DESIGN AND DEVELOPMENT OF A SOCIAL ROBOTIC HEAD – DOROTHY

DAI DONGJIAO

NATIONAL UNIVERSITY OF SINGAPORE

2010

DESIGN AND DEVELOPMENT OF A SOCIAL ROBOTIC HEAD – DOROTHY

DAI DONGJIAO

(B.Eng.(Hons.), NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF ENGINEERING

DEPARTMENT OF MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2010

Acknowledgements

My Master of Engineering has come to the end. At this moment, I would like to express my grateful thanks to a number of people. The project was not possible without their supervision and invaluable support.

First and foremost, my most sincere thanks go to my supervisor, A/P Marcelo H. Ang Jr (Associate Professor, Department of Mechanical Engineering, National University of Singapore), for his professional supervision, continuous encouragement and unreserved support throughout the project. My interest in the fields of robotics and emotion modeling started when I developed and worked with Doris and Dorothy, as part of my projects in both undergraduate and graduate program and they have become my research interests since then.

Secondly, I would like to express my special thanks to my co-supervisor, Prof Budiman Sastra (Adjunct Research Professor, Division of Engineering and Technology Management, National University of Singapore), for his professional guidance and invaluable advices all along. This project could not be completed smoothly without his full support. His ideas and concepts have had a remarkable influence on my entire research in the field of robotics.

Next, my heartfelt thanks go to my co-supervisor Han Boon Siew (Senior Research Officer, Institute for Infocomm Research, A*Star). Thanks for helping me with the innumerable facets of my work, mentoring me throughout my master program and giving me the opportunity to gain exposure to cutting-edge technologies in robotics. I cannot overemphasize the help he offered and I have really learnt much more beyond the project itself.

Fourthly, special thanks are due to A/P John-John Cabibihan (Associate Professor, Department of Electrical & Computer Engineering, National University of Singapore), who guided me and helped a lot in the design of the skin for Doris and Dorothy.

Next, special appreciation goes to the technicians Ms. Ooi-Toh Chew Hoey, Ms. Tshin Oi Meng and Mr. Sakthiyavan S/O Kuppusamy (Mechatronics & Control Lab 1, Department of Mechanical Engineering, National University of Singapore). Thanks for their assistance in financial affairs and fabrication support.

Last but not least, I am deeply indebted to my project partners Yan Haibin. It was a pleasure cooperating with her for this social robotics project.

Table of Contents

Summary

Nowadays, more and more robots are created with intention to interact and communicate with humans autonomously while following social rules. The objective of this project is to build a robotic head that is able to emulate facial expressions incorporating actuators and appropriate control algorithms and at the same time, can speak the input text typed.

Our social robot head is called Dorothy (Gift of God). Dorothy is constructed within the anthropomorphic domain. It is able to express five basic emotions: happiness, sadness, surprise, anger and fear as well as advanced emotions.

Facial features play an important role in expressing emotional states. Dorothy' face is comprised of 2 eyebrows, 2 eyeballs, 2 eyelids and 2 lips, which are the essential components that serve the emotions. Its eyebrows utilize the four-bar mechanism. The eyeballs are very compact with everything hidden behind the eyeballs. Eyelids also contribute a lot to the expressions of emotions. They enable the eyes to open and close at various degrees and blink as well. The mouth consists of two lips that are actuated by two micro servos. A prototype was built to examine the feasibility of facial features' mechanism before the fabrication. In terms of degree of freedom, Dorothy has 9 DOFs in total, 2 for eyebrows, 4 for eyeballs, 1 for eyelids and 2 for mouth.

As for the hardware for controlling Dorothy, 9 Hitec HS-65HB micro servos are used as the actuators to generate facial expression. The control board is SSC-32, which is very a compact servo control board available. In terms of the software, users can control Dorothy via a GUI. A scenario was predefined for a human user commanding Dorothy. There are three modules in the software architecture of Dorothy. MOUTH MOTION (MM) is to convert the input text to corresponding mouth shapes via two stages - text-to-phonemes and phoneme-to-viseme. SPEECH (SP) is to convert the input text to sound track. FACIAL EXPRESSION enables Dorothy to show the proper emotion as assigned. In brief, Dorothy is able to speak the input text out with the correct mouth shapes, at the same time, show the corresponding emotions at different stages of the scenario.

Upon the completion of mechanical structure and electronic control, a questionnaire was conducted to examine the capability of Dorothy. The result shows that Dorothy is capable to accomplish the mission as described in the predefined scenario.

Recommendations for future work include integration of multimedia functionalities, skin improvement and it learning ability.

List of Tables

List of Figures

Chapter 1 – Introduction

Robotics is an evolving technology. Human beings have been constructing automatic machines for thousands of years; the development of robotics has been exploited at the end of the last century. After decades of hype and disappointment, (1980s and 1990s) robots are at last moving out of the shop-floor, finding their way into our homes and offices, hospitals, museums and other public spaces, in the form of self-navigating vacuum cleaners, lawn mowers, window washers, toys, medical surgical, etc [1]. Nowadays, robotics technology is developing at an accelerating pace all over the world, opening up new possibilities for automating tasks and enriching the lives of humans. From the automobile assembly line, automatic home vacuum cleaners to humanoid robot receptionists, robotics is playing a more and more important role in our world.

1.1 Robots & Social Robots

Robotics is the science and technology of designing, making, and applying robots, including theory from many contributing fields [2]. The products of robotics are robots. A robot is a computer controlled machine which is able to do tasks on its own. It is usually an electromechanical system, which, by its appearance or movements, conveys a sense that it has intent or agency of its own [3]. A robot can be a mechanical or virtual, artificial agent (e.g. an avatar in a virtual world). According to its functionality and main features, robots can be classified as industrial robots, mobile robots, androids, autonomous robots, humanoid robot and social robots. The classification is not absolute. One robot can belong to multiple categories concurrently.

An industrial robot is an automatically controlled, reprogrammable, multipurpose manipulator programmable with three or more axis of motion [2]. Typical industrial robot applications include welding, painting, pick and place, packaging and palletizing, product inspection, and testing, all accomplished with high endurance, speed, and precision. A mobile robot is an automatic machine that is capable of locomotion in a given environment. Mobile robots have the capability to move around in their environment and are not fixed to one physical location. Android is a robot or synthetic organism designed to look and act likes a human. Androids are humanoid robots built to aesthetically resemble a human. Autonomous robots are robots that can perform desired tasks in unstructured environments without continuous human guidance. Many kinds of robots have some degree of autonomy. Different robots can be autonomous in different ways. A humanoid robot is a robot with its overall appearance, based on that of the human body, allowing interaction with made-for-human tools or environments. In general humanoid robots have a torso with a head, two arms and two legs, although some forms of humanoid robots may model only part of the body, for example, from the waist up. Some humanoid robots may also have a 'face', with 'eyes' and 'mouth'.

A social robot is defined as an autonomous robot that interacts and communicates with humans or other autonomous physical agents by following social behaviors and rules attached to its role [2]. Social robots are the agents that deserve to have special human-robot interaction (HRI) systems to be accepted by humans as natural partners. As humans, we not only strive to understand ourselves, but we also turn to technology to enhance the quality of our lives [4]. From an engineering perspective, we try to make these technologies natural and intuitive to use and to interact with. As our technologies become more intelligent and more complex, we still want to interact with them in a familiar way. We tend to ascribe human features to our computers, our cars, and other gadgets for this reason, and their interfaces resemble how we interact with each other more and more. All these inspire human to create social robots, the most anthropomorphized agents that enrich our lives. Nowadays, social robots are receiving much interest in the robotics community. In-depth knowledge of social robots is very important and must be acquired by researchers and engineers before designing any social robots. It will help to keep them on the right track when developing robots.

The most important goal for social robots lies in their social interaction capabilities. A sociable robot must be able to communicate and interact with humans to certain degree, understand and even relate to humans in a personal way. It should be able to understand humans and itself in social terms as well. We, in turn, should be able to understand it in the same social terms - to be able to relate to it and to empathize with it. Such a robot can adapt and learn throughout its lifetime, incorporating shared experiences with other individuals into its understanding of self, of others, and of the relationships they share [4]. In short, a sociable robot is socially intelligent in a humanlike way, and interacting with it is like interacting with another person. At the pinnacle of achievement, they could befriend us as we could.

Socially interactive robots can be used for a variety of purposes: as research platforms, as toys, as educational tools, or as therapeutic aids. The common, underlying assumption is those humans prefer to interact with machines in the same way that they interact with other people. Socially interactive robots operate as partners, peers or assistants, which means that they need to exhibit a certain degree of adaptability and flexibility to drive the interaction with a wide range of humans. Socially interactive robots can have different shapes and functions, ranging from robots whose sole purpose and only task is to engage people in social interactions to robots that are engineered to adhere to social norms in order to fulfill a range of tasks in human-inhabited environments. Some socially interactive robots use deep models of human interaction and pro-actively encourage social interaction. Others show their social competence only in reaction to human behavior, relying on humans to attribute mental states and emotions to the robot. Regardless of function, building a socially interactive robot requires consideration of the human in the loop: as designer, as observer, and as interaction partner.

Robots have limited perceptual, cognitive, and behavioral abilities compared to humans. Thus, for the foreseeable future, there will continue to be significant imbalance in social sophistication between human and robot. As with expert systems, however, it is possible that robots may become highly sophisticated in restricted areas of socialization, e.g., infant-caretaker relations. Differences in design methodology mean that the evaluation and success criteria are almost always different for different robots. Thus, it is hard to compare socially interactive robots outside of their target environment and use. Socially interactive robots must address important issues imposed by social interaction [5].

- **Human-oriented perception**: A socially interactive robot must proficiently perceive and interpret human activity and behavior. This includes detecting and recognizing gestures, monitoring and classifying activity, discerning intent and social cues, and measuring the human's feedback.
- **Natural human–robot interaction**: Humans and robots should communicate as peers who know each other well, such as musicians playing a duet. To achieve this, the robot must manifest believable behavior: it must establish appropriate social expectations, it must regulate social interaction (using dialogue and action), and it must follow social convention and norms.
- **Readable social cues**: A socially interactive robot must send signals to the human in order to: (1) provide feedback of its internal state; (2) allow human to interact in a facile, transparent manner. Channels for emotional expression include facial expression, body and pointer gesturing, and vocalization.
- **Real-time performance**: Socially interactive robots must operate at human interaction rates. Thus, a robot needs to simultaneously exhibit competent behavior, convey attention and intentionality, and handle social interaction, all in a timely fashion.

Robots in individualized societies exhibit a wide range of social behavior, regardless if the society contains other social robots, humans, or both. Breazeal [4] defines four classes of social robots in terms of how well the robot can support the social model that is ascribed to it and the complexity of the interaction scenario that can be supported as followings:

- **Socially evocative.** Robots that rely on the human tendency to anthropomorphize and capitalize on feelings evoked when humans nurture, care, or are involved with their "creation".
- **Social interface.** Robots that provide a "natural" interface by employing human-like social cues and communication modalities. Social behavior is only modeled at the interface, which usually results in shallow models of social cognition.
- **Socially receptive.** Robots that are socially passive but that can benefit from interaction (e.g. learning skills by imitation). Deeper models of human social competencies are required than with social interface robots.
- **Sociable.** Robots that pro-actively engage with humans in order to satisfy internal social aims (drives, emotions, etc.). These robots require deep models of social cognition.

Complementary to this list we can add the following three classes which can be considered a different classification:

- **Socially situated.** Robots that are surrounded by a social environment that they perceive and react to. Socially situated robots must be able to distinguish between other social agents and various objects in the environment.

- **Socially embedded.** Robots that are: (a) situated in a social environment and interacts with other agents and humans; (b) structurally coupled with their social environment; and (c) at least partially aware of human interactional structures (e.g., turn-taking).
- **Socially intelligent.** Robots that show aspects of human style social intelligence, based on deep models of human cognition and social competence.

In brief, all robot systems, socially interactive or not, must be designed in every aspect, including sensing (sound localization, vision system, facial emotion recognition system), cognition (planning, decision making, computational intelligence), perception (navigation, obstacle avoidance , environment sensing), action (mobility, manipulation, gestures), human–robot interaction (user interface, input devices, feedback display) and architecture (control, electromechanical, system) [6].

1.2 Motivations

In order to study social robots, we come up with a research platform. For social robots to assist humans in their daily life effectively, the capability for adequate interaction with human operators is a key feature. Gestures are expressed by the movement of torso and limbs. Facial expressions result from motions or positions of facial features. These facial features are the organs of vision, auditory, speaking, and olfactory. In most cases, it is sufficient for us to understand and get acquired with each by the senses of vision, auditory, speaking and olfactory. Social robots should process similar human characteristics to a certain degree like what we do in the context of human communication. That is to say, social robots should be able to sense as we can. Moreover, human infants seem to have a preference for faces, and it appears that even newborns possess an 'innate' ability to spot basic facial features, not to mention the adults. Hence, we select the robot head (face) as our primary research platform of social robots at the present stage. Now many research projects are also focusing on the development of social robot heads worldwide.

1.3 Objective

The primary goal of this project is to develop a complex robot head Dorothy (meaning the Gift of God) that is capable to interact with humans through facial expressions and speech. Dorothy is not designed to perform tasks. Instead, she is designed to be a robotic creature that can interact physically, affectively, and socially with humans in order to ultimately learn from them. These skills help it to cope with a complex social environment, to tune its responses to the human, and to give the human social cues so that he/she is better able to tune him/herself to Dorothy. At the present stage, Dorothy is used predominantly for research. Utilizing Dorothy, who is endowed with personal qualities, we will have a better understanding of what features and dimensions of a robot head most dramatically contribute to people's perception of it sociability. In addition, we can make use of Dorothy to

- identify the latest multimedia technologies that are necessary for social interaction, such as face recognition, speech, facial displays, emotional expressions, knowledge of people's status and etiquette rules;
- integrate these multimedia technologies into a multimodal interface that can help us to enhance Human-Robot Interaction (HRI) from the social interaction perspective;
- and evaluate the user's acceptance of such an anthropomorphic interface in a specific context.

1.4 Dissertation Outline

This dissertation elaborates how Dorothy has been designed and built from scratch. The dissertation is structured as follows:

Firstly, an extensive research covering appearance and emotions of social robots has been carried out. As for the appearance of social robots, three important terms, anthropomorphism, "Uncanny Valley" and embodiment are discussed in details. It also focuses on critical issues in human-robot interaction area - emotions and facial expressions.

Moreover, it reviews the representative social robot heads to date, from which we can get inspiration for Dorothy. All these propose a research framework to study human aspects of robotic system design.

After a comprehensive overview of social robotics, it introduces the design approach that guides the entire design process of Dorothy all the time, including Dorothy's appearance, personality and capabilities.

Thirdly, it describes the mechanical design of the head and the accompanying neck joint. Design details of eyebrow, eyelid, eyeball, nose, mouth are given. It also describes frame construction as well as skin fabrication process. The multi-perspective simulations of Dorothy are presented in the last place.

Next, it covers all of controlling Dorothy: actuators, microcontroller, power system and programming. In terms of hardware, it includes fundamentals, selection criteria and mounting technique of servo motors as well as the microcontroller used in Dorothy - Lynmotion SSC-32, which is a very popular and powerful controller suitable for robot control using RC servo motors. For both actuators and microcontroller, power system is vital because not only it gives the motive power to devices but also most practical problems are caused by power issues. On top of that, software is equally important in contributing to the capability of human-robot interaction of Dorothy. It elaborates the algorithm for controlling Dorothy. Three modules that control three functionalities are expounded in detail. Eventually, controlling Dorothy is integrated into a user-friendly graphic interface.

Based on parameters obtained from human-robot interaction tests, we are able to evaluate the appearance of Dorothy as well as its performance. The current development and future prospects of research on social robotic heads are discussed. Lastly, summary for the whole thesis and the main opportunities for Dorothy in the future are given.

Chapter 2 – Related Work

2.1 Appearance of Social Robots

The appearance of robots has a substantial influence on the assumptions people have about specific applications and behaviors. Therefore, the appearance of the robot should match the expectations a user has, or the designer of robots should guarantee that the form of a robot matches its functions. In this context DiSalvo [7] suggests to consider a) an amount of robot-ness to emphasize the robot machine capabilities and to avoid false expectations, b) an amount of human-ness such that the subjects feel comfortable, and c) a certain amount of product-ness such that the robot is also seen as an appliance.

The design of a robot's head is an important issue within human-robot interaction (HRI) because it has been shown that the most non-verbal cues are mediated through the face. Especially for the head design there is an ongoing discussion if it should look like a human head or if a more technical optimized head construction should be developed. The advantage of latter is that there is no restriction according to the design parameters like head size or shape. This fact reduces the effort for mechanical construction. On the other hand, if realistic facial expressions should be used to support communication between a robot and a person, human likeness could increase the performance of the system as humans are more inclined to interact with their fellows. The physiognomy of a robot changes the perception of its human-likeness, knowledge, and sociability. Therefore, people avoid negatively behaving or looking robots and prefer to interact with positive robots. Furthermore, an expressive face indicating attention and imitating the face of a user makes a robot more compelling to interact with.

2.1.1 Classification

Fong et al. [5] distinguishes between four broad categories of the robot's aesthetic form: anthropomorphic, zoomorphic, caricatured, and functional. An anthropomorphic appearance is recommended to support a meaningful interaction with users, because many aspects of nonverbal communication are only understandable in similarity to a human-like body. There are three degrees in anthropomorphism: humanoid, android and human-likeness. Robots with an anthropomorphic appearance possess high degree of human-likeness; this property entitles them to be social robots. Robots with a zoomorphic appearance are intended to behave like their animal counterparts. Zoomorphic is to soothe the fear of humanlike-ness and in most cases, they are created for entertaining purpose. Robots with a caricatured appearance are used to focus on very specific attributes. Many Caricatured robots are in virtual forms instead of embodied agents because it is more expressive to convey in books or movies. Finally, functional robots are designed in a technical/functional manner to illustrate their ultimate functions. Functional robots, in most case, we would rather call them machines, are in the corresponding mechanical forms in order to maximum its functionality.

