Maxwell GPU, and a quad-core ARM Cortex-A57 CPU running at up to 1.43 GHz. The board provides sufficient computational power to support onboard, real-time execution of vision-based processing, interaction recognition, and high-level control algorithms. Camera, microphone, speaker and the boards are interfaced with the Jetson Nano via USB connections.

Table 3.7: Elements connected to the Jetson Nano

| Device | Type of Connection |
|---|---|
| Arduino | USB |
| ESP32 | USB |
| ESP32 for sonars | USB |
| Camera | USB |
| Microphone | USB, through an Aux-to-USB converter |
| Speaker | USB |
| Power Supply | PW20V, through a driver that lowers voltage to 5 V |

In addition, a TP-Link USB wireless network adapter is connected to the Jetson Nano to enable communication with an external PC. This connection is used both during development, for debugging and monitoring purposes, and in the final system architecture as part of the algorithmic pipeline, allowing data exchange and coordination between the robot and the external computing environment.

Accelerometer and sonars are not directly connected to the Jetson Nano, as they interface with the ESP32s instead.

## 3.5. Sensors

The following section details all sensors employed by BlackWing to infer the actor's intentions and intended interaction.

### 3.5.1. Camera

A USB webcam is mounted on the head of the robot and acts as its primary visual sensor. The camera provides a native resolution of 1080p at 15 fps and is connected directly to the Jetson Nano.

For the purposes of this project, the effective resolution was intentionally reduced to decrease computational load and enable real-time execution of the recognition algorithms.

### 3.5.2. Accelerometer

The robot is equipped with an *ADXL345* accelerometer to measure linear acceleration along three axes. The sensor supports selectable measurement ranges of $\pm 2g$, $\pm 4g$, $\pm 8g$, and $\pm 16g$, with a maximum sampling frequency of 3200 Hz. It is connected to the main ESP32.

In this project, a sampling frequency of 1000 Hz was used, as the accelerometer data were primarily employed for threshold-based event detection rather than continuous signal analysis.

### 3.5.3. Sonars

BlackWing is equipped with four *HC-SR04* ultrasonic sensors, each featuring a beam angle of approximately 15° and a maximum measurable distance of up to 3 m, with an accuracy of ±3 mm. The sensors are managed by a dedicated ESP32.

In this project, the operational range was limited to 70 cm. This configuration was sufficient to detect intrusions into the robot's peripersonal space, without the need for precise distance measurements at longer ranges. This is necessary because at that distance the position of a person cannot be reliably obtained with the camera only.

### 3.5.4. Microphone

An omnidirectional microphone captures audio signals for interaction and environmental monitoring. It has a sensitivity of $-37 \pm 3$ dB and an effective pickup range of approximately 3 m.

## 3.6. Speaker

BlackWing is equipped with a speaker that allows it to accompany physical movements with sound output, enhancing expressiveness and disambiguating emotional behaviors. The speaker is connected directly to the Jetson Nano, which handles playback of audio files in `.mp3` format.

## 3.7. Power Supply

The robot is powered by two lithium-ion batteries, each with a nominal voltage of 11.1 V and a capacity of 5200 mAh. One battery exclusively powers the top section of the robot, supplying the ESP32 and the head and wing servos through a voltage regulator that steps the voltage down to 5 V. The second battery supplies the Jetson Nano, the Arduino Mega, and the wheel motors through two identical motor drivers, which distribute power to the three wheels.

Under intensive operation, the robot's autonomy is approximately 1-2 hours.

A simplified electrical schematic of the power distribution architecture is provided in figure 3.7.

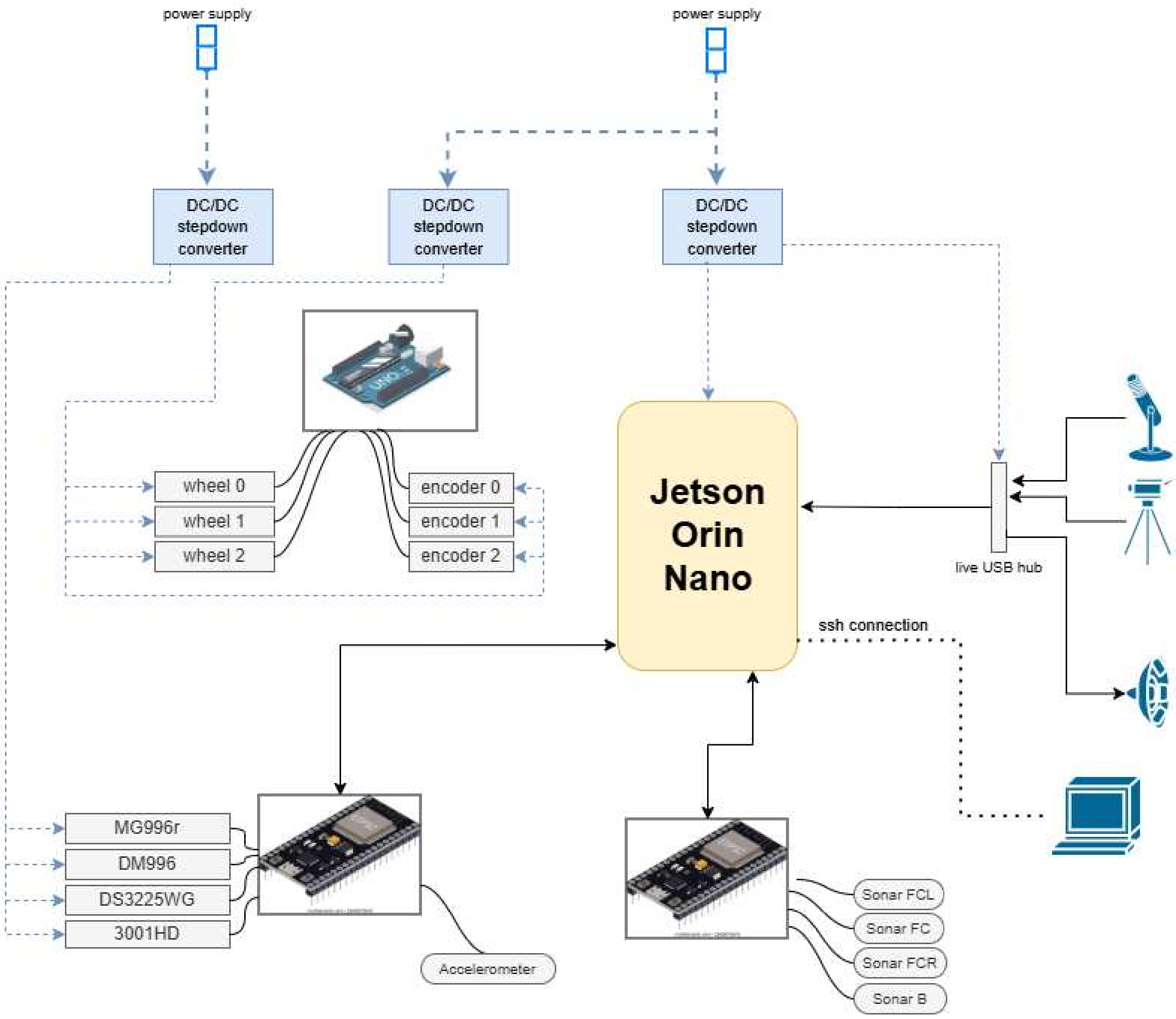



Figure 3.7: Qualitative drawing of robot architecture

# 4 | Software Architecture

## 4.1. General Software Architecture

In this project, high-level computation and decision-making are handled by a Jetson Nano. Its primary role is to collect data from all sensors, process the incoming inputs in real time, and update the internal emotional state of the robot. Each input contributes to the variation of one or more emotional variables. The system continuously checks whether any emotional threshold has been reached. When this occurs, the Jetson Nano sends a command to the main ESP32, which controls the motors responsible for expressive movements, and to the Arduino board that manages wheel motion, in order to trigger the corresponding emotional behavior.

