Research on an Intelligent Transformable Phone Robot: BaBi

（変形可能な知的ロボット BaBi に関する研究）

September 2021

LIU Jiansheng

Abstract

Robots are becoming more widespread. In recent years, products and researches that utilize smartphones as parts of robots have been actively promoted. With the increases in the interface boards and sensing, a smartphone with great computing power can perform complex tasks, and it has been a familiar device in our life. However, the availability of humanoid robotics which provide multimodal interactions and varied express, lags behind the smartphones. Therefore, we can use smartphones to facilitate the robot applications development and reduce hardware costs.

This paper proposes an intelligent transformable phone robot named BaBi and develops its prototype. BaBi equips a smart phone, transforms from the smart phone to a movable robot when calling, "open", and performs more varied express to extend current smart phone functionality to be far-field, context-driven and multimodal interactions. This paper proposes a new robot form that can establish both autonomous mobility and portability and presents the implementation of the proposed transformable robot. We have developed a semiconductor chip that controls wheels and servomotors of the proposed robot shape to reduce the robot footprint about the same as a smartphone. A preliminary questionnaire survey has been conducted to investigate the effectiveness of the proposed robot form. BaBi is a portable, transformable, movable, intelligent partner.

ABSTRACT

This paper also presents an interactive method for designing robot facial expressions and motions of the phone robot BaBi. Designing an objective function to evaluate facial expressions and motions suitable for BaBi with an original structure is quite hard although users can easily imagine favorite motions. Therefore, the proposed method employs Interactive Evolutionary Computation (IEC) in addition to general optimization. It is an IEC process when users make candidate selections, while it is a non-Evolutionary Computation (non-IEC) process when users do not make any selection, resulting in being a fusion of IEC and non-IEC. Experimental results have shown that the proposed method produces the robot motions effectively.

Key Words: Phone Robot, Multimodal Interactions, Transformable Robot, Interactive Evolutionary Computation, Non-Interactive Evolutionary Computation, Robot Motion Design

Contents

Contents

Chapter 1 Introduction

1.1 Background and Objective of the research

Due to the development of artificial intelligence, machine learning, big data processing and pattern recognition technologies, intelligent robots are receiving more attention than ever before. Industrial robotics have been applied to various industries, work instead of human, and have improved production efficiency largely [1]. However, intelligent robotics are still under development in early research phases with limited trickle down towards products [2] because they demand human-level intelligence, varied express, multimodal communication, and interactions, resulting in great technical difficulties. Intelligent robots need to be ubiquitous and low cost

Human-machine interaction is a key technology of intelligent robotics. Besides verbal languages, people use many kinds of interactions such as facial expressions and gestures. However, in the real world most of products only consider new sensors, and do not support dynamic feed-back beyond 2D displays, resulting in dry conversations. It is very hard to establish an emotional sympathy between those kinds of boring relationship resulting in failure at building long-term intimacy.

Figure 1: BaBi Robot

On the other hand, smartphone has great computing power and can do very complex tasks with the increases in the sensing, and the recent availability of interface boards. It has become a very familiar device in our life. For example, we wake up from sleep with alarm, exchange text messages, take pictures and read news. Humanoid robotics provide varied express and multimodal interactions; however, their availability lags behind the latest smartphones. Therefore, smartphones can be used to reduce hardware costs and facilitate the development of robot applications [3-6]. Phone robots have changed the emotional frame where people interact with robots and phones and explore the boundaries of human-robot interaction in daily life.

It is possible to realize the robot's mind using a smart phone, because low price is crucial when developing a community robot partner. Because smartphone has been equipped with various sensors, it is enough for robot to just be equipped using cheap range sensors. In order to display the facial expression of the robot, the smartphone's display can be used.

Phone-based robots have been widely developed. They are divided into relatively large robots (Emoto by Carnegie Mellon University, USA [7], CALLY by Simon Fraser University, Canada [8], etc.) and small robots (RoBoHoN by Sharp[1], MobBon by Cevinius[2], etc.), but there is no phone robot having the portability at the same level as a smartphone.

In this paper, we developed an intelligent transformable phone robot as an intelligent assistant partner: BaBi, which has both autonomous mobility and portability at the same level as a smartphone. When closing, BaBi is a rectangular box with a smart phone size. We consider it as a moonlight box as shown in Fig. 1. When we open it, it will provide us more services and more convenient functions to bring us humanized use. Therefore we consider BaBi likes a treasure box, when opening it we can obtain a lot of goods from it and so call it moonlight box that is "月光宝盒" in Chinese. It automatically open to transform from the moonlight box to a robot when its owner calls, "open," and then it performs varied express and multimodal interactions with people as an intelligent partner to extend current smart phone functionality to be far-field, context-driven. BaBi is compact and portable, transformable, does not add more burdens. People operates phone through voice conversation,

[1] RoBoHoN, https://robohon.com/

[2] MobBob, http://www.cevinius.com/mobbob/

it can free hands and does not delay doing anything else. BaBi is interesting and funny.

It is very time consuming and hard to design a satisfying face and body motion for the robot to perform rich express such as pleasure, sadness, anger, surprise. It is effective to apply a machine learning method to help generate robot motions. However, designing an objective function to evaluate motions is quite hard although users can easily imagine favorite motions, resulting in solution evaluation performed with human preference, experience, and knowledge.

Interactive Evolutionary Computation (IEC) is an approach where solution evaluation must be performed by human users [9]. There have been various proposed IEC applications such as robot gestural expression [10], 3D face graphics [11]. However, in IEC approach users are required to make evaluation in every generation resulting in a serious issue of user's fatigue.

This paper proposes a method to solve the problem of the three-dimensional robot motion design of the robot phone BaBi based on an approach combing IEC and non-IEC [12, 13, 14]. A Case-Based Reasoning (CBR) approach [15] is used for user preference estimation in order to reduce the burden on users of evaluation. The proposed method allows users to add candidates to a case base when finding satisfying candidates and evaluates candidates' scores according to the case base. It applies a Different Evolution (DE) method [16] to generate new candidates. Users may alternately push the buttons to generate robot motion without making any selection of candidates. The method will make a selection with the scores when users do not make any selection of candidates. Users can select or not select. It is an IEC process when users make selections, while it is a non-IEC process when users do not make any selection, resulting in being a fusion of IEC and non-IEC. Adding more candidates to the case base can speed up generating the desired motion and promote the non-IEC progress and reduce the user workload.

The main contributions of this paper are as follows:

The new robot form that can establish both autonomous mobility and portability was proposed.

Implementation of the proposed transformable robot was shown. We developed a semiconductor chip that controls wheels and servomotors of the proposed robot shape. The chip makes it possible to reduce the robot footprint about the same as a smartphone.

A three-dimensional robot motion design method of the robot phone BaBi based on an approach combing IEC and non-IEC is proposed.

Preliminary questionnaire survey was conducted to investigate the effectiveness of the proposed robot form. Experimental results have also shown that the non-IEC approach improves generating the desired motion and reduce the user workload, and the method by fusing non-IEC and IEC approach promote generating the most desirable solution with less user workload.

1.2 Organization of the thesis

This paper is organized as follows:

Chapter 1 presents the background of this paper and the target problems. The features, originality, and main contributions of the proposed robot compared to existing phone robots are described. The differences between the previous methods and the IEC and non-IEC fusion method for designing the facial expressions and motions of the robot are also described.

Chapter 2 begins with the related research. First, it gives a brief outline of the related research for the communication robots and phone robots. Then, IEC methods used in previous studies and this study are explained.

In Chapter 3, the key idea and design for the proposed robot BaBi are described.

Chapter 4 presents the implementation. First, the implementation architecture of the proposed robot is presented. Then, the functions of the robot are described, where we prototyped a set of interactions and scenarios that users would be able to engage with.

Chapter 5 describes the IEC and non-IEC fusion method for designing the facial expressions and motions of the robot, where the basic ideas of the proposed method, the structure and process flow, chromosome representations, estimation of user evaluation value by case-based reasoning, and user interface are presented.

Chapter 6 presents the evaluation results. The evaluated usability results show a high tendency of positive results, where 67% - 90% of the participants very agree or agree with the scenarios. For ICE and no-ICE fusion method, experimental results have shown that the proposed method helps users to design robot motions with less workload.

In Chapter 7， the results of this study were summarized for Chapters 1 to 6 described so far.

Chapter 2 Literature Review

2.1 The description of Communication Robot

In recent years, many companies and researchers have been developing many different kinds of robots, which are expected to contribute to our lives. These robots should be able to communicate with people, however the communication capabilities of current robots are poor when compared to humans.