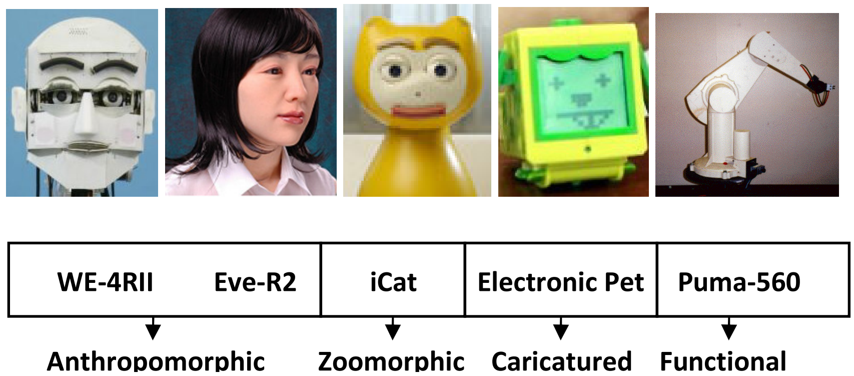


Figure 1 Four Different Robot Aesthetic Form

There is another approach to classify robots' appearance. McCloud [8] proposed a triangle (Fig 2) that illustrates the three categories and their relationship in an illustrative "map" of anthropomorphism that applies to robotic heads to date. The three sides of the triangle (realism/objective, iconic and abstract) embrace the primary categorizations for robots employing anthropomorphism to some degree. Most are 'real-life' robots although several fictional robots have been included. Functionality has no bearing on the classification in this context.

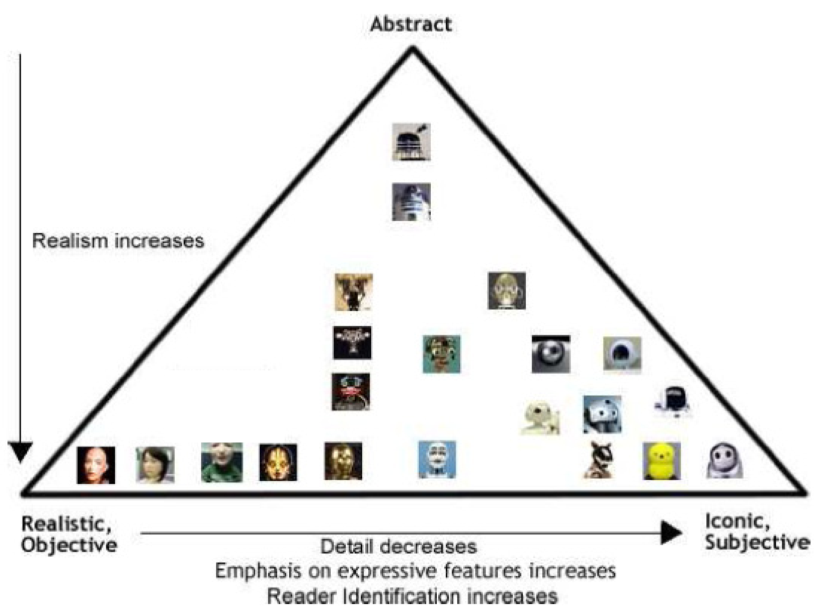


Figure 2 Classified Robot Appearance Triangle [8]

The apex of the triangle is a robot with abstract appearance. The “Abstract” corner refers to more mechanistic functional design of the robot with minimal human-like aesthetics. The left corner at the bottom is realistic (objective) while the right corner is iconic (subjective). “Human” correlates to an as-close-as-possible proximity in design to the human head. “Iconic” seeks to employ a very minimum set of features as often found in comics that still succeed in being expressive. From top to bottom, realism decreases. From left to right, the trend is from objective to subjective. Based on their realism and objectivity, each robot can be located in a specific point in the triangle. This triangle is very useful for human factors study of robotics as well as determining the appearance of robots before building them.

Another analysis of robots’ appearance focuses on the trend from machine to human [9]. This classification is based on the definition of mechanoid and humanoid adopted by Gong and Nass [10] and Android from Mac-Dorman and Ishiguro [11].

Mechanoid is a robot that is relatively machine-like in appearance and has no overtly human-like features. Humanoid is not realistically human-like in appearance and readily perceived as a robot by human interactants. However, it will possess some human-like features, which are usually stylized, simplified or cartoon-like versions of the human equivalents, including some or all of the following: a head, facial features, eyes, ears, eyebrows, arms, hands, legs. Android exhibits appearance (and behavior) which is as close to a real human appearance as technically possible.

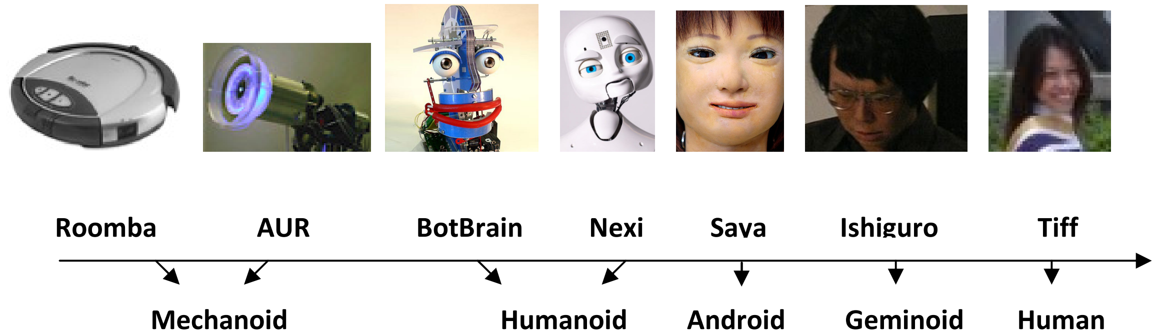


Figure 3 - Appearance of Robots: From Mechanoid to Human

Based on the experiments done by Michael L. et al [9], it concludes that a robot behavior or feature is rated by humans as less liked or approved of than a robot's overall appearance might suggest, there will inevitably be a degree of disappointment. Most participants preferred the humanoid robot appearance overall, except for a few individuals who favored a robot with a mechanical appearance. It also implies that differences in robot appearance lead to marked differences in perceived robot personality.

2.1.2 Anthropomorphism

There is a very important term in social robots – anthropomorphism, which comes from the Greek word anthropos meaning man, and morphe meaning form/structure. Our natural tendency to anthropomorphism, grounded in Theory of Mind and related psychological mechanisms, is crucial to our interactions with robots. Physical appearance of robots can trigger animistic, even empathetic, responses on the part of human beings. Other factors are more subtle, e.g. various aspects of the language used by the artifice, and of the thinking-processes apparently going on. Robotics promises to alter how people think about themselves. Unlike AI programs, robots are physical entities moving around in a physical world. This makes them more humanlike in various ways [12].

Anthropomorphism is the tendency to attribute human characteristics to inanimate objects, animals and others with a view to helping us rationalize their actions [13]. It entails attributing humanlike properties characteristics, or mental states to real or imagined nonhuman agents and objects. According to the Three-Factor-Theory of Anthropomorphism by Epley et al. [14] the extent to which people anthropomorphize is determined by three factors as below.

- **Elicited Agent Knowledge**: Knowledge about humans in general or self knowledge serve as a basis for induction primarily because such knowledge is acquired earlier and is more richly detailed than knowledge about nonhuman agents or objects.
- **Effectance Motivation**: Effectance describes the need to interact effectively with one´s environment. Sociality Motivation describes the need and desire to establish social connections with other humans.
- **Sociality Motivation**: refers to the attribution of a human form, human characteristics, or human behavior to nonhuman things such as robots, computer, and animals.

Duffy [13] argues a robot has to have a certain degree of anthropomorphic attributes for meaningful social interaction. Humans are experts in social interaction. Thus, if technology adheres to human social expectations, people will find the interaction enjoyable, feeling empowered and competent. Many researchers, therefore, explore the design space of anthropomorphic (or zoomorphic) robots, trying to endow their creations with characteristics of intentional agents. For this reason, more and more robots are being equipped with faces, speech recognition, lip-reading skills, and other features and capacities that make robot– human interaction “human-like” or at least “creature like”.

2.1.3 The Uncanny Valley

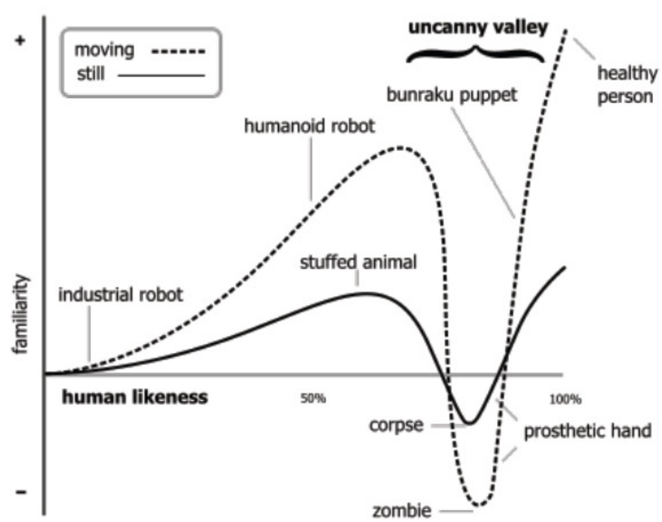


Figure 4 Uncanny Valley [15]

Proposed by roboticist Masahiro Mori [15] in 1970, the uncanny valley is a hypothesis regarding a robot's lifelikeness in the field of robotics. The theory holds that when robots and other facsimiles of humans look and act almost like actual humans, it causes a response of revulsion among human observers. Mori's hypothesis states that as a robot is made more humanlike in its appearance and motion, the emotional response from a human being to the robot will become increasingly positive and empathic, until a point is reached beyond which the response quickly becomes that of strong revulsion. However, as the appearance and motion continue to become increasingly human-like, the emotional response becomes positive once more and approaches human-to-human empathy levels. This area of repulsive response aroused by a robot with appearance and motion between a "barely human" and "fully human" entity is called the uncanny valley. The name captures the idea that a robot which is "almost human" will seem overly "strange" to a human being and thus will fail to evoke the empathic response required for productive human-robot interaction. Hypothesized emotional response of human subjects is plotted against anthropomorphism of a robot, following Mori's statements [Fig 4]. The uncanny valley is the region of negative emotional response towards robots that seem "almost human". Movement of the robot amplifies the emotional response (dotted curve in Fig 4).

2.1.4 Embodiments

Embodiment is another important term in social robotics. The widely accepted meaning of embodiment in the fields of Artificial Intelligence (AI) and Robotics is physical instantiation, or more simply, bodily presence [16]. A physically embodied robot, thus, should have both an actual physical shape and embedded sensors and motors. Investigating the effects of physical embodiment of social robots in human-robot interaction is essential in design social robots because it questions whether or not physical embodiment is required for the successful social interaction between human and social robots. There are authors that believe that a social robot is not required to exist within a physical body, others agree in setting both restrictions to the requirements on the social robot [17].

In this context, there are two types of robots: physically embodied and disembodied. The embodied robots are tangible while the disembodied robots are virtual agents. K.M. Lee et al [16] from Communication University of Southern California did two experiments to investigate the importance and effect of embodiment in social robotics. Experiment 1 is to learn the effects of physical embodiments of social robots. Two conclusions are drawn from this experiment. One is that people evaluate a physically embodied social agent more positively than a disembodied social agent. The other one is that physical embodiment yields a greater sense of social presence in human-agent interaction. Experiment 2 aims to study the significance of physical embodiment. The conclusions are physical embodiment with no possibilities of tactile interaction decreases an agent's social presence and social agents are more socially attractive to lonely people. Especially, experiment 2 helped to make a solid conclusion about the effects of touch input capability in human-robot interaction by separating two nesting component of physical embodiment: (1) visual; (2) touch.

Breazeal, C. [18] believes that The embodied systems have the advantage of sending para-linguistic communication signals to a person, such as gesture, facial expression, intonation, gaze direction, or body posture. These embodied and expressive cues can be used to complement or enhance the agent's message.

2.2 Emotions and Facial Expressions of Social Robots

2.2.1 Human Emotions and Facial Expressions

Emotion plays a crucial role in the cognition of human beings and other life forms, and is therefore a legitimate inspiration for providing situated agents with adaptability and autonomy [19]. However, there is no unified theory of emotion and many discoveries are yet to be made in its applicability to situated agents. One function of emotion commonly identified by psychologists is to signal to other cognitive processes that the current situation requires an adaptation. The human face is a very complex system, with more than 44 muscles whose activation can be combined in non-trivial ways to produce thousands of different facial expressions [20]. Several theorists argue that a few select emotions are basic or primary—they are endowed by evolution because of their proven ability to facilitate adaptive responses to the vast array of demands and opportunities a creature faces in its daily life. The emotions of joy, sadness, surprise, anger, fear are often supported as being basic from evolutionary, developmental, and cross-cultural studies [21]. Each basic emotion is posited to serve a particular function (often biological or social), arising in particular contexts, to prepare and motivate a creature to respond in adaptive ways. They serve as important reinforces for learning new behavior. In addition, emotions are refined and new emotions are acquired throughout emotional development. Social experience is believed to play an important role in this process [18] [21]. Besides basic emotions, the rest can be considered as advanced emotions. Advanced emotions comprise basic emotions. Table 1 shows the description of basic emotions and their corresponding trigger factors.

Expressions of emotion are used to transfer the effectiveness message when it occurs in a social context or in a human-to-human communicating assuming that the facial expressions are communicative signals to transfer mostly psychological message in human-to-human communication. The study of facial expressions is broadly interested in different disciplines and strongly associated with human body kinetics. We used the following biological observations to make artificial models of human visual action and facial expressions [22]. Table 2 describes the facial expressions of basic emotions.

Emotions	**Trigger Factor**
Happiness	Happiness is the only emotion which is always a positive one. One can feel happy in anticipation of an event, while experiencing a pleasant moment; the relief of pain or of fright may make one feel happy; it can also arise because one is content.
Sadness	Sadness is the emotion that generally lasts the longest. It is a passive feeling. A sad person does not suffer physical pain but disappointment, loss of something important.
Surprise	Surprise is the briefest emotion. It is a reaction to a sudden, unexpected event. It lasts until one has evaluated the event. It should be differentiated from startle.
Anger	Anger can be aroused from frustration, physical threat, or when psychologically hurt or violated morally.

Fear	Expressions of fear and of surprise are very similar. Fear arises from persons, objects, situations, real or imaginative, that seem dangerous.

Table 1 - Basic Emotions and Corresponding Trigger Factors [23]

Emotions	Facial Expressions
Happiness	Corners of lips are drawn back and up, the mouth may or may not be parted, with teeth exposed or not, a wrinkle runs down from the nose to the outer edge beyond the lip corners, the cheeks are raised, the lower eyelid shows wrinkles below it
Sadness	The inner corners of the eyebrows are drawn up, the skin below eyebrow is triangulated, with the inner corner up, the upper eyelid inner corner is raised, the corners of the lips are down or lip is trembling
Surprise	The brows are raised, so that they are curved and high, the skin below the brow is stretched, horizontal wrinkles go across the forehead, the eyelids are opened, the jaw drops open so that the lips and teeth are parted, but there is no tension or stretching of the month
Anger	Vertical lines appear between the brows, the lower lid is tensed and may or may not be raised, the upper lid is tense and may or may not be lowered by the action of the brow, the eyes have a hard stare and may have a bulging appearance, the lips are pressed firmly together, with the corners straight or down.
Fear	The brows are raised and drawn together, the wrinkles in the forehead are in the center, not across the entire forehead, the upper eyelid is raised, exposing sclera, and the lower eyelid is tensed and drawn up, the mouth is open and the lips are either tensed slightly and drawn back or stretched and drawn back

Table 2 Basic Emotions and Corresponding Facial Expressions

2.2.2 Facial Action Coding System

More research has already been conducted in the area of non-verbal communication between a robot and a human that include facial expressions that focus on the communication task. Researchers have been fascinated by various facial expressions that social robots can achieve.

Proposed by EKman and Fresen [24] in 1978, Facial Action Coding System (FACS) is a system to analyze humans' facial expressions and movements. Ekman's Facial Action Coding Systems (FACS) can be used to determine the control points of a face so that most robots faces express emotion in accordance with Ekman and Frieser's FACS system. It defines 46 action units - a contraction or relaxation of one or more muscles. It is a common standard to systematically categorize the physical expressions of emotions and it has proven useful to psychologists and animators. The human face is a very complex system, with more than 44 muscles whose activation can be combined in non-trivial ways to produce thousands of different facial expressions. One concept of non-verbal interaction is mainly based on FACS, which consequently describes the motions of the skin, eyes, and neck. The results of FACS are extended with information concerning body pose and the influence on man-machine communication. The possibility to express emotions is therefore mainly based on the availability of the so called action units which have to be combined to express a specific emotional state. Fig 5 lists common action units for facial expressions while Fig 6 shows the action units defined in CrazyTalk® [25].

AU No.	Appearance Changes
1	Inner Brow Raiser
2	Outer Brow Raiser
4	Brow Lowerer
5	Upper Lid Raiser
6	Cheek Raiser & Lid Compressor
7	Lid Tightener
9	Nose Wrinkler
10	Upper Lid Raiser
12	Lip Corner Puller
15	Lip Corner Depressor
17	Chin Raiser
20	Lip Stretcher
25	Lips Part
26	Jaw Drop

Fig. 2: A close-up of actions units defined in FACS.

Figure 5 Examples of Facial Actions [24]

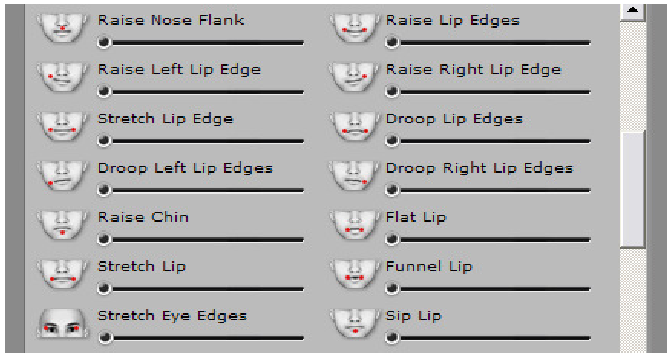


Figure 6 Examples of Action Units in Crazy Talk® [25]

2.3 Successful Social Robot Heads

There are mainly two types of robot heads: Mechanical Model and Optical Model, alternatively virtual / disembodied Agents and Physically Embodied Agents. Some robot heads are in the middle of virtual and physical agents. The robot heads in the pictures below are the models studied before building our robot. For virtual faces, most of them are available for sale in the market [26].

