Universiteit Leiden
The Netherlands

Data Science and AI

Optimizing Multimodal Emotion Expression
for an Appearance-Constrained Robot

Maurits Koppers

Supervisors:
Dr.ir. Joost Broekens & PhD candidate. Bernhard Hilpert

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

01/07/2025

Abstract

This thesis investigates how appearance-constrained robots can express joy, sadness, anger, and fear through combinations of light, sound, and movement. The study involved analyzing existing video data to select effective behavior parameters, implementing and optimizing these behaviors on an mBot robot, and conducting an online user study to test emotional recognition and perceived intensity.

Results showed that three out of four target emotions were recognized significantly above chance, especially anger, which was clearly recognized even without sound. Sound contributed positively by improving recognition of fear and anger and increasing perceived intensity for joy and anger. Participants also tended to group emotions into general categories like positive and negative, suggesting that robots can convey emotional tone even when specific labels are less clear.

Future work should focus on improving sound design, testing with larger and more diverse participant samples, using a larger surface area to reduce unintended behavior, and applying the method to other appearance-constrained robots to assess generalizability.

Contents

1 Introduction

Robots are becoming more and more important in our society [7]. From lawn mowing to complex industrial operations. As technology advances, robots are expected to take on even more significant roles in various sectors, including healthcare, education, and public services [2]. As robots become more common, it's increasingly important that people can understand and interact with them in an intuitive way. Effective human-robot interaction (HRI) is therefore a key research area, particularly as it relates to the interpretability of robot behavior.

Interpretability in this context refers to the degree to which humans can understand a robot's intentions, actions, and internal states. Making robots more interpretable leads to more natural and fluid interactions, increasing user trust, cooperation, and acceptance [35]. One powerful way to improve interpretability is through emotional expression. Emotions make it easier for humans and other species to communicate and it helps us understand what others' feelings and intentions are [12]. In robots, affective signals such as light, sound, and movement can serve as effective substitutes for human-like expressions [1, 4].

Humans naturally communicate emotions through facial expressions, vocal tones, and body language [14]. These ways of expressing emotion are a very important part of how we interact with others. It helps us share complex feelings quickly and clearly. However, many robots, especially those designed for functional tasks, do not have anthropomorphic features such as faces or limbs. Designing robots with full humanoid features is expensive and not always necessary. These appearance-constrained robots, often used in fields like logistics, cleaning, or education, don't have the physical features needed to express emotions in human-like ways.

This challenge creates an opportunity for further research: how can non-humanoid robots express emotions clearly and understandably? Prior research has shown that non-verbal modalities such as light, movement, and sound can successfully influence perceived emotion in robotic systems [25]. Studies have explored each modality individually, or sometimes in pairs. Researchers found that faster motion correlates with high arousal, red or blue light conveys certain affective states, and sound patterns can suggest happiness, sadness, or fear [12, 27, 30, 37].

Yet despite these insights, most studies are descriptive or correlative in nature and focus on one or two modalities. Recent work by Fernando Vargas [12], which will be discussed more thoroughly later on, made a substantial contribution to this field. Vargas identified correlations between parameters using a large dataset, which he collected through a user study. However, his findings remained correlational, and no attempt was made to optimize these parameters for clearer or more consistent emotional expression.

1.1 Research Objective and Scope

The goal of this research is to find and evaluate optimized combinations of light, sound, and movement that help non-humanoid robots express emotions more clearly. This should make robots more emotionally understandable, socially intuitive, and better suited for natural interaction with people.

This thesis builds on the work of Vargas [12], who explored how different behavioral parameters relate to perceived emotions using a large video dataset. While his research focused on identifying correlations, the aim here is to go a step further by creating and testing optimized expressions for specific emotions.

The project initially planned to include all six of Ekman's basic emotions (joy, sadness, anger, fear, surprise, and disgust) [15] . However, early testing showed that surprise and disgust were often hard to express or distinguish clearly. This finding is supported by the work of Jack et al. [22], who demonstrated that the recognition of certain emotions, especially disgust and surprise, varies significantly across cultures and is less consistent than for other emotions. Because of this, the scope was narrowed to four emotions that participants consistently recognized better: joy, sadness, anger, and fear.

The research consists of three main stages. First, Vargas' dataset will be analyzed to identify the most promising video examples for each emotion. Second, an appearance-constrained robot will be programmed with optimized settings based on this analysis. Third, a user study will be conducted in which participants are divided into two groups: one views robot behaviors with sound, and the other without sound. This between-subjects setup allows for direct evaluation of the added effect of sound on emotional recognition and intensity.

The ultimate goal of this research is to contribute to the development of robots that are emotionally intelligible, socially intuitive, and better suited for natural interaction with humans, particularly when traditional expressive features are unavailable.

2 Background and Related Work

This chapter provides the necessary background and theoretical concepts needed for this thesis. It is divided into five sections: affective computing, emotion theories, appearance-constrained robot design, expression modalities, and the use of multiple modalities together.

2.1 Foundations of Affective Computing

The way people interact with robots becomes increasingly important as they move from industrial settings into everyday life. In environments like schools and hospitals, robots are no longer just tools but social actors that share physical and emotional space with humans. To make these interactions smoother and more intuitive, robots must not only behave functionally but also communicate their intentions and internal states in ways that people can easily interpret [17, 18, 36].

Affective computing, introduced by Rosalind Picard [31], is the field that studies how machines can recognize, simulate, and respond to human emotions. Although early related works in human-robot interaction mainly focused on how humans express emotions through facial cues and gestures [20, 33, 34], more recent research has begun to explore other modalities such as movement [1,11, 16, 23], light [24, 39], sound [4, 40], and their combinations [36, 37]. For instance, Song and Yamada [36] demonstrated how variations in vibration, color, and sound pitch can communicate emotions like joy or sadness in minimalist robots.

In human-robot interaction, emotional cues help to increase interpretability and predictability. It allows people to form accurate mental models of what a robot is doing and why [17]. This becomes more important when users are unfamiliar with the robot's technology and must rely on its behavior to understand the robot's intentions. Studies show that affective behavior in robots can boost user engagement, cooperation, and trust. Without such cues, interactions can be confusing, inefficient, or even uncomfortable [18].

Historically, the clearest example of affective expression on robots have come from humanoid robots, which have faces, bodies, and gestures that are similar to humans [26]. These designs take advantage of the fact that people better recognize anthropomorphic traits. However, many functional robots lack those traits.

Designing emotional behavior through these modalities is not straightforward. Facial expressions are relatively universal. For example, a smile is widely understood to indicate happiness, while downward-turned lip corners suggest sadness [15]. These signals are rooted in human biology and have been studied extensively. In contrast, the emotional meaning of light patterns, sound effects, or movement patterns are less standardized. A pulsating blue light might suggest calm to one person but sadness to another. The interpretation of such signals is often context-dependent and shaped by prior experiences and expectations [17, 21].

Importantly, robots do not need to mimic human emotions exactly. Instead, the goal is to create emotionally intelligible behaviors that make the robot's intentions more transparent and relatable. As Law et al. [26] argue, non-humanoid robots offer a unique opportunity to investigate how humans

assign emotional meaning based on cues like movement style or lighting, without relying on facial mimicry.

2.2 Emotion Theories for Robotic Expression

Designing emotionally expressive robots requires a clear understanding of what emotions are and how they can be modeled in artificial systems. Psychological emotion theories offer this foundation and help guide which emotions to express, how to simulate them, and how to evaluate user perception. Emotion theories are typically divided into two major categories: categorical and dimensional models.

2.2.1 Categorical Theories of Emotion

Categorical emotion theories suggest that emotions are distinct, biologically driven states that people can universally recognize and clearly distinguish from one another [15]. According to this perspective, emotional experiences are hardwired into the human brain and expressed through consistent patterns of behavior. One of the most influential figures in this field is Paul Ekman, who identified six basic emotions: joy, sadness, anger, fear, disgust, and surprise. Each of these emotions is associated with unique physiological responses, especially facial expressions [15]. Ekman's model builds on Darwin's hypothesis that emotions serve an evolutionary purpose and that their expressions, especially through the face, are universally shared across cultures [13].

In addition to Ekman's theory, other categorical models have expanded the framework of basic emotions. For instance, Plutchik's emotion wheel organizes eight primary emotions, which are organized into four opposing pairs: joy vs. sadness, trust vs. disgust, fear vs. anger, and surprise vs. anticipation. For instance, fear (avoidance) and anger (confrontation) are considered opposites because they often trigger contrasting responses to threats [28].

Another approach, Parrott's hierarchical model, categorizes over 100 emotions into a tree-like structure. At the top are six primary emotions, love, surprise, joy, sadness, anger, and fear, which branch into more specific secondary emotions and tertiary emotions. While Plutchik's and Parrott's models offer more detail, Ekman's six basic emotions remain the most widely used in affective computing and human-robot interaction [28].

In the field of human-robot interaction, using categorical models has practical advantages. Bretan et al. [9] found that predefined emotion sets make it easier to design and test emotion-specific robot behaviors in user studies. Similarly, Löffler et al. [27] highlights that categorical models work particularly well for applications in situations where emotions need to be recognized quickly and easily, such as in interactive or therapeutic scenarios.