In the robot software development, what we most want to achieve is the function of human-computer interaction (HRI) as shown in Fig. 2, including these key technologies of voice conversation, emotional expression and motion expression. The key technologies to realize the voice conversation include speech recognition, conversation generation

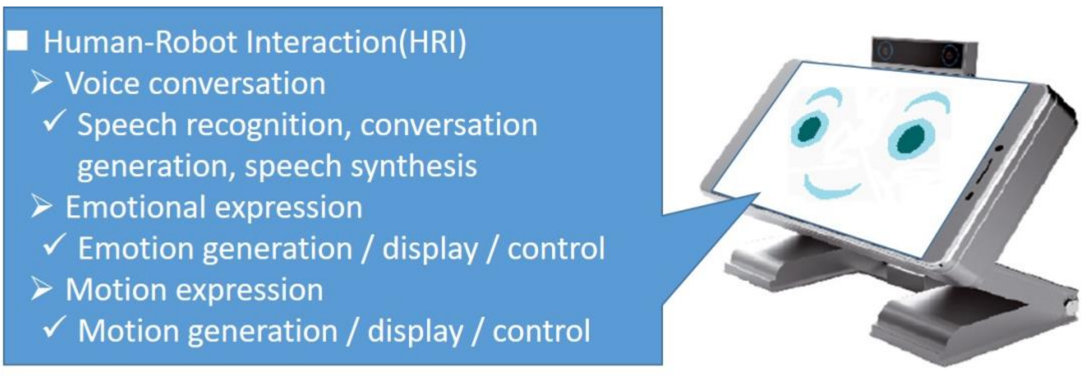


Figure 2: The Big Picture in Robot Software Development

and speech synthesis, while those to realize the emotional expression are emotion generation / display / control and those to realize the motion expression are motion generation / display / control.

Recently, the utilization of Robot Operating System (ROS) [17] has been attracting attention in robot software development. ROS is a type of middleware and software framework that runs on existing operating systems. ROS aims to promote worldwide collaborative development of robot software. The first release version was released in 2010. It has been developed and released as open-source software, and many people from all over the world are participating in the development.

The OSs on which ROS runs are mainly Linux such as Ubuntu and Linux Mint, and some functions are also supported on mac OS, Windows, and Android. It provides hardware abstraction, low-level device control, general-purpose function implementation, message communication among processes, package management, and tools and libraries for software development and execution. Designed as a distributed computing system, it performs various calculations required for robots in parallel by multiple processes. Each process can run on a single computer or on multiple computers and operate as one large software system, by sending and receiving processing results to and from each other. ROS supports many programming languages such as C ++ and Python, as well as Lisp and Java. In addition to mobile robots and robot arms, robots equipped with ROS include humanoids, autonomous vehicles, unmanned aerial vehicles (drones), and autonomous unmanned submersibles.

Most communication robots are developed using ‘model-based’ techniques. They have their own models (scenarios) of communication and can only communicate with people based on the models. Although these techniques are currently suitable to achieve the context sensitive communication, such as verbal communication, the variety of behaviors is physically limited.

Recently, the rate of elderly people is increasing in the population. Elderly people need to be cared about giving interactive communication support. Since the trained caregivers are not enough, other solutions have to be found to solve this problem. The introduction of human-friendly robot partners in the community is one possible solution.

Community-centric robot partners need interaction with human. The interaction can be verbal, and also non-verbal, including gestures and facial expressions as well. Gestures and facial expressions are important in natural communication.

The relationship between communication and emotion was studied [18] and the robot partner’s emotional based communication was also investigated [19]. In these proposed methods, the gestural and facial communication with the robot partner were performed by using emotional model, where the robots performed gesture recognition and face classification through an emotional model to communicate with people.

Chapter 2 Literature Review

It is possible to develop a facial and gestural expression system, which uses the robot's emotional state as an input. Facial expression is an important element of non-verbal communication. The robots also express gestures, however, these gestures were preprogrammed in advance. In the gesture generation, Laban theory was applied for a simulator robot's gesture analysis and estimating its emotional state in [20].

During the past half-century, HRI systems have become an active topic in the field of robotics. Many researchers have been keen to develop an intelligent robotic system to mimic and understand human emotion and behavior. Key to such systems is the outward expression or communication of a robot, which is essential to conveying intent or inner emotional state [21]. For example, an autonomous mobile robot which can interact with humans via hand gestures was proposed in [22].

Systems enabling robots to interpret human emotion are also being investigated. A system has been proposed for recognizing human emotion based on facial expression [23], where an approach has been extensively studied. However, facial expression is not the only means by which human feeling is revealed. Voice [24] and body movement [25] also figure prominently in expression of emotion. Machine learning methods are now being used to identify and classify emotional states [26] which measured position, velocity, and acceleration of certain body joint movements of the design team members. According to it, because of the complexity of human motion the collection of large amounts of joint data did not always lead to improved system performance, and the selection of joints to track needed to be studied further. One of the most known well techniques for describing and interpreting human movement is Laban Movement Analysis (LMA). It was showed that the Effort and Shape variables in LMA can be used effectively to encode emotional states [27]. This suggested the possibility of using LMA to generate robot behavior capable of revealing its emotional state. A robot with expression capabilities could communicate more successfully [28] that concluded that a robot in physical form is better suited for interaction with humans than a bodiless computer program.

The following shows some examples for developed communication robots in details.

Figure 3: Talking-Ally Robot [29]

(1) Talking-Ally [29]

As shown in Fig. 3, a social robot utilizing spontaneous speech was systematically shaped who was planning the conversation, while gaining attention coming from the addresser, refreshing the conversation as well as error-handling.

In Talking-Ally [29], a similar mechanism was utilized where the resource of hearership and addressivity were considered by connecting with entrusting behaviors for enhancing the conversation engagement.

This study has discussed about the concept for novel utterance generation mechanism and performance for human-robot interactions.

Talking-Ally has a pretty design with a pretty existence such as babies or small animals. It is possible for Talking-Ally to perform turning the face, nodding, as well as tilting the body backward or forward and so on.

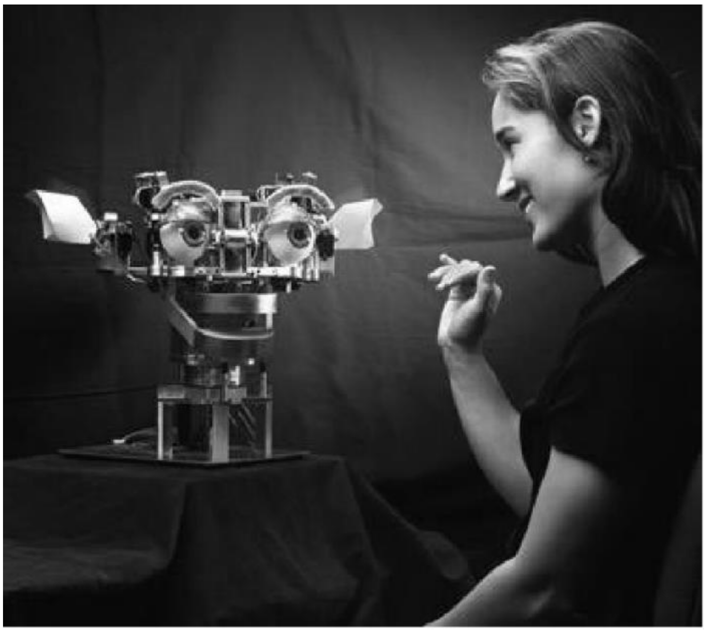


Figure 4: Kismet Robot [30]

Talking-Ally has a hearer's gaze recognition system by using an eye-gaze tracking system (faceLAB 5.0).

(2) Kismet [30]

Berazeal et al. studied on integrating attention, perception, emotions, drives, expressive acts and behavior arbitration for a robot called Kismet as shown in Fig. 4, which was designed for interacting socially with humans [30].