Figure 7 - Selected Virtual Faces

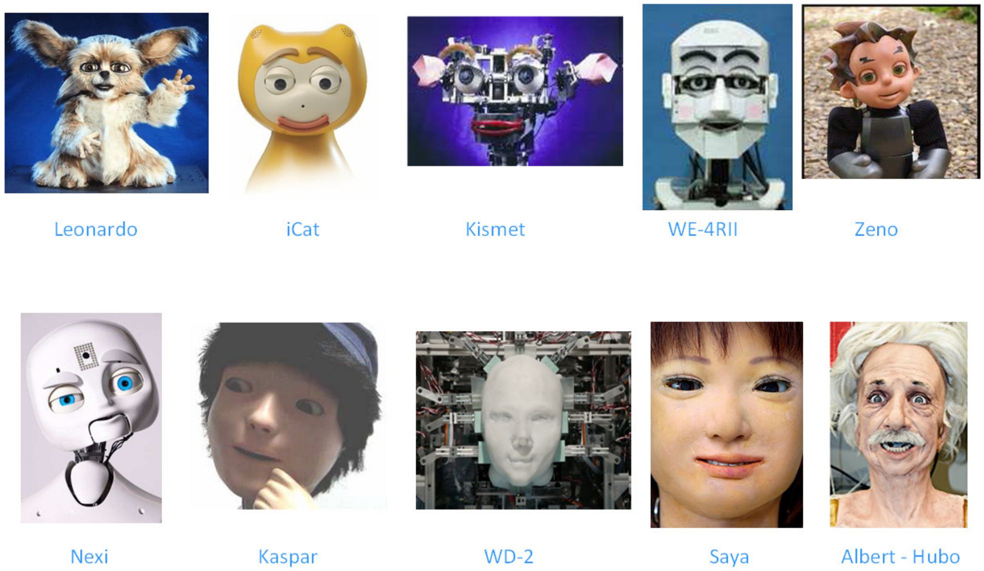


Figure 8 - Selected Physically Embodied Faces

Based on the study in Chapter 2.2, our survey focus are physically embodied faces. 7 robot heads are selected as our main study targets. They are lab platforms, prototypes and first constructive attempts of social robots and commercial robots with certain interaction capacity.

2.3.1 iCat

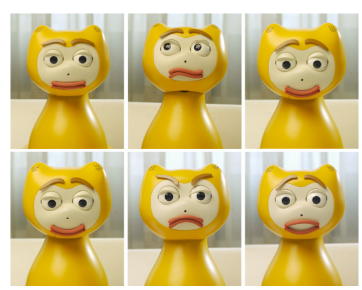


Figure 9 iCat

The iCat Research Platform [27] [28], developed by Philips Research (Eindhoven, the Netherlands), has been one of the most successful commercialized robot heads so far. It has more natural appearance, high computational power, friendly interfaces and advanced on-board communication devices. It is a research platform for studying human-robot interaction. iCat is a plug & play desktop user-interface robot that is capable of mechanically rendering facial expressions, in other words, without an onboard processor which is controlled by a PC (laptop or desktop) via a USB cable. This capability makes the robot ideal for studying human-robot interaction. The robot has been made available to stimulate research in this area further and in particular to stimulate research topics such as social robotics, human-robot collaboration, joint-attention, gaming, and ambient intelligence.

2.3.2 Kismet

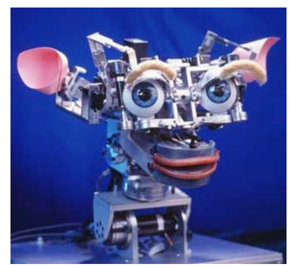


Figure 10 Kismet

Kismet [21] [29] [30] has undoubtedly been the most influential social robot. It is an anthropomorphic robotic head with facial expressions. Developed in the context of the Social Machines Project at MIT, it can engage people in natural an expressive face-to-face interaction. Kismet is an expressive anthropomorphic robot that engages people in natural and expressive face-to-face interaction. The robot has been designed to support several social cues and skills that could ultimately play an important role in socially situated learning with a human instructor. Kismet adopts six basic emotions: anger, disgust, fear, joy, sorrow and surprise. They are often supported as being basic from evolutionary, developmental and cross-cultural studies. Kismets' facial movements are created through movements of the ears, eyebrows, eyelids, lips, jaw, and head. This robotic head has 15 DOFs in total.

2.3.3 WE-4RII

Figure 11 WE-4RII

WE-4RII [31] [32] is the abbreviation for Waseda Eye No.4 Refined II. It is the latest one in WE-series Emotion Expression Humanoid Robots developed since 1995. Part of this research was conducted at the Humanoid Robotics Institute (HRI), Waseda University. And part of this was supported by a Grant-in-Aid for the WABOT-HOUSE Project by Gifu Prefecture. WE-4RII has 59 DOFs and a lot of sensors that serve as sense organs (Visual, Auditory, Cutaneous and Olfactory sensation) for extrinsic stimuli. WE-4RII uses the Six Basic Facial Expressions of Ekman in the robot's facial control, and has defined the seven facial patterns of "Happiness", "Anger", "Disgust","Fear", "Sadness", "Surprise", and "Neutral" emotional expressions. The strength of each emotional expression is variable by a fifth-grade proportional interpolation of the differences in location from the "Neutral" emotional expression.

2.3.4 Zeno

Figure 12 Zeno

Zeno [33] is the first of his kind. It’s a member of RoboKind™–cute, animated characters brought to life through Hanson Robotics ’breakthrough technology. It is able to see, hear, talk, remember and even walk and perform amazing stunts. Its face is so soft like human that it can show emotions, just like you - happy, sad, puzzled, and lots more. It operates independently and can act even a few of antics. Once linking to a PC wirelessly, it can have complete conversations with human. Plus, if it is connected to internet through a PC, it can keep learning and growing smarter.

2.3.5 Nexi

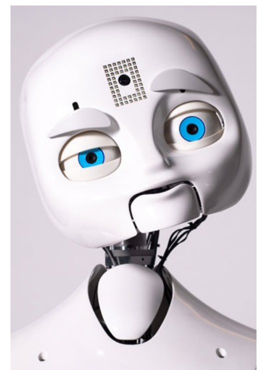


Figure 13 Nexi

MDs [34] is created by the personal robots group of MIT Media Lab. The expressive head and face are designed by Xitome Design with MIT. The neck mechanism has 4 DoFs to support a lower bending at the base of the neck as well as pan-tilt-yaw of the head. The head can move at human-like speeds to support human head gestures such as nodding, shaking, and orienting. The 15 DOF face has several facial features to support a diverse range of facial expressions including gaze, eyebrows, eyelids and an articulate mandible for expressive posturing. Perceptual inputs include a color CCD camera in each eye, an indoor Active 3D IR camera in the head, four microphones to support sound localization, a wearable microphone for speech. A speaker supports speech synthesis.

2.3.6 Kaspar

Figure 14 Kaspar

KASPAR [35] [36] is a child-sized humanoid robot developed by the Adaptive Systems Research Group at the University of Hertfordshire. KASPAR has 8 degrees of freedom in the head and neck and 6 in the arms and hands. The face is a silicone mask, which is supported on an aluminum frame. It has 2 DOF eyes fitted with video cameras, and a mouth capable of opening and smiling. Similar to Zeno, Karspar gives a good example of the child-size robot with soft skin. But Kaspar looks like a real child much more because of its size and silicone mask, which provides a good guidance about the material I may use for my robotic face surface.

2.3.7 Einstein

Albert-Hubo [37] is a humanoid robot based on HUBO, but with Einstein's face on top of it. The robot head, "Einstein", was developed by Hanson Robotics [38]; a company specialized in making robot faces. Its skin is a special material that is often used at Hollywood – Frubber. It deforms in a skin-like manner contributing to the realism of the robot expressions. The head is actuated by 31 servo motors, 27 of them controlling the expressions of the face and 4 controlling the neck. While the robot is able to simulate the actions of all major muscle groups in the face and neck, there are some important differences in the way the human muscles and the robot servo motors actuate the face. In contrast to human muscles, these servos can both pull and push loads and thus each motor can potentially simulate the action of 2 individually controlled muscle groups [20].

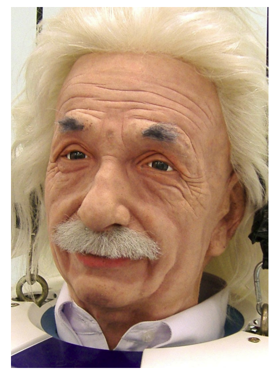


Figure 15 Einstein (Robot)

2.3.8 Summary of Social Robotic Head

To sum up, most of social robots look like human to certain degree. The guideline of social robots' appearances, capabilities and performance varies mainly as the purpose of the research or applications. Table 13 summarizes the appearance and capabilities of social robots mentioned above. From the table, we can see that most of social robots possess anthropomorphic property and equipped by the ability of human – it can listen, speak, see, and show emotion via facial expressions and gestures.

Name	**Appearance**	**Skin**	**Body**	**Abilities**
Leonard	Zoomorphic	Fur	√	Facial recognition, visual tracking, facial
iCat	Anthropomorphic	Hard	√	Recognizing objects & faces, facial expressions, listening, speaking
Kismet	Anthropomorphic	No	x	Vision, auditory, vocalization, facial expressions
WE-4RII	Anthropomorphic	Hard	√	Visual, auditory, tactile, olfactory, facial
Zeno	Anthropomorphic	Soft	√	Vision,auditory,speaking,face remember, facial expressions, gestures
Nexi	Anthropomorphic	Hard	√	Vision, auditory, vocalization,
Kaspar	Anthropomorphic	Silicon	√	Vision, auditory, vocalization,
Saya	Anthropomorphic	Soft	√	facial expressions, gestures
WD-2	Anthropomorphic	Septo	x	facial expressions
Einstein	Anthropomorphic	Frubbe r	√	Vision, auditory, vocalization, facial expressions, gestures

Table 3 Social Robot Summary - Appearance & Abilities

Chapter 3 – Our Design Approach

Based on the study of social robotics and successful social robotic heads, we come up with an approach to designing Dorothy's appearance and defining its emotions and facial expressions.

3.1 Dorothy's Appearance

Faces help humans to communicate, regulate interaction, display (or betray) our emotions, elicit protective instincts, attract others, and give clues about our health. Several studies have been carried out into the attractiveness of human faces, suggesting that symmetry, youthfulness and skin condition are all factors. The primary design concern is the appearance of our robot face, e.g. whether it has physical facial features or virtual parts and what Dorothy should look like, human, animal or other figure.

Based on the robot head survey and analysis in Chapter 2, the appearance of Dorothy should process high anthropomorphism in accordance with sociability but at the same time, without falling into uncanny valley. An anthropomorphic robotic head should have similar features with a human head. This study showed that the presence of certain features, the dimensions of the head, and the number of facial features greatly influence the perception of humanness in robot heads. Some robots are much more successful in the portrayal of humanness than others. This success is due, at least in part, to the design of the robot's head. From these findings we have created an initial set of guidelines for the design of humanoid robot heads. Specifically, we have identified features and dimensions that can be used to modulate how humanlike a robot head will be perceived. These findings should serve as a connection between ongoing robot research and the social robot products of the future. Hence, Dorothy has the same facial features as human but different facial dimensions. DiSalvo [7] proposed the following guidelines for this kind of robot head.

- **Wide head, wide eyes**: To retain a certain amount of robot-ness, by making the robot look less human, the head should be slightly wider than it is tall and the eye space should be slightly wider than the diameter of the eye.
- **Features that dominate the face**: The feature set, from browline to bottom of mouth, should dominate the face. Proportionally, less space should be given to forehead, hair, jaw or chin. This distribution is in contrast to a human s and combined with the size of the head, will clearly state the form of the head as being robot-like.
- **Complexity and detail in the eyes**: Human eyes are complex and intricate objects. To project humanness a robot must have eyes, and the eyes should include some complexity in surface detail, shape of the eye, eyeball, iris, and pupil.
- **Four or more features**: The findings from our study show that the presence of a nose, a mouth, and eyebrows, greatly contribute to the perception of humanness. To project a high level of humanness in a robot these features should be included on the head.

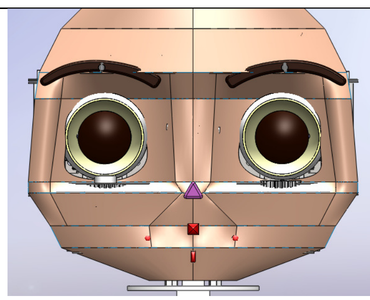
- **Skin**: For a robot to appear as a consumer product it must appear finished. As skin, or some form of casing is necessary to achieve this sense of finish. The head should include a skin or covering of mechanical substructure and electrical components. The skin may be made of soft or hard materials.
- **Humanistic form language**: The stylized appearance of any product form is important in directing our interaction with it. To support the goal of a humanoid robot the head shape should be organic in form with complex curves in the forehead, back head and cheek areas.

Robert D. Green [40] also proposed the best face proportions for social robots based on experiments. The result shows that narrower-set eyes are preferred in human features while wider-set eyes in robotics features.

Based on these design guidelines, Dorothy, correspondingly, has a wide head with facial features dominate the face, especially the eyes. Its facial features include eyebrow, eyeballs (including eyelid and eyelash), nose and mouth, completely the same as human being. Dorothy also has elastic skin that will definitely enhance its facial looks and expressions. The skin is modeled in accordance with the concave-convex of human being as well. Ironically, Dorothy has an enchanting and lovely face like a doll below.

Figure 16 - Appearance Reference of Dorothy [41]

Figure 17 is the 2D simulation of Dorothy including the measurements of its face: the percentage of the forehead region, feature region, and chin region, the size of the eyes, the distance between the eyes, and the width of the mouth.



Dimension & Weight	
Height	130 mm
Width	140 mm
Depth	80 mm
Circumference	540 mm
Weight	500 g

Figure 17 Dimension of Dorothy

3.2 Dorothy's Emotions and Facial Expressions

Most robot faces express emotion in accordance with FACS system. We will employ this system for our analysis as well. The head should be designed taking into account that the different head expressions must be easily recognized by a human being. It can express not only basic but also advanced emotions as listed below.

- Happiness
- Sadness
- Surprise
- Anger
- Anticipation
- Fear

Facial expressions for Dorothy are created by the intricate, coordinated movement of motors located eyebrows, eyelids, eyeballs, mouth and neck. In order to express the emotions mentioned above, Dorothy must be capable of the following facial feature movements.

- Eyebrow – Frown, Horizon
- Eyelid – Open , Close
- Eyeball – Turn Left, Right, Up, Down
- Mouth – Open and Close
- Neck – Pan, Tilt

In terms of the Action Units, Table 4 lists the AUs that are used by Dorothy. These AUS comprise the basic emotions.

No.	**Action Unit**		**No.**	**Action Unit**
Eyebrow			Eyelid	
1	Horizontal Left Brow		7	Left Eye Fully Open
2	Raise Left Inner Brow		8	Left Eye Half Open
3	Raise Left Outer Brow		9	Left Eye Fully Close
4	Horizontal Right Brow		10	Right Eye Fully Open
5	Raise Right Inner Brow		11	Right Eye Half Open
6	Raise Right Outer Brow		12	Right Eye Fully Close
Eyeball			Mouth	
13	Left Eye Staring Front		26	Raise Lip Corners Most
14	Left Eye Turn Left		27	Raise Lip Corners Mediate

15	Left Eye Turn Right		28	Linearize Lip Corners
16	Left Eye Turn Up		29	Droop Lip Corners Most
17	Left Eye Turn Down		30	Droop Lip Corners Mediate
18	Right Eye Staring Front		31	Linearize M Point
19	Right Eye Turn Left		32	Droop M Point Mediate
20	Right Eye Turn Right		33	Droop M Point Most
21	Right Eye Turn Up			
22	Right Eye Turn Down			

Table 4 Action Unit List Used by Dorothy

Table 5 is the summary of basic emotions and their corresponding action units. In line with the algorithm (elaborated in Chapter 5), Table 5 is split into two portions: non-talking face and talking face.

Index	Emotion	No Talking Face Description	Talking Face Description
1	Neutral	1+4+7+8+11+13+18+28+31	1+4+7+8+11+13+18+talking
2	Happines	1+4+7+10+13+18+25+26+31(33)	1+4+7+10+13+18+25 + talking
3	Sadness	2+5+8+11+13+18+25+29+31	2+5+8+11+13+18+25 + talking
4	Surprise	2+3+4+7+10+13+18+25+29+33	2+3+4+7+10+13+18+25 + talking
5	Anger	3+6+7+10+15+19+23+30+33	3+6+7+10+15+19+23 + talking

Table 5 Six Typical Facial Expressions Organized by AUs

Table 6 shows the viseme simulation of Dorothy's mouth shape in CrazyTalk and SolidWorks respectively as well as the action unit index for the corresponding shape.

The lip synching includes all the viseme needed for talking. More details provided in Chapter 5.