Categorical emotions are mostly used, but this approach also faces some challenges. Ekman's theory is based on facial expressions and vocal cues, which appearance-constrained robots often don't have. As a result, applying his framework to robots that rely on abstract modalities like light, sound, or movement is not always straightforward. Additionally, research across different cultures has raised doubts about how universal these expressions really are. For example, Jack et al. [22] suggest that people from different cultures understand emotions in different ways. Their studies showed that recognition of certain emotions, particularly disgust and surprise, varied significantly across cultures,

while joy, sadness, anger, and fear were more consistently interpreted. These findings challenge the idea that all six of Ekman's basic emotions are universally recognized.

2.2.2 Dimensional Theories of Emotion

Instead of seeing emotions as separate categories, dimensional theories describe them as points in a continuous space. A well-known example is Russell's Circumplex Model of Affect, which organizes emotions along two dimensions: valence and arousal [32]. This model helps explain why some emotions seem similar or overlap. For example, surprise and fear are both high arousal, but are different in valence [9].

Dimensional approaches are especially useful in HRI for appearance-constrained robots. In those cases, cues like light or motion don't clearly point to specific emotions but can still give a general emotional impression [12]. For instance, Song and Yamada [36] found that people judged a robot's behavior based on how energetic or pleasant it seemed, rather than using clear emotion labels. It supports the use of valence–arousal as a useful way to understand emotion perception.

A related model is the PAD model (Pleasure, Arousal, Dominance), which adds a third factor: dominance. It refers to the degree of control a person feels [28]. Although commonly used in simulation studies, dominance is harder to express in simple robot behaviors [12].

Despite their strengths dimensional models have, they don't offer clear rules for designing behavior connected to specific emotion names, which can make them less useful in settings like therapy.

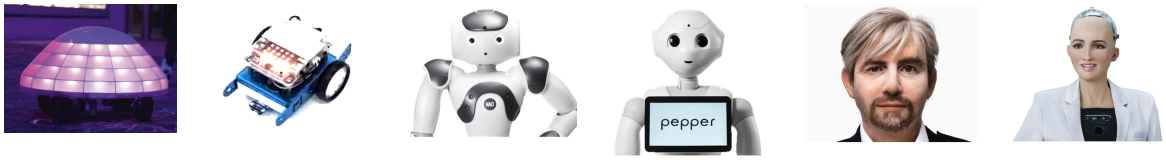
2.3 Appearance-Constrained Robot Design

In human-robot interaction, a robot's physical appearance fundamentally influences how it can express emotion. As mentioned earlier, appearance-constrained robots lack anthropomorphic features. Bethel and Murphy describe these robots as being designed for function, not for expressing emotions [6]. Common examples of appearance-constrained robots are Woody and mBot (Figure 2.1).

To clarify the boundary between humanoid and non-humanoid robots, Epley et al. [44] introduced the concept of morphological similarity, which refers to how closely a robot's observable features resemble a human. Robots with high morphological similarity lead people to expect more human-like behavior. However, if these expectations aren't met the interaction can feel unnatural. This effect is known as the "uncanny valley", a concept in robotics and animation that describes discomfort or unease when something looks or acts almost human, but not quite [29].

Building on the idea of morphological similarity, Ferrari et al. [41] offers a classification system that divides robots into three morphological categories: Minimal similarity (e.g., Roomba and mBot), Moderate similarity (e.g., Nao and Pepper), and High similarity (e.g., Geminoid DK and Sophia) (Figure 2.1).

Using this framework, it becomes clear that appearance-constrained robots fall into the minimal similarity category. In such cases, when traditional signals are missing, affective recognition and expression become more context-dependent [19].



a) Woody [21] b) mBot c) Nao [45] d) Pepper [46] e) Geminoid DK [47] f) Sophia [48]

Figure 2.1: Examples of robots with minimal similarity (a, b), moderate similarity (c, d) and high similarity (e, f).

2.4 Expression Modalities in Appearance-Constrained Robots

Appearance-constrained robots use movement, light, and sound as their main ways to express emotion. Each of these modalities are able to convey emotional meaning. However, how well they work depends on the situation, how the signals are tuned, and whether they are used together. This thesis follows the approach of Vargas [12], who studied these three modalities in detail using the same type of robot. As such, the explanations in this section are based on his analysis and highlight the main ideas from his work.

2.4.1 Movement

Even without complex mechanics, robot motion can be a rich source of emotional expression. Research shows that speed and acceleration are linked to arousal, while direction, smoothness, and shape of the trajectory influence valence [12]. For instance, smooth, rounded, upward movements are commonly associated with joy, while slow, downward, or contracting motions convey sadness or fear [27].

Emotional movements in robots often take inspiration from patterns observed in nature. For instance, fast and upward movements are usually linked to joy, while slow, heavy, or inward movements suggest sadness. These patterns reflect things we see in the natural world, like a healthy plant growing upward or a wilting plant curling down. Additionally, how stable a robot's movement appears also matters. Smooth, predictable movement is generally seen as more positive, whereas unstable movement is linked to negative emotions [12].

Lastly, animation techniques like exaggeration and anticipation can make even simple robots feel more expressive. In short, movement is one of the most effective ways to express emotion [12].

2.4.2 Light

Light is a simple but effective way for robots to express emotion, especially when they don't have faces or bodies. Features like color, brightness, and blinking speed can influence emotional interpretation. For example, warm colors like red and orange often feel urgent or intense, while cool colors like blue and green are seen as calm and positive [12].

However, emotional interpretation of color is not universal and can be context-dependent. Studies show that people may associate the same color with different emotions depending on its use or cultural background [36]. Moreover, Löffler et al. [27] found that light alone was less effective than

motion or sound at conveying emotion in isolation. However, when combined with motion, emotion recognition accuracy significantly improved.

2.4.3 Sound

Sound, particularly non-linguistic utterances like tones or beeps, can carry emotional meaning. Rising tones are often associated with tension or anger, while falling tones suggest sadness or resignation [36]. Other parameters, such as tone slope, duration, and repetition, also shape how emotions are perceived.

However, sound alone can be misinterpreted. Jeong et al. [42] found that participants often confused affective tones with system alerts or functional robot cues. Despite this, sound has specific strengths: it appears to enhance recognition of sadness, possibly due to its association with lower energy and downward pitch patterns. In contrast, sound did not improve recognition for joy and it even reduced confidence for fear classification [27].

2.5 Multimodal Emotional Expression in Robots

While modalities like movement, light, and sound can communicate affective cues individually, research shows that emotional clarity improves when they are combined.

Early work by Bethel & Murphy [6] identified motion, posture, orientation, color, and sound as effective modalities for non-humanoid robots. They found that motion and sound worked especially well at close range, while visual elements like color contributed to emotional tone.

Building on this, Song & Yamada [36] systematically tested color, sound, and vibration on a simple robot. They showed that single modalities often failed to clearly convey emotions. However, when modalities were combined it significantly improved recognition for anger and sadness. Happiness was notably harder to express. Their follow-up study Song & Yamada [37] confirmed that adding motion to light-based expressions reduced misclassification.

Löffler et al. [27] used an appearance-constrained robot to compare unimodal and multimodal designs. They found that the combination of color and motion was most effective in conveying joy. Sound alone was most effective for expressing sadness, while motion was the dominant cue for fear. The study also showed that when multiple modalities convey the same emotional message, both recognition accuracy and user confidence increase.

Ghafurian et al. [18] found that MiRo, a zoomorphic robot without a face, could express 11 emotions using posture, movement, and light. Recognition was significantly above chance for most expressions. Tsiourti et al. [43] warned that inconsistent cues (e.g., sad gestures with a happy voice) confuse users and lower trust, highlighting the importance of alignment across modalities.

Vargas [12] found that behavioral parameters, especially motion speed, strongly influence perceived arousal in appearance-constrained robots. While there were also statistically significant effects for valence, dominance, and some categorical emotions, these were much weaker. His models showed that although emotional perception can be predicted from parameters, only arousal was consistently

expressed with high reliability. The study suggests that appearance-constrained robots are effective at signaling intensity, but struggle to clearly convey specific emotions like joy or sadness.

3 Research Question

This chapter introduces the motivation behind the study by discussing a research gap. It then presents the main research question, followed by three sub-questions and their related hypotheses.

3.1 Research Gap

Although prior work has explored emotion expression through non-verbal modalities in robots, most studies have remained descriptive or correlational, especially for appearance-constrained robots. Research by Vargas [12] and others (e.g., Song & Yamada, Löffler et al.) [27, 36, 37] confirm that these modalities affect how emotions are perceived, but only a few have gone further by systematically selecting and fine-tuning the underlying parameters.

3.2 Main Research Question and Sub-Questions

This thesis addresses the following main research question:

MRQ: How can combinations of light, sound, and movement be optimized in appearance-constrained robots to accurately and consistently communicate the emotions of joy, sadness, anger, and fear to human observers?

To answer this overarching question, the following sub-questions are investigated:

SQ1: Can the target emotions (joy, sadness, anger, and fear) be selected by users at rates significantly above chance when expressed through optimized robot behaviors?

SQ2: Does the addition of sound improve participants' ability to select the robot's emotional expressions, compared to expressions using only light and movement?

SQ3: Does the presence of sound lead to higher perceived emotional intensity on the target emotion, compared to the no-sound condition?