It was based on a model for human visual search behavior to design a visual attention system. The attention system integrated perceptions including color saliency and motion detection, with influences and habituation effects from the robot's behavioral and motivational state to

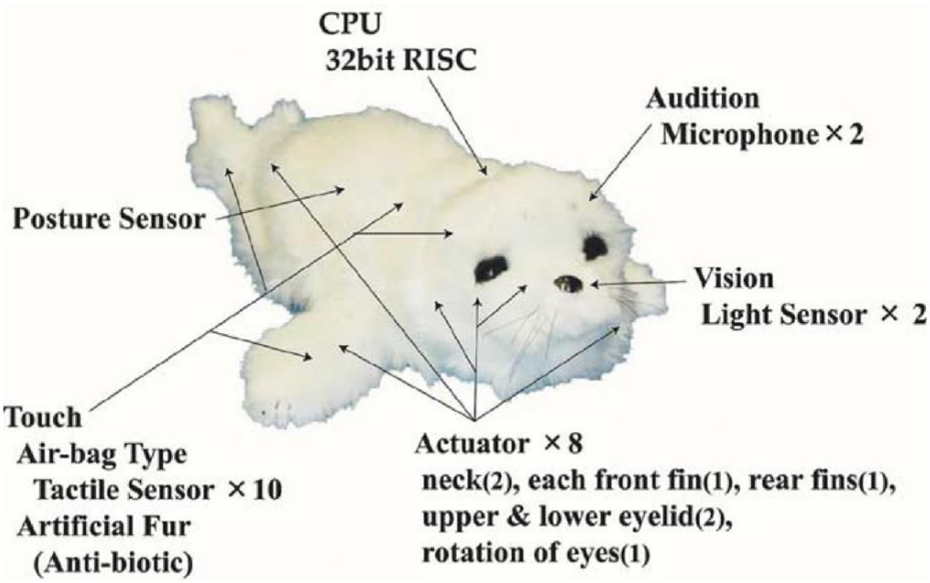


Figure 5: Seal Robot Paro [31]

create an attention activation map of context dependent. This activation map was used for directing eye movements and satiating the drives for the motivational system.

Kismet was a robot designed to interact with humans socially. It has an active vision system and it is possible for Kismet to display a variety of facial expressions.

Kismet used many visual feature detectors (motion, color, and face detectors) and combined them with a habituation function for producing an attention activation map.

The attention process influenced eye control and the robot's internal behavioral and motivational state.

(3) Paro [31]

Seal robots as shown in Fig. 5 have been applied for assisting the activity for the elderly at someday service centers [31].

This study has investigated the effects on elderly people for seal robots, by evaluating the mood of elderly people using face scales and questionnaires. The face scales identified the appropriate facial illustration to express moods. They measured the changes in reaction to stress of the elderly by using urinary tests. The stress in nursing staff was also investigated. They used a “burnout scale” to evaluate their mental state.

They provided the day service center the seal robots for five weeks and obtained a result that interacting with the robots has improved the feelings on elderly people, it has been showed by urinary tests that the ability of the people for overcoming stress has also been improved, and the stress levels in the nursing staff has decreased since less supervision was required by the elderly people when they interacted with the robots.

The seal robots have been judged to be useful for the elderly at institutions like the day service center.

(4) Elderly care support services based on a cloud robotics platform [32]

This study has discussed an implementation for elderly care support services using a cloud robotics platform that facilitated cooperation between IoT devices and focused on communication robots [32]. The study considered two issues to affect the care support service of elderly that are technical issues about the cloud robotics platform, and issues with the services operation.

The cloud robotics platform processed the sensing data that was received form IoT devices when connected via the network. It collected and analyzed this information for making decisions and actuating IoT devices according to the results and kept down the robots' cost.

The experimental results presented that the communication robot was responded favorably to by the care recipients, they had enjoyed the daily conversations while finding its presence reassuring, and the robot allowed the nursing care staffs to watch over their elderly constantly, result in being popular with the nursing staff.

(5) Elderly support by robot and sensors with cloud robotics [33]

This study was subject to two people (an 87-year-old female and 104-year-old female), and four nursing, and conducted a preliminary survey [33].

They checked the living conditions for the two elderly subjects using used a rhythm sheet of resident life, and also used a record sheet to analyze the state for the staff members to observe the contents on their work. It also investigated quantitatively a variety of psychological stress, including tension felling, depression, vigor, anger, confusion and fatigue by using the Visual Analogue Scale on the Profile for Mood States.

The frequency for nurse calls (NC) had been investigated, which is considered as one of the main factors to cause stress to care workers.

The robot and sensors were introduced into the care for the 77-year-old subject for four nights that evaluated their effects. Real-time sensing data was obtained, and they were sent to the robotics cloud platform by the data I/O devices. It also sends the conversations between the robot and the subject to the cloud platform by the robot I/O device. The system subsequently analyzed and controlled them.

(6)R100 Robot [34]

A personal robot R100 as shown in Fig. 6 was developed with autonomic actions [34]. The robot used the obstacle detection result by image recognition and the distance information by the ultrasonic sensor to automatically move to a vacant area. When discovering people, it autonomously approached and talked to him.

The robot made interaction with people, and by a sound direction detection function it turned to the direction when called. It approached the people and talked to him after detecting people by image recognition and measuring distances by stereo vision. It can also be given instructions by speech recognition.

Its reaction varies depending on the situations. By emotion model, feeling for each person, sleepiness and other internal states, the reactions changed according to partners, situations, and time.

R100 had a variety of information communication functions such as audio remote control, messages, mail transmission, and internet browser display.

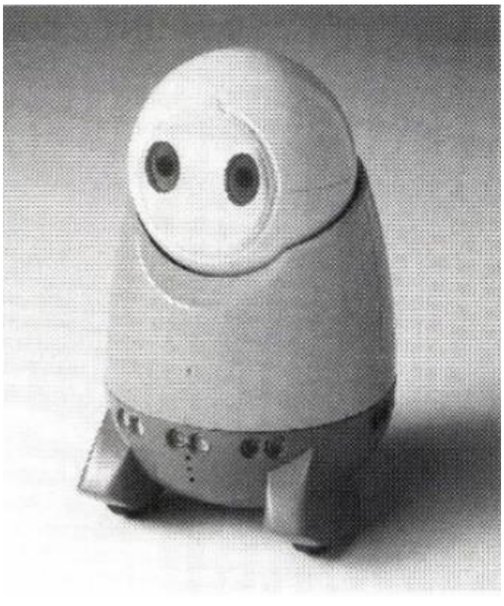


Figure 6: R100 Robot [34]

(a) Humanoid robot (b) Cart robot

Figure 7: Chat Robots for Assisting Elderly's Shopping [35]

(7) Chat robots for assisting elderly's shopping [35]

The rapid advancement of robotics has enabled the robots to serve physical help for elderly people in everyday life. This study investigated what types of robot elderly was willing to go shopping with [35].

It investigated the effect into two factors: conversation and robot-type. Conversation means the no purpose talking which human often do. About robot-type, it investigated two robot-types: humanoid and cart robot as shown in Fig. 7.

A field experiment was conducted in a real supermarket where 24 elderly participants shopped with robots. The experimental results revealed that they preferred a conversational humanoid as a shopping assistant partner.

(8) Robovie robot [36]

In a study comparing a real robot and an agent on a computer, it was found that a robot existing in a real space is more suitable for the interaction with an object in a real space than an agent on a computer [36]. The presences of such communication robots in real spaces make experiences in interaction more realistic and possibly lead to learning support causing to be interested.

From the viewpoint of robot presence and interesting, this study outlined the possibilities for communication robots related to learning.

They developed a robot called Robovie with human-like motion and cognitive abilities as shown in Fig. 8. In Robovie various gestures used for communication with humans are possible with four degrees of freedom arm, three degrees of freedom head, and two degrees of freedom of eyeballs.

It is possible for Robovie to autonomously interact with humans by various sensors such as microphones, contact sensors, ultrasonic sensors, stereo cameras, omnidirectional sensors.

The developed robot was used for English education at an elementary school for 2 weeks among 228 students. The children actively integrated with the robot during the first week, but the idle time of the robot

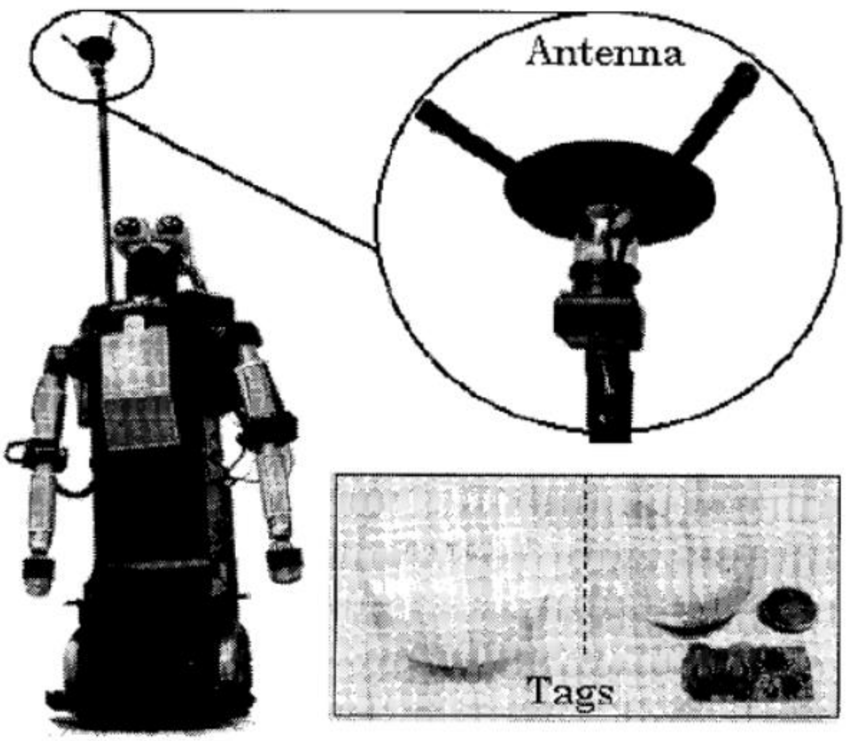


Figure 8: Robovie Robot [36]

gradually increased. It can be said that this showed the limits of current interactive robots and research issues.