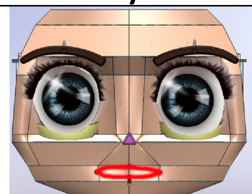
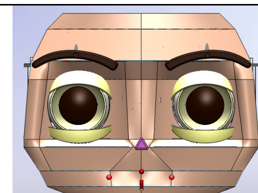
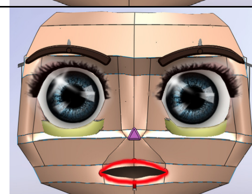
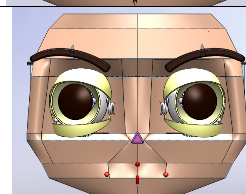
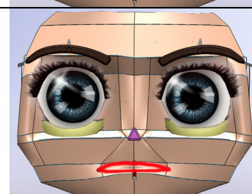
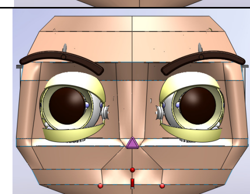
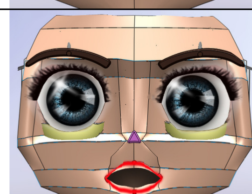
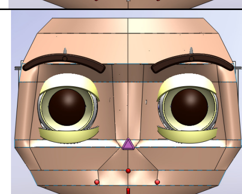
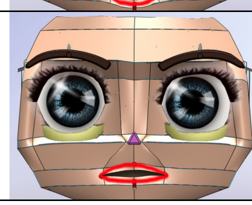
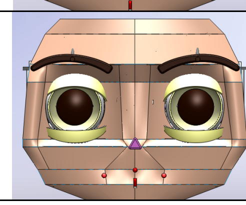
Lip Synching	CrazyTalk	Solidworks	Mouth Action
None			28+31
Ah / Ch_J / Er / K_G			30+33
B_M_P / F_V			29+31
Oh			29+32
EE / Ih/N_NG / R /S_Z / T_L_D / Th / W_OO			30+33

Table 6 Viseme (Mouth Shape)

In addition, to achieve vivid facial expressions, the motion of facial features is not enough. The expressive behavior of robotic faces is generally not life-like, which reflects limitations of mechatronic design and control. For example, transitions between expressions tend to be abrupt, occurring suddenly and rapidly, which rarely occurs in nature. To minimize the limitation of mechatronic system, elastic skin is employed. More details of skin is introduced the Section 4.6.

3.3 First Version Doris

Figure 18 Doris - First Version of Robot Head [42]

Prior to Dorothy, the first version of robot Doris was created. Doris is a robot head driven by 6 micro servo motors with 11 degrees of freedom that allows multiple facial expressions to be preformed. It is made of acrylic (skeleton) and silicon (skin). Doris is capable to show four basic emotions: happiness, sadness, surprise and anger. Currently, the robot has function of facial expressions.

Chapter 4 - Mechanical Construction of Dorothy

This section describes the mechanical hardware that constitutes Dorothy. Mechanism for generating facial expressions, e.g. mechanical designs of eyebrow, eyelid, eyeball and mouth, frame construction and skin fabrication are presented.

4.1 Mechanical Design of Facial Features

4.1.1 Eyebrow

The motion of Dorothy' eyebrow is inspired by humans'. The motion path of human eyebrows is very complicated. To simplify it, only basic emotions' eyebrow paths are considered. As a rule of thumb, we find that each motion of simple eyebrows can be broken down into two sub-paths. One is the motion in vertical direction and the other is along the curvature direction. Without deformation involved, we come out with three different designs of eyebrows.

"Eyebrow #1" [Fig 19] is the simplest one among these three. It directly utilizes the shaft output of motors and creates the motion along the curvature of eyebrows themselves. The single path makes it easy to fabricate and control.

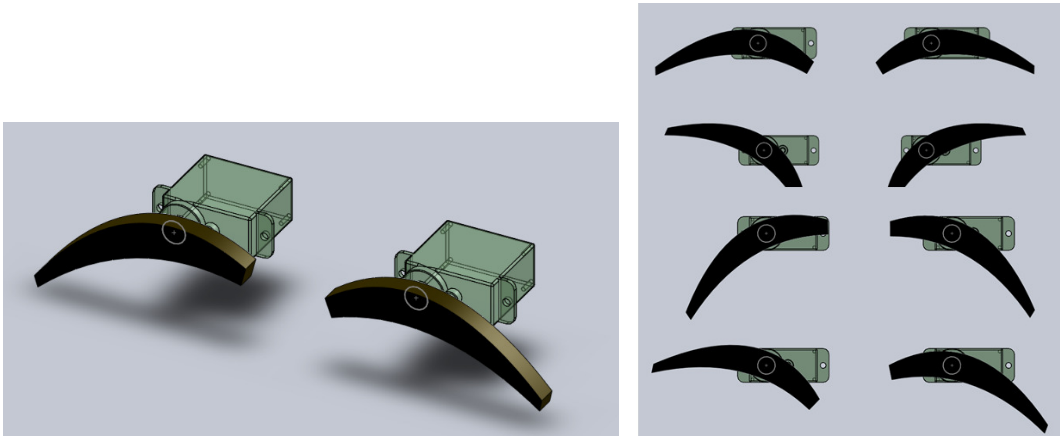


Figure 19 - Eyebrow #1 in SolidWorks (Left: Trimetric View; Right: Front View)

The second design "Eyebrow #2" is cam mechanism [43]. Using one actuator, the two eyebrows translate and rotate along the profile of the cam [Fig 20]. The outcome of the simulation is not good enough. Moreover, Eyebrow #2 needs much space beneath the eyebrows for the cam profile to rotate. How to keep the roller in the track is another problem that must be considered. Eyebrow #2 is therefore not used.

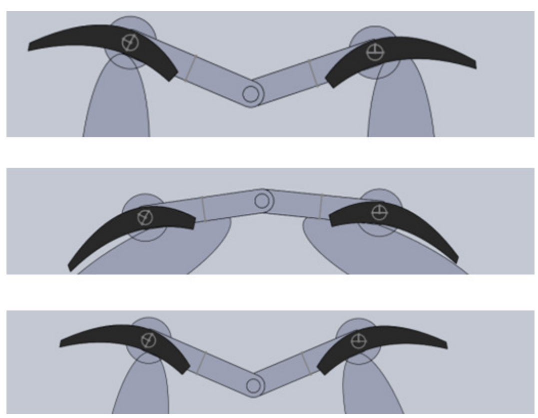


Figure 20 Eyebrow #2 in SolidWorks (Front View)

Eyebrow #3 is implemented by the four-bar mechanism, which is consists of a crank, a connecter, a rocker, and a frame. This mechanism results in more calculation in comparison of the previous two, but with the help of animation software, the calculation becomes precise and easy.

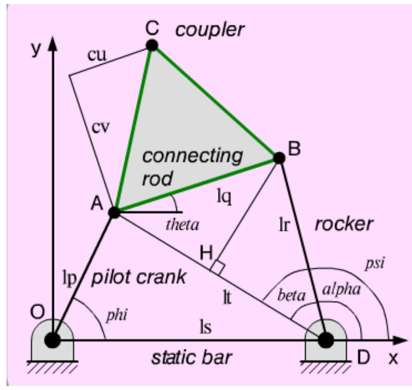


Figure 21 Standard Four-Bar Linkage

Each eyebrow system is actuated by one micro servo motor. The motion profile of eyebrows results from the length and ratio of four bars. In order to optimize it, we do a simulation in the software following the optimization criteria.

- Fixed parameter: static bar AD=40 mm, the fix length is selected based on the width of the whole face, which is 140 mm.
- Dimension range of crank, rocker and connecting rod: 20 ~ 30 mm
- Symmetric sway, e.g. crank length = rocker length
- Large sway angle, e.g. connecting rod is shorter than crank and rocker
- Not too much shift from original position

The software used for eyebrow four-bar linkage simulation is "Geometry Expressions v2.1 Demo" [44]. It allows users to assign different lengths of the four bars and will calculate the instantaneous angle automatically.

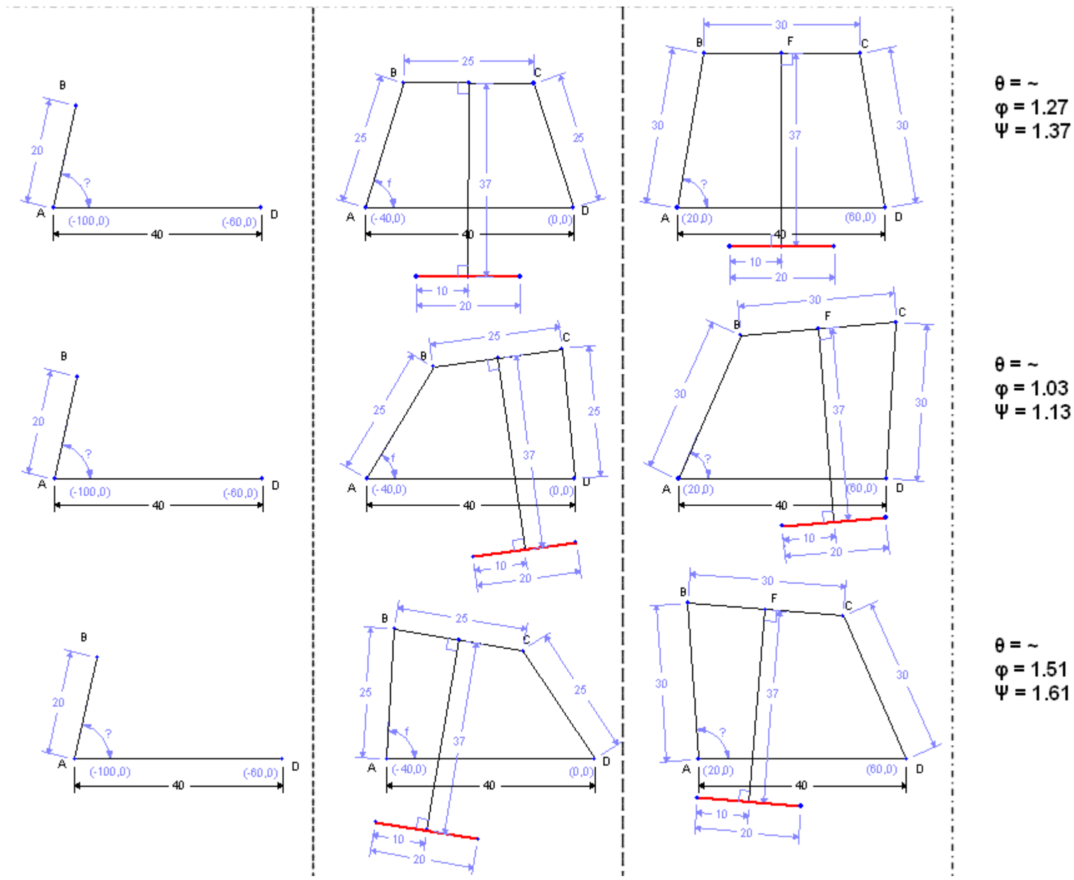


Figure 22 Simulation of Eyebrow Four-Bar Linkage Mechanism
Same length of rank, connecter and rocker

Fig 22 shows the simulation of same length of rank, connecter and rocker. When rank=connecter=rocker=20mm, the linkages cannot be completed. When they equal to 25mm, the motion of eyebrows are not obvious. It is even worse when the lengths of three linkages equal to 30 mm.

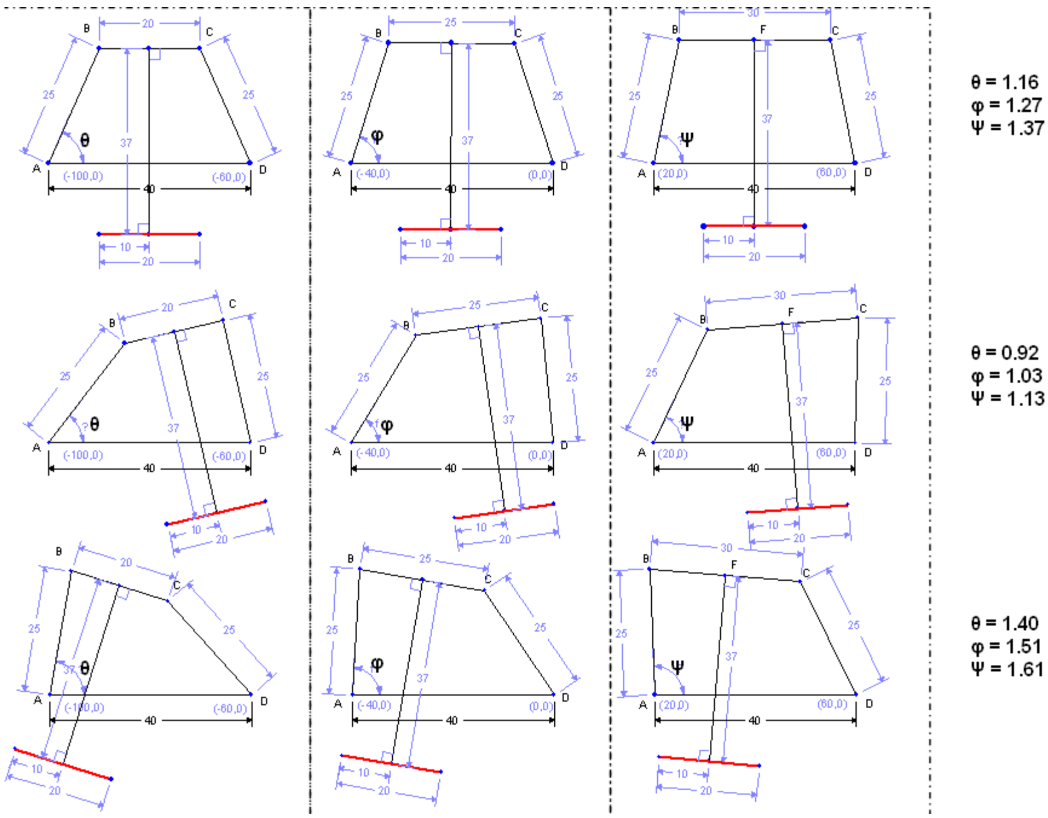


Figure 23 Simulation of Eyebrow Four Bar Linkage Mechanism
Same length of rank and rocker, different length of connector

In Fig 23, rocker and rocker have the same length of 25 mm for symmetric reason. The length of connector changes as 20mm, 25mm, 30mm. The first length gives the most obvious motion. Hence, the length selected is crank : connecter : rocker : frame = 25mm: 20mm: 20mm: 40mm. The swing angles are calculated in Fig 24.

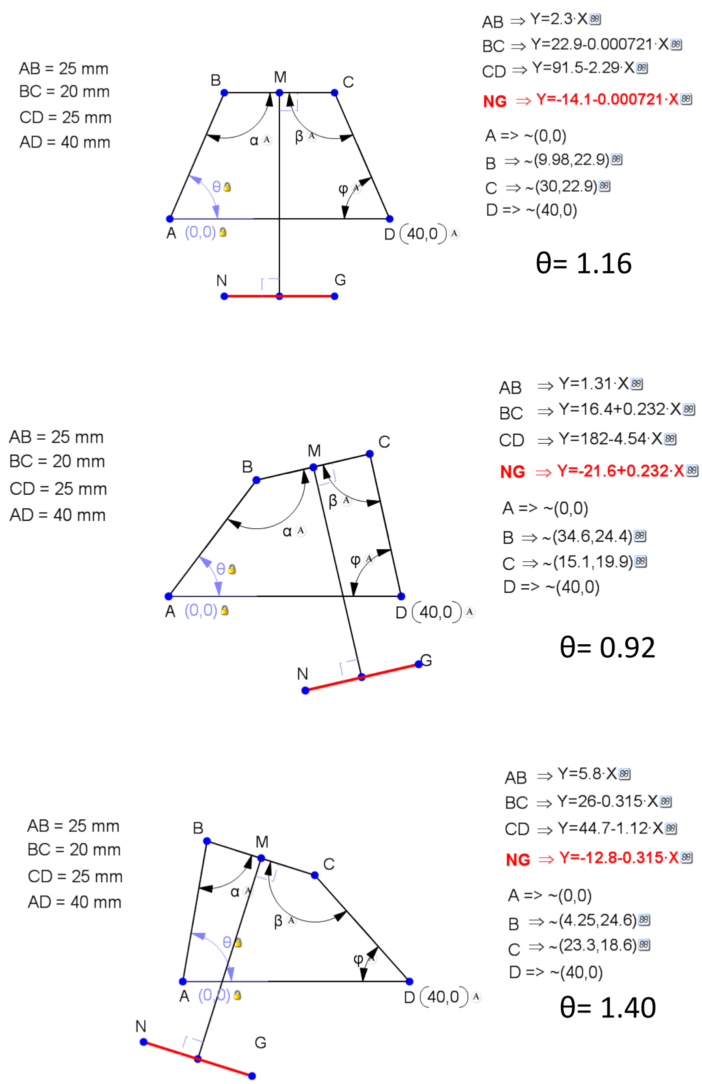


Figure 24 Simulation of Four-Bar Linkage
Crank : Connecter : Rocker : Frame = 25mm: 20mm: 20mm: 40mm.

In practical, the decisive factor is upon their performances. Based the analysis of each design above, expressions of Eyebrow 3 are superior to the other two. Hence, Eyebrow #3 was selected as the final design of eyebrows.

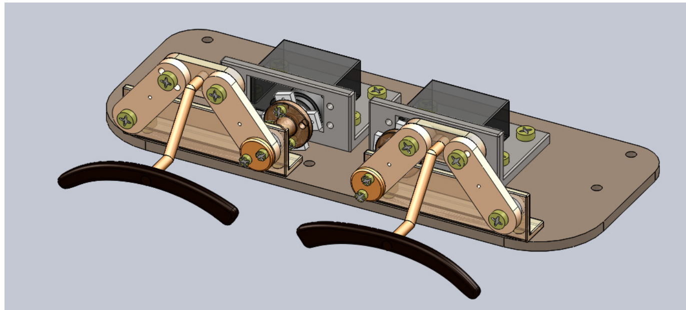


Figure 25 Eyebrow #3 in SolidWorks

4.1.2 Eyelid

Eyelids contribute a lot to the expressions of emotions. They enable the eyes to open and close at various degrees. One micro motor mounted at the back on eyeball plate, is used to drive eyelids. The core mechanism of eyelid is timer belt and gear [Fig 26]. With them, we can drive the eyelids using one motor only. The concern of eyelid is whether we should employ double eyelids or upper single eyelid only. Fig 27 is the illustration of double eyelids and single upper eyelid. How to determine is based on the complete facial expressions simulation.