3.3 Hypotheses

To evaluate the effectiveness of the robot's emotional expressions, three main hypotheses are defined, each corresponding directly to one of the sub-questions introduced in Section 3.2. These hypotheses target recognition accuracy and intensity as perceived by users in both sound and no-sound conditions. Recognition (H1 and H2) corresponds to the forced-choice emotion selection, while intensity (H3) is based on Likert-scale ratings. Each main hypothesis is further supported by detailed micro-hypotheses, which are presented in Appendix A.

H1: Optimized combinations of light, sound, and movement enable participants to recognize the target emotions (joy, sadness, anger, and fear) at rates significantly above chance level (1 out of 12 = 8.3%).

H2: Participants in the sound condition show significantly higher recognition accuracy than those in the no-sound condition for the target emotions (joy, sadness, anger, and fear).

H3: Participants in the sound condition rate the emotional intensity of the robot behaviors significantly higher than participants in the no-sound condition for the target emotions (joy, sadness, anger, and fear).

4 Method

All code used for data analysis, visualization, and robot behavior control is publicly available at [https://github.com/Maurits319/Maurits-Koppers---Bachelor-Thesis-2025]. The repository also contains a PDF version of the full survey used in the user study.

4.1 Overview

This study combines data analysis, robot programming, and a user experiment to investigate how non-humanoid robots can express emotions through light, sound, and movement. The process began with the selection of promising robot behaviors based on an existing video dataset. These data were then analyzed to determine key parameter values. The identified parameters were implemented and further refined with an mBot robot through an optimization session. A between-subjects user study was then conducted via an online survey to evaluate how well participants could recognize the intended emotions in videos of the optimized behaviors. The collected quantitative and qualitative survey data were analyzed to evaluate the study's hypotheses and to interpret the findings.

4.2 Robot and Behavior Architecture

This study used the mBot as the robot for implementing and testing expressive emotional behaviors (image b in Figure 2.1). It is a low-cost educational robot developed by Makeblock. The mBot includes two DC motors, RGB LEDs, a buzzer, and several onboard sensors, and is programmable through the Arduino IDE via its mCore control board.

To ensure the robot had an appearance-constrained design, two visual features that typically give the mBot a face-like appearance were removed. The ultrasonic sensor, which is often seen as eyes, was not needed for navigation and was therefore removed. Additionally, a red and a blue LED that always stay on when the robot is activated, were covered with white duct tape to avoid unintended visual cues. This helped ensure that only the RGB LEDs used for affective expression were visible.

To maintain consistency with Vargas' thesis [12], I used the same codebase to control the mBots. The robot's expressive behavior was defined using three base behaviors, each corresponding to a different modality of affective expression: Wander (movement), Blink (light), and Beep (sound). Each base behavior includes several parameters (see Table 4.1), but only the three most influential ones per behavior are examined in this thesis, as well as in Vargas' original study.

Modality	Parameter	Description
Wander (Movement)	wander_speed	Base speed of the robot while wandering (in % of motor power).
	wander_roundness	Controls how rounded or sharp the turns are.
	wander_cycle_rate	Frequency of switching between moving forward and turning (in cycles per second, or Hz).
Blink (Light)	blink_temperature	Target temperature of the lights during the blink behavior, controlling the color.

	blink_slope	Determines how the brightness of the lights changes over time.
	blink_cycle_rate	Rate at which the blink cycles occur (in Hz).
Beep (Sound)	beep_pitch	Base pitch (frequency) of the beep sound in Hz.
	beep_slope	Determines how the pitch of the beep sound changes over time.
	beep_cycle_rate	Rate at which the beep cycles occur (in Hz).

Table 4.1: An overview of the parameters used to control the robot's emotional expressions across three modalities: movement (Wander), light (Blink), and sound (Beep).

Each base behavior followed a cyclical pattern: wander alternated between forward and turning segments, blink alternated between lights-on and lights-off periods, and beep alternated between sound and silence. The parameter values used for each emotion were initially derived through data analysis (see Section 4.3), and later refined during an in-person optimization session (see Section 4.4).

4.3 Initial Data Analysis

To identify promising robot behaviors for emotional expression, this study used a dataset created by Vargas (2024) [12], which includes 512 videos of an mBot robot displaying different combinations of movement, light, and sound. Each video was rated by human participants on multiple emotional dimensions, resulting in a rich dataset of perceived emotion scores.

Two main metrics were used to select promising videos: intensity and purity.

- **Intensity**: Measures how strongly a specific emotion was perceived in a video, based on ratings from 0–5. For each video, all participant ratings for the target emotion were added together and then divided by the number of participants who rated the video. This gives the average perceived intensity of the target emotion per video.
- **Purity**: Measures how dominant one emotion was over all others. For each video, purity was calculated by dividing the participant's rating for the target emotion by the sum of all their emotion ratings for that video. These ratios were then averaged across all participants to obtain a single purity value for each video.

For each of Ekman's six basic emotions (joy, sadness, anger, fear, disgust, and surprise), the top 20 videos were selected twice: once based on highest average intensity and once based on highest average purity. From these subsets, the four most promising videos per emotion were selected by comparing the two lists of 20 videos (based on intensity and purity) and choosing the ones that ranked highly in both. This reduced the dataset to 24 high-quality examples (4 per emotion, 6 emotions).

This selection process formed the first optimization step of the study. From the original dataset of 512 videos, a subset of 24 videos was created that best represented each of the six basic emotions. These selected videos were considered strong candidates for each emotion and formed the foundation for further implementation and testing. To identify consistent behavioral features, the four selected

videos per emotion, along with their corresponding parameter values, were analyzed to find recurring patterns in light, movement, and sound settings. It is important to note that at this stage, no decisions were made yet to exclude the emotions of disgust and surprise from further analysis.

4.4 Implementation and Refinement of Emotional Behaviors

After selecting the 24 most promising videos from Vargas' dataset and analyzing them for patterns, the next step was to implement and optimize emotional behaviors on an mBot. The goal of this phase was to optimize the behavior parameter values through manual adjustments and human feedback in order to improve the clarity and accuracy of the emotional expressions.

4.4.1 Physical Setup

The robot was assembled using the same appearance-constrained configuration described in Section 4.2. To ensure consistency during video creation, a simple filming area was used: a white cardboard surface of approximately 1×1 meter, bordered with black tape to keep the robot within frame (see Figure 4.1). A line sensor ensured the robot stayed within this boundary during each take.

Figure 4.1: Filming setup with mBot on a 1×1 meter white cardboard surface. The black tape marks the boundary used to keep the robot within the camera frame. A top-mounted camera and ensured consistent positioning and behavior during recording.

4.4.2 Optimization Procedure

For each of the four target emotions (joy, sadness, anger, and fear), the initial parameter implementation was based on the values identified during the data analysis phase. For each emotion, several slightly different parameter configurations were manually created by adjusting the light, movement, and sound settings. Each configuration was recorded as a separate video, and a reference table was maintained to track the corresponding parameter values.

These implementations were reviewed together with two other individuals, who provided feedback. Based on this input, another round of refinement was conducted, producing four new variations per

emotion. This iterative process was repeated until smaller and smaller ranges were found for each parameter, aiming to optimize the emotional expressions.

Once these parameter ranges were determined, four videos were recorded for each emotion. Due to stochastic elements in the behavior code, each video differed slightly in aspects such as the robot's starting position and movement pattern. The random elements were intentionally implemented to make the robot's behavior look natural, instead of fully preprogrammed. As a result, some videos appeared more expressive or visually appealing than others. Therefore, four recordings were made for each emotion, which allowed for a comparison between videos to identify the most expressive one.

All four videos per emotion were then edited to synchronize movement, light, and sound, which will be explained in the next section. After this editing step, the best video for each emotion was selected as the final version for use in the user study.

4.4.3 Synchronizing Sound and Motion

The Wander (movement) and Blink (light) base behaviors were initially recorded separately from the Beep (sound) behavior. This was done to reduce the background noise caused by the robot's motors, allowing participants to focus more clearly on the sound itself. Each initial video consisted of 20 seconds of movement and light, and 20 seconds of sound. A Python script was used to combine these segments, aligning all three modalities into a single 20-second video.

4.5 User Study Design

To evaluate whether the optimized robot behaviors effectively conveyed the intended emotions, an online user study was conducted using Qualtrics (full survey available here). The survey included an opening statement that provided participants with important information about their rights and the task description.

Immediately following the introduction, participants were asked to complete an informed consent form. This form contained a series of yes/no questions confirming that they had read the study information, understood their rights, and voluntarily agreed to participate. The form also clarified that all responses would remain anonymous and that no personal or identifiable data would be collected.

The study followed a between-subjects design, with participants randomly assigned to one of two conditions using Qualtrics' randomizer function to ensure even distribution. Group A (n = 20) viewed videos that included synchronized sound, light, and movement. Group B (n = 27) viewed the same videos but without the sound component. This design enabled the evaluation of Hypothesis 2 and 3.

Each participant viewed four videos, one for each of the target emotions (joy, sadness, anger, and fear). After watching each video, participants answered three types of questions: a Likert-scale rating of 12 emotions, a forced-choice question asking them to select the best-matching emotion, and an open-ended text box to describe their interpretation.