(9) Embodied communication focusing on physical expression [37]

Figure 9: Subject Turning his Eyes to the Robot [37]

This study proposed a framework of embodied communication focusing on physical expression [37]. It aimed at clarifying the mechanism of communication based on the relationship which emerges by the entrainment resulted from mutual physical actions.

Psychological experiments were focused on a correlation between

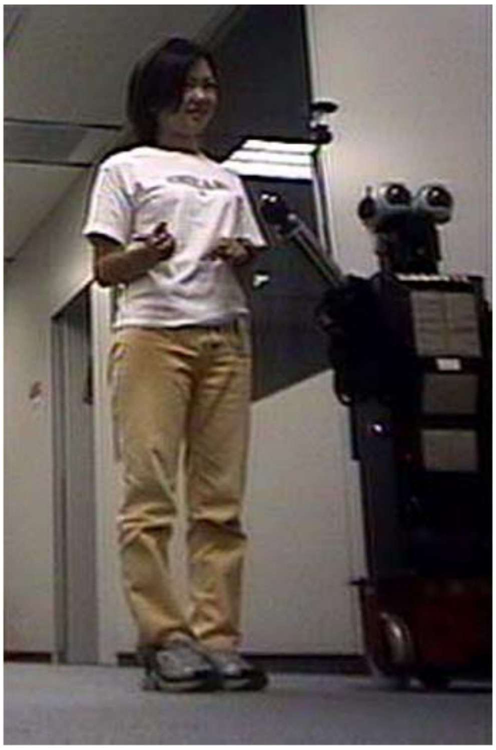


Figure 10: Subject Turning his Eyes to the Same Direction with the Robot [37]

physical expression and utterance understanding in human-human and human-robot interactions. The results showed that subjects communicated with humans smoothly only when the robot practiced sufficient physical expression entrained subjects' actions.

Fig. 9 showed an example of subject's doing physical expression where the subject was turning his eyes to the robot [37]. Fig. 10 showed an example of subject's doing physical expression where the subject was turning his eyes to the same direction with the robot. Fig. 11 showed an example of entrained physical expression in human-human interaction

(10) Communicative robots with infra-red radiation monitoring system [38]

The impacts of robots and ICT in care settings had not been sufficiently evaluated, particularly in relation to workers' workload. The purpose of

Figure 11: Entrained Physical Expression in Human-Human Interaction [37]

this study was to evaluate quantitatively the effects of introducing communicative robots on night shift duties of nursing facility workers [38].

The subjects were five late-night care works, looking after elderly people to whom communicative robots with an infra-red radiation monitoring system were introduced.

It investigated fatigue level using the "method for checking subjective symptoms". Assessments were conducted at the beginning of the night shift (16:30), before dinner (20:00), before nap (00:00), after nap (02:00) and at the end of the night shift (09:50) before and 4 weeks after

Figure 12: Monitoring System of Care Recipients's Movement by an Infra-red Radiation Camera [38]

introduction. The effects of the system on the nighttime work burden and frequency of accidents were compared.

During the fourth week, compared with the pre-introduction phase, the total fatigue level was improved before nap, after nap, and the end of the night shift.

Along with this improvement, the accidents especially the trauma accidents decreased. The introduction of a communicative robot with a monitoring system significantly improved the total fatigue level of midnight nursing care workers and alleviated the nighttime work workload.

Fig, 12 showed care recipients' movement on the bed such as sitting upright, turning off light and leaving bed were monitored by an infra-red radiation camera, where messages were sent to the nurse center as well as to mobile devices of care workers.

Fig.13 showed the conversation robot COTA coupled with the monitoring system that produced appropriate pre-programmed response.

Fig. 14 showed the silhouette imege capured by infra-red rediation camera for that in Fig. 13.

Figure 13: Conversation Robot COTA Coupled with the Monitoring System [38]

Figure 14: Silhouette Imege Capured by Infra-red Rediation Camera for That in Fig. 13 [38]

2.2 Systemic-functional approaches on Phone Robot

In the intelligent robotic systems, it is necessary to provide an intelligent voice interaction function that can assist and support people as an intelligent assistant partner. Siri (by Apple) [39] and Cortana (by Microsoft) [40] are such intelligent assistants. However, in human-machine interaction, there is a lack of means for communication and interaction, where they are limited inside or on the surface of phone.

Table I gives a brief outline of the related researches below.

Examples of more phone robots can be seen such as Emoto (by Carnegie Mellon University, USA) [7], CALLY (by Simon Fraser University, Canada) [8], RoBoHoN (by Sharp) [1], MobBon (by Cevinius) [2], Smartpet (by BANDAI)[3], ROMO (by Kickstarter)[4], Tovbot's Shimi (by Georgia Tech’s Center, USA)[5] as shown in Fig. 15-21.

TABLE I: FEATURES OF SMARTPHONE-BASED ROBOTS

Robots	Physicality	Transformability	Portability (compactness)	Autonomous mobility
Siri,Cortana [11-12]	✗	✗	✓	✗
Emoto [7]	✓	✗	✗	✓
CALLY [8]	✓	✗	✗	✓
RoBoHon[1]	✓	✗	✓	✗
MobBob[2]	✓	✗	✓	✓
Smartpet[3]	✓	✗	✗	✓
Romo[4]	✓	✗	✗	✓
Tovbot’s Shimi [5]	✓	✗	✗	✓
The,proposed robot; BaBi	✓	✓	✓	✓

---

[3] Smartpet, http://sp.asovision.com/

[4] ROMO, https://www.kickstarter.com/projects/peterseid/romo-the-smartphone-robot-for-everyone

[5] Tovbot's Shimi, http://mshci.gatech.edu/hg/item/137351

Chapter 2 Literature Review

(1) Emoto

In Emoto it is reimagined how we relate to smart home devices as shown in Fig. 15 [7]. The robotic platform gives body and expression to an AI of a future operating system. We imagined it in an ecosystem less about individual apps, but more so the skillsets of the AI sidekick.

Emoto has explored some of the following concepts + interactions.

Figure 15: Emoto [7]

When you get home, everything else has its place: the coat on the hanger, shoes on the rack, and keys by the door. But where is your phone? It is in your hand and on your mind, in bed, at dinner, or on the bathroom. It is referred to placemaking as the process of finding/making a place of belonging for an idea, device, or service in someone's life. Placemaking includes considering people's behaviors and habits, not just physical where and how. Intentional placemaking contributes to how you frame interaction.

There might not be a specific place for your phone, but you'll always need to charge. The robot's magnetic mount is also a wireless charger. It is not just a touch screen when leaving your phone on as Emoto; the responsive AI character enables far-field interactions, so skillsets are accessible hands free. It extended sensing for the home-assistant. Computer vision enables more contextual awareness such as object tracking for smart filming.

Facial recognition allows for tailored behavior and information for each person as the phone becomes a shared device for the space. Emoto is responsive to speech and sound input. Using the phone like a home-assistant means more interactions begin with declared, and purposeful intent, and augmenting phone capabilities. Emoto can catch your attention for a timely notification or dismiss one because the robot knows you saw it. Notification is been handling with contextual awareness. When not being actively engaged, the robot is a clock and calendar. It fits with phones having become the default alarm clock and wristwatch. Motors and computer vision allow you to video call handsfree. The camera's viewport is option ably controllable from the other caller's end phone as a character. Nonverbal communication (gestures, eye gaze, and body language) can go both ways between you and your phone.

Another person might know better than to interrupt conversation with irrelevant info. Can we hold personal AI to the same standards? We can try because Emoto is open to learning. The levels of access to skills and information for each person on the phone are determined by how they relate to the robot sidekick.

(2) CALLY

CALLY has been developed as a robot cellphone. And been explored the roles for gestural and facial expressions of robotic in human computer interaction as shown in Fig. 16 [8]. In CALLY non-verbal anthropomorphic affect features were introduced as media to build emotional intimacy between a product and a user. Two ideas of social robot application which were generated from workshop for initial participatory and brainstorming design have been presented with their implementations and usage scenarios. From the pilot test it was learned that the bodystorming ideation technique and prototyping enabled participants to take part in generating new ideas more actively on design of robotic products.