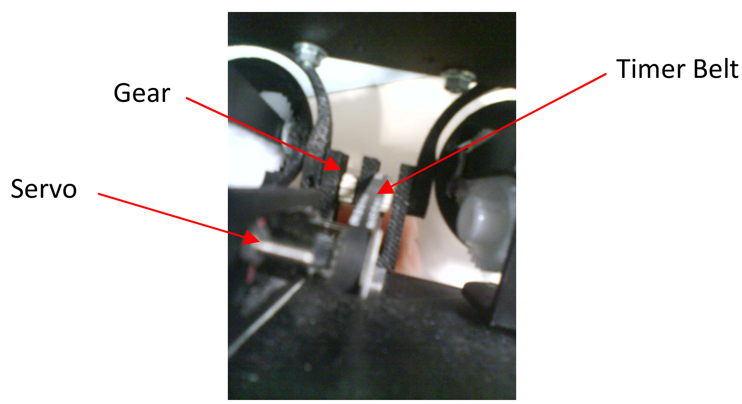


Figure 26 Mechanism of Eyelids of Dorothy

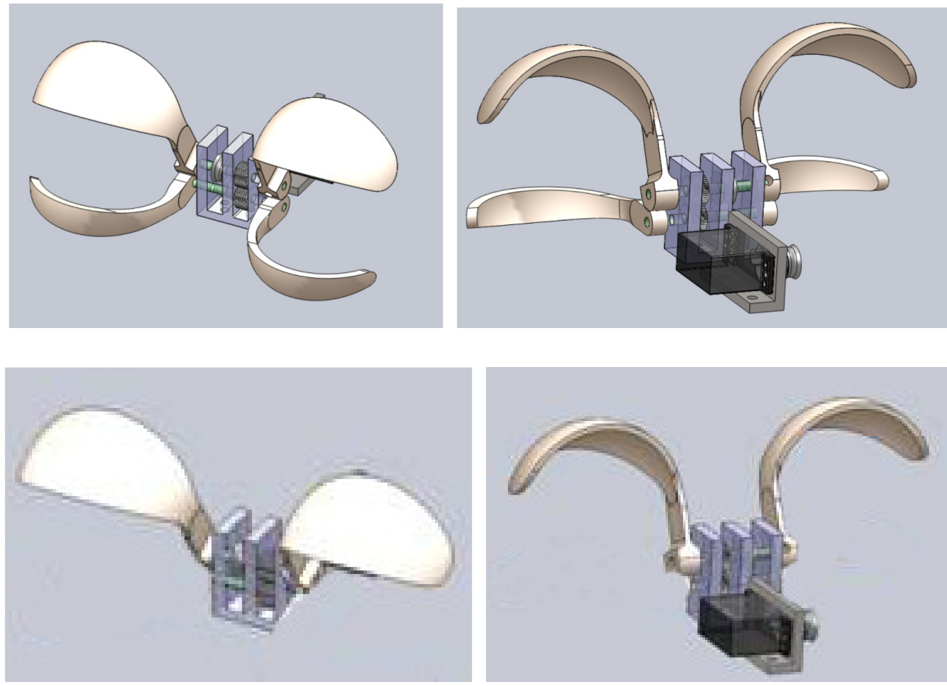


Figure 27 Simulation of Eyelid in SolidWorks (Double Eyelids & Upper Eyelid Only)

4.1.3 Eyeball

Eyeballs are designed to enhance the interaction with people based on human anthropomorphic data. Equipped with actuated eyes, social robots are able communicate with human via eye contact. The outcomes expected for eyes are to pan and tilt separately. In this context, four eyeball designs are proposed.

Eyeball #1 is a friction driven mechanism. Two eyeballs are actuated by frictional force due to the rolling of a cylinder that touches the eyeballs from behind as in Fig 28. When the roller move horizontally or rotate, the two eyeballs will undergo synchronized motions (Panning is achieved when the cylindrical bar move left and right, Tilt is achieved when the cylindrical bar rotates). It is the inverse mechanism of ball mouse. Eyeball #1 works tolerably well on the condition that its disadvantage can be overcome. There are many factors attributing to the imprecision of friction-driven, e.g. material of eyeballs and roller, contacting area, magnitude and direction of friction at contacting area, roller speed, and friction at supporting pillars. Eyeball #1 has been applied to Doris [42], three major problems include: (1) not enough fiction at contacting area, (2) the cylinder roller is not strictly straight, (3) eyeballs are not heavy enough for stable rotation. Solving these problems may create other problems. In short, due to the imprecise of the friction control, the movement of eyeballs is not accurate. We thereby not implement this design.

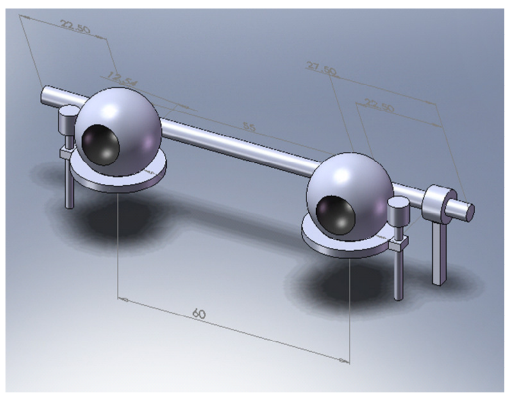


Figure 28 Eyeball #1 - Friction Driven Eyeballs

Eyeball #2 is wire-pulling mechanism, inspired by the design of Probo’s eyes [45] [Fig 29]. The five DOF eyes module consist of two hollow eyeballs supported in an orbit. This mechanism can controlled easily and precisely, however, it needs much more space at the back for the pulling portion to move. Because we want our head as compact as possible, Eyeball #2 is not adopted.

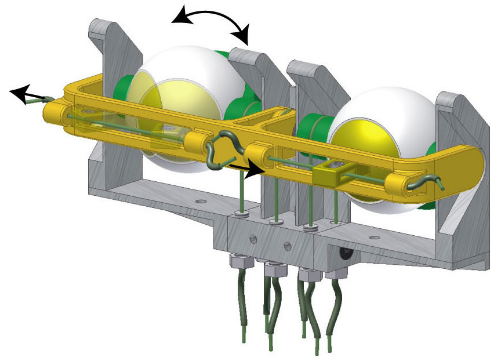


Figure 29 Wire-Pulling Mechanism of Eyeball of Probo [45]

Eyeball #3 is a two-ring mechanism. The two rings contain eyeball inside and control the rotation in horizontal and vertical direction respectively. The disadvantage of Eyeball #3 is that it leads to visible mechanical parts, which will reduce the anthropomorphism to a great degree. Hence, we did not choose the design.

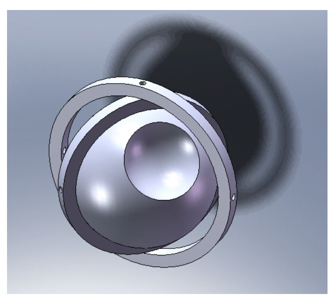


Figure 30 Ring Design of Eyeball

Eyeball #4 is a very compact design. Most components are hidden behind the big eyeballs. Two micro servos and four gears are used to drive the two half-sphere eyeballs. The four gears are exactly the same, two for pan motion while the other two for tilt. Utilizing transmit of same size gears, the motors actuate the eyeballs in the same angle with the same speed as itself.

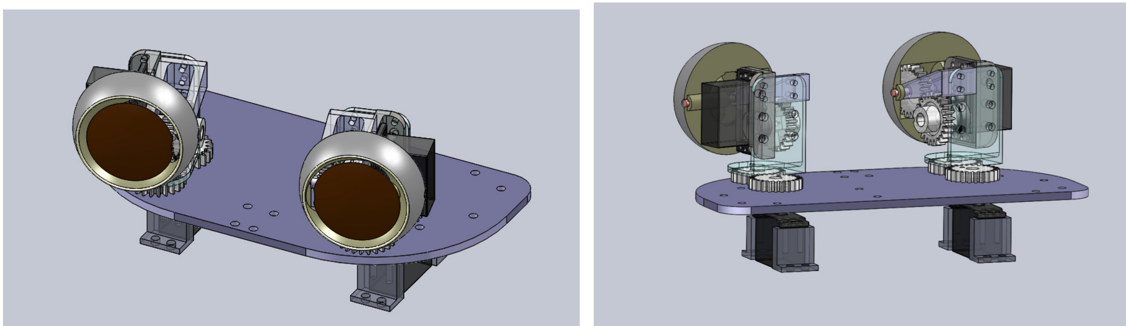


Figure 31 EyeBall #4 Mechanism
Left: Front Trimetric View; Right: Back Trimetric View

4.1.4 Nose

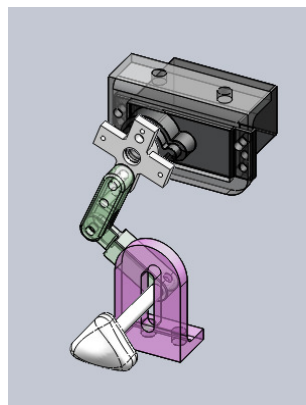


Figure 32 Design of Nose

The nose has been designed to enhance the performance of facial expressions. However, it shows that it doesn't take much part in the expression of facial features in simulation. Hence, the linkage-driven nose design has been discarded in the final version of Dorothy. Instead, Dorothy has an ornamental and static nose.

4.1.5 Mouth

The motion of mouth has an important role in the formulation of various expressions and speaking capability of a social robot. We now present the mechanical design of the mouth so that its mouth movement can be used to produce the sound used in speech. The model of mouth is used to identify the range of movement and response times required to imitate human lip movement during speech.

Basically, there are two approaches to design the mouth. One is wire-pulling mechanism that shapes the mouth by pulling certain control point on the skin [46]. With this approach, the mouth must have an elastic skin. Many robot heads with high anthropomorphism use this method, Albert Hubo [37] and Saya [5], as shown in Fig 34.

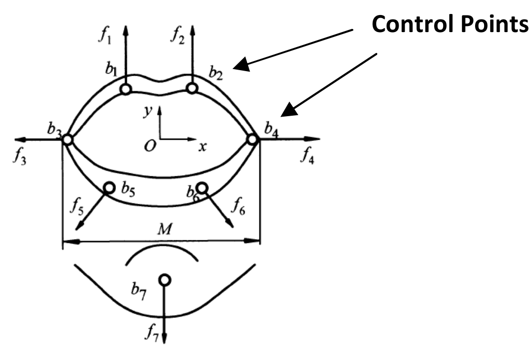


Figure 33 Illustration of Control Point Mechanism of Mouth [47]

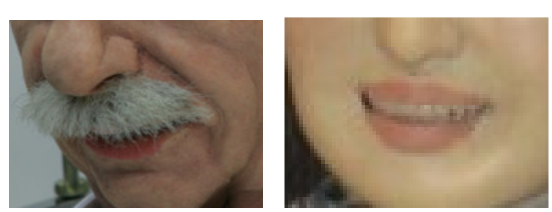


Figure 34 Mouth of Einstein (Left) and Saya (Right)

The other approach is to speak by changing the shape of extruding lips [48]. Different positions of upper and lower lips create different shapes of the mouth. The typical examples include iCat [27], Sparky [49], Kismet [29] and so on [Fig 35].

Figure 35 Mouth of iCat, Sparky, Kismet (Left to Right)

Inspired by these two ideas, we come up with a special mouth design, which has extruding lips that are driven by control points – A, B, C and D. The upper lip is controlled by three control points, one in the center and the other two at the lip corners. The central one is fixed. The corners are synchronically actuated by two identical gears driven by one micro servo. The lower lip is shaped by the control point at the center of the lower lip.

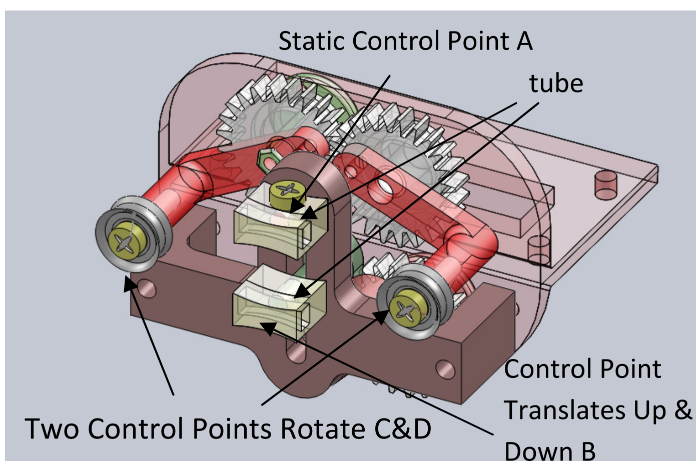
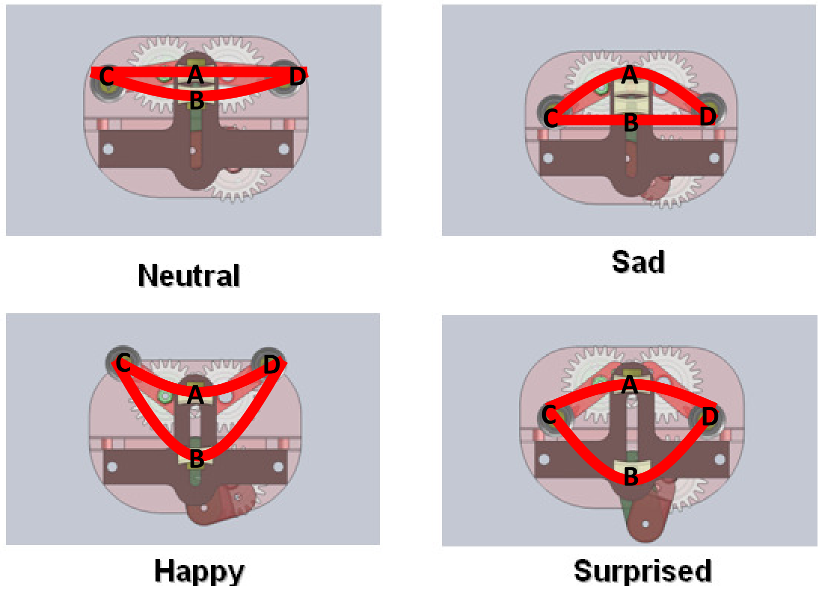


Figure 36 Revised Mouth Design in SolidWorks

In order to make the curve of lips smooth, two rubber tube and two pulleys are added. These four controls points together with the elastic lips outline the shape of the mouth.



A – Control Point A
B – Control Point B
C – Control Point C
D – Control Point D

Figure 37 Lip Shapes of Different Emotions

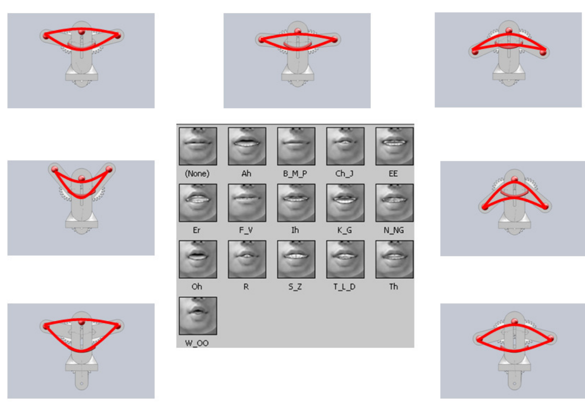


Figure 38 Talking Mouth Shape Simulation in SolidWorks

4.1.6 Degree of Freedom

Degree of freedom is one of the most significant indexes in robotics. Generally, it implies the quantity of actuators and reflects the motion capability of a robot. In robotics, degree of freedom is defined as the number of independent parameters required to specify the position and orientation of an object. The number of degrees of freedom is to the total number of independent driven joints. A machine may operate in two or three dimensions but have more than three and six degrees of freedom.

Dorothy has in total 9 independent driven joints generated from 9 independent servo motors. That is to say, the total degrees of freedom are 9. In comparison of other robotic faces in the literature review, Dorothy ranks in the middle for her degree of freedom.

Name	No. of Features	No. of Motors	DOFs
Eyebrow	2	2	2
Eyeball	2	4	4
Eyelid	2	1	1
Lip	2	2	2
Nose	1	0	0
Total	**9**	**9**	**9**

Table 7 Degrees of Freedom of Dorothy

4.2 Frame Construction

The main frame of Dorothy is made by fused deposition modeling (FDM). Fused deposition modeling (FDM) is an additive manufacturing technology commonly used for modeling, prototyping, and production applications [50]. FDM works on an "additive" principle by laying down material in layers. A plastic filament or metal wire is unwound from a coil and supplies material to an extrusion nozzle which can turn on and off the flow. The nozzle is heated to melt the material and can be moved in both horizontal and vertical directions by a numerically controlled mechanism, directly controlled by a computer-aided manufacturing (CAM) software package. The model or part is produced by extruding small beads of thermoplastic material to form layers as the material hardens immediately after extrusion from the nozzle. Several materials are available with different trade-offs between strength and temperature properties. As well as acrylonitrile butadiene styrene (ABS) polymer, the FDM technology can also be used with polycarbonates, polycaprolactone, polyphenylsulfones and waxes. A "water-soluble" material can be used for making temporary supports while manufacturing is in progress. Marketed under the name WaterWorks by Stratasys, this soluble support material is quickly dissolved with specialized mechanical agitation equipment utilizing a precisely heated sodium hydroxide solution. Moreover, plastic pillars are used as the supporting items as well. Standard screw and nuts are used to fasten.