4.6 User Study Data Analysis

The data collected through the online survey included both quantitative and qualitative components. The quantitative analysis focused on recognition accuracy and perceived emotional intensity, based on

participants' responses to Likert-scale and forced-choice questions. The qualitative analysis included participants' open-ended text responses, which were summarized using a Large Language Model (LLM) to identify recurring patterns in their answers.

4.6.1 Data Preparation

A total of 73 responses were collected through the Qualtrics platform using a between-subjects design. After excluding incomplete submissions, 47 valid responses remained: 20 from participants in the sound condition and 27 in the no-sound condition. Only these complete cases were included in the analyses.

4.6.2 Tools Used

Data analysis was primarily conducted in Python, which was used for data cleaning, visualization, and the execution of statistical tests. Additionally, SPSS was used to perform a MANOVA (Multivariate Analysis of Variance) on the emotional intensity ratings.

4.6.3 Recognition Accuracy Analysis

Recognition accuracy data was collected using participants' answers to the forced-choice question, in which participants selected the one emotion label that best matched the video they had just seen. To evaluate whether the robot's emotional expressions were successfully recognized, two types of statistical analyses were conducted:

- **Binomial tests** were used to determine whether the recognition accuracy for each target emotion was significantly above chance level (8.3%, or 1 out of 12 options), corresponding to Hypotheses 1.
- **Proportion z-tests** were used to compare the recognition accuracy between the sound and no-sound conditions, corresponding to Hypotheses 2.

4.6.4 Emotional Intensity Analysis

Participants also rated each of the 12 emotions on a 5-point Likert scale after watching each video. The rating for the target emotion in each video was used to evaluate perceived intensity. Two types of analyses were conducted:

- **Independent-samples t-tests** were used to compare the perceived intensity of each target emotion between the sound and no-sound conditions, corresponding to Hypotheses 3.
- **A MANOVA** was conducted to explore general patterns in intensity ratings, testing the main effects of Emotion and Condition and their interaction.

4.6.5 Qualitative Analysis Using a Large Language Model

In addition to the closed-ended questions, participants provided open-ended descriptions of the robot's emotional behavior after each video. These responses were grouped by target emotion and split by condition. A Large Language Model (GPT-4o by OpenAI) was then used to generate short summaries of the participant responses. The resulting summaries highlight common patterns in how participants perceived the robot's behavior and emotional expression, which complements the quantitative findings.

5 Results

5.1 Overview

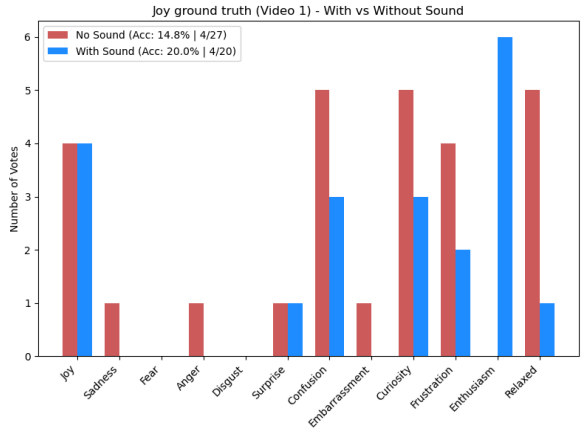
This section presents the results of the study, which was designed to evaluate how participants perceived the emotional expressions of a robot. The results are structured around the hypotheses defined earlier and are divided into two main parts: the forced-choice recognition task and the perceived emotional intensity ratings. For each part, relevant figures and tables are presented alongside appropriate statistical analyses. The final parameter values used for the optimized robot behaviors are provided in Appendix B.

5.2 Forced-Choice Recognition Accuracy

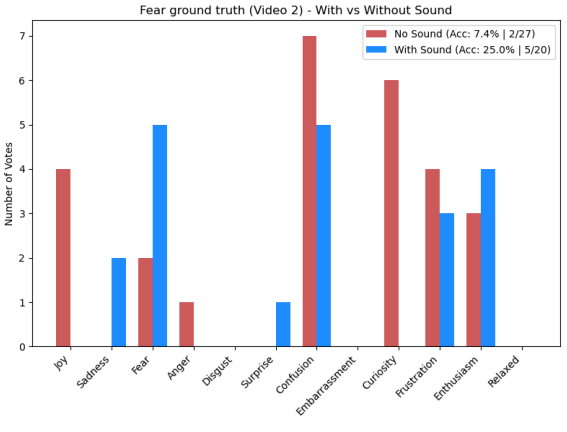
In the forced-choice recognition task, participants viewed four video clips and selected which of twelve emotion labels best described the robot's behavior for each video. Each clip corresponded to a target emotion (joy, fear, anger, or sadness). The recognition accuracy was then computed based on whether participants selected the correct target emotion.

5.2.1 Response Distribution Plots

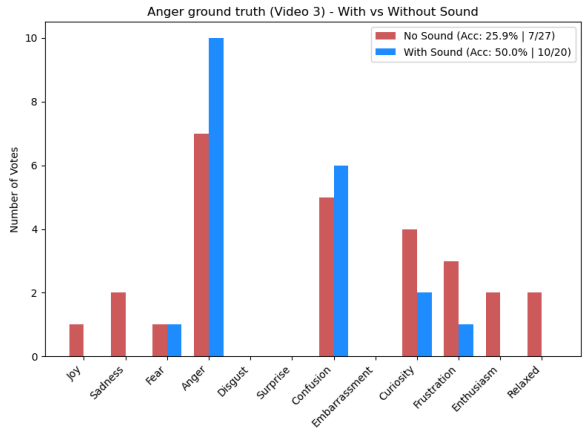
The plots in Figure 5.1 display the distribution of responses for each condition in the forced-choice recognition task.



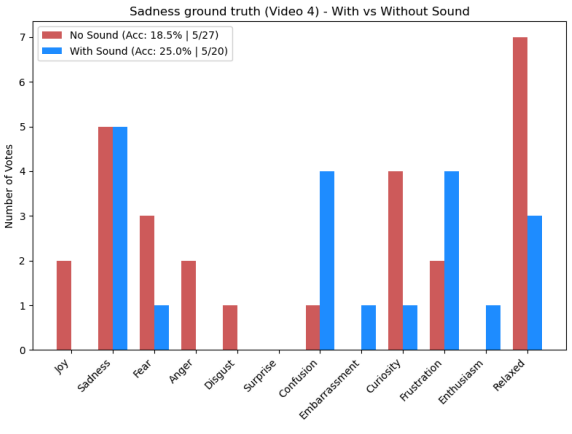
a) Plot for Joy: accuracy = 20.0% (4/20) for sound, 14.8% (4/27) for no sound



b) Plot for Fear: accuracy = 25.0% (5/20) for sound, 7.4% (2/27) for no sound



c) Plot for Anger: accuracy = 50.0% (10/20) for sound, 25.9% (7/27) for no sound



d) Plot for Sadness: accuracy = 25.0% (5/20) for sound, 18.5% (5/27) for no sound

Figure 5.1: Forced-choice recognition plots for the four target emotions (joy, fear, anger, and sadness), comparing participant selections between the sound and no-sound conditions. Each plot displays the number of votes per emotion category and the corresponding accuracy rate for the intended (ground truth) emotion.

5.2.2 Accuracy Against Chance (Hypothesis 1)

To determine whether participants recognized the intended emotion above chance level (8.3%), binomial tests were conducted for each target emotion in both the sound and no-sound conditions. The results are shown in Table 5.1.

Emotion	Condition	Accuracy	Correct / Total	p-value	Significance
Joy	Sound	20.0%	4 / 20	0.080	Not Significant
	No Sound	14.8%	4 / 27	0.184	Not Significant
Fear	Sound	25.0%	5 / 20	0.022	**Significant**
	No Sound	7.4%	2 / 27	0.670	Not Significant
Anger	Sound	50.0%	10 / 20	< 0.0001	**Significant**
	No Sound	25.9%	7 / 27	0.0056	**Significant**
Sadness	Sound	25.0%	5 / 20	0.022	**Significant**
	No Sound	18.5%	5 / 27	0.069	Not Significant

Table 5.1: Results of binomial significance tests for forced-choice emotion recognition accuracy. For each target emotion, the number of correct responses, overall accuracy, and corresponding p-values are reported separately for the sound and no-sound conditions. It is also indicated whether values are statistically significant ($p < 0.05$), meaning that recognition performance is significantly above the chance level of 1 out of 12 options (8.3%).

5.2.3 Condition-Based Accuracy Comparison (Hypothesis 2)

In addition to testing against chance, Hypothesis 2 examines whether recognition accuracy is significantly higher in the sound condition compared to the no-sound condition. Proportion z-tests were conducted for each emotion, and the results are shown in Table 5.2.