Communication between the different control units is implemented through serial interfaces. The Arduino board, which controls the lower part of the robot (wheel motion), communicates with the main ESP32 via USB serial connection. This interface is used to transmit low-level motion parameters such as wheel speed and movement duration.

The ESP32 is responsible for controlling the upper part of the robot, including the actuators used for expressive movements. High-level behavioral commands are received from the Jetson Nano through serial communication, and translated into predefined motor control sequences.

The overall control architecture follows a centralized structure, where the Jetson Nano performs high-level perception, emotional state evaluation, and decision-making, while the microcontrollers execute low-level control tasks.

The system runs on Ubuntu Linux on the Jetson Nano.

## 4.2. Input acquisition and elaboration

Blackwing integrates multiple sensors to detect different types of human interaction and environmental stimuli. Each sensor requires a dedicated acquisition and preprocessing procedure, and the outputs of these modules are aggregated by a centralized input manager running on the Jetson Nano.

Each sensor module operates as an independent thread and writes a keyword representing the detected event into a dedicated text file. In addition to these sensor-driven inputs, the system also supports manual commands sent through the terminal, allowing debugging, testing, and temporary overrides of the robot's behavior. Terminal commands are handled by the input manager in the same way as sensor events, ensuring seamless integration within the overall event-handling pipeline.

Communication between sensor modules, the terminal, and the input manager is implemented through file-based inter-process communication. Although not typically adopted in high-frequency robotic control systems, this approach was chosen for its simplicity, modularity, and ease of debugging. Each sensing module can therefore operate independently without requiring complex synchronization mechanisms, which simplifies system integration and maintenance.

The input manager polls the sensor files at 10 Hz (every 0.1 s), resulting in a maximum communication latency of 100 ms. This delay is acceptable for human–robot interaction scenarios, where stimuli occur at relatively low frequency and strict real-time constraints are not required.

At each polling cycle, the manager reads the content of the sensor files and, if a non-empty keyword is detected, identifies the corresponding source module and invokes the registered callback function with the detected event. Most inputs are treated as instantaneous events: once processed, the associated file is cleared to prevent repeated triggering. However, some interactions, such as *like/dislike*, represent persistent states rather than discrete events. These are handled with additional checks to ensure that the keyword remains active only for as long as the corresponding condition is truly detected.

### 4.2.1. Input from Camera: Like and Dislike Detection

The camera mounted on the robot's head allows Blackwing to capture images of the surrounding environment. The acquired frames are processed on the Jetson Nano using the YOLO object detection framework, which identifies people and selected objects by generating bounding boxes around them. In order to meet the computational constraints of the platform, the lightweight *YOLOv4-tiny* model was adopted.

Although YOLO is primarily designed for object detection, the generated bounding boxes were used to isolate regions of interest for color analysis. For each detected region, the image is converted from the BGR color space to HSV using the OpenCV function `cv2.cvtColor`. The HSV representation was selected because it provides greater robustness for color segmentation under varying lighting conditions.

The robot was designed to associate a positive reaction with the color orange and a negative reaction with the color blue. Color detection is performed through threshold-based segmentation using empirically determined HSV ranges:

- **Orange**: lower = [5, 100, 100], upper = [20, 255, 255]
- **Blue**: lower = [100, 150, 50], upper = [140, 255, 255]

For each processed frame, the percentage of pixels within the bounding box that fall inside the specified color range is computed. If more than 5% of the pixels correspond to orange or blue, the system writes the keyword *"like"* or *"dislike"* to a dedicated file for the input manager (see example in Figure 4.1 ).


(a) A person dressed in "blue", as seen from the robot point of view.

(b) A person dressed in "orange", as seen from the robot point of view.

Figure 4.1: Camera frames from BlackWing as seen after Yolo elaboration.

Since visual stimuli may remain in the field of view for extended periods, a cooldown interval of 30 seconds was introduced to prevent repeated triggering while the same object remains visible.

In addition to color-based evaluation, the camera module also detects the presence of a person in the scene by monitoring the *person* class provided by the YOLO model. When a person is detected, a corresponding flag is written to a separate file. This information is used by the sonar processing module to determine whether a detected approach is likely to be caused by a human.

The HSV thresholds were determined empirically through experimental tuning under typical indoor lighting conditions. Despite the improved robustness of the HSV color space, the method remains sensitive to significant illumination changes and to objects containing similar color distributions.

### 4.2.2. Input from Sonars: closeness detection

The ultrasonic sensors are used to detect the presence of nearby obstacles and to identify approaching movements toward the robot. These sensors operate in cooperation with the camera module in order to distinguish human approach from static environmental objects.

The sensors are managed by a dedicated ESP32 board. The implemented firmware supports up to seven ultrasonic sensors; however, only four were used in this project: three positioned at the front of the robot (front-left, front-center, and front-right) and one positioned at the rear for safety purposes.

The ESP32 continuously measures the distances detected by each sensor and transmits the data to the Jetson Nano via serial communication. The rear sensor is used exclusively for collision prevention: if an obstacle is detected at a distance lower than 50 cm, the information is used to inhibit backward motion and avoid potential collisions.

The front sensors are used to detect approaching objects. The algorithm monitors the variation of the measured distances and identifies a potential approach when a decreasing trend is observed. When this condition occurs, the system checks whether the camera module has detected the presence of a person in the scene.

More specifically, the firmware maintains a short history of the measurements from the frontal sensors and evaluates whether at least one of them is detecting a consistent distance reduction. If an approaching movement is detected, the system verifies the content of the file containing the latest camera detections. If the file reports the presence of a person, the keyword *"closeness"* is written to a dedicated file for the input manager and maintained for a short time to ensure reliable detection. The keyword is removed as soon as either the approaching condition or the person detection is no longer satisfied.

In parallel, the rear sensor is continuously monitored to detect potential collision risks and to inhibit backward motion when necessary.

### 4.2.3. Input from Microphone: noise detection

The audio signal is acquired and processed directly on the Jetson Nano. The microphone operates at a sampling rate of 48000 kHz, allowing the system to capture sudden acoustic events with sufficient temporal resolution.

The system continuously computes the root mean square (RMS) value of the audio signal over a fixed time window and compares it to a predefined threshold. The threshold was empirically determined in order to distinguish normal environmental noise and speech from sudden loud events such as hand claps, footsteps, or doors closing.

When the RMS value exceeds a threshold of 1000, the keyword *"noise"* is written to a dedicated file for the input manager. Since the detection is intended for short and transient acoustic events, no cooldown or debounce mechanism was implemented. The event is naturally self-limiting, as the RMS value rapidly returns below the threshold once the impulse noise has ended.

No frequency-domain filtering or spectral analysis is applied. The detection relies solely on the amplitude of the time-domain signal, which was considered sufficient for identifying abrupt high-intensity sounds in the intended human-robot interaction environment.

### 4.2.4. Input from Accelerometer: caress and hit detection

The accelerometer is managed by the main ESP32, which performs onboard processing to detect physical interactions with the robot. Sensor acquisition and event classification are executed within a dedicated task running on the microcontroller.