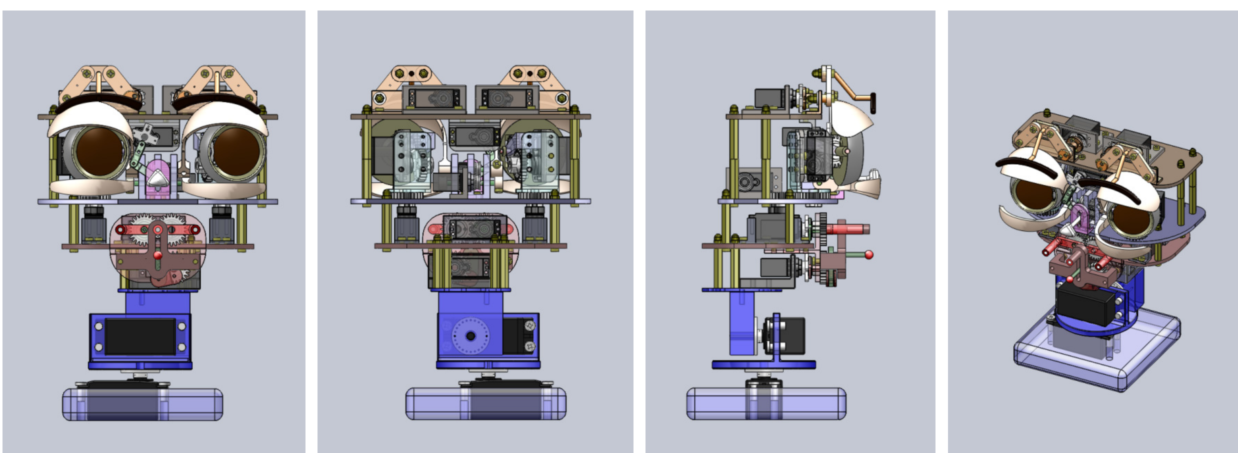
4.3 Dorothy's Skin

For Dorothy to appear as a good research platform it must appear finished. Skin or other forms of casing is necessary to achieve completion. Dorothy has a soft skin covering of mechanical substructure and electrical components for better facial expressiveness. The material used for injection is EcoFlex SuperSoft 0030, which is a kind of platinum silicon rubber compound. It is translucent and can be colored by adding dye materials. It is also very stretchable and can elongate up to 900% until breaks. The supplier of EcoFlex Series is Smooth-On (US) [51]. We ordered it from distributor in Australia, Rowe Trading.

The mold is made by rapid prototyping by ME Design Studio (National University of Singapore). It took two hours to complete the injection and two days to cure the material.

4.4 Graphical Simulation in SolidWorks

Dorothy was designed in SolidWorks® 2010. Fig 40 is the multi-directional view of Dorothy in the absence of Skin in SolidWorks® 2010.



Front View **Back View** **Side View** **Trimetric View**

Figure 39 Dorothy in SolidWorks
From Left to Right: Front View, Back View, Side View and Trimetric View

2D simulation of Dorothy surfaced is also done in SolidWorks® 2010. Initially, two versions of Dorothy were created, double eyelids or single eyelid. Fig 41 is the simulation six basic emotions of the covered face (Note: Only Upper Eyelids (Left), Both upper and lower eyelids (Right)).

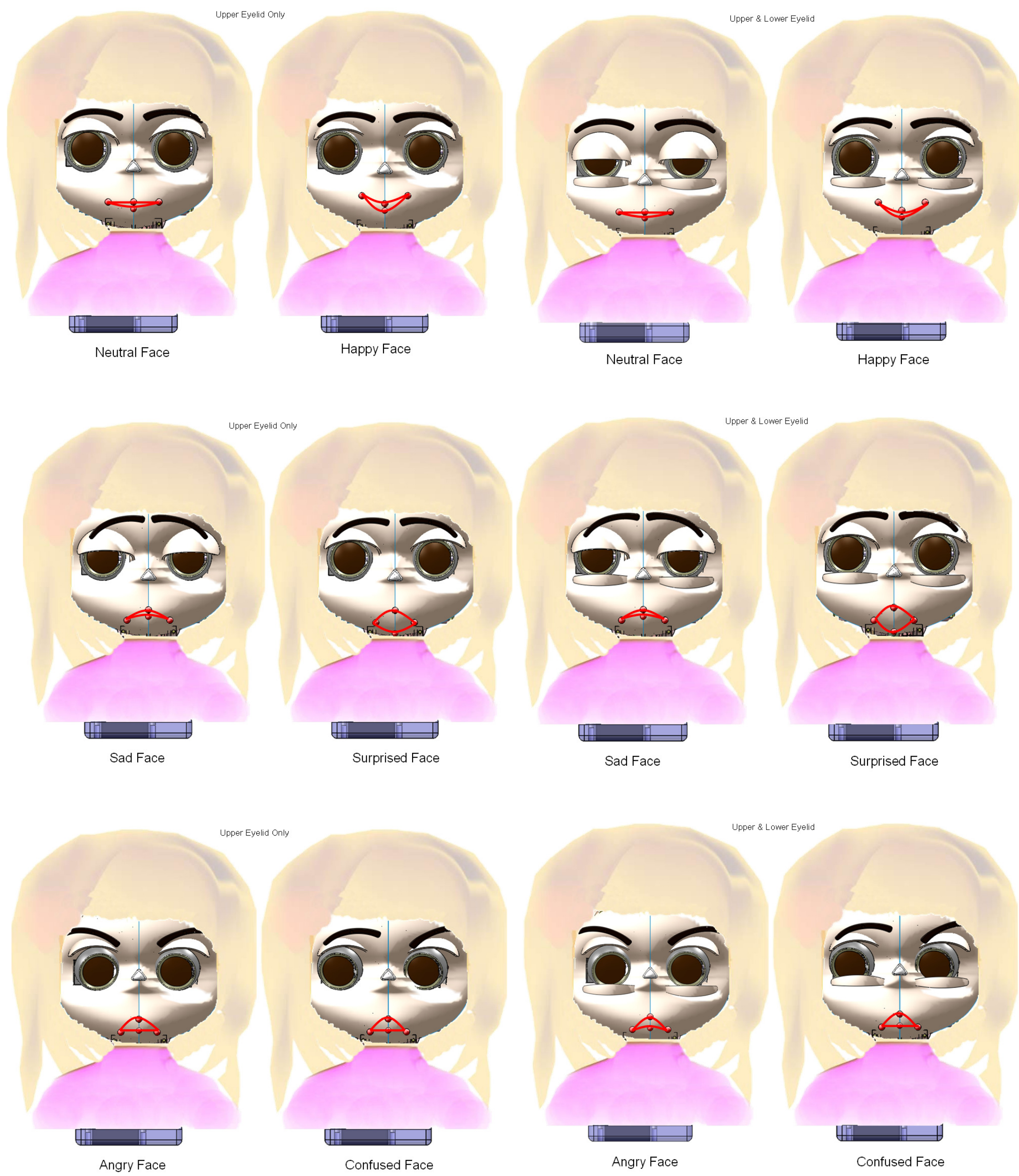


Figure 40 Facial Expressions of Dorothy with and without lower eyelid

From the simulation, we can see that the lower eyelids don’t play an important part in the facial expressions. Hence, Dorothy has upper eyelid only.

Chapter 5 – Dorothy Control Architecture

This chapter describes the architecture of controlling Dorothy, including actuators, microcontroller, power system, algorithm and serial communication.

5.1 Actuators

5.1.1 Fundamentals of Servo Motors

A servo [52] is a small actuator with sensor and circuitry built-in. It has a positional shaft that can be arranged in a number of angled positions via a coded signal. The position of the shaft changes as it receives different signals. A servo motor operates on the principal of "proportional control”, which means that the motor will only run as hard as necessary to accomplish the task at hand. If the shaft needs to turn a great deal, the motor will run at full speed. If the movement is small, the motor will run more slowly. A control wire sends coded signals to the shaft using "pulse coded modulation." With pulse-coded modulation, the shaft knows to move to achieve a certain angle, based on the duration of the pulse sent via the control wire. For example, a 1.5 millisecond pulse will make the motor turn to the 90-degree position. Shorter than 1.5 moves it to 0 degrees, and longer will turn it to 180 degrees.

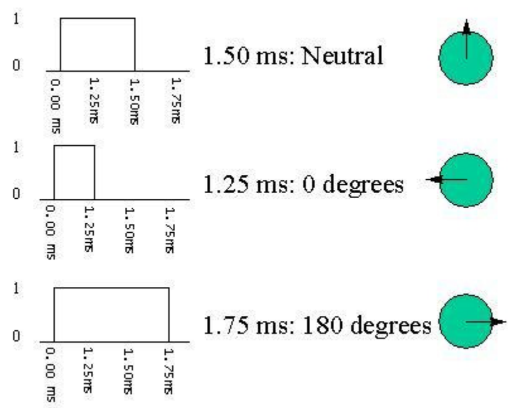


Figure 41 Illustration of Servo Pulse

Despite their small size, servo motors are powerful but do not consume much energy. Servos can operate under a range of voltages, typically from 4.8V to 6V. In most cases, unless there is a battery voltage/current/power limitation, micro servos should operate at 6V to gain higher torque. In terms of current, servo current draw is very unpredictable.

5.1.2 Servo Motor Selection

The servos selected for Dorothy must meet the following criteria.

1. Because the average of human blinking rate is 0.3-0.4s, only servos with speed <0.3s will be taken into account.
2. The torque requirement stems from the force required to drag the Dragon skin material. Assume a Dragon Skin Rod with 1 cm^2 cross-section and 1 cm long. It will elongate 100% under 22 psi. The force needed for 100% elongation (1 cm) is 22 psi ≈ 15.18 N. The torque required for 1 cm elongation is 15.18 N * 1 cm = 1.55 kg.cm. If the original length is 2 cm, the torque required for 1 cm elongation is 0.774 kg.cm.

Other properties of the skin material can be roughly calculated because most likely they will never run over the limits.

<table>
<tr><th>Properties</th><th colspan="2">Limitations</th><th>Comments</th></tr>
<tr><td>Mixed Viscosity</td><td colspan="2">23,000 cps</td><td>”Thick” enough</td></tr>
<tr><td>Shore A Hardness</td><td colspan="2">10A</td><td>Less than Rubber Band SAH=25</td></tr>
<tr><td>Tensile Strength</td><td colspan="2">327.75 N/cm2</td><td>Much more than enough</td></tr>
<tr><td>Elongation at Break %</td><td colspan="2">1000%</td><td>Much more than enough</td></tr>
<tr><td>Die B Tear Strength</td><td colspan="2">1825.8 kg/m</td><td>Much more than enough</td></tr>
<tr><th colspan="4">Unit Conversion</th></tr>
<tr><td colspan="2">1 psi ≈ 0.69 N/cm2
• 475 psi ≈ 327.75 N/cm222
• psi ≈ 15.18 N/cm2</td><td colspan="2">1 pli ≈ 17.9 kg/m
• 102 pli ≈ 1825.8 kg/m</td></tr>
</table>

Table 8 Other Properties of Dragon Skin Material

3. Due to the dimension constrain of the head, the servo should be in small size and very compact. In the family of servos, micro/mini servo motors, whose dimension range is 20++ (mm)*10++ (mm)*20++ (mm) (L*W*H), are the best choices.
4. The weight is the minor factor that taken into account. Normal micro/mini servo that is less than 20g/each will do.
5. Last but not least, the money issue should be considered as well. Actuators are most costly in building robots. Our budget is $500 for actuators. Hence, the budget of one micro servo motor is ($500 / 9 servos = $55.6).

Based on the selection criteria and the research of micro servos available in the markets, we decide use Hitec HS-65HB as the actuators of Dorothy. HS 65-HB can operate 180° when given a pulse signal ranging from 600usec to 2400usec. Table 16 lists the specifications of HS-65HB. The specifications of HS 65-HB are fully meet the above-mentioned criteria.

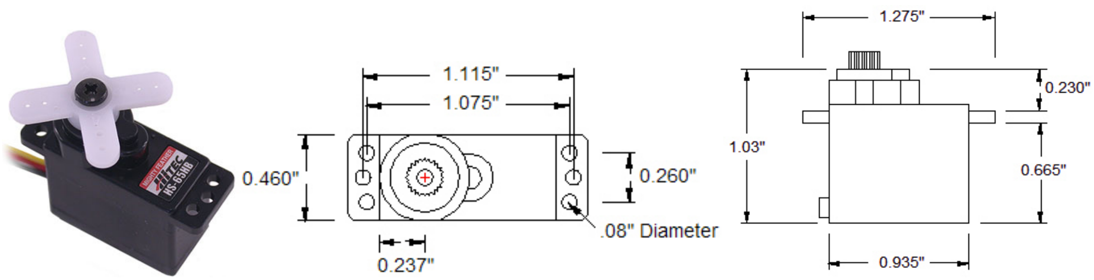


Figure 42 Hitec HS-65HB

Control System	+Pulse Width Control 1500usec Neutral
Required Pulse	3-5 Volt Peak to Peak Square Wave
Operating Voltage	4.8-6.0 Volts
Operating Temperature Range	-20 to +60 Degree C
Operating Speed (4.8V)	0.14sec/60° at no load
Operating Speed (6.0V)	0.11sec/60° at no load
Stall Torque (4.8V)	25 oz/in. (1.8kg.cm)
Stall Torque (6.0V)	31 oz/in. (2.2kg.cm)
Operating Angle	45 Deg. one side pulse traveling 400usec
360 Modifiable	No
Direction	Clockwise/Pulse Traveling 1500 to 1900usec
Current Drain (4.8V)	5.3mA/idle and 400mA no load operating
Current Drain (6.0V)	6.6mA/idle and 500mA no load operating
Dead Band Width	4usec
Motor Type	3-Pole Neodynium Magnet
Potentiometer Drive	Indirect Drive
Bearing Type	Top Ball Bearing
Gear Type	Karbonite Gears
Connector Wire Length	6" (150mm)
Dimensions	0.92" x 0.45"x 0.94" (23.6 x 11.6 x 24mm)
Weight	.39oz. (11.2g)

Table 9 Specifications of HS-65HB

5.1.3 Attachment and mounting of servos

Due to the small size of micro servo motors, we need to take into account the attachment and mounting problems of these mini servos. Fig 44 and Fig 45 shows the common mounting brackets for micro servos and servo mounting bracket for Dorothy respectively.

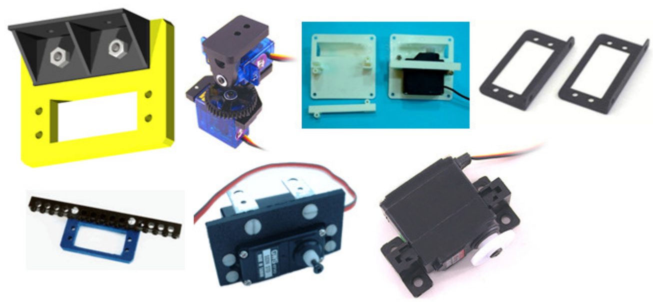


Figure 43 Common Mounting Brackets for Micro Servos

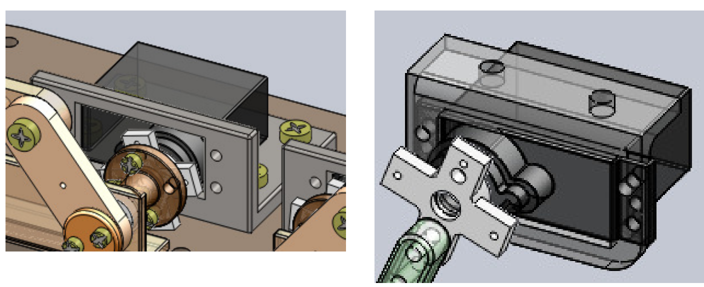


Figure 44 Mounting Brackets of HS 65-HB in Dorothy

5.2 Microcontroller

The microcontroller used in Dorothy is SSC-32 [53], which we find is the most suitable controller available. It is a small pre-assembled servo controller with many good features.

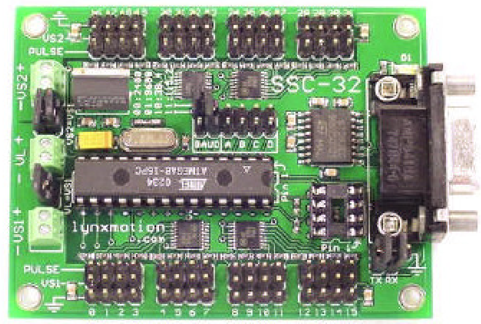


Figure 45 SSC-32 Servo Controller from LynxMotion

It has high resolution (1uS) for accurate positioning, and extremely smooth movement. The range is 0.50mS to 2.50mS for a range of about 180°. The motion control can be immediate response, speed controlled, timed motion, or a combination. In addition, a unique feature of group move allows any combination of servos to begin and end motion at the same time, even if the servos have to move different distances. This is a very powerful feature for creating complex walking gaits for multi servo walking robots. The standard commands are according to the following rules.

Servo Group Move Example 1: "#5 P1600 #10 P750 T2500 <cr>"
Servo Group Move Example 2: "#5 P1600 #17 P750 S500 #2 P2250 T2000 <cr>"

In example 1, it will move servo 5 to position 1600 and servo 10 to position 750. It will take 2.5 seconds to complete the move, even if one servo has farther to travel than another. In example 2, it follows the three rules below.

1. All channels will start and end the move simultaneously.
2. If a speed is specified for a servo, it will not move any faster than the speed specified (but it might move slower if the time command requires).

3. If a time is specified for the movement, then the movement will take at least the amount of time specified (but might take longer if the speed command requires).

These specifications pertain to firmware version 2.01XE:

- Microcontroller = Atmel ATMEGA168-20PU
- EEPROM = 24LC32P (Required for 2.01GP)
- Speed = 14.75 MHz
- Internal Sequencer = 12 Servo Hexapod (Alternating Tripod)
- Serial input = True RS-232 or TTL, 2400, 9600, 38.4k, 115.2k, N81
- Outputs = 32 (Servo or TTL)
- Inputs = 4 (Static or Latching, Analog or Digital)
- Current requirements = 31mA
- PC interface = DB9F
- Microcontroller interface = Header posts
- Servo control = Up to 32 servos plug in directly
- Servo type supported = Futaba or Hitec
- Servo travel range = 180°
- Servo resolution = 1uS, .09°
- Servo speed resolution = 1uS / Second
- Servo motion control = Immediate, Timed, Speed or Synchronized.
- PC board size = 3.0" x 2.3"
- VS current capacity = 15 amps per side, 30 amps max

5.3 Power Systems

During the testing period, the power of Dorothy is drawn from a DC power supply. The voltage applied to servo power is 4.8 - 6.0 volts. The working current is about 100 mA. The current limit is 3500 mA. In the demo stage, Dorothy is powered by 4A batteries.

There are two power supply inputs on SSC-32 broad. The logic supply (VL) powers the microcontroller and it supports circuitry through a 5vdc regulator. The servo supply (VS) powers the servos directly. In single supply mode (default) the jumper VS1=VL will provide power to the VL 5vdc regulator from the VS terminal. This improves 4A battery use as long as the voltage does not drop too much. But if it does drop, the voltage to the microcontroller is interrupted and the SSC-32 resets. Hence, instead of using the VL-VS jumper, we connect a separate power to the VL input. This isolates the servo and logic supplies so that one cannot affect the other.