Emotion	Sound Accuracy	No-Sound Accuracy	p-value	Significance
Joy	20.0%	14.8%	0.3200	Not Significant
Fear	25.0%	7.4%	0.0470	**Significant**
Anger	50.0%	25.9%	0.0447	**Significant**
Sadness	25.0%	18.5%	0.2957	Not Significant

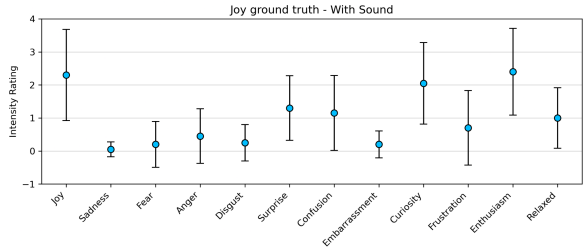
Table 5.2: Results of z-tests for condition-based differences in recognition accuracy. For each target emotion, recognition accuracy is shown separately for the sound and no-sound conditions. The table includes p-values from z-tests comparing the proportions and indicates whether the differences are statistically significant ($p < 0.05$).

5.3 Emotion Intensity Ratings

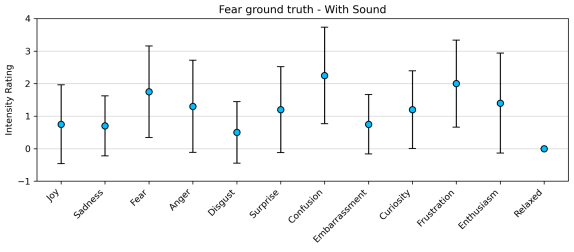
In addition to the forced-choice task, participants also rated the intensity of all twelve emotion categories for each video on a 5-point Likert scale. These ratings were used to analyze whether the presence of sound influenced the perceived emotional intensity of the robot's behavior.

5.3.1 Error Bar Plots of Intensity Ratings

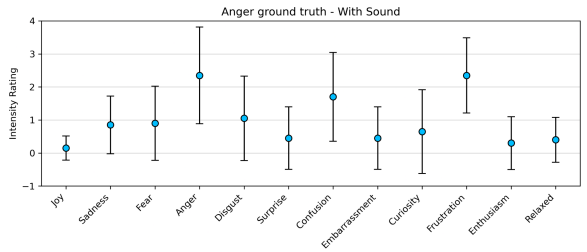
The plots below show the average intensity ratings for each emotion option, across the four target emotions. Error bars indicate the standard deviation across participants. There are 8 plots in total: four for the sound condition shown in Figure 5.2 and four for the no-sound condition shown in Figure 5.3.



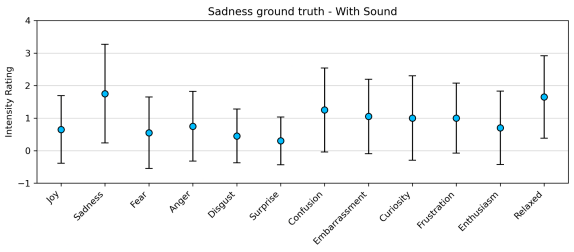
a) Plot for Joy



b) Plot for Fear

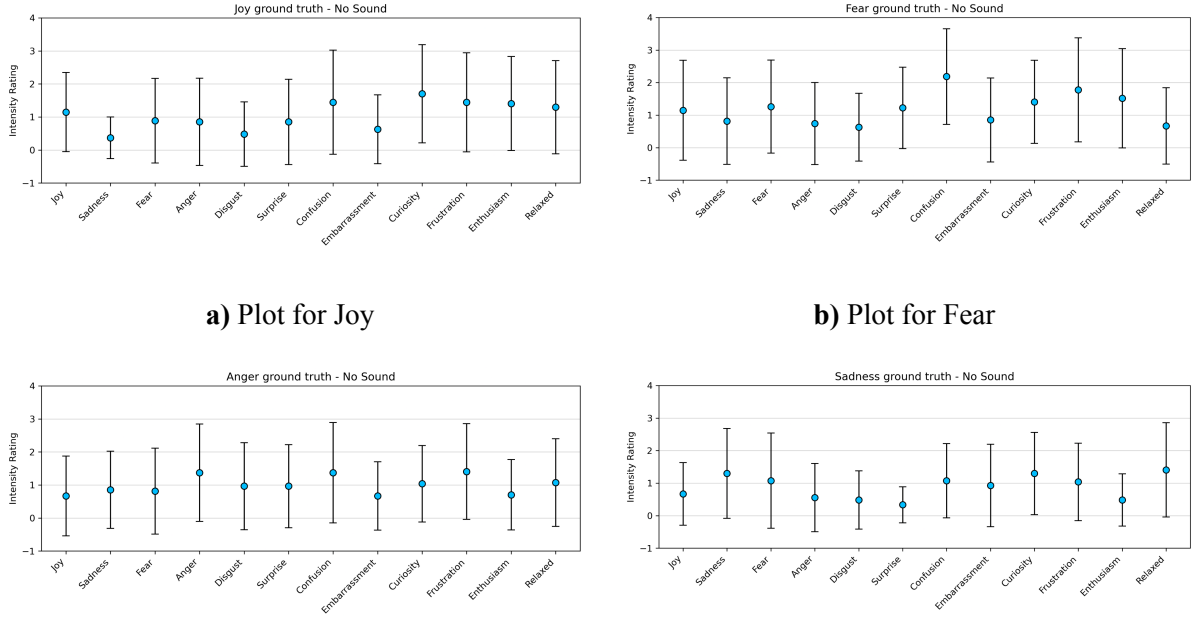


c) Plot for Anger



d) Plot for Sadness

Figure 5.2: Error bar plots showing mean emotional intensity ratings (Likert 0–4) per emotion label, for the four target emotions (joy, fear, anger, and sadness) in the **sound** condition. Each dot represents the average rating across participants for a specific emotion label, with error bars indicating ±1 standard deviation.



a) Plot for Joy

b) Plot for Fear

c) Plot for Anger

d) Plot for Sadness

Figure 5.3: Error bar plots showing mean emotional intensity ratings (Likert 0–4) per emotion label, for the four target emotions (joy, fear, anger, and sadness) in the **no sound** condition. Each dot represents the average rating across participants for a specific emotion label, with error bars indicating ±1 standard deviation.

5.3.2 Condition-Based Intensity Comparison (Hypothesis 3)

To statistically evaluate Hypothesis 3, independent-samples t-tests were performed. A total of four tests were conducted, one for each target emotion. These tests compared the mean intensity ratings between the sound and no-sound conditions. Table 5.3 shows the results, indicating whether the differences were statistically significant.

Emotion	Sound Mean Rating	No-Sound Mean Rating	p-value	Significance
Joy	2.30	1.15	0.0025	**Significant**
Fear	1.75	1.26	0.1238	Not Significant
Anger	2.35	1.37	0.0144	**Significant**
Sadness	1.75	1.30	0.1495	Not Significant

Table 5.3: Results of independent-samples t-tests for perceived emotional intensity ratings. For each target emotion, the mean intensity rating is shown separately for the sound and no-sound conditions. The table also reports the corresponding p-values and whether the difference in ratings between the conditions is statistically significant ($p < 0.05$).

5.4 MANOVA

To gain deeper insight into the emotional intensity ratings, a MANOVA (Multivariate Analysis of Variance) was conducted. The within-subject factor was emotion (12 levels, corresponding to the full set of rated emotions), and the between-subject factor was condition (sound vs. no-sound).

This analysis complements the earlier hypothesis testing (binomial, z-tests, and t-tests), which focused on recognition accuracy and intensity ratings for specific target emotions. In contrast, the MANOVA examines general patterns across all emotions.

First, the results of a Wilks' Lambda multivariate test are presented to examine the overall main effects and the interaction between condition and emotion. These results are shown in Table 5.4.

Effect	p-value	Significance
Condition	0.383	Not Significant
Emotion	0.031	**Significant**
Emotion × Condition	0.648	Not Significant

Table 5.4: MANOVA Results (Wilks' Lambda) for the Effects of Condition and Emotion on Perceived Emotional Intensity

Next, univariate test results using the sphericity assumed assumption are presented to explore which specific emotions showed significant differences in perceived intensity. The results are shown in Table 5.5 and Table 5.6

Emotion	p-value	Significance
Joy	< 0.001	**Significant**
Sadness	< 0.001	**Significant**
Fear	< 0.001	**Significant**
Anger	< 0.001	**Significant**
Disgust	0.004	**Significant**
Surprise	< 0.001	**Significant**
Confusion	< 0.001	**Significant**
Embarrassment	0.014	**Significant**
Curiosity	< 0.001	**Significant**
Frustration	< 0.001	**Significant**
Enthusiasm	< 0.001	**Significant**

Relaxed	< 0.001	**Significant**

Table 5.5: This table shows the results of univariate Manova tests for each individual emotion, examining whether the intensity ratings significantly differ across emotions, regardless of sound condition. All tests are based on the "Sphericity Assumed" correction. Significant p-values indicate that participants rated the intensity of that emotion as significantly different from others.

Emotion	p-value	Significance
Joy	0.003	**Significant**
Sadness	0.344	Not Significant
Fear	0.054	Not Significant
Anger	0.050	**Significant**
Disgust	0.847	Not Significant
Surprise	0.117	Not Significant
Confusion	0.693	Not Significant
Embarrassment	0.522	Not Significant
Curiosity	0.376	Not Significant
Frustration	0.022	**Significant**
Enthusiasm	0.040	**Significant**
Relaxed	0.217	Not Significant

Table 5.6: This table presents the univariate Manova results for the interaction between emotion and sound condition. It tests whether the presence of sound had a different effect on intensity ratings depending on the specific emotion shown. All results use the "Sphericity Assumed" correction. Significant values suggest that the sound condition influenced how a particular emotion was perceived in terms of intensity.