Figure 16: CALLY [8]

(3) RoBoHon

RoBoHoN is a Japanese robot phone, where the mobile phone has a cute little robot which can talk and walk as shown in Fig. 17 [1]. The head has a small projector that can share photos. It's not as heavy (390g) and as big (19.5cm high) as ordinary mobile phones, while an ordinary mobile phone doesn't look like a cute robot.

On it there are almost all the common functions which a mobile phone should have, such as LTE, Wi Fi, face and voice recognition. It has a little small screen, which has a resolution of 320 x 240 and 2 inches. On it there are also functions which an ordinary mobile phone doesn't have, such as sitting, standing, dancing, and raising hands. This is a robot after all. It is possible for RoBoHoN to be a more intuitive and friendly "Siri" to make more humane human-computer interaction.

(4) MobBob

Figure 17: RoBoHoN[1]

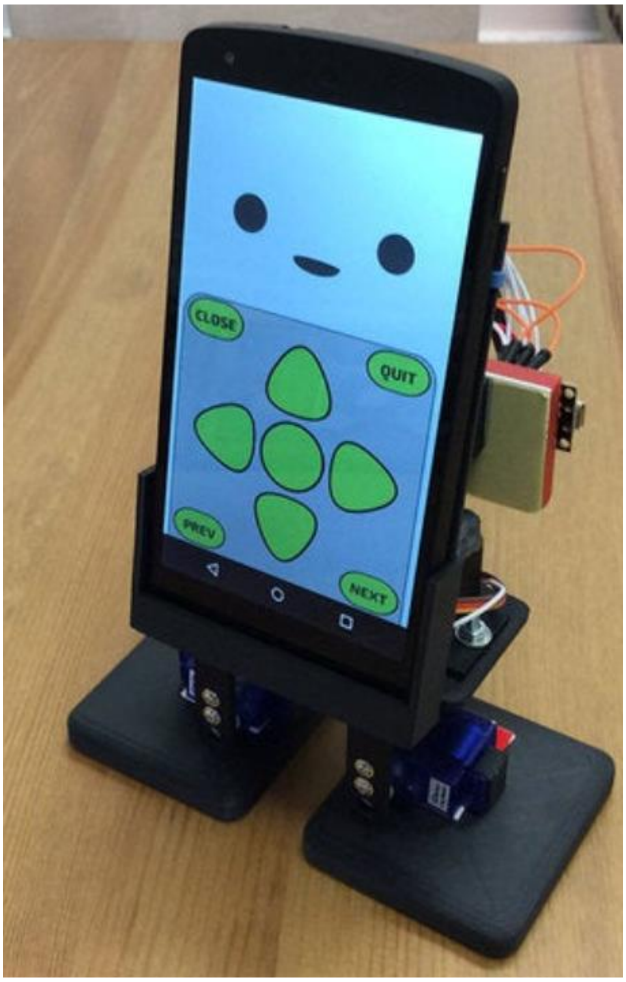


Figure 18: MobBob[2]

MobBob is 3D-printed robot that is controlled by a smart phone, which was inspired by the Biped robot BoB that is popular and 3D printable as shown in Fig. 18[2]. The developer used the Unity game engine to write MobBob’s app, and his features are including speech, computer vision for colored ball detection and face detection, speech recognition for commands and conversation, interaction by a touch screen interface, using screen face to emote, using movement to emote, walking around. The phone uses Bluetooth to communicate with MobBob’s body. The robot doesn’t have many parts, and it is possible to build it easily and cheaply.

(5) Smartpet

Japan's Bandai's Smartpet can make your smartphone around you anytime, anywhere as shown in Fig. 19[3].

Figure 19: Smartpet[3]

Figure 20: Romo[4]

At the 2012 Tokyo International Toy show, Smartpet can transform your smartphone and iPod touch into a robot pet dog. After transformation, the robot dog can walk around (maximum speed: three steps per second), sit up and wag its tail.

This app can display all kinds of cartoon pictures on the touch screen of mobile phone. As long as you can think of it, it will realize your wish as much as possible. Users in Japan are able to spend about $100 to get the product, with a choice of white or black.

Smartpet can also recognize voice and gesture commands. When the user touches it, it can make more than 100 reactions according to the user's stimulation. This app enables you to interact with the robot dog. You can give it virtual food and play virtual games. At the same time, it can complete its own personality growth. Although there is a certain gap between this app and Sony Aibo in terms of function, it is quite popular and creative in the exhibition.

(6) Romo

Romo does not cost a fortune where users are not required a computer science degree to program as shown in Fig. 20[4]. The designers of Romo made efforts for making programming fun, as well as making it easy to have a Romo robot in home on the bank account.

Programming Romo for taking a picture when Romo sees a face or getting angry when someone is too close is very as easy as using an smartphone. Romo’s brain is served as by a smartphone and anyone can control it by a computer or compatible device in the world from anywhere. The robot base with a NiMH battery is built. The battery can be used for 2 hours on a full charge, and we can recharge it via any mini-USB cable. Like other iPhone docks without moving or making funny expressions, the smartphone can also be charged by Romo. It consists of hardware and software. The hardware includes the electric board, wheel, smartphone,

Figure 21: Tovbot's Shimi[5]

lithium battery, interface and so on. The software is composed of some mobile applications, and developers can develop applications to extend the robot functions.

There are three applications Romo has been developing: Romo remote, Romo kart and drag and drop programming module. The Romo remote remotely control the robot. The situation around Romo can be viewed through the camera on smartphone that realize the spy function. The Romo kart is an attack application, through which attack commands can be sent to robots. Using the drag and drop programming module, programing can be done directly on smartphone and it can be watched that

Romo executes the commands. We can use smartphones, remote control toys and so on to achieve more advanced evolution.

(7) Tovbot's Shimi

Google I/O gave Tovbot's Shimi as shown in Fig. 21 the first public appearance[5]. Perfect coordination is performed by three Shimis. A group of robot researchers from IDC in Israel, Georgia Tech and MIT Media Lab have developed Tovbot earlier, to promote a new field for personal robots which clean floors or pool, and interact with people on an almost personal human level. A news on the website of Georgia Tech has showed a musical companion that Georgia Tech's Center developed for recommends songs, Music Technology, dances to the beat. Based on listener feedback it can also keep the music pumping.

Robots are loved by Google, where there are real robots including cloud-based robots, self-driving cars and so on that are supposed to be confidential. Therefore, it's not a surprise when a bunch of robots are invited by Google to a conference of developers. Shimi is one of them that is a musical robot and smartphone-enabled. Georgia Tech developed it and Google I/O unveiled it today.

Basically, Shimi is a robot with interactive speaker function. When a smartphone is docked on the bot, it will play songs and make recommend tracks and it also dances to the beat. On it there are remained other speaker robot hybrids such as the musical robot OLogic AMP and the "Bieber Bot" TOSY shown at CES. At Google I/O, there are three Shimi bots that performed together.

Using image recognition technics and the phone camera, it can be performed by Shimi tracking position of speakers by optimal sound and a user around room. When a beat is tapped or clapped, user’s musical library will be searched by Shimi for tracks matching the rhythm. A team (led by Professor Gil Weinberg, who was a director at Georgia Tech’s Center for Music Technology) have developed the robot. Weinberg and his colleagues have a plan that commercialize Shimi through Tovbot (a new startup company, based in Atlanta).

Future works for Shimi is developing the robot that is able to recognize gestures, including shaking user’s head when disliking a sound or waving user’s hand for skipping tracks or changing the volume. Tovbot also plans for allowing developers to create some new behaviors on the robot with an API.

When listening to music, new recommendations or a buddy for dancing along to new music is always desired by users. A musical companion robot called Shimi has been developing by a team from the Center for Music Technology of Georgia Tech. Musical suggestions and dancing among many other functions have been incorporated on the robot.

It is possible for Shimi to do unprecedented musical interaction. In Shimi there is a dock for android phones. By using the dock, it is possible for Shimi to tap into one of musical apps that you have and access the camera and mic. It is possible for Shimi to perform functions with a wide range such as using music making apps on the phone for composing its own dance beat while Shimi “shimmies” to it. It is also possible for the robot to analyze beats that are clapped by the users and match the song from the musical library best fitting that beat.

The robot is very accommodating. It is possible for Shimi to get to know users’ musical tastes and suggest new songs or artists according to users’ current library and collected user’s feedback. Shimi can also track users’ position to make sure its speakers are in an optimal direction. Developers are still trying to come up with other ways innovative to interact. Some apps are also been developing where Shimi is allowed to perform functions such as turn up the volume or skip songs when looking at the user’s hand or head motions.