When SSC-32 is connected to the power supply and SSC-32 is turned on, the green LED will be on steady. It will remain on until it has received a valid serial command, then it will go out and only blink when receiving serial data. The green LED is not a power indicator, but a status indicator. If you notice the servos turn off, or stop holding position when moving several servos at one time. This indicates the SSC-32 has reset.

Beside the voltage issue, the only other thing that can cause this effect is a poor power delivery system. If the wires carrying the current are too small, or connections are made with stripped and twisted wire, or cheap plastic battery holders are used, the same problem may occur. 99% of practical problems with the SSC-32 are power supply related. When noticing erratic or unstable servo movements, we need to look at the power delivery system, e.g. current limitation.

5.4 Algorithm

The key of controlling Dorothy is the algorithm that makes Dorothy expressive and talk-able in a very natural way. Based on the human factor research, a scenario has been set as the first achievement for Dorothy to implement.

Scenario: When text is entered, Dorothy is supposed to speak them out and at the same time, express corresponding emotions according to the text input.

In this context, we create an input text protocol to synchronize facial expressions. The input command must be typed in the follow pattern with emotion indicator included. The input contains two parameters: words and emotion states. The words are anything that we want Dorothy to say. The emotion states are assigned to each word and will remain the same until a new emotion state is given. That is to say, each word has one emotion state.

Text Sample: Hello, I am very #1A lovely robot. #0 But I am very #2B sleepy #0 now.

Explanation of the above example: in the absence of emotion marker denoted by # followed by 2 numerals, the default state is neutral face which is applied to “Hello, I am very”. Upon reaching the marker “#1A” (refer to Fig 47 for codes), a happy face is presented for the following speech segment “lovely robot”. Similarly, upon reaching the marker “#0”, neural face is resumed.

Emotion Index (EI)	Emotion State (ES)	Intensity	
		A	B
#0	Neutral	#0 (by default)	
#1	Happy	#1A	#1B
#2	Sad	#2A	#2B
#3	Surprise	#3A	#3B
#4	Angry	#4A	#4B

Figure 46 Control Dorothy Command

Emotion states are presented by emotion index. There are five emotion index (EI) which refers to five emotion states (ES) - neutral, happiness, sadness, surprise, angry. The neural state is also the default state. In other words, if there is no emotion state is assigned; Dorothy will implement the neutral state. Except neutral state, the rest four basic emotions all have two intensity levers for each, indicated by capital letter "A" and "B". "A" is more intensive than "B". For example, #1A means very happy while #1B means slightly happy. The degree of intensity "very" and "slightly" is defined based on human factor study and mechanical capability of Dorothy. Each facial feature has its neutral position and limit position. "Very" is defined as the 90% of the angle limitation, whereas "slightly" is the 30% of the angle limitation respect to neutral position. The formula is as below.

VERY=neutral position ± (limitation - neutral position) *0.9
SLIGHTLY= neutral position ± (limitation - neutral position) *0.3

For example, the neutral position of left eyeball is 512. Its upper limitation is 900. Then the "very" upper position is 512 + (900 - 512) * 0.9=816.2. The calculated angels are used for facial features except mouth. Based on this rule, we can obtain all the angles for these 9 emotion states. Table 10 shows the angles for the 9 emotion states, these angles apply for eyebrows, eyelids, eyeballs at any time and non-speaking mouth only.

Facial Expression							Eye Motion				
		Angle (Non - Speaking Face)					Eyeball				Eyelid
ID	Location	Neutral	Happiness	Sadness	Surprise	Anger	Up	Down	Left	Right	Blink
#1	Left Eyebrow	1500	1500	1600	1350	1300	-	-	-	-	-
#2	Right Eyebrow	1500	1500	1300	1550	1600	-	-	-	-	-
#3	Eyelid	1500	1400	1600	1700	1500	-	-	-	-	-
#4	Left Eyeball (up/down)	1500	1500	1500	1500	1500	1650	1400	-	-	-
#5	Left Eyeball (left / right)	1500	1500	1500	1500	1500	-	-	1600	1450	-
#6	Right Eyeball (up/down)	1500	1500	1500	1500	1500	1650	1400	-	-	-
#7	Right Eyeball (left / right)	1500	1500	1500	1500	1500	-	-	1550	1400	-
#8	Upper Lip	1500	1250	1600	1600	1500	-	-	-	-	-
#9	Lower Lip	1500	1600	1300	1600	1300	-	-	-	-	-
Notes: P: position S: speed T: time											

Table 10 Angles for 9 Emotion States

There are mainly three modules to process the input; they are Month Motion (MM), Speech (SP) and Facial Expression (FE). Mouth motion is to actuate the lip movement via two stages text-to-phoneme and lip-sync. This module converts input text to the corresponding mouth shape. Speech is to convert input text to sound track. Facial expression is to generate the facial expressions of Dorothy. These three modules are not independent. They are connected by the time frame. As described in the predefined scenario, the mouth motion must match the corresponding speech. At the same time, proper facial expressions are shown. Fig 48 is the block diagram illustrating the algorithm clearly. More details of each module will be described in the following paragraphs.

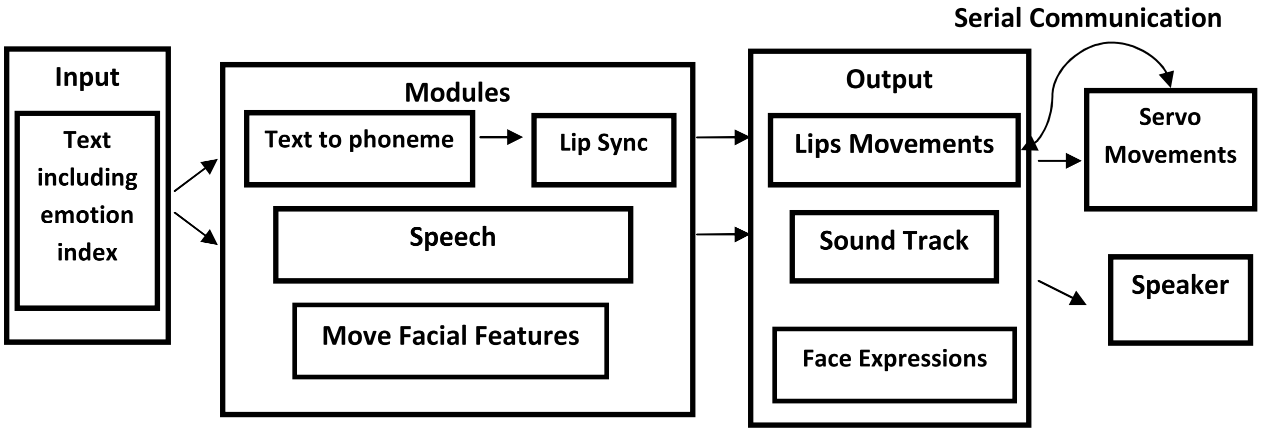


Figure 47 Control Dorothy Block Diagram

5.4.1 Mouth Motion

The mouth motion module consists of two stages: Text-To-Phoneme (TTP) and Lip Synchronization. TTP, as the name suggests, means convert from text to phonemes. The sounds of a language are called phonemes. In a language or dialect, a phoneme is the smallest segmental unit of sound employed to form meaningful contrasts between utterances. Each language has a set that is slightly different from that of other languages. English phonemes can be classified into consonant phonemes vowel phonemes. If the vocal cords are vibrating at the same time, as in the case of the voiced fricatives VV, DH, ZZ, and ZH. Vowels are usually produced with a relatively open vocal tract and a periodic sound source provided by the vibrating vocal cords. They are classified according to whether the front or back of the tongue is high or low, whether they are long or short, and whether the lips are rounded or unrounded. In English all rounded vowels are produced in or near the back of the mouth (UW, UH, OW, AO, OR, AW). Speech sounds which have features in common behave in similar ways. For example, the voiceless stop consonants PP, TT, and KK should be preceded by 50-80 millisecond of silence, and the voiced stop consonants BB, DD, and GG by 10-30 millisecond of silence.

Speech synthesis systems use two basic approaches to determine the pronunciation of a word based on its spelling, a process which is often called text-to-phoneme conversion (phoneme is the term used by linguists to describe distinctive sounds in a language) [54]. The simplest approach to text-to-phoneme conversion is the dictionary-based approach, where a large dictionary containing all the words of a language and their correct pronunciations is stored by the program. Determining the correct pronunciation of each word is a matter of looking up each word in the dictionary and replacing the spelling with the pronunciation specified in the dictionary. The other approach is rule-based, in which pronunciation rules are applied to words to determine their pronunciations based on their spellings. This is similar to the "sounding out", or synthetic phonics, approach to learning reading. Each approach has advantages and drawbacks. The dictionary-based approach is quick and accurate, but completely fails if it is given a word that is not in its dictionary. As dictionary size grows, so too does the memory space requirements of the synthesis system. On the other hand, the rule-based approach works on any input, but the complexity of the rules grows substantially as the system takes into account irregular spellings or pronunciations. (Consider that the word "of" is very common in English, yet is the only word in which the letter "f" is pronounced [v].) As a result, nearly all speech synthesis systems use a combination of these approaches.

From TTP, a list of phonemes is obtained. The next step is to convert phonemes to viseme and eventually to the motion of mouth. Followed pre-processing of input command, the input text has been split into phonemes and then parse them to the next stage - lip synchronization. Lip-synchronization or Lip -synch for short, is a technical term for matching lip movements with voice. The term here refers to matching lip movements of Dorothy facial expressions. In lip-synch; the most important term is viseme. A viseme is a representational unit used to classify speech sounds in the visual domain. The term viseme was introduced based on the interpretation of the phoneme as a basic unit of speech in the acoustic/auditory domain. This is, however, at variance with the accepted definition of the phoneme as the smallest structural unit that distinguishes meaning within a given language - as a cognitive abstraction that is not bound to any sensory modality. A "viseme" describes the particular facial and oral positions and movements that occur alongside the voicing of phonemes.

Phonemes and visemes do not always share a one-to-one correspondence; often, several phonemes share the same viseme. In other words, several phonemes look the same on the face when produced. However, there could be differences in timing and duration during actual speech in terms of the visual 'signature' of a given gesture that cannot be captured with a single photograph. Conversely, some sounds that are hard to distinguish acoustically are clearly distinguished by the face. For example, acoustically speaking English /l/ and /r/ could be quite similar (especially in clusters, such as 'grass' vs. 'glass').

Yet visual information can show a clear contrast. This is demonstrated by the more frequent mishearing of words on the telephone than in person. Some linguists have argued that speech is best understood as bimodal (aural and visual), and comprehension can be compromised if one of these two domains is absent. The comprehension of speech by visemes alone is known as speech reading or "lip reading".

The list of viseme determines the parameters of mouth motion. Turing Angle, speed and time of servos are set to obtain exact mouth shape with respect to the corresponding viseme. Servo motion control will be discussed in the next session serial communication.

5.4.2 Speech

This module is to convert the text into sound track with selected voices, speed, pitches and other parameters. The speech machine used in our project is Microsoft SAPI 5.3 [55]. Importing the SAPI 5.3 library allows to use its function – spVoice.speak() to implement Dorothy's speaking capability. The key issue here is the sound delay of mouth movement.

5.4.3 Facial Expression

The emotion module is to implement the capability of facial expression of Dorothy. This is the primary ability of Dorothy and even Doris. Different from Doris, Dorothy's facial expressions are not controlled by clicking buttons in a GUI, instead, its emotion state is assigned in the input based on the words entered. To control Dorothy's facial expressions, emotion parameters need to map with the parameters actuating the facial mechanism. The emotion parameters consist of six basic emotion states: happiness, sadness, surprise, anger and neutral. Each word element has its emotion state (ES), indicated by emotion index (EI). The default ES is neutral, #0. The ES will not be changed until a new ES is assigned. The new ES is applied to the following word until another new ES command is given.

Each emotion state consists of the facial expressions of emotion states are predefined in the programs. In total, there are 9 facial expressions - they are neutral face, very happy face, slightly happy face, very sad face, slightly sad face, very surprising face, slightly surprising face, very angry face and slightly angry face. Pictures are the preview of these emotions.

Up to here, three outputs can be obtained, lip motion, voice and facial expressions. The next stage is to synchronize them based on the timeline. The ultimate output is the combination of them and eventually we can achieve the goal scenario set at the very beginning.

In order to achieve these outputs, three functions are added into the sample code of TTSApps. The first step is to generate an etag (emotion tag) list on a basis of word. When the programs scan the word beginning with symbol #, it will change the emotion state (represented by etag) for the next word. Each emotion state will remain until new emotion state symbol # is detected. The second step is to assign the emotion state to each word. This function is inserted in the SEPI_WordBoundary. When the program detects a word boundary, it will get the etag from the etag list generated in step 1. In other words, each word has its own emotion state now. The final step is to make the face motion based on the emotion state. The command string has been inserted just after each viseme picture is downloaded. In this way, Dorothy has both a physical face and virtual face, and they are synchronized with the error less than 0.001s. Up to here, Dorothy is able to achieve the goal scenario.

5.5 Serial Communication

Dorothy uses standard serial communication protocol connected to PC. It uses wireless serial communication to receive the command from a computer. In telecommunication and computer science, serial communication is the process of sending data one bit at a time, sequentially, over a communication channel or computer bus. This is in contrast to parallel communication, where several bits are sent together, on a link with several parallel channels. Serial communication is used for all long-haul communication and most computer networks, where the costs of cable and synchronization difficulties make parallel communication impractical.

The serial communication architecture used in Dorothy is RS-232. In telecommunications, RS-232 (Recommended Standard 232) is a standard for serial binary single-ended data and control signals connecting between a DTE (Data Terminal Equipment) and a DCE (Data Circuit-terminating Equipment). It is commonly used in computer serial ports. A similar ITU-T standard is V.24.

The wireless devices used are AIR$_{CABLE}$ S$_{ERIAL}$ 3X and AIR$_{CABLE}$ USB3 [49]. The two devices are paired in advance through a series of setting and communicate with each other via Bluetooth. The AIR$_{CABLE}$ USB3 runs all the Bluetooth connection software on the module and presents only a virtual COM port to the PC. There is no Bluetooth software required that must be installed on the PC. This product will help integrators to save big on support costs. It automatically makes Bluetooth connections to medical devices, GPS receivers, bar code scanners, credit card readers, and more. The AIR$_{CABLE}$ Serial3, the pre-installed cable replacement application, like other applications used in mobile phones, allows zero-configuration use as "service-slave", allowing incoming serial connections from other Bluetooth devices. With the push of a button it configures in a secure one-to-one slave or master mode.

Figure 48 AirCable BlueTooth Pair Connectors
Left: AIR$_{CABLE}$ USB3; Right: AIR$_{CABLE}$ SERIAL 3X

The cSerial class member functions include:

cSerial::cSerial() - constructor

cSerial::~cSerial() - disconstructor

cSerial::Open(int nPort = 5, int nBaud = 57600) - This member function is used to open the serial port. It takes two integer arguments. The first argument contains the port number where the valid entries are 1 through 5. The default port is COM5. The second argument is the baud rate. Valid values for this argument are 2400, 4800, 9600, 57600 and 19200. This function returns TRUE if successful. Otherwise, it returns a value of FALSE. For LynxMotion communication, the default baud rate is 57600. The key function for open port is called CreateFile.

CreateFile(szPort, GENERIC_READ | GENERIC_WRITE, 0, NULL, OPEN_EXISTING, FILE_ATTRIBUTE_NORMAL | FILE_FLAG_OVERLAPPED, NULL);

CSerial::Close() - While the program will automatically close the serial port for you, this function has been added just in case there is a reason that you need to explicit close the port. The key function is called CloseHandle.

CloseHandle(m_hIDComDev);

CSerial::SendData(const char *, int) - This function writes data from a buffer to the serial port. The first argument it takes is a const char* to a buffer that contains the data being sent. The second argument is the number of bytes being sent. This function will return the actual number of bytes that are successfully transmitted. The key function is WriteFile.

WriteFile(m_hIDComDev, txBuff, (DWORD)(txBuffWork-txBuff), &dwBytesWritten, &m_OverlappedWrite);

CSerial::ReadDataWaiting(void) - This function simply returns the number of bytes that waiting in the communication port's buffer. It basically allows you to "peek" at the buffer without actually retrieving the data.

CSerial::ReadData(void*, int) - This function reads data from the port's incoming buffer. The first argument that it takes is a void* to a buffer into which the data will be placed. The second argument is an integer value that gives the size of the buffer. The return value of this function contains the number of bytes that were successfully read into the provided data buffer. The key function is ReadFile.

ReadFile(m_hIDComDev, vbuffer, dwBytesRead, &dwBytesRead, &m_OverlappedRead);

Moreover, a time delay function is used to smooth the motion. Hence, through RS-232 serial communication, PC is able to command servo's movement.

5.6 User Interface

Users can use GUI to control Dorothy. The GUI is written in C++ Window 32 API. There are mainly three parts on the GUI: input, control and simulation. The input box is for input command to be typed in. The control part includes control buttons and history record. The simulation area shows a graphic Dorothy that is synchronized with real Dorothy.