5.5 Open-Ended Response Summaries

To complement the quantitative data from the forced-choice and Likert-scale questions, Table 5.7 presents summaries of participants' open-ended responses. These summaries were generated using a Large Language Model (GPT-4o by OpenAI) and are grouped by target emotion and experimental condition (sound vs. no-sound). The prompt used to generate these summaries was carefully designed to minimize bias, and is included in Appendix C.

Emotion	Condition	LLM Summary of Participant Descriptions
Joy	Sound	Participants most often described the robot as enthusiastic, happy, and curious, frequently referencing its active movements, cheerful or high-pitched sounds, and colorful lights. Several also interpreted the robot as relaxed or playful, with some noting a childlike or exploratory energy. A few responses mentioned confusion or frustration, particularly when the robot appeared to repeat motions or bump into barriers, but the general tone was upbeat and energetic.
Joy	No Sound	This version elicited a more mixed impression. Some participants perceived the robot as joyful, curious, or confident—citing its fast movement and bright lights—while others saw it as confused, frustrated, or even panicked, especially due to its repetitive motion patterns and apparent inability to escape the environment. Descriptions often reflected a robot that was energetically exploring but unsure or overwhelmed.
Fear	Sound	Participants predominantly interpreted the robot as anxious, afraid, or in panic, with frequent references to its fast, erratic movements, flashing blue lights, and high-pitched, alarming sounds. Some also described confusion or malfunctioning behavior. A few respondents noted enthusiasm or excitement, but the general tone conveyed a sense of urgency, stress, or disorientation.
Fear	No Sound	Reactions were diverse, with interpretations ranging from joyful and curious to confused, panicked, or stressed. Many noted the robot's rapid, irregular movements and flashing lights, which led some to perceive excitement and others to describe fear or nervousness. Overall, the robot was often seen as energetic but lacking clear direction, producing a mix of positive curiosity and overwhelmed confusion.
Anger	Sound	Most participants identified the robot as angry, frustrated, or annoyed, emphasizing its red lights, sharp movements, and intense sounds. Several also mentioned confusion or aimlessness, suggesting the robot was struggling to find a way out or was reacting to a problem. Some found the robot's behavior aggressive or alarming, while others saw it as searching or defensive in tone.
Anger	No Sound	Interpretations ranged from angry and frustrated to curious, relaxed, or confident. The red light and fast back-and-forth movements led many to identify confusion or goal-seeking frustration, though others saw a robot calmly and purposefully exploring. Emotional clarity varied widely, with some describing a chaotic or indecisive robot and others seeing mission-driven behavior.
Sadness	Sound	The robot was commonly described as sad, confused, or lost, often based on its slow, cautious movement and lower-pitched or strange sounds. Some participants interpreted its behavior as frustrated or depressed, while a few saw it as relaxed or resigned. Emotional tones leaned toward low-energy states, with occasional comments about the robot seeming embarrassed or disoriented.
Sadness	No Sound	This version also elicited strong impressions of confusion, sadness, and slowness, though some participants saw hints of curiosity, relaxation, or even embarrassment. The red lights contributed to varied interpretations—some saw anger or frustration, while others contrasted the color with the robot's gentle behavior to suggest melancholy. Overall, the robot was often viewed as emotionally subdued or gently searching.

Table 5.7: Presents the summaries of participants' open-ended responses for each of the four emotion videos (joy, fear, anger, and sadness) under two conditions: with sound and without sound.

6 Discussion

This chapter begins by discussing general patterns and observations from the results, followed by a detailed analysis of the three main hypotheses. The chapter concludes with suggestions for future research.

6.1 General Patterns and Observations

This subsection discusses four main topics: the distinctiveness of the emotional videos, the lack of a general sound effect on perceived intensity, patterns in alternative emotion selections, and the pattern that certain emotions tend to cluster together.

6.1.1 Distinctiveness of Emotional Videos

The MANOVA results showed a significant main effect of Emotion (Wilks' Lambda, $p = 0.031$), meaning that participants rated the twelve emotions differently in terms of intensity. The follow-up univariate tests (Table 5.5) confirmed that all twelve emotion labels were significantly distinct from one another. This suggests that the videos clearly conveyed different emotional expressions. While this does not necessarily indicate how accurately the emotions were conveyed, it does show that each video has a unique behavioral pattern. This is a useful finding in itself, as it confirms that the robot's behavior varied across the emotion categories.

6.1.2 No General Sound Effect on Intensity Across Emotions

Although sound was expected to influence perceived intensity, the MANOVA showed no significant main effect for Condition ($p = 0.383$) and no significant Emotion × Condition interaction ($p = 0.648$). This means that adding sound did not lead to an overall increase or decrease in perceived emotional intensity. Instead, the effect of sound depended on the specific emotion. Some emotions were influenced by sound, but not all.

6.1.3 Patterns in Alternative Emotion Selections

Besides the statistical analyses, an inspection of the forced-choice and Likert-scale data showed some interesting patterns. There was a frequent appearance of confusion as a selected emotion, particularly in the no-sound condition. Participants often interpreted the robot's behavior as uncertain, especially for emotions like joy and fear. This suggests that sound contributes to making the robot's behavior less confusing.

In addition to confusion, participants often selected curiosity and frustration, especially in the no-sound condition. This may be because the robot frequently bumped into the tape at the edge of the area, which people interpreted as exploratory behavior or as the robot getting stuck. This pattern also appeared in the open-ended descriptions that were summarized using a Large Language Model (LLM).

Another interesting finding was that in the joy condition, enthusiasm was most frequently selected by participants in the sound condition, while it was not even selected once in the no-sound condition. This indicates that the presence of sound had a strong influence on how energetic or excited the

robot's behavior was perceived. Since enthusiasm and joy are closely related, it is reasonable that participants interpreted the expression as enthusiastic.

6.1.4 Emotional Grouping

The recognition and intensity results in this study show a mix of significant and non-significant outcomes. However, a closer look at the data visualization shows an underlying structure: emotions tend to cluster along positive and negative dimensions. For example, emotions such as joy, curiosity and enthusiasm were often selected together, while anger and frustration formed a more negative cluster.

This pattern suggests that even though specific emotion labels were not always accurately recognized, participants were often able to interpret the general emotional tone of the robot's behavior. This aligns with dimensional emotion theories discussed in Section 2.2.2. In this experimental setup, participants may have found it easier to distinguish positive from negative emotions rather than making distinctions between similar emotions.

It is also important to consider the limitations of the survey design. Participants were asked to choose from 12 emotion labels, presented in a randomized order for each question. A different structure, such as a two-step approach where participants first decide between positive and negative, followed by selecting a more specific label might have led to higher recognition accuracy. Additionally, the videos lacked contextual information, which is often important for interpreting emotional expressions, as confirmed by Angel-Fernandez and Bonarini [1].

6.2 Interpreting Hypothesis 1 – Recognition Above Chance

To evaluate whether participants could select the intended emotion above chance level (8.3%), binomial significance tests were conducted for each of the four target emotions (joy, fear, anger, and sadness) under both sound and no-sound conditions. The results provide mixed support for Hypothesis 1.

For the anger condition, recognition was significantly above chance in both the sound (50.0%, $p < 0.0001$) and no-sound (25.9%, $p = 0.0056$) conditions. This indicates that the robot's expression of anger was clear and, based on the forced-choice responses, it was the most successfully recognized emotion overall. This aligns with findings from Song & Yamada [36], who showed that anger is particularly effectively conveyed when light, sound, and motion are combined.

Sadness was recognized significantly above chance in the sound condition (25.0%, $p = 0.022$), but not in the no-sound condition (18.5%, $p = 0.069$). This means that the addition of sound is crucial for helping participants interpret the emotion as sadness. This finding is consistent with Löffler et al. [27], who found that sound is particularly effective for expressing sadness.

For the fear condition, recognition was significant in the sound condition (25.0%, $p = 0.022$), but clearly not in the no-sound condition (7.4%, $p = 0.670$). This large difference suggests that motion and light alone may not be sufficient to convey fear effectively in this setup. Interestingly, this finding is in contrast with Löffler et al. [27], who found that sound did not improve fear classification and even reduced participant confidence.

Finally, joy was not recognized above chance in either condition. Accuracy was relatively low in both the sound (20.0%, p = 0.080) and no-sound (14.8%, p = 0.184) conditions. The data suggest that this may be due to many participants selecting enthusiasm instead of joy in the sound condition. Enthusiasm is typically more goal-oriented and associated with anticipation of a future outcome, whereas joy is often experienced when a goal has been achieved or in a state of calmness and contentment [13]. This distinction suggests that reducing the robot's movement speed might increase the recognition accuracy of joy.

6.3 Interpreting Hypothesis 2 – Effect of Sound on Recognition Accuracy

To evaluate if participants in the sound condition would recognize the intended emotion more accurately than participants in the no-sound condition, proportion z-tests were performed for each of the four target emotions. The results show partial support for this Hypothesis 2.

For both the fear and anger conditions, recognition accuracy was significantly higher in the sound condition. For fear, accuracy increased significantly from 7.4% to 25.0% (p = 0.0470). A similar pattern was observed for anger, where accuracy increased from 25.9% to 50.0% (p = 0.0447). These results indicate that sound plays an important role for these two emotions.