These phone robots tried to expand far-field interactions with smartphones; however, our proposed phone robot BaBi is different from them at the points of a smartphone compact size and a transformable phone from a smartphone to a robot.

2.3 Current study on Interactive Evolutionary Computation

Metaheuristics are commonly used to optimize problems for which the objective function is a black box, where the steepest descent methods and quasi-Newton methods using gradients cannot be used. By using the metaheuristics, not only the convexity and differentiability of the objective functions are not required, but also it is possible to optimize even the problems that have a global multimodal objective function and noises in the design variables and objective functions. Even the problems with objective functions that are difficult to implement on a computer can be optimized and can also be applied to human-in-the-loop optimization using humans as an evaluator.

2.3.1 Evolutionary Computation

Evolutionary computation (EC) is a part of artificial intelligence (AI) and machine learning [41] [42]. The EC solves computational problems using biological evolution principles. It is possible for EC to solve a computational problem requiring following some convergence criteria on the problem. The optimization processes of EC attempt to improve an organism's ability to live longer in unstable environments and rapidly changing. EC processes include natural selection, reproduction and survival. In the processes of natural selection, all living creatures are the result from adaptation to environmental needs on earth. If a living creature has successfully adapted to such an environment, then it has a higher level of fitness because it has a greater chance of surviving. The entities with higher survival of fitness have more chances for spreading their genotype to generation of future by reproduction.

EC employs a number of procedures based on evolutionary techniques and principles, which is the soul of EC.

Among these searching methods to find the problem solution, the natural evolution approach is the best option by far. The methods of EC are categorized into two main types: evolutionary algorithms (EAs) and swarm intelligence (SI).

The EA include the following subfields:

(1) Differential Evolution (DE)

DE algorithm is a stochastic optimization method that has been invented for minimization of an objective function under constrained certain [43]. DE can find the global true solution with fewer control parameters and faster convergence rate

(2) Differential Search Algorithm (DSA)

DSA uses the migration of stable motion of super organisms to solve the optimization problems of real valued numbers [44]. It is a multi-way method where the mid-point is used in journey of the migration of super organism.

(3) Evolutionary Strategy (ES)

ES uses mutation and selection for searchers, and it applies normal parameters of problems for representations [45]. In ES, generations' evolution uses an iterative manner that is stopped by termination criterion. For numerical optimization as a particular strategy, covariance matrix adaptation evolution strategy (CMA-ES) is one of the highest-performance algorithms at present [46].

(4) Genetics Programming (GP)

It is possible for biological evolution methods to discover or evaluate specific tasks of user on computers [47] [48]. GP can optimize a computer population to perform user tasks. Although it is impossible for GP to solve all complex problems, excellent outcomes have been produced by GP in computing.

(5) Evolutionary Programming (EP)

EP does not follow any constant pattern that is different from GP following fixed structures [43]. In EP all parents become generators because of extension of membership concept.

(6) Genetics Algorithm (GA)

GA uses a heuristic search of natural evolution characteristics to process natural selection [48] [49]. Automatic electronic circuit generating is a application of GA.

(7) Gene Expressing Programming (GEP)

In GEP, complex trees are used for making adaptable programs where all properties are almost changed [47]. It is possible for GEP to distribute genetic information using its genome, resulting in calling it genotype. Due to the adapting ability to the new environment, GEP is also classified as phenotype.

The SI include the following subfields:

(1) Ant Colony Optimization (ACO)

ACO is a meta-heuristic procedure for optimization. It was developed to find an optimal path of a graph based on observations where ants are seeking sources and colonies of food [50] [51]. In the observations, if one of the colony members discovers a short path where it has already benefited and tested from the path, other members converge to that path, resulting in concentrated ants' rally.

(2) Bees Algorithm (BA)

Like ACO, BA is also based on observations where bees search for food and share secret paths as food sources with other members [52]. When the food source does not have enough high quality or richness, the bees will stop advertising the food source. Therefore, BA uses global and neighborhood searches simultaneously for continuous and combinatorial optimizations. It is a search algorithm based on population.

(3) Cuckoo Bird Search Algorithm (CSA)

Cunning cuckoos will lay their eggs in other birds' nests that have similar patterns on their eggs to exploit the host birds for protecting all eggs to increase the chance to become chicks for Cuckoos' eggs [53]. CSA was inspired by the way and was developed as a simple optimization algorithm.

(4) Particle Swarm Optimization (PSO)

Inspired by fishes and birds, in PSO candidate solutions' population corresponds to swarm for particles changing their position based on some formulas for guiding the whole swarm for a possibly optimized and uncertain solution [53]. PSO is a computational method that is iterative and meta-heuristic to solve a problem. With respect to a quality measure it improves the solution possible correct, resulting in searching a very large possible solution space, even without assumption about the problem.

Swarm tries to move toward the center while trying to match its neighbors' speed and avoid collisions.

2.3.2 Interactive Evolutionary Computation

Interactive Evolutionary Computation (IEC) is an approach where solution evaluation mush be performed by human users [9] [54] [55]. The IEC is suitable when designing an objective function cannot be implemented on a computer and human must participate in the optimization loop as an evaluator. IEC approaches have been applied to hearing aid fitting [9], vector graphics images generation [56], and designing collective behavior of humans [57]. It has been combined to generative adversarial networks (GANs) for image generation [58].

IEC uses subjective evaluation from a real human in its optimization approach, where a real human is used for replacing the fitness function to get the problem converge to a real human's subjective evaluation.

The IEC involve a human user to optimize tasks in an IEC optimization. The IEC optimization framework includes three parts that are a real human, an IEC algorithm, and a target system to optimize.

The IEC main issue is the user fatigue when fitness is given by a real human. The problem of user fatigue occurs in the IEC optimization process, where it is necessary for a real human to give feedback repeatedly from the subjective evaluation to an IEC algorithm. This problem limits the IEC optimization application.

In IEC approach, some approaches to solve the problem of user's fatigue have been proposed. Ono and Nakayama have proposed a fusion of IEC and non-IEC for two-dimensional barcode decoration [12, 14]. They also applied their method to a problem of designing graph-structured image processing pipeline [13]. Suga proposed a human-machine hybrid evaluation to increase the diversity to reduce the burdens on both the robot and experimental subjects [59].

A main characteristic of IEC is optimization with subjective evaluation by human. Therefore, each EC algorithm has an IEC version when a real human's evaluation replaces the fitness function. Several EC approaches are used in IEC, including interactive genetic programming, interactive genetic algorithms (IGA), human based genetic algorithm, interactive evolution strategy, interactive differential evolution (IDE), interactive particle swarm optimization, etc. The IEC characteristics are shown as follows:

(1) To obtain the optimum it is necessary for optimization using IEC algorithm to use less generations and less individuals.

(2) It is difficult to get the global optimum. The global optimal solution is not unique.

(3) The fitness is a discrete and relative value.

(4) The fitness has noise. Since the subjective evaluation by a real human

is used for IEC optimization, user fatigue increases with the number of individuals and evaluation generation times

We need to design an IEC algorithm that has less generations and less individuals. A promising solution is to use machine learning techniques to build a human model, and replace evaluation of a real human using the relations between variables of search and their fitness.

When between two individuals the difference cannot been distinguished by a real human, IEC considers the individuals are almost same when conducting an IEC evaluation. This is why it is hard to difficult to get the global optimum. The global optimal solution is not unique.

The fitness is obtained by the comparisons of individuals with a result that the fitness is a relative value in IEC optimization. For the fitness in IEC with a continuous and absolute value, in the earlier generations the fitness on an individual is a worse value while in the later generations it is a better value. The fitness in IEC with a continuous and absolute value reduces the selective pressure in the IEC algorithm resulting in influencing the IEC algorithm convergence.

The fitness in IEC optimization with discrete value has noise. Because noise cannot be avoided in IEC evaluation, the IEC algorithm is sensitive to noise resulting in poor evaluation performance. Therefore, to improve its optimization ability it is necessary for the IEC algorithm to apply a noise filter processing.

2.3.3 Applications of IEC to Robotics

IEC has also been applied for robot motion design, where gesture expression based on LMA and IEC [59, 60], and interactive robotic behavior design by using IEC [61, 62] have been proposed.

For robotic expression at emotional states, Inthiam el al. applied a use on bodily movement [59], where a set of articulators at head, arms and trunk have been equipped on the robot for realizing such expression. A model taking normal human movements has been constructed. For motions of human upper body, a depth camera recorded observation. To generate corresponding robotic movement, the LAM method was applied to analyze the obtained visual data to obtain parameters. To evaluate and verify the proposed system effectiveness, they used the robot upper body to conduct experiments.