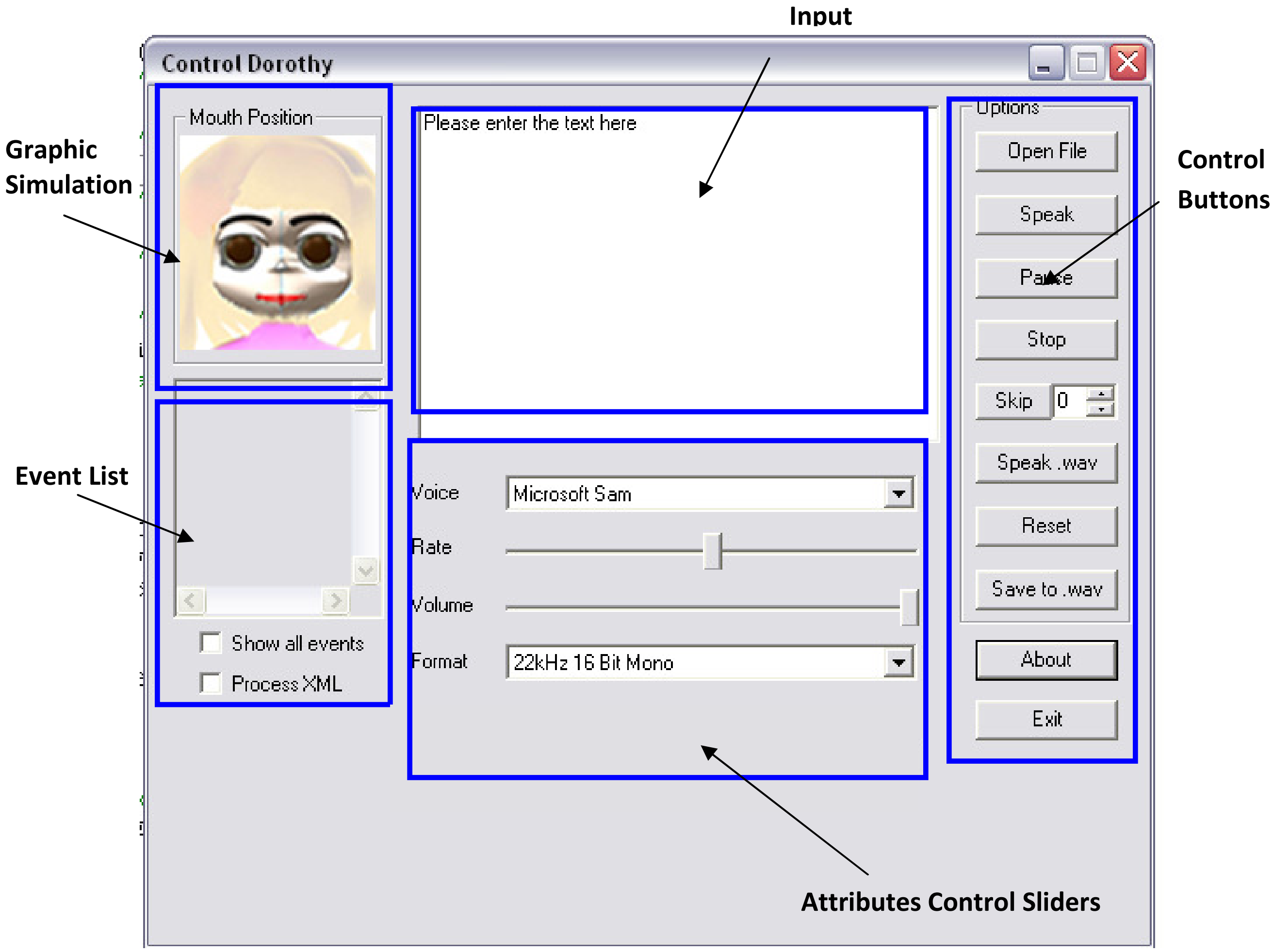
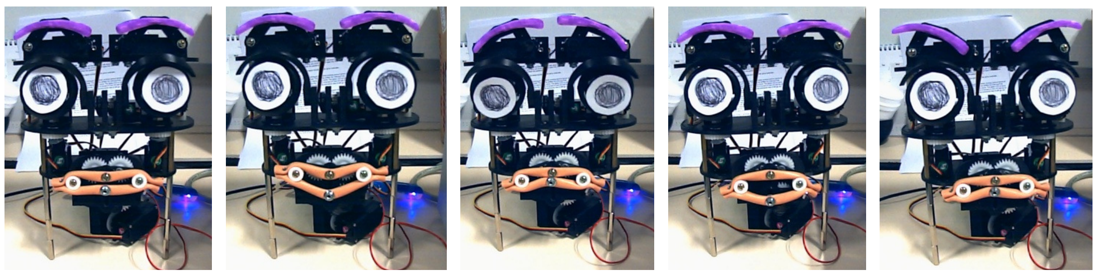


Figure 49 User Interface (GUI) for Controlling Dorothy

Chapter 6 – Performance Evaluation & Discussion

This chapter evaluates Dorothy's performance and investigate its sociability. Figure 51 are the pictures of Doris' five basic emotion expressions: neutral, happiness, sadness, surprise and anger.



Neutral Happiness Sadness Surprise Anger

Figure 50 Basic Facial Expressions of Dorothy

Emotion recognition tests were carried out based on subjects' judgments of emotions expressed by Doris. A multiple-choice questionnaire was devised where subjects were asked to label emotional expressions from pictures of Dorothy expressing happy, sad, surprise and angry. 20 subjects filled out the questionnaire. Half of the subjects have technical background in robotics while the rest do not have. There were four pages in the questionnaire. Each page had a large color image of Dorothy displaying one of four expressions. The subjects could choose the best match from seven possible labels (anger, fear, surprise, sadness, disgust, happiness, anticipation). In a follow-up section of the questionnaire, they were asked to specify on a five-point scale how confident they were of their answers. On top of that, subjects are required to evaluate the facial movements of Dorothy in a news-reading scenario. They will give Dorothy a grade based on their own acceptation of its facial movements. The result shows that 34% of the subjects selected "Medium", 45% select "Good" and 21% select "Very Good" respectively.

The approach we use to evaluate the result is to take a weighted sum of answers where the weight of each answer was its confidence degree. Other researchers have also carried out similar experiments for their robots. For instance, Breazeal has evaluated performance of Kismet using both color images and video clips. We took image-based results to obtain more comparable values against ours. Since she considered different statistics for children and adults, we used their average values. Table 11 shows correct recognitions for Kismet, Doris and Dorothy. It is to be noted that, from the statistics perspective, the comparison is not precise because the quality and quantity of examples choices are quite different. Nevertheless, it should give an indication on the performance to a certain extent and how well Dorothy performs.

Emotions	Kismet	Doris	Dorothy
Anger	76	---	88
Disgust	71	---	70
Fear	47	---	75
Happiness	82	78	90
Sadness	82	82	92
Surprise	82	65	80
Anticipation	---	60	---

Table 11 Performance of Different Robotic Faces

Chapter 7 – Conclusion and Future Work

Robotics is an attractive science and technology discipline. The noticeably improved quality of life and ever-increasing “grey” population (elderly) have created an urgent demand for social robots.

We presented the design and fabrication steps of our social robotic head platform, namely Dorothy, which is developed to be used in research, studied focusing on human robot interaction. Dorothy’s face region consists of a 9-DOF mechanism. The eyes (include eyebrows and eyelids) have 7 DOF forming a vision system with two cameras placed like the eyeballs. The mouth has a 2-DOF mechanism. The ranges of motion of each joint are similar to those of humans. Hence, kinematic constraints extracted from anthropomorphic data are achieved. The actuators are selected to satisfy dynamic constrains. The fast delivery mold (FDM) is used as the main manufacturing process. It allows designing complicated parts and even mechanisms as a whole. The process is faster than the traditional techniques. Tests are performed to investigate the response of people to different appearances. Evaluation results are acceptable.

This work opens up more challenges. In the future, when Dorothy evolves (both in hardware and software) it will be used as a test bed for investigating human robot interaction in various social contexts. Looking forward, social robots will play an ever larger role in our world, working for and in cooperation with humans. The key to the success of social robots will be close and effective interaction between humans and robots. There is still much that can be done in each of these modules, which involves scientists and researchers from all disciplines of engineering, computer science and humanities.

Appendix A

Survey on Dorothy's Facial Expressions and Speaking Ability

Your Name: ______________________________ Sex: M / F

<u>There are two sections in this survey: Part A and Part B. It takes about 10 mins to complete this survey.</u>

Part A is about the recognition of Dorothy's emotions.
Dorothy will show 10 different emotions in turn.
Please circle the one that you think describes its emotion best.
The same emotions may be shown more than one time.

<u>Emotion 1</u>

a. Happiness b. Sadness c. Surprise d. Anger e. Fear f. Disgust g. Not sure

<u>Emotion 2</u>

a. Happiness b. Sadness c. Surprise d. Anger e. Fear f. Disgust g. Not sure

<u>Emotion 3</u>

a. Happiness b. Sadness c. Surprise d. Anger e. Fear f. Disgust g. Not sure

<u>Emotion 4</u>

a. Happiness b. Sadness c. Surprise d. Anger e. Fear f. Disgust g. Not sure

<u>Emotion 5</u>

a. Happiness b. Sadness c. Surprise d. Anger e. Fear f. Disgust g. Not sure

<u>Emotion 6</u>

a. Happiness b. Sadness c. Surprise d. Anger e. Fear f. Disgust g. Not sure

<u>Emotion 7</u>

a. Happiness b. Sadness c. Surprise d. Anger e. Fear f. Disgust g. Not sure

<u>Emotion 8</u>

a. Happiness b. Sadness c. Surprise d. Anger e. Fear f. Disgust g. Not sure

Emotion 9

a. Happiness b. Sadness c. Surprise d. Anger e. Fear f. Disgust g. Not sure

Emotion 10

a. Happiness b. Sadness c. Surprise d. Anger e. Fear f. Disgust g. Not sure

Part B is the evaluation of facial movements of Dorothy. Dorothy will read a piece of news. Please circle the option that you think about the performance of Dorothy's facial movements.

a. Very bad b. Bad c. Medium d. Good e. Very good f. Excellent

Please write your comments here

Thank you very much for your time!

Bibliography

1. Christensen, R.G.a.H., *European Robotics Platform – EUROP*. 2005.
2. Wikipedia. *Robotics*. [cited 2010 June 20]; Available from: http://en.wikipedia.org/wiki/Robotics.
3. Wikipedia. *Robots*. [cited 2010 June 20]; Available from: http://en.wikipedia.org/wiki/Robots.
4. Breazeal, C.L., *Design Social Robots*. 2002, London, England: The MIT Press.
5. Terrence Fong, I.N., Kerstin Dautenhahn, *A survey of socially interactive robots.* Robotics and Autonomous Systems, 2003. **42**: p. 143-166.
6. S. S. Ge, "Social Robotics: Integrating Advances in Engineering and Computer Science," Proceedings of the 2007 Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI) Interbational Conference, pp. xvii-xxvi, Mae Fah Luang University, Chang Rai, Thailand, May 9-12, 2007.
7. Carl F. DiSalvo, F.G., Jodi Forlizzi, Sara Kiesler, *All Robots Are Not Created Equal: The Design and Perception of Humanoid Robot Heads.* Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques, 2002. **6**: p. 321-326.
8. McCloud, S., *Understanding Comics: The Invisible Art.* Kitchen Sink Press, 1993.
9. Michael L. Walters · Dag S. Syrdal, K.D.R.t.B., Kheng Lee Koay, *Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion.* Auton Robot, 2008. **24**: p. 159-178.
10. Gong, L., & Nass, C., *When a talking-face computer agent is half-human and half humanoid: human identity and consistency preference.* Journal of Human Communication Research, 2007. **33**(2): p. 163-193.
11. MacDorman, K., & Ishiguro, H., *The uncanny advantage of using androids in cognitive and social science research.* Interaction Studies, 2006. **7**(3): p. 297-337.
12. Boden, M.A., *Robots and Anthropomorphism.*
13. Duffy, B.R., *Anthropomorphism and the social robot.* Robotics and Autonomous Systems, 2003. **42**.
14. N. Epley, A.W., and J. T. Cacioppo, *On Seeing Human: A three-factor theory of anthropomorphism.* Psychological Review, 2007. **114**(864-886).
15. Wikipedia. *Uncanny Valley*. [cited 2010 May 09]; Available from: http://en.wikipedia.org/wiki/Uncanny_valley.
16. Kwan Min Lee, Y.J., Jaywoo Kim, Sang Ryong Kim, *Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction.* Int. J. Human-Computer Studies, 2006. **64**: p. 962–973.
17. Kerstin Dautenhahna, B.O., Tom Quick, *rom embodied to socially embedded agents – Implications for interaction-aware robots.* Cognitive Systems Research, 2002. **3**: p. 397-428.
18. Breazeal, C., *Emotion and sociable humanoid robots.* Int. J. Human-Computer Studies, 2003. **59**.

19. Clément Raïevsky, F.M. *Improving Situated Agents Adaptability Using Interruption Theory of Emotions*. in *10th international conference on Simulation of Adaptive Behavior: From Animals to Animats*. 2008. Osaka, Japan: Springer-Verlag Berlin, Heidelberg.
20. Tingfan Wu, N.J.B., Paul Ruvulo, Marian S. Bartlett, Javier R. Movellan, *Learning to Make Facial Expressions.* IEEE 8TH INTERNATIONAL CONFERENCE ON DEVELOPMENT AND LEARNING 1, 2009.
21. Ekman, P., Oster, H., *Review of research 1970 to 1980*, in *Emotion in the Human Face.*, P. Ekman, Editor. 1982, Cambridge University Press: Cambridge UK. p. 147-174.
22. Do Hyoung Kim, H.S.L., Myung Jin Chung, *Biologically Inspired Models and Hardware for Emotive Facial Expressions.* IEEE International Workshop on Robots and Human Interactive Communication, 2005: p. 679-685.
23. Pelachaud, C., *Emotion Expressiveness Embedded in Representation Languages for ECAs.* 2003.
24. Ekman, F. *FACS - Facial Action Coding System*. 1978; Available from: http://www.cs.cmu.edu/afs/cs/project/face/www/facs.htm.
25. Reallusion. *CrazyTalk6*. 2010; Available from: http://www.reallusion.com/crazytalk/.
26. Robot, C. *Robots*. [cited 2010 May 15]; Available from: http://www.communistrobot.com/robots.php.
27. Research, P. *iCat, user-interface robot, Experimentation platform for Human-Robot Interaction research*. 2005; Available from: www.research.philips.com/robotics.
28. N.V., K.P.E., *Robotics-iCat.*
29. Cynthia Breazeal, B.S., *How to build robots that make friends and influence people*.
30. Laboratory, M.A.I., *Kismet.*
31. Hiroyasu Miwa, K.I., Munemichi Matsumoto, Massimiliano Ze~ca,Hideaki Takanobu's, Stefan Roccella, Maria Chiara Carrozza, Paolo Dario, Atsuo Takanishi, *Effective Emotional Expressions with Emotion Expression Humanoid Robot WE-4RII.* Proceedings of 2004 IEEEiRSJ International Conference an Intelligent Robots and Systems, 2004: p. 2203-2208.
32. Laboratory, H.R.T.T., *Emotion Expression Humanoid Robot WE-4RII (Waseda Eye No.4 Refined II).*
33. Robotics, H. *The Zeno Robot*. Available from: http://www.zimbio.com/The+Zeno+Robot.
34. Lab, M.I.T.M., *MDS.*
35. Mike Blow, K.D., Andrew Appleby, Chrystopher L. Nehaniv, David C. Lee, *Perception of Robot Smiles and Dimensions for Human-Robot Interaction Design.*
36. Group, A.S.R., *Kaspar.*
37. Jun-Ho Oh, D.H., Won-Sup Kim, Il Young Han, Jung-Yup Kim, Ill-Woo Park, *Design of Android type Humanoid Robot Albert HUBO.* Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006: p. 1428-1433.
38. Robotics, H. *Hanson Robotics*. Available from: http://hansonrobotics.wordpress.com/.
39. Braun, K.B.a.T., *A Humanoid Head for Assistance Robots*, in *Climbing and Walking Robots -Proceedings of the 8th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines*, G.S.V.a.M.A.H. M.O. Tokhi, Editor. 2006, Springer Berlin Heidelberg. p. 977-984.
40. Robert D. Green, K.F.M., Chin-Chang Ho, Sandosh Vasudevan, *Sensitivity to the proportions of faces that vary in human likeness.* Computers in Human Behavior, 2008. **24**: p. 2456-2474.

41. Masayuki ARAKAWA, T.S., Naoko ARIGA, *THE INTERFACE DESIGN OF ROBOT WITH HUMAN EMOTIONAL CHARACTERISTICS*, in *International Association of Socirties of Design Research*. 2007: HongKong.
42. Tiffany, D.D., *Design of a “Robotic Face” to Emulate Facial Expressions*. BEng Thesis 2009: National Univerity of Singapore.
43. Mechanisms101. *Cam Shaft*. [cited 2009 Aug 15]; Available from: http://www.mechanisms101.com/camshaft.html.
44. Expressions, G. *Geometry Expressions 2.0*. [cited 2010 Jan 16]; Available from: http://www.geometryexpressions.com/.
45. Kristof Goris, J.S., Bram Vanderborght, Dirk Lefeber, *The huggable robot Probo: design of the robotic head*.
46. Fumio Hara, K.E., and Shingo Shirata, *LIP-CONFIGURATION CONTROL OF A MOUTH ROBOT FOR JAPANESE VOWELS.* IEEE International Workshop on Robot and Human Communication, 1997: p. 412-418.
47. Meng Qingmei, W.W., Zhong Yusheng, Song Ce, *Research and Experiment of Lip Coordination with Speech for the Humanoid Head Robot-"H&FrobotIII*, in *8th IEEE-RAS International Conference on Humanoid Robots* 2008: Daejeon, Korea. p. 603-608.
48. Hideki, K., *Mechanism and Driving Method for Robot Lips.* Journal of the National Institute of Information and Communication Technology, 2006. **53**(2).
49. Mark Scheeff, J.P., Kris Rahardja, Scott Snibbe, and Robert Tow, *Experiences with Sparky, a Social Robot.*
50. Wikipedia. *Fused Deposition Modeling*. [cited 2010 May 06]; Available from: http://en.wikipedia.org/wiki/Fused_deposition_modeling.
51. Smooth-On. *Ecoflex Series*. Available from: www.smooth-on.com.
52. Society, S.r. *What is a Servo*. Available from: http://www.seattlerobotics.org/guide/servos.html.
53. Frye, J., *SSC-32 Manual*. 2010.
54. Wikipedia. *Speech synthesis*. [cited 2010 April 3]; Available from: http://en.wikipedia.org/wiki/Speech_synthesis.
55. Microsoft. *Microsoft Speech API (SAPI) 5.3*. [cited 2010 May 5]; Available from: http://msdn.microsoft.com/en-us/library/ms723627(VS.85).aspx.
56. Inc., W.C. *AIRcable Serial3*. Available from: http://www.aircable.net/serial.html.