In contrast, for joy and sadness, the differences in recognition accuracy between the sound and no-sound conditions were not statistically significant. Joy recognition increased slightly with sound (from 14.8% to 20.0%, p = 0.3200), and sadness also didn't improve much (from 18.5% to 25.0%, p = 0.2957). As discussed earlier, many participants in the sound condition selected enthusiasm rather than joy. If some of those participants had instead chosen joy, the difference in recognition rates might have reached statistical significance. The lack of statistical significance for sadness could be due to the light and movement cues already being clear enough that the addition of sound added little value, or the sound design was ineffective and failed to improve emotional recognition.

These findings partially align with Löffler et al. [27]. In their study, sound helped participants recognize sadness more easily, but this effect was not found in this study. Both studies agree that sound did not improve recognition of joy. However, Löffler et al. found that sound actually lowered participants' confidence when recognizing fear, while the current study found the opposite.

6.4 Interpreting Hypothesis 3 - Effect of Sound on Perceived Emotional Intensity

To evaluate whether participants in the sound condition would rate the robot's emotional expressions as more intense than those in the no-sound condition, independent-samples t-tests were conducted for each target emotion. These tests compared the mean intensity ratings between the sound and no-sound conditions. The results provide partial support for Hypothesis 3.

For joy, participants rated the expression as significantly more intense (p = 0.0025) when sound was included (M = 2.30) compared to the no-sound condition (M = 1.15). Similarly, for anger, mean intensity ratings were significantly higher (p = 0.0144) in the sound condition (M = 2.35) than in the no-sound condition (M = 1.37). These findings suggest that sound significantly boosts the perceived intensity of joy and anger.

In contrast, for fear, the difference in intensity ratings between the two conditions was not statistically significant ($p = 0.1238$), with mean ratings increasing from 1.26 in the no-sound condition to 1.75 in the sound condition. Similarly, for sadness, the difference was also not significant ($p = 0.1495$), with mean ratings increasing from 1.30 in the no-sound condition to 1.75 in the sound condition. Although both trends suggest slightly higher perceived intensity when sound is added, the effects were not strong enough to reach statistical significance.

Once again, these results are partially in contrast with Löffler et al. [27], who found that sound was the main driver for communicating sadness and helped participants recognize it more easily. However, our results indicate that the addition of sound did not significantly improve the perceived intensity of sadness. Interestingly, for fear, no significant difference in intensity was found, which aligns with Löffler et al., but this is in contrast with the earlier discussed forced-choice recognition results (Section 6.3), where sound did significantly improve recognition of fear.

6.5 Suggestions for Future Research

This study provides valuable insights into emotional expression in appearance-constrained robots, but there are several opportunities for future research to build upon these findings.

First, the robot operated in a limited 1×1 meter area, which often caused it to bump into the boundaries. As discussed earlier, participants sometimes interpreted this unintended behavior as part of the emotional expression. Since this was a side effect of the limited space rather than a design choice, future studies should consider using a larger area to minimize such distractions.

Second, the survey structure could be improved. Participants were asked to choose from twelve emotion labels presented in a random order, which may have made the task overly complex. A two-step approach by first selecting between positive or negative valence, then choosing a specific emotion might improve accuracy.

Third, the sound design used in this study could be completely reworked. The current audio setup was chosen to maintain consistency with the earlier thesis by Vargas, on which this study is based. However, future research could focus on optimizing sound independently while keeping movement and light in the same style. As discussed, sound can play a crucial role in emotional expression, and improving its design could add significant value.

Fourth, a larger and more diverse participant pool would improve the generalizability and statistical reliability of the results. Including people of different ages, cultural backgrounds, genders, and levels of experience with robots would make the findings more valuable. A larger and more varied sample also reduces the risk of bias.

Finally, future studies could explore whether these findings apply to other appearance-constrained robots. This would help determine whether the findings are specific to the mBot or generalizable across other low-morphology robots such as the Roomba.

7 Conclusion

This thesis explored how combinations of light, sound, and movement can be optimized in appearance-constrained robots to express the emotions of joy, sadness, anger, and fear. The research involved several stages: analyzing existing data to select promising behavior parameters, optimizing these behaviors on a physical robot, and conducting a user study to evaluate how well participants recognized the intended emotions.

The findings show that optimization of the three modalities had a strong impact on emotional recognition. Three out of the four target emotions (anger, sadness, and fear) were recognized significantly above chance level. Anger was especially well recognized, even without sound. When comparing the sound and no-sound conditions, there was no general increase in perceived emotional intensity across all emotions. Instead, the impact of sound varied by emotion. It significantly improved recognition accuracy for anger and fear, and increased perceived intensity for anger and joy, while having little to no effect on sadness.

In conclusion, this study demonstrates that even with limited physical features, appearance-constrained robots can effectively communicate emotions when their behavior is carefully designed.

References

1. Angel-Fernandez, J. M., & Bonarini, A. (2016). Robots showing emotions: Emotion representation with no bio-inspired body. *Interaction Studies, 17*(3), 408–437. https://doi.org/10.1075/is.17.3.06ang
2. Aymerich-Franch, L., & Ferrer, I. (2021). Socially assistive robots' deployment in healthcare settings: A global perspective. *International Journal of Humanoid Robotics, 18*(1). https://doi.org/10.1142/S0219843623500020
3. Berns, K., & Zafar, Z. (2018). Emotion based human-robot interaction. *MATEC Web of Conferences, 161*, Article 01001. https://doi.org/10.1051/matecconf/201816101001
4. Bethel, C. L., & Murphy, R. R. (2006). Affective expression in appearance-constrained robots. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/1121241.1121299
5. Bethel, C. L., & Murphy, R. R. (2006). Auditory and other non-verbal expressions of affect for robots. In *Proceedings of the AAAI Fall Symposium: Aurally Informed Performance* (pp. 1–5). Association for the Advancement of Artificial Intelligence.
6. Bethel, C. L., & Murphy, R. R. (2008). Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 38*(1), 83–92. https://doi.org/10.1109/TSMCC.2007.905845
7. Blessing, E., & Klaus, H. (2024). The impact of robotics on society and civilization. *Robotics Engineering*.
8. Breazeal, C., & Brooks, R. (2004). Robot emotions: A functional perspective. In J.-M. Fellous & M. A. Arbib (Eds.), *Who needs emotions? The brain meets the robot* (pp. 271–310). Oxford University Press.
9. Bretan, M., Hoffman, G., & Weinberg, G. (2015). Emotionally expressive dynamic physical behaviors in robots. *International Journal of Human–Computer Studies, 78*, 1–16. https://doi.org/10.1016/j.ijhcs.2015.01.006
10. Cass, A. G., Striegnitz, K., & Webb, N. (2018). A farewell to arms: Non-verbal communication for non-humanoid robots. In *Proceedings of the Workshop on Natural Language Generation for Human–Robot Interaction* (pp. 22–26). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-6905
11. Cauchard, J. R., Zhai, K. Y., Spadafora, M., & Landay, J. A. (2016). Emotion encoding in human-drone interaction. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 263–270). https://doi.org/10.1109/HRI.2016.7451761
12. Corte Vargas, F. (2024). *Designing emotionally expressive behaviors for an appearance-constrained robot: Evaluating the affective interpretation of motion, light and sound* (Master's thesis, Delft University of Technology).
13. Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray. https://doi.org/10.1037/10001-000
14. de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. Cognition & Emotion, 14(3), 289–311. https://doi.org/10.1080/026999300378824
15. Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068
16. Embgen, S., Luber, M., Becker-Asano, C., Ragni, M., Evers, V., & Arras, K. O. (2012). Robot-specific social cues in emotional body language. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'12)* (pp. 1019–1025). IEEE. https://doi.org/10.1109/ROMAN.2012.6343883