Woo et al. proposed a system of the expression of gestural and facial among robot partners [60]. Many elements such as embodiment and speech are necessary for interaction with a robot. For natural communication of the robot, it is important to use gestures and facial expressions. Based on LMA and IEC, a gesture expression system was proposed. They discussed the facial expression generation and conducted experiments for their proposed method and showed the experimental results.

Suga et al. described how to implement IEC for interactive robotic behavior design [61], where for a human-robot communication system IEC has been applied to a real robot, and IEC's application has been simulated. On the experimental subjects and robot, large burdens were led. To overcome the problem, an approach using human-machine hybrid evaluation was proposed, where the number of interactions within the genetic pool was increased while the diversity was increased. On the experiment the WAMOEBA-3 (a communication robot) was used, and the robot interacted with human assessors. The experimental results showed that the motions desired by assessors has been learnt by the robot, and they considered that it is suitable to use the IEC to the system of the communication robot.

Virčíková et al. presented humanoid robotics using artificial intelligent techniques, where an interactive robotic behavior was designed [62]. How the interaction between social robot and human was focused on. At the first step, the robot acquired knowledge by asking user to interactively cooperate with the robot. On the approach the robot learnt by interacting freely with objects and humans. The NAO humanoid robot was used in the experimental setup. At the second step, pleasant motions or dances were generated by using the technique IEC. Users can obtain the new possibilities of motions and create a behavior of robot based on the IEC approach not having any programming knowledge.

The applications to control of the IEC have recently increased. The first application to control of the IEC was the control of an insect-style robot with six legs by IGA in 1992 [63].

A similar application to control of a robot with eight legs has introduced the IEC [64]. An obstacle avoidance control for a Khepera miniature robot has been proposed to use IEC approach [65]. Their first evolving phase used the interaction between user's visual evaluation and the GA, while their second evolving phase used the interaction between a human evaluation model and the GA to reduce user evaluation fatigue. The human evaluation model was obtained by the user evaluation on the first phase.

Chapter 3 Research Procedures and Analysis

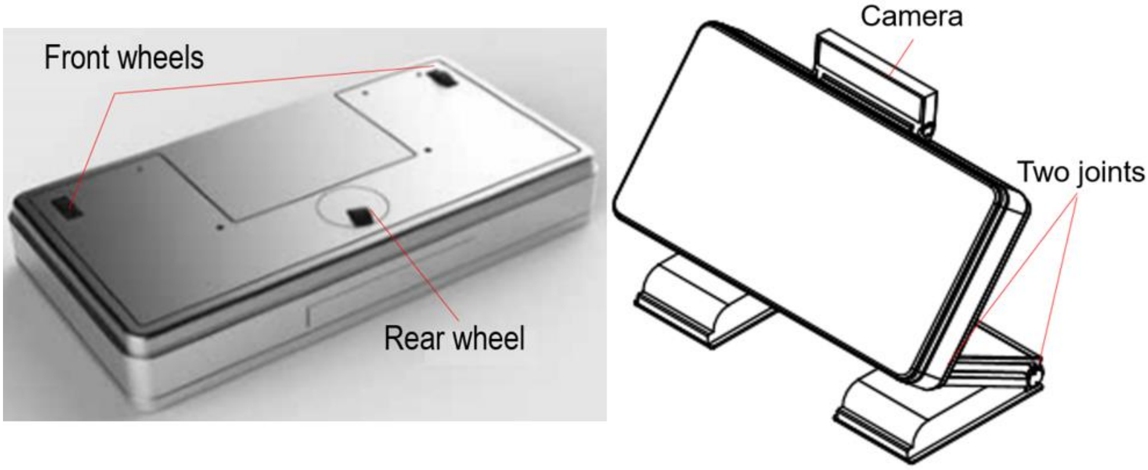


Figure 22: BaBi's Three Feet Wheels and Two Joints and Camera

Figure 23: BaBi's Facial Expressions

Chapter 3 Research Procedures and Analysis

3.1 Key Idea

The proposed robot BaBi has three small wheels (two front wheels and one rear wheels) and two joints as shown in Fig.22 to achieve transformability that can establish both autonomous mobility and portability almost equivalent to a smartphone. Because BaBi equips a smartphone, all functions and sensors the smartphone has are available on BaBi and those can be combined with BaBi's functions. In particular, voice control and monitor are used to realize voice interaction and facial expressions. There is also a camera at the head of BaBi as shown in Fig. 22 that can be used to take pictures and recognize them.

3.2 Hardware Architecture Design

3.2.1 Whole Architecture

As shown in Fig. 22 there are wheels under its feet and it moves by them. The two front wheels are the driving wheels and the rear wheel is the no-driving wheel. When the robot is viewed from the front, it looks like it has two legs, indicating the robot's appearance. In fact, by installing a rear wheel between the two legs on the back side, three points (the appearance, stability and reduction of parts) are achieved simultaneously. We control it mainly by voice interaction and also by touch screen. When opened, BaBi shows a face on the screen and speaks with people. When speaking, its eyes and mouth move like humans to emote and express feelings as shown in Fig. 23.

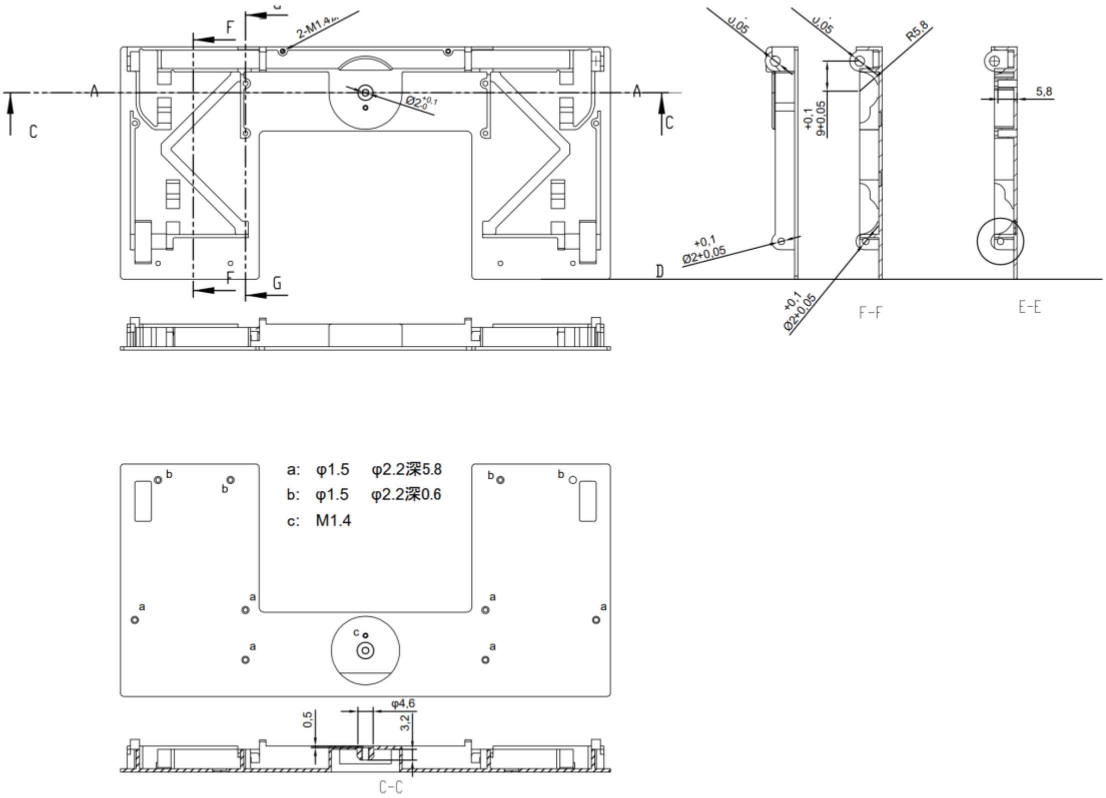


Figure 24: Leg (Foot) Module

3.2.2 Leg Module

BaBi has three wheels under its feet and it moves by them. When viewed from the front, it looks like it has two legs, indicating the robot's appearance. By installing a rear wheel between the two legs on the back side, three points (the appearance, stability and reduction of parts) are achieved simultaneously. Fig. 24 shows the engineering drawings of the leg module.

3.2.3 Body Module

The robot BaBi uses two joints to achieve transformability that can establish both autonomous mobility and portability almost equivalent to a smartphone. It can automatically open to transform from a box to a movable robot. Fig. 25 shows the engineering drawings of the body module.

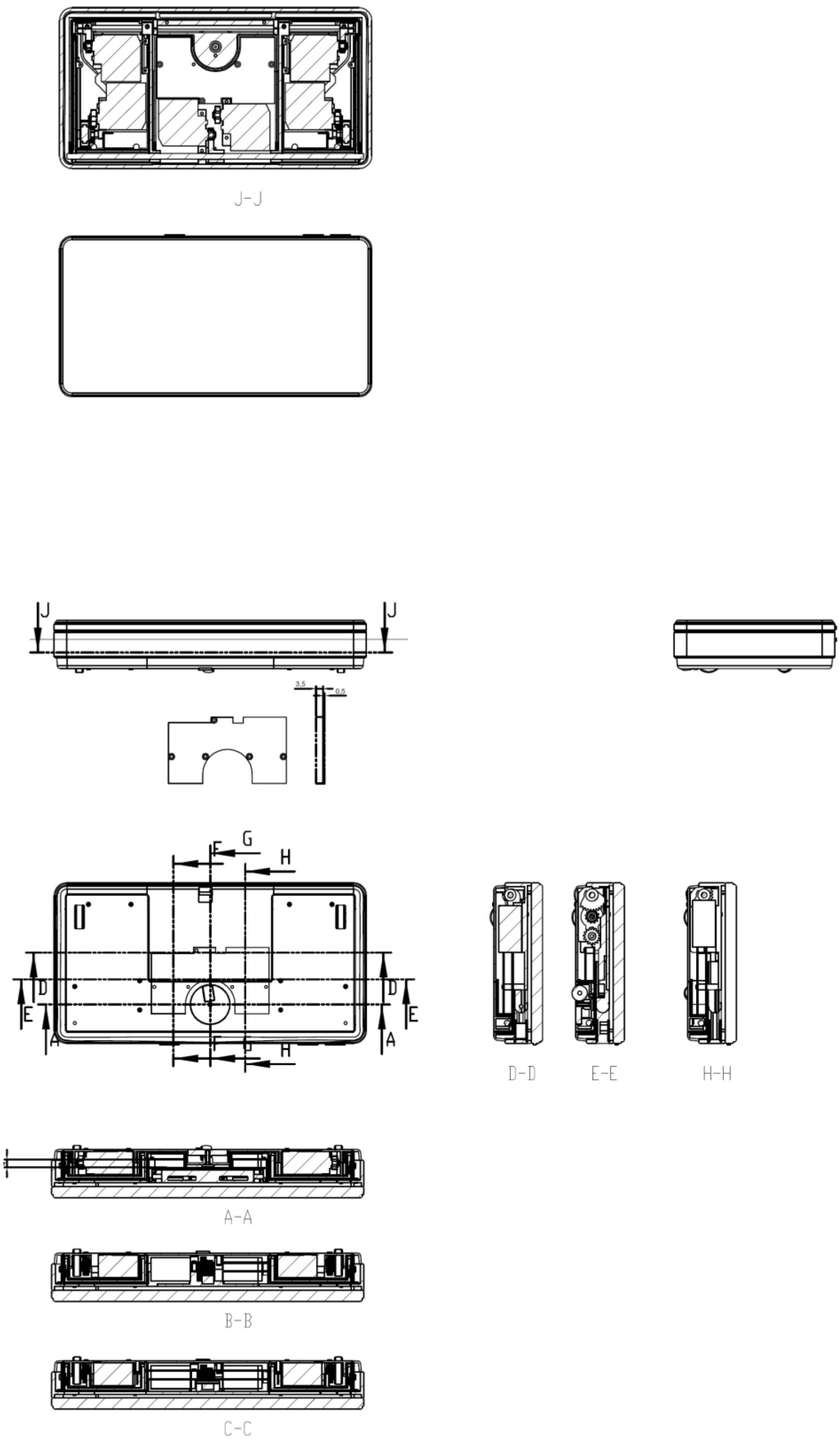


Figure 25: Body Module

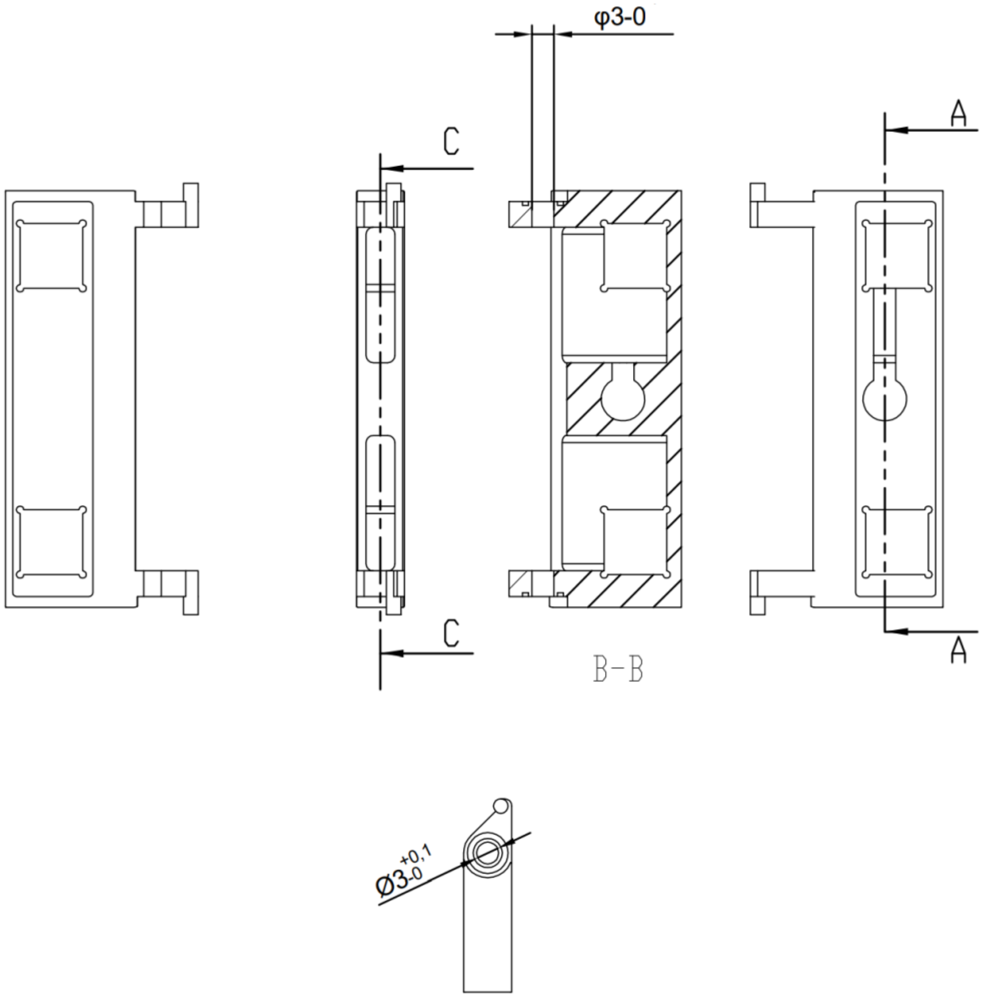


Figure 26: Head Module

3.2.4 Head Module and Camera Module

There is a camera at the head of BaBi that can be used to take pictures and recognize them. The head camera can be lifted out from or inside the box. Fig. 26 shows the engineering drawings of the head module, and Fig. 27 shows that of the camera lifting module to realize the functions.

3.3 Control Design

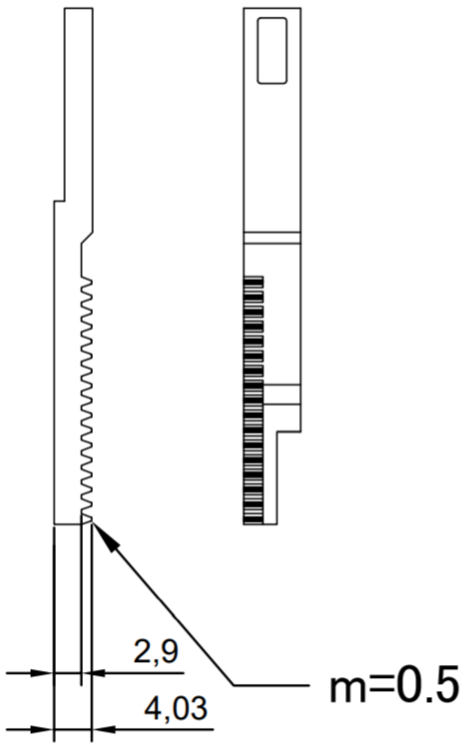


Figure 27: Camera Lifting Module

We have developed a semiconductor chip that receives commands from the smartphone via Bluetooth, and controls the motors to rotate to make BaBi's acts.

BaBi has five motors where two are for controlling the rotation of the front wheels, two is for opening and closing the two joints and one is for controlling the head camera to be lifted out from or inside the box. We can make BaBi walk by send commands to rotate the back wheels, and make BaBi