The sampling frequency is determined by the task loop timing. A delay of 5 ms is introduced at the end of each iteration using `vTaskDelay(pdMS_TO_TICKS(5))`, resulting in a nominal sampling rate of approximately 200 Hz.

Interaction detection is based on the variation of the measured signal between consecutive samples. Let $s$ be the current sensor reading and $s_{\text{last}}$ the previous one. The variation is computed as:

$$\Delta s = s - s_{\text{last}}$$

and the absolute value $|\Delta s|$ is used as an indicator of motion intensity.

Two thresholds are defined to classify the interaction:

- **Caress**: $|\Delta s| > 8$
- **Hit**: $|\Delta s| > 20$

Low-intensity and smooth variations, typically generated by gentle physical contact, are interpreted as a *caress*. Conversely, high-intensity and abrupt variations caused by shaking or striking the robot are classified as a *hit*.

To prevent multiple detections of the same physical interaction, a cooldown mechanism is implemented. After a *caress* event, the acquisition task is suspended for 1 second. After a *hit*, a longer cooldown of 3 seconds is applied. This blocking delay reduces repeated triggering caused by residual vibrations or sustained contact.

When an interaction is detected, the ESP32 sends the corresponding label (*"caress"* or *"hit"*) to the Jetson Nano via serial communication. A dedicated process on the Jetson receives the message and writes the detected event to a specific text file, making it available to the input manager for further processing.

### 4.2.5. Input Manager

The input manager, implemented in `input_manager.c`, serves as a central hub for collecting events from all sources. It runs as a dedicated thread, continuously monitoring both the sensor output files and the terminal input, and forwards any detected events to a user-provided callback function.

Four main input channels are defined: *audio*, *sonar/closeness*, *vision/like-dislike*, and *touch*. Each channel corresponds to a dedicated text file updated by the respective sensor module. The input manager reads these files periodically and forwards non-empty signals to the callback, effectively decoupling input acquisition from the main control logic.

Special handling is applied to the *like/dislike* channel to prevent spurious repetitions: events are only forwarded if the content differs from the last sent value or if a minimum time interval has elapsed. A short delay is also introduced at the end of each polling iteration to reduce CPU usage and avoid busy-waiting.

The module exposes a simple API: `input_initialize` spawns the aggregation thread and provides the callback interface. This design allows the rest of the system to react asynchronously to events without concerning itself with the details of how each input is collected or processed.

## 4.3. Emotion values updates and threshold checks

In the main program (`main.c`), each event received through the callback function provided to `input_initialize` triggers an update of the internal emotional state. The system follows an *event-driven* architecture: emotional variables are updated only when a new interaction keyword is received, without any periodic scanning or time-based polling.

Upon reception of an event, the program updates the current levels of all emotions according to the parameters described in Table 2.1 in Chapter 2. After the update, the function `threshold_check()` evaluates whether any emotional value has exceeded one of the three predefined intensity thresholds associated with each emotion.

If one or more thresholds are crossed, the emotion with the highest intensity level is selected for expression. In the unlikely case of equal intensity, only one is expressed randomly. Since threshold evaluation is performed immediately after each event, emotional reactions are generated as soon as an interaction causes a threshold to be exceeded.

Triggering an emotion corresponds to calling a function with the same name, which contains the sequence of actions and commands required to execute the corresponding expressive behavior.

As introduced in the hardware architecture (Chapter 3), the expression of an emotion involves multiple hardware components, each responsible for a different aspect of the robot's behavior:

- the main **ESP32 board** controls head and wing movements
- the **Arduino board** controls the wheelbase motion
- the **Jetson Nano** manages the playback of the MP3 audio tracks

In the absence of new events, emotional values remain unchanged.

### 4.3.1. Emotion expressions through ESP32

On the main program running on the Jetson Nano, each emotional expression function invokes `send_command(const char *msg)`, which transmits the name of the triggered emotion to the main ESP32 controlling head an wing through a serial interface.

Upon receiving the command string from the Jetson Nano, the ESP32 maps the emotion label to the corresponding motion routine responsible for controlling the head and wing actuators.

Each routine defines motion parameters for the involved servomotors, including:

- target speed
- acceleration and deceleration
- minimum speed
- execution state (true or false)

Based on these parameters, the motors execute predefined angle trajectories stored in local header files. These trajectories are implemented as sequences of position targets and executed using the function:

`run_servo_sequence(...)`.

This architecture separates high-level behavioral decisions (Jetson Nano) from low-level motion execution (ESP32), ensuring modularity and real-time responsiveness.

### 4.3.2. Emotions expression through Arduino

Locomotion commands are generated by the Jetson Nano and transmitted to the Arduino board controlling the ground wheels through serial communication using a JSON-formatted message. The message contains the desired platform velocities expressed in the robot's local reference frame:

- forward velocity $V_x$,
- lateral (strafe) velocity $V_y$,
- angular velocity $\omega$.

The robot reference frame is defined with the $x$-axis pointing forward and the $y$-axis pointing to the left. All velocity values are expressed in local motion units, defined with respect to this coordinate system and internally mapped by the Arduino to motor control signals.

An example of the transmitted command is:

```
{
  "v_x": 30,
  "v_y": 0,
  "w": -10
}
```

Upon reception, the Arduino computes the angular velocity required for each wheel using the inverse kinematic model of the omnidirectional platform. The robot is equipped with three omni-wheels arranged in a symmetric 120° configuration around the chassis.

The relationship between platform velocities and wheel speeds is given by:

$$\begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{R} & \frac{L}{R} \\ -\frac{\sqrt{3}}{2R} & \frac{1}{2R} & \frac{L}{R} \\ \frac{\sqrt{3}}{2R} & \frac{1}{2R} & \frac{L}{R} \end{bmatrix} \begin{bmatrix} V_x \\ V_y \\ \omega \end{bmatrix}$$

where:

$W_i$ is the speed command for wheel $i$,

$R$ is the wheel radius,

$L$ is the distance between the robot center and each wheel.

The computed wheel speeds are then converted into motor power values and applied to the motor drivers. Speed saturation and safety limits are enforced at the Arduino level to prevent excessive commands and ensure stable and safe operation.

### 4.3.3. Emotion expression through sound

Audio feedback is generated directly by the Jetson Nano in order to complement the visual and motion-based emotional expressions. When an emotion is triggered, the corresponding expression function calls `play_track(...)` to reproduce the audio file associated with that emotional state.

Each emotion is mapped to a predefined sound stored locally on the Jetson Nano. The audio files are organized in a dedicated directory and selected by passing the corresponding filename to the playback function.

The use of local audio playback avoids additional communication latency and simplifies the integration of multimodal emotional feedback.

The full code is linked in Appendix A.

## 4.4. Emotion Expression

This section presents the final, integrated expressions of Blackwing's emotional states, combining the motor, wheel, wing, and audio commands described in the previous sections. Each expression is designed to convey a distinct emotional state in a manner that is both realistic and easily interpretable by a human observer.

The design of the robot's movements draws inspiration from animation principles as well as Laban's movement analysis, together with the observation of the animal kingdom.

These frameworks guided the speed, intensity, and coordination of the head, wings, and wheelbase to create motions that are expressive and natural.

The following subsections describe individual emotional expressions, detailing the coordinated movements of each robot component and the associated sounds. Tables are provided to summarize the specific actions for the head, wings, wheelbase, and audio output, providing a clear reference for each emotion.

### Annoyance

Annoyance is expressed through subtle, deliberate movements aimed at conveying mild irritation. Blackwing slowly rotates its head upwards and in a circular motion, simulating an eye-roll gesture, while the fully open wing follows the same tilting movement, accentuating the head's motion. The auditory cue is a soft sigh, derived from a turtle sound (Tab 4.1).

Table 4.1: Annoyance expression. The table summarizes coordinated movements and sounds for each robot component.

| **Robot Part** | **Movement** |
|---|---|
| Wing | Moves slowly from front to back and then returns to the front |
| Head | Rotates slowly from left to right and tilts backward |
| Wheelbase | Static |
| Sound | Sigh |

### Anger

Anger is conveyed through rapid, forceful wing movements and pronounced head oscillations. The wing spreads widely, evoking an aggressive posture, while a high-pitched screech, inspired by dolphin sounds, serves as the corresponding auditory expression (Tab 4.2).

Table 4.2: Anger expression. Summary of movements and sounds for each robot component.

| **Robot Part** | **Movement** |
|---|---|
| Wing | Leans forward, opening and closing very quickly |
| Head | Moves rapidly from left to right, with shaking motion |
| Wheelbase | Moves quickly from left to right, giving the impression of shaking |
| Sound | High pitched, screeching sound |

## Rage

Rage builds upon anger with added intensity, including abrupt forward motion to simulate loss of control and heightened aggression. The auditory component consists of forceful, aggressive vocalizations inspired by llama sounds (Tab 4.3).

Table 4.3: Rage expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Moves backward and forward, opening and closing aggressively |
| Head | Shakes both up and down and side to side |
| Wheelbase | Moves rapidly from left to right with shaking, including a forward leap |
| Sound | Aggressive high pitch sounds |

## Interest

Interest is expressed by orienting the wings upward in a closed position, analogous to attentive animal ears, while the head raises slightly to direct the gaze toward stimuli of interest (Tab 4.4).

Table 4.4: Interest expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Fully closed and pointing upwards |
| Head | Slightly raised |
| Wheelbase | Static |
| Sound | Sound similar to “uh uh” |

## Anticipation

Anticipation manifests as a forward-leaning wing posture, combined with gradual forward motion and subtle nodding of the head, signaling readiness and expectation of forthcoming events (Tab 4.5).

Table 4.5: Anticipation expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Fully closed, pointing upwards, and slightly forward |
| Head | Initially looks upwards, then nods |
| Wheelbase | Moves slowly forward |
| Sound | Sound similar to “uh uh” |

## Vigilance

Vigilance is characterized by alert wing posture with minor oscillations, rapid head movements scanning the environment, and slow backward wheel movement, simulating careful monitoring for potential threats (Tab 4.6).

Table 4.6: Vigilance expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Nearly fully closed, with slight shaking |
| Head | Rapidly looks left and right, as well as up and down |
| Wheelbase | Moves slowly backwards |
| Sound | None |

## Boredom

Boredom is expressed through slow backward tilting of the head and wings, evoking disengagement, accompanied by a yawning sound to indicate reduced interest (Tab 4.7).

Table 4.7: Boredom expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Slowly tilts backward, almost fully open |
| Head | Slowly tilts backward |
| Wheelbase | Static |
| Sound | Produces a yawn |

## Disgust

Disgust is communicated by retreating movements, head and wing rotations to avoid visual contact with the offending stimulus, and auditory cues inspired by lynx vocalizations, reflecting aversion and withdrawal (Tab 4.8).

Table 4.8: Disgust expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Fully closed and leaning backward |
| Head | Looks to the left, away from the actor |
| Wheelbase | Moves backward while rotating 90° to the right |
| Sound | Disgusted sounds |

## Loathing

Loathing combines rapid backward motion, side-to-side head oscillations, and repeated wing flapping to signal strong aversion. The corresponding sound intensifies the sense of disgust (Tab 4.9).

Table 4.9: Loathing expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Opens and closes very quickly, leaning backward |
| Head | Moves rapidly from side to side |
| Wheelbase | Moves quickly backward |
| Sound | Repeated disgusted sounds |

## Apprehension

Apprehension is signaled by partially closed wings and rapid scanning head movements, accompanied by slow backward motion, reflecting cautious assessment of potential threats (Tab 4.10).

Table 4.10: Apprehension expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Nearly fully closed |
| Head | Moves quickly side to side, as well as up and down |
| Wheelbase | Moves very slowly backward |
| Sound | None |

## Fear

Fear is expressed through rapid retreat, rhythmic wing movement simulating labored breathing, and initial scream sounds followed by heavy breathing, conveying acute distress (Tab 4.11).

Table 4.11: Fear expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Rhythmic open-and-close motion; wings bent backward |
| Head | Looks slightly upward |
| Wheelbase | Moves quickly backward, then stops after a few seconds |
| Sound | Emits a scream, followed by heavy breathing |

## Terror

Terror involves freezing and protective postures, including shielding with wings, partial rotation of the body, and brief auditory screams, conveying overwhelming fright (Tab 4.12).

Table 4.12: Terror expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Slightly open, initially all backward, then moves all forward |
| Head | Raises slightly, then lowers to hide itself |
| Wheelbase | Rotates quickly almost fully on itself |
| Sound | Short, high pitched scream |

## Serenity

Serenity is represented by slow oscillatory wing movements, subtle wheel adjustments, and gentle auditory cues, reflecting calm contentment (Tab 4.13).

Table 4.13: Serenity expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Slightly open, slowly tilts forward and backward |
| Head | Looks straight ahead |
| Wheelbase | Moves slowly forward and backward |
| Sound | Produces a soft, happy mumbling noise |

## Joy

Joy is conveyed through rapid wing oscillations, slight head elevation, and small rotational wheel movements, embodying playful excitement (Tab 4.14).

Table 4.14: Joy expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Opens and closes quickly |
| Head | Slightly raised |
| Wheelbase | Performs small 45° rotations in both directions |
| Sound | Laughter |

## Ecstasy

Ecstasy is communicated via rapid full-body rotation, energetic wing movements, and expressive laughter, portraying exuberant happiness (Tab 4.15).

Table 4.15: Ecstasy expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Opens and closes quickly |
| Head | Slightly raised |
| Wheelbase | Performs a full rotation at relatively high speed |
| Sound | Laughter |

## Pensiveness

Pensiveness is expressed by slow retreat, downward gaze, and partially closed wings, emphasizing contemplative withdrawal (Tab 4.16).

Table 4.16: Pensiveness expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Slightly closed, fully bending forward |
| Head | Looks down slowly |
| Wheelbase | Moves slowly backward |
| Sound | None |

## Sadness

Sadness involves slow backward movement, downward and sideways head motion with small shakes, and forward-leaning wings, conveying dejection (Tab 4.17)

Table 4.17: Sadness expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Slightly closed, fully bending forward |
| Head | Slowly looks down and to the right, followed by small shakes |
| Wheelbase | Moves backward and slightly to the left |
| Sound | None |

## Grief

Grief is expressed through small periodic wing movements, downward head motion with intermittent lifts, and turning followed by straight movement, accompanied by sobbing sounds, signaling intense sorrow (Tab 4.18).

Table 4.18: Grief expression. Summary of movements and sounds for each robot component.

| **Robot Part** | **Movement** |
|---|---|
| Wing | Small periodic movements, bending forward |
| Head | Looks down with small periodic upward movements |
| Wheelbase | Turns around, then moves straight |
| Sound | Produces sobbing sounds |

## Distraction

Distraction is conveyed by closing the wings and rapid lateral head scanning, while the robot remains stationary, indicating attentional shift (Tab 4.19).

Table 4.19: Distraction expression. Summary of movements and sounds for each robot component.

| **Robot Part** | **Movement** |
|---|---|
| Wing | Closed |
| Head | Looks left, then right |
| Wheelbase | Static |
| Sound | None |

## Surprise

Surprise is expressed through rapid head movements in multiple directions, sudden backward wheel motion, and closed wings, reflecting unexpected stimuli (Tab 4.20).

Table 4.20: Surprise expression. Summary of movements and sounds for each robot component.

| **Robot Part** | **Movement** |
|---|---|
| Wing | Closed |
| Head | Quickly looks up and down, as well as left and right |
| Wheelbase | Performs a short, sudden movement backward |
| Sound | None |

## Amazement

Amazement combines slow backward movement, upward head tilt, and fully extended wings with an accompanying “wow” sound, conveying awe and fascination (Tab 4.21).

Table 4.21: Amazement expression. Summary of movements and sounds for each robot component.

| Robot Part | Movement |
|---|---|
| Wing | Fully open, slowly bends backward |
| Head | Slowly tilts upward |
| Wheelbase | Moves slowly backward |
| Sound | Produces a “wow” sound |

### Acceptance,Trust, Admiration

These emotions, although included in Plutchik’s model, are currently omitted due to the robot’s non-anthropomorphic design and the intrinsic difficulty of conveying and interpreting them, even for human observers, when gestures only are considered. Nevertheless, the control architecture is flexible, allowing for straightforward integration of these or other complex emotions in future implementations.

# 5 | Survey-Based Evaluation of Emotional Perception

An online survey was conducted to evaluate the effectiveness of the robot's emotional expression system. The primary objective was to assess whether the emotions displayed by the robot could be correctly recognized by participants. Accurate perception of emotional states is a fundamental requirement for natural and effective human–robot interaction.

A secondary objective of the survey was to evaluate the perceived naturalness and realism of the robot's expressive behavior. In addition to emotion recognition accuracy, participants were asked to rate the quality of the interaction in terms of how natural and believable the robot's responses appeared.

A further objective of the study was to investigate whether variations in emotional intensity were correctly perceived. For selected emotions, participants were shown multiple video clips representing increasing levels of intensity, according to Plutchik's emotions wheel [30], and were asked to identify the perceived intensity level. This evaluation aimed to assess whether the proposed model was able to convey not only distinct emotional categories, but also meaningful gradations within the same emotional state.

The survey focused exclusively on the observation of the robot's expressive outputs. The processes related to user input interpretation and internal response generation were not included in this evaluation, as their assessment would require controlled in-person interaction rather than an online study.

## 5.1. Survey Design

Participants accessed the study through an online link. Upon access, each participant was randomly assigned to one of seven short survey variants. Each variant focused on a different target emotion. Participants were allowed to repeat the survey if they wished to observe additional emotional expressions. Multiple submissions from the same participant were treated as independent observations.

The survey consisted of two main sections.

In the first section, participants were shown a short video clip depicting a brief interaction between a human and the robot. The robot expressed an emotion at intensity level 2, corresponding to one of the seven primary emotions represented in Plutchik's wheel. Participants were asked to select which emotion was being expressed from a list of the seven main categories.

In addition to emotion recognition, participants evaluated the interaction using a Likert-scale questions (1–5), rating the coherence of the robot's behavior with respect to the human's actions.

In the second section, participants were presented with three video clips representing different intensity levels (level 1, level 2 and level 3) of a single emotion, different from the one shown in the first section. For each clip, participants were asked to identify the perceived intensity level of the expressed emotion. The survey was conducted online and was open for a period of five days and the entire survey required only a few minutes to complete (Full questionnaire is linked in Appendix B).

A total of 221 responses were collected through the survey. Since participants were randomly assigned to one of the survey variants and were allowed to complete multiple surveys voluntarily, the number of responses obtained for each emotion was not uniform. Table 5.1 summarizes the demographic characteristics of the sample.

Table 5.1: Demographic characteristics of the collected responses (N = 221)

| Variable | N | (%) |
|---|---|---|
| **Age** | | |
| 18–24 | 68 | 30.77% |
| 25–34 | 80 | 36.20% |
| 35–44 | 6 | 2.71% |
| 45–54 | 1 | 0.45% |
| 55–64 | 54 | 24.43% |
| 65+ | 12 | 5.3% |
| **Gender** | | |
| Male | 121 | 54.75% |
| Female | 99 | 44.8% |
| Other / Prefer not to say | 1 | 0.45% |
| **Robot familiarity** | Mean = 2.43 ± 1.24 (1–5 scale) | |

The participant sample was predominantly young to middle-aged adults, with the largest proportions in the 18–24 (31%) and 25–34 (36%) age groups. The gender distribution was slightly skewed toward males (55%), while a small number of responses were reported as female (45%) or other/prefer not to say (0.5%).

Participants' familiarity with robots, measured on a 1–5 scale, was moderate overall (Mean = 2.43, SD = 1.24), indicating that the sample included both novices and more experienced users.

It is important to note that, because participants could complete multiple survey variants voluntarily, these statistics refer to responses rather than unique individuals. The sample provides a broad representation of the target population for evaluating the robot's emotional expressions, though some age and familiarity categories are less represented.

## 5.2. Results

The following results were obtained at the end of the survey. Due to the limited sample size, results are primarily reported using descriptive statistics.

### Accuracy Table

In Table 5.2 the results of the first question are showed.

Table 5.2: Accuracy table per emotion

| **Emotion Clip** | **Accuracy(%)** |
|---|---|
| Anger | 0,71 |
| Anticipation | 0,34 |
| Disgust | 0,26 |
| Fear | 0,67 |
| Joy | 0,46 |
| Sadness | 0,17 |
| Surprise | 0,31 |
| **Average** | **0,42** |

Overall recognition accuracy ranged from 17% to 71%, with an average of 42%. The highest recognition rates were observed for Anger and Fear, while Sadness showed the lowest accuracy.

## Confusion Matrix

The distribution of participant responses is shown in the confusion matrix in Table 5.3. A confusion matrix shows the relationship between the categories presented (rows) and the categories selected by participants (columns). Values on the main diagonal represent correct classifications, while off-diagonal values indicate misclassifications. Row totals do not sum exactly to 100% because participants were also allowed to select a "None of the above" option, which is not included in the matrix.

Table 5.3: Confusion matrix per emotion

| Shown/Perceived | **Anger** | **Anticipation** | **Disgust** | **Fear** | **Joy** | **Sadness** | **Surprise** |
|---|---|---|---|---|---|---|---|
| **Anger** | 0,71 | 0 | 0 | 0,26 | 0 | 0 | 0 |
| **Anticipation** | 0 | 0,34 | 0 | 0 | 0,03 | 0 | 0,58 |
| **Disgust** | 0,17 | 0,08 | 0,26 | 0,14 | 0 | 0 | 0,31 |
| **Fear** | 0,16 | 0 | 0 | 0,67 | 0,09 | 0 | 0,03 |
| **Joy** | 0,11 | 0 | 0,14 | 0,11 | 0,46 | 0,08 | 0,11 |
| **Sadness** | 0 | 0,04 | 0,27 | 0,50 | 0 | 0,17 | 0 |
| **Surprise** | 0 | 0 | 0 | 0,62 | 0 | 0 | 0,31 |

The confusion matrix highlights several patterns in participants' emotion recognition. Overall, Anger and Fear were recognized with relatively high accuracy (0.71 and 0.67, respectively), indicating that these expressions were comparatively distinctive. Surprise also showed moderate recognition (0.31), although it was frequently misclassified as Fear (0.62), suggesting a strong perceptual overlap between these two emotions.

Lower recognition rates were observed for Anticipation, Disgust, Joy, and Sadness. In particular, Anticipation was most often perceived as Surprise (0.58), while Disgust was frequently confused with Surprise (0.31) and Anger (0.17). Sadness showed very low correct identification (0.17) and was predominantly misclassified as Fear (0.50) and Disgust (0.27). Similarly, Joy was distributed across multiple categories, indicating a lack of a clearly distinctive perceptual pattern.

Overall, the results suggest that emotions associated with high arousal (e.g., Fear, Anger, Surprise) were more consistently recognized, whereas lower-arousal or more ambiguous emotions (e.g., Anticipation, Sadness, Disgust, Joy) showed substantial perceptual overlap.

## Perceived Coherence

Participants rated the interaction on a 5-point Likert scale in terms of perceived realism and coherence between the robot's behavior and the human's actions. The results are showed in tab 5.4.

Table 5.4: Mean and Standard Deviation table for action-reaction coherence, according to the participants of the survey

| Emotion Clip | Mean | Std. Dev. |
|---|---|---|
| Anger | 4,41 | 1,05 |
| Anticipation | 4,00 | 1,00 |
| Disgust | 3,34 | 1,33 |
| Fear | 3,9 | 1,27 |
| Joy | 4,11 | 0,99 |
| Sadness | 3,29 | 1,27 |
| Surprise | 3,72 | 1,09 |
| **Total** | **3,84** | **1,19** |

The average coherence score was 3.84/5. Overall, the responses indicate a moderately positive perception of interaction quality. Surprisingly, since the interaction was designed and acted to be coherent with the emotion intended to be expressed, these results seem to confirm that the robot's behavior was in most cases well designed for the situation, in apparent contrast with some of the low values obtained in the recognition test.

## Intensity Recognition

The second part of the survey evaluated whether participants were able to distinguish between different intensity levels of an emotional expression. Results are shown in tab 5.5.

Table 5.5: Accuracy table per intensity level

| Intensity Level | Accuracy (%) |
|---|---|
| Level 1 (low) | 0,58 |
| Level 2 (medium) | 0,51 |
| Level 3 (high) | 0,63 |
| **Total** | **0,57** |

The confusion matrix (Table 5.6) illustrates the relationship between the intensity level presented in the videos (rows) and the intensity level perceived by the participants (columns). The grey cells along the main diagonal represent the proportion of correct classifications for each intensity level. Again, row totals do not sum exactly to 100% because participants could also select a “None of the above” option, which is not included in this representation.

Table 5.6: Confusion matrix per intensity level

| Shown/Perceived | L1 | L2 | L3 |
|---|---|---|---|
| **L1** | 0,79 | 0,13 | 0,03 |
| **L2** | 0,13 | 0,47 | 0,35 |
| **L3** | 0,06 | 0,29 | 0,62 |

Overall, the results indicate that participants were generally able to distinguish between different levels of emotional intensity. However, most misclassifications occurred between the two highest intensity levels (L2 and L3). This pattern suggests that, although participants perceived the overall progression of emotional intensity, the boundaries between adjacent higher-intensity levels were not always clearly differentiated.

## 5.3. Possible reasons for ambiguity in emotion recognition

Several confusions observed in the results can be explained by similarities in movement dynamics, arousal level, temporal characteristics, and functional meaning of the emotional reactions. The following section has the purpose of assessing the most relevant cases.

### Anger-Fear

Anger and Fear were both recognized with relatively high accuracy(0.71 and 0.67 respectively), but some confusion between them was observed. One possible explanation is that both emotions were expressed through high-energy, rapid movements and strong body activation. In addition, both emotions represent coherent defensive reactions to a perceived threat or negative stimulus. While Anger reflects an approach-oriented defensive response and Fear an avoidance-oriented one, this distinction may be less evident when only movement intensity and speed are considered, leading to perceptual overlap.

### Disgust-Surprise

Disgust was frequently confused with Surprise. Both emotions can be triggered by sudden, unexpected stimuli and often involve immediate, reflex-like reactions. Although Disgust may also develop gradually, it is commonly elicited by the sudden appearance of an aversive object or situation. From a perceptual perspective, this can make it resemble Surprise, differing mainly in emotional valence. In this sense, Disgust may be interpreted as a negatively valenced form of Surprise.

### Fear-Surprise

A strong confusion was also observed between Surprise and Fear. Fear is often elicited by unexpected and potentially threatening events (e.g., a sudden stimulus or "jump scare"), which initially produce a startle-like response similar to Surprise. Because early behavioral reactions to unexpected events share similar temporal and kinematic characteristics, participants may have interpreted surprised reactions as fearful ones, especially when threat cues were ambiguous.

### Anticipation

Anticipation showed relatively low recognition accuracy and was most often perceived as Surprise. One reason might be the conceptual ambiguity of the term itself. In Plutchik's model [30], Anticipation includes both Interest (positive engagement) and Vigilance (cautious monitoring), which differ substantially in affective tone. This broad conceptual range may have made the intended emotion unclear. In the video, the actor revealed one of the robot's preferred object (an orange scarf), which could have been interpreted simply as a "pleasant unexpected event", leading participants to categorize the reaction as Surprise rather than Anticipation.

### Joy

Joy also showed moderate accuracy (0.46%) and was distributed across multiple perceived categories. As a high-arousal emotion, it involves energetic and dynamic movements, which may have contributed to occasional confusion with other high-activation emotions such as Anger or Fear. Additionally, the relatively abrupt onset of the reaction may have increased its similarity to Disgust and Surprise. A small number of responses classified Joy as Sadness; given the low frequency, these cases likely reflect isolated interpretation variability rather than a systematic perceptual overlap.

### Sadness

Sadness showed the lowest recognition accuracy and was most frequently confused with Fear and Disgust. Although Sadness is typically a low-arousal emotion characterized by reduced movement and slower dynamics, these emotions share a common functional component: withdrawal from a negative stimulus. In Fear and Disgust this withdrawal reflects avoidance of threat or aversion, while in Sadness it represents disengagement or loss of motivation. From an observer's perspective, however, these different internal states may appear similar in terms of reduced approach behavior or increased distance from the stimulus.

Another possible reason for the ambiguity is the low perceptual salience of Sadness. Unlike high-arousal emotions, which are conveyed through clear cues such as speed, intensity, and abrupt changes in movement, Sadness relies on more subtle indicators such as decreased activity and subdued posture. In the absence of strong distinctive cues, participants may have relied on the general negative valence of the reaction, leading to its interpretation as Fear or Disgust.

Overall, the results suggest that the behavioral expression of Sadness lacked sufficiently distinctive markers and overlapped perceptually with other avoidance-related negative emotions.

## 5.4. Chapter Summary

In summary, the survey results indicate that the robot's emotional expressions were generally interpretable by participants, although recognition accuracy varied substantially across emotions. Participants reported moderate-to-positive evaluations of interaction realism and behavioral coherence with the situation. The intensity analysis further suggests that differences in emotional strength were perceptible, even though some overlap between adjacent levels was observed.

Recognition performance was higher for high-arousal emotions such as Anger and Fear, which are characterized by dynamic and salient movement patterns. In contrast, lower-arousal or conceptually ambiguous emotions—including Anticipation, Disgust, Joy, and especially Sadness—showed lower accuracy and were frequently misclassified. These confusions often occurred between emotions sharing similar activation levels, temporal dynamics, or functional meanings, such as avoidance-related responses (e.g., Sadness–Fear, Sadness–Disgust) or reactions to unexpected events (e.g., Fear–Surprise, Anticipation–Surprise). The results suggest that when distinctive kinematic or contextual cues were limited, participants relied primarily on general arousal and valence, leading to perceptual overlap between categories.

The intensity analysis revealed that participants could reliably distinguish low from high activation levels, but confusions were common between adjacent high-intensity states, particularly between Level 2 and Level 3. This indicates that subtle gradations in emotional strength were more difficult to perceive than coarse differences in activation.

Overall, the findings suggest that the architecture successfully conveys broad affective states, but its expressivity remains limited for low-arousal emotions, conceptually ambiguous categories, and fine-grained intensity variations. These aspects represent key areas for future refinement in behavioral design and in the coordination of multi-modal expressive cues.

# 6 | Conclusions

## 6.1. Summary and main contributions of this work

This thesis addressed the design and implementation of an affective architecture for a non-anthropomorphic robot capable of interacting with a human actor in an improvisational context. The main objective was to enable autonomous, coherent, and expressive behavior based on internal emotional dynamics, moving beyond predefined and scripted interaction strategies.

To achieve this goal, an emotion-driven control architecture was developed and integrated on the *Blackwing* platform. The proposed system models eight emotions based on Plutchik's framework and combines three internal components: emotions, personality, and mood. Emotional states evolve over time as a function of multimodal sensory inputs, including vision, proximity, touch, and sound, as well as internal parameters that regulate thresholds, temporal persistence, and long-term affective bias. When emotional values exceed predefined thresholds, the dominant state is translated into coordinated movements and sound, allowing the robot to express its internal condition through its morphology.

The architecture integrates perception, emotional processing, and behavioral generation within a centralized control system and enables differentiated responses to similar external conditions depending on the internal state. This allows the robot to exhibit behavior that is adaptive over time and suitable for unscripted interaction scenarios.

The main contributions of this work can be summarized as follows. First, an affective control architecture integrating emotions, personality, and mood was designed to support coherent and state-dependent behavioral generation. Second, the proposed system was fully implemented and integrated on a real robotic platform, enabling real-time operation in an interactive scenario. Third, a set of expressive behaviors was developed to exploit the communicative potential of a non-anthropomorphic morphology. Finally, the perceptual effectiveness of the proposed approach was assessed through a user study, providing an initial experimental evaluation of the system.

## Possible future developments

Future work could focus on extending the emotional model through adaptive or learning-based mechanisms, increasing the richness and variability of the expressive behaviors, and evaluating the system in longer and more complex interaction scenarios. Additional studies involving a larger and more diverse participant pool would be necessary to further assess the effectiveness and robustness of the proposed architecture.

Another relevant aspect that would benefit from further investigation concerns the robot's localization and spatial awareness. If the robot were able to estimate its own position and that of the human actor within the environment, the relative distance and spatial relationship between them could become an additional expressive channel for interaction.

The current proximity detection system, based on sonar and camera information, provides limited performance and was mainly adopted due to the computational constraints of the onboard platform. The adoption of a more powerful embedded computer would enable the integration of AI-based perception methods, such as pose estimation algorithms (e.g., YOLO-based approaches), allowing not only the detection of a person but also the interpretation of body posture and movement. Such information could represent a highly informative input for the emotional and behavioral system.

Overall, this work demonstrates the feasibility of integrating emotional modeling, personality, and mood within a control architecture for non-anthropomorphic robots. The proposed approach contributes to the development of more natural and engaging human–robot interaction.

# Bibliography

[1] B. Academy. The world wide history of improvisational theatre - blackbox academy, 2024. URL https://blackboxacademy.ca/the-world-wide-history-of-improvisational-theatre/.

[2] A. P. Association. Personality, 2025. URL https://www.apa.org/topics/personality. Consulted on Dec 10th.

[3] A. P. Association. Apa dictionary of psychology, 2025. URL https://dictionary.apa.org/mood. Consulted on Dec 10th.

[4] A. P. Association. Apa dictionary of psychology, 2025. URL https://dictionary.apa.org/personality. Consulted on Dec 10th.

[5] C. Bell. *Essays on the Anatomy of Expression in Painting - Sir Charles Bell - Google Libri*. Longman, Hurst,Rees, and Orme, Paternoster Row, 1806. URL https://books.google.it/books?id=8_RZAAAAYAAJ&printsec=frontcover&dq=bell+Essay+on+the+Anatomy+of+Expression+in+Painting&hl=it&sa=X&ei=NBYdUZ3xFIjMswbRl4G4BA&ved=0CDgQ6AEwAQ#v=onepage&q&f=false.

[6] L. Braida. *Carlo Goldoni and the Construction of Authorship*, pages 139–179. Brill, 09 2022. ISBN 978-3-031-03897-6. doi: 10.1007/978-3-031-03898-3_5.

[7] S. J. Burton, A. A. Samadani, R. Gorbet, and D. Kulić. *Laban movement analysis and affective movement generation for robots and other near-living creatures*, volume 111, pages 25–48. Springer Verlag, 2016. doi: 10.1007/978-3-319-25739-6_2.

[8] O. Crick and S. Costola. The dramaturgy of commedia dell'arte. *The Dramaturgy of Commedia dell'Arte*, 12 2021. doi: 10.4324/9781003102373.

[9] K. Danziger. Thomas dixon. from passions to emotions: The creation of a secular psychological category. cambridge: Cambridge university press, 2003. 287 pp. £47.50 (cloth). isbn 0-521-82729-9. *Journal of the History of the Behavioral Sciences*, 40, 06 2004. doi: 10.1002/jhbs.20028.

[10] S. Dewan, S. Agarwal, and N. Singh. Laban movement analysis to classify emotions

from motion. In *Laban movement analysis to classify emotions from motion*, page 9, 04 2018. doi: 10.1117/12.2309451.

[11] T. Dixon, editor. *Thomas Brown: Selected Philosophical Writings*. Imprint Academic, 2010.

[12] T. Dixon. Emotion: The history of a keyword in crisis. *Emotion Review*, 4:338–344, 2012. ISSN 17540739. doi: 10.1177/1754073912445814. URL `/doi/pdf/10.1177/1754073912445814?download=true`.

[13] P. Ekman. *Basic Emotions*, chapter 3, pages 45–60. John Wiley and Sons, Ltd, 1999. ISBN 9780470013496. doi: https://doi.org/10.1002/0470013494.ch3. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013494.ch3`.

[14] F. Espositi and A. Bonarini. My robot body: A workshop to promote body awareness and digital literacy through embodiment with robots. *12th COnference on Computation,Communication,Aesthetics& X*, 2024. doi: 10.34626/2024_xcoax_010.

[15] espressif. Esp32-wroom-32 datasheet, 2026. URL `https://documentation.espressif.com/esp32-wroom-32_datasheet_en.pdf`. Consulted on Jan 25th.

[16] E. e.VS. Theory - eurolab e.v., 2025. URL `https://www.laban-eurolab.org/lbms/theory/`. Consulted on December 10th.

[17] S. C. for Affective Sciences-UNIGE. The geneva emotion wheel - swiss center for affective sciences - unige, 2025. URL `https://www.unige.ch/cisa/gew/`. Consulted on December 10th.

[18] P. Frassinetti. *Fabula Atellana : saggio sul teatro popolare latino / Paolo Frassinetti*. Istituto di Filologia Classica, Genova, 1953.

[19] L. R. Goldberg. Goldberg.am.psych.1993. *American Psychology*, 1 1993.

[20] C. Izard. More meanings and more questions for the term "emotion". *Emotion Review*, 2:383–385, 10 2010. doi: 10.1177/1754073910374670.

[21] L. John. Principles of traditional animation applied to 3d computer animation. Technical report, Pixar, 1987.

[22] J. Lecoq. The moving body (le corps poétique). *SKENE Journal of Theatre and Drama Studies*, 1997.

[23] N. A. Limited. Tipi di personalità | 16personalities, 2025. URL `https://www.16personalities.com/it/tipi-di-personalita`. Consulted on Dec 13th.

[24] D. McCaw. *The Laban Sourcebook*. Routledge, 2011.

[25] J. McCosh. Psychology: the motive powers, emotions, conscience, will; : Mccosh, james, 1811-1894 : Free download, borrow, and streaming : Internet archive, 1880. URL `https://archive.org/details/psychologymotive00mcco/`.

[26] S. Meisner, D. Longwell, S. Pollack, A. M. Cianciulli, and D. Padoa. *Sulla recitazione / Sanford Meisner con Dennis Longwell ; introduzione di Sidney Pollack ; edizione italiana a cura di Anna Maria Cianciulli ; traduzione di Olga Dalia Padoa ; con la collaborazione di Anna Maria Cianciulli*. D. Audino, Roma, 2007. ISBN 978-88-7527-034-6.

[27] T. Nakata, T. Mori, and T. Sato. Analysis of impression of robot bodily expression. *Journal of Robotics and Mechatronics*, 12 2001.

[28] A. official Store. Arduino mega 2560 rev3 – 54 digital i/o, 16 analog inputs, 256kb flash — arduino official store, 2026. URL `https://store.arduino.cc/products/arduino-mega-2560-rev3?utm_source=google&utm_medium=cpc&utm_campaign=EU-Pmax&gad_source=1&gad_campaignid=22591755262&gbraid=0AAAAACbEa85eKmmAveYrahSH-r5ZqHwzW&gclid=CjwKCAiAssfLBhBDEiwAcLpwfi5UuOA7qCknVJC3_zovpA9tqPd-xB-fsdx0d43aCAv8noFt7H_JhxoCFdgQAvD_BwE`. Consulted on Jan 25th.

[29] D. J. Pittenger. Measuring the mbti... and coming up short. *ResearchGate*, 1993.

[30] R. Plutchick. The nature of emotions. Technical report, American Scientist, 2001.

[31] Plutchik and Robert. *THE EMOTIONS: Facts,Theories, and a New Model*. New York, Random House, 1968.

[32] Robohub. When robots become art: Interview with yi-wei keng - robohub, 2025. URL `https://robohub.org/when-robots-become-art-interview-with-yi-wei-keng/`. Consulted on Dec 21th.

[33] F. Ross. Carl jung personality types. *the SAP*, 2026. URL `https://www.thesap.org.uk/articles-on-jungian-psychology-2/about-analysis-and-therapy/typology/`.

[34] M. Scheeff, J. Pinto, K. Rahardja, S. Snibbe, and R. Tow. *Experiences with Sparky, a Social Robot*, pages 173–180. Kluwer Academic Publishers, 1 2006. doi: 10.1007/0-306-47373-9_21.

[35] V. Shuman, K. Schlegel, and K. Scherer. (pdf) geneva emotion wheel rating study. *ResearchGate*, 8 2015. URL `https://www.researchgate.net/publication/280880848_Geneva_Emotion_Wheel_Rating_Study`.

[36] V. Spolin. *Improvisation for the Theatre*. Northwestern University Press, 1963.

[37] strangefacemasks. Set di maschere neutre in stile lecoq | maschere strangeface, 2026. URL `https://strangefacemasks.com/it/products/lecoq-style-neutral-mask-set?srsltid=AfmBOoqeWZwm6eRHXvaQgLEN7gK8_Us2nACwHJm9QDq8IXbNPNvWeIcG`. Consulted on Jan 4th.

[38] M. Szuster. Theater without a script - improvisation and the experimental stage of the early mid-twentieth century in the united states. *Text Matters*, 9:374–392, 11 2019. ISSN 2084574X. doi: 10.18778/2083-2931.09.23.

[39] T. T. Times. The gaze of the robot: Oriza hirata's robot theatre - the theatre times, 2025. URL `https://thetheatretimes.com/the-gaze-of-the-robot/`. Consulted on Dec 21th.

[40] Treccani. Atellana - enciclopedia - treccani, 2025. URL `https://www.treccani.it/enciclopedia/atellana/`. Consulted on December 20th.

[41] H. Yanco. *Proceedings of the seventh annual ACMIEEE international conference on Human-Robot Interaction*. ACM, 2012. ISBN 9781450310635.

# A | Appendix A

All files are stored on Github at the following link:
`https://github.com/EleonoraRighi/BLACKWING.git`

# B | Appendix B

Link to questionnaire: `https://blackwingtesting.netlify.app`

All data and figures resulting from the survey up to February 18th are stored in a dedicated folder at the GitHub link written before (see Appendix A).

# List of Figures

# List of Tables

# List of Symbols

| Variable | Description[SI unit] |
|---|---|
| $MAXSPEED$ | maximum velocity of the robot [m/s] |
| $W_i$ | angular speed of wheel $i$ [internal] |
| $V_j$ | linear speed along the $j$ axis [%MAXSPEED] |
| $\omega$ | angular speed around z axis [%MAXSPEED] |
| $R$ | wheel radius[m] |
| $L$ | distance between robot center and each wheel[m] |
| $N$ | Number of Responses |
| $std.dev.$ | standard deviation |
| $L1$ | level 1 of intensity |
| $L2$ | level 2 of intensity |
| $L3$ | level 3 of intensity |
| $Likertscale$ | Evaluation scale with values from 1 to 5 |
| $Accuracy$ | percent value of correct emotion recognition |

# Acknowledgements

First and foremost, I would like to express my sincere gratitude to Professor Andrea Bonarini for giving me the opportunity to carry out this thesis under his supervision. His guidance and support proved invaluable on every level, from technical insight to creative direction and beyond.

I would also like to thank everyone at AIRLab for the welcoming and stimulating environment, with special thanks to Dr. Erica Panelli for her constant availability and willingness to help whenever needed.

A special acknowledgment goes to Mattia Olari, who gave BlackWing its distinctive voice. His dedication, spending an entire day with me, a synthesizing app and countless animal sound videos, made a unique and memorable contribution to this project.

I am also deeply grateful to my lifelong friends, Laura and Matilde, who patiently endured months of conversations about feathery robots and were kind enough to show enthusiasm for every work-in-progress video I shared.

Finally, I would like to thank my family for their unwavering support and belief in me, not only throughout this thesis but throughout my entire life. Their encouragement, patience, and love have been the foundation of every step I have taken, giving me the confidence to pursue my goals and overcome challenges. They have celebrated my successes, supported me through setbacks, and inspired me to keep striving for excellence. Without their steadfast presence and guidance, this achievement would not have been possible.

I am truly and deeply grateful.