17. Frederiksen, M. R., & Stoy, K. (2019). A systematic comparison of affective robot expression modalities. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. https://doi.org/10.1109/IROS40897.2019.8967846
18. Ghafurian, M., Lakatos, G., & Dautenhahn, K. (2022). The zoomorphic MiRo robot's affective expression design and perceived appearance. *International Journal of Social Robotics, 14*, 945–962. https://doi.org/10.1007/s12369-021-00832-3
19. Gunes, H., & Churamani, N. (2023). Affective computing for human–robot interaction research: Four critical lessons for the hitchhiker. In *Proceedings of the 32nd IEEE RO-MAN: Robot and Human Interactive Communication* (pp. 1565–1572). https://doi.org/10.1109/RO-MAN57019.2023.10309450
20. Hashimoto, T., Hitramatsu, S., Tsuji, T., & Kobayashi, H. (2006). Development of the face robot SAYA for rich facial expressions. In *SICE-ICASE International Joint Conference 2006* (pp. 5423–5428). IEEE. https://doi.org/10.1109/SICE.2006.315751
21. Hoggenmueller, M., Chen, J., & Hespanhol, L. (2020). Emotional expressions of non-humanoid urban robots: The role of contextual aspects on interpretations. In *Proceedings of the 9th ACM International Symposium on Pervasive Displays* (pp. 87–95). https://doi.org/10.1145/3393712.3395341
22. Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. Proceedings of the National Academy of Sciences, 109(19), 7241–7244. https://doi.org/10.1073/pnas.1200155109
23. Knight, H., & Simmons, R. (2014). Expressive motion with X, Y and Theta: Laban Effort Features for mobile robots. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '14)* (pp. 267–273). https://doi.org/10.1109/RO-MAN.2014.6926323
24. Kobayashi, K., Funakoshi, K., Yamada, S., Nakano, M., Komatsu, T., & Saito, Y. (2011). Blinking light patterns as artificial subtle expressions in human-robot speech interaction. In *Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 181–186). https://doi.org/10.1109/RO-MAN.2011.6005124
25. Kraack, K. (2024). A multimodal emotion recognition system: Integrating facial expressions, body movement, speech, and spoken language. *arXiv preprint* arXiv:2412.17907. https://doi.org/10.48550/arXiv.2412.17907
26. Law, T., de Leeuw, J., & Long, J. H. (2021). How movements of a non-humanoid robot affect emotional perceptions and trust. *International Journal of Social Robotics, 13*(8). https://doi.org/10.1007/s12369-020-00711-3
27. Löffler, D., Tscharn, R., & Lindeman, R. W. (2018). Multimodal expression of artificial emotion in social robots using color, motion and sound. In *Proceedings of the 3rd International Conference on Intelligent Human Systems Integration* (pp. 127–132). https://doi.org/10.1145/3171221.3171261
28. Mohammed, S. N., & Abdul Hassan, A. K. (2020). A survey on emotion recognition for human–robot interaction. *Journal of Computing and Information Technology, 28*(2), 125–146. https://doi.org/10.20532/cit.2020.1004841
29. Mori, M. (1970). Bukimi no tani [Uncanny valley]. Energy, 7(4), 33–35. (Authorized English translation by MacDorman & Minato republished in Mori, M. (2012). The uncanny valley. IEEE Robotics & Automation Magazine, 19(2), 98–100.)
30. Novikova, J., & Watts, L. A. (2014). A design model of emotional body expressions in non-humanoid robots. In *HAI 2014 – Proceedings of the 2nd International Conference on Human-Agent Interaction* (pp. 353–360). Association for Computing Machinery. https://doi.org/10.1145/2658861.2658892

31. Picard, R. W. (1997). *Affective computing*. MIT Press.
32. Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178. https://doi.org/10.1037/h0077714
33. Saldien, J., Goris, K., Vanderborght, B., Vanderfaeillie, J., & Lefeber, D. (2010). Expressing emotions with the social robot Probo. *International Journal of Social Robotics, 2*(4), 377–389. https://doi.org/10.1007/s12369-010-0067-6
34. Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., & Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence, 166*(1–2), 140–164. https://doi.org/10.1016/j.artint.2005.03.005
35. Siau, K. L., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal, 31*(2), 47–53.
36. Song, S., & Yamada, S. (2017). Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. https://doi.org/10.1145/2909824.3020239
37. Song, S., & Yamada, S. (2018). Designing expressive lights and in-situ motions for robots to express emotions. In *Proceedings of the 6th International Conference on Human-Agent Interaction* (pp. 222–228). https://doi.org/10.1145/3284432.3284458
38. Tabari, N., Zadrozny, W., & Reddy, C. K. (2018). Emotion detection in text: A review. *arXiv*. https://doi.org/10.48550/arXiv.1806.00674
39. Terada, K., Yamauchi, A., & Ito, A. (2012). Artificial emotion expression for a robot by dynamic color change. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication* (pp. 314–321). https://doi.org/10.1109/ROMAN.2012.6343772
40. Thiessen, R., Rea, D. J., Garcha, D. S., Cheng, C., & Young, J. E. (2019). Infrasound for HRI: A robot using low-frequency vibrations to impact how people perceive its actions. In *Proceedings of the 2019 ACM/IEEE International Conference on Human–Robot Interaction (HRI)* (pp. 11–18). https://doi.org/10.5555/3378680.3378685
41. Ferrari, F., Paladino, M. P., & Jetten, J. (2016). Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. International Journal of Social Robotics.
42. Jeong, E., Kwon, G. H., & So, J. (2017). Exploring the taxonomic and associative link between emotion and function for robot sound design. In 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI) (pp. 641–643). IEEE. https://doi.org/10.1109/URAI.2017.7992692
43. Tsiourti, T., Kühn, P., Becker-Asano, C., & Röcker, C. (2019). Multimodal integration of emotional signals from voice, body, and context: Effects of incongruence on emotion recognition and attitudes towards robots. International Journal of Social Robotics, 11(3), 451–468. https://doi.org/10.1007/s12369-019-00524-z
44. Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864
45. Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., Marnier, B., Serre, J., & Maisonnier, B. (2009). Mechatronic design of NAO humanoid. In 2009 IEEE International Conference on Robotics and Automation (pp. 769–774). IEEE. https://doi.org/10.1109/ROBOT.2009.5152516
46. Lafaye, J., Gouaillier, D., & Wieber, P.-B. (2014). Linear model predictive control of the locomotion of Pepper, a humanoid robot with omnidirectional wheels. In 2014 IEEE-RAS

International Conference on Humanoid Robots (pp. 336–341). IEEE. https://doi.org/10.1109/HUMANOIDS.2014.7041381

47. Vlachos, E., & Schärfe, H. (2015). Towards designing android faces after actual humans. In G. Jezic, R. J. Howlett, & L. C. Jain (Eds.), Agent and multi-agent systems: Technologies and applications (pp. 109–119). Springer International Publishing. https://doi.org/10.1007/978-3-319-19728-9_9

48. Riccio, T. (2021). Sophia Robot: An emergent ethnography. TDR: The Drama Review, 65(3), 42–77. https://doi.org/10.1017/S1054204321000319

Appendix A – Micro Hypotheses

ID	Detailed Micro-Hypotheses for RQ1
H1.1	There is no significant recognition of ‘joy’ above chance level when expressed using light, sound, and movement.
H1.2	There is no significant recognition of ‘sadness’ above chance level when expressed using light, sound, and movement.
H1.3	There is no significant recognition of ‘anger’ above chance level when expressed using light, sound, and movement.
H1.4	There is no significant recognition of ‘fear’ above chance level when expressed using light, sound, and movement.
H1.5	There is no significant recognition of ‘joy’ above chance level when expressed using only light and movement.
H1.6	There is no significant recognition of ‘sadness’ above chance level when expressed using only light and movement.
H1.7	There is no significant recognition of ‘anger’ above chance level when expressed using only light and movement.
H1.8	There is no significant recognition of ‘fear’ above chance level when expressed using only light and movement.

Table A.1: Detailed Micro-Hypotheses for RQ1

ID	Detailed Micro-Hypotheses for RQ2
H2.1	Recognition accuracy for ‘joy’ is not significantly higher in the sound condition than in the no-sound condition.
H2.2	Recognition accuracy for ‘sadness’ is not significantly higher in the sound condition than in the no-sound condition.
H2.3	Recognition accuracy for ‘anger’ is not significantly higher in the sound condition than in the no-sound condition.
H2.4	Recognition accuracy for ‘fear’ is not significantly higher in the sound condition than in the no-sound condition.

Table A.2: Detailed Micro-Hypotheses for RQ2

ID	Detailed Micro-Hypotheses for RQ3
H3.1	The perceived intensity of ‘joy’ is not significantly higher in the sound condition than in the no-sound condition.
H3.2	The perceived intensity of ‘sadness’ is not significantly higher in the sound condition than in the no-sound condition.
H3.3	The perceived intensity of ‘anger’ is not significantly higher in the sound condition than in the no-sound condition.
H3.4	The perceived intensity of ‘fear’ is not significantly higher in the sound condition than in the no-sound condition.

Table A.3: Detailed Micro-Hypotheses for RQ3

Appendix B – Used mBot Parameters

	Joy	**Fear**	**Anger**	**Sadness**
wander_speed	85	100	82	25
wander_roundness	0.6	0.3	0.10	0.1
wander_cycle_rate	1.6	5.5	0.55	1
blink_temperature	0.55	0.2	0.99	0.95
blink_slope	1	0	1	0
blink_cycle_rate	2	4.5	1.7	1
beep_pitch	900	750	100	350
beep_slope	1.3	1	0	0
beep_cycle_rate	1.4	4	0.7	0.5

Table B.1: Optimized Robot Parameters for Emotional Expressions

Appendix C – Prompt for LLM Summaries

I have collected free-text responses from a user study in which participants described what emotion they thought a robot was expressing in a video. The videos differed in design, and participants were exposed to different versions. Your task is to generate brief summaries of how participants described the robot's behavior for each group of responses.

Instructions:

- You will receive grouped responses, each from a different version of a video (e.g., Version A of Video 1, then Version B of Video 1, etc.). The responses are in English.

For each group:

- Provide a short paragraph summarizing the most commonly mentioned emotions, perceived behaviors, and interpretations.
- Focus on clarity, emotional language, and the general impression participants had of the robot's behavior.
- Do not infer or rely on any assumptions about the robot's intended emotion or the conditions under which the video was shown.

Label the output like this:
Video 1 – Version A
Summary: …

Video 1 – Version B
Summary: …

(Repeat for all 4 videos)

Please keep each summary concise but informative (~3–5 sentences). Let me know when you're ready to begin with the first group of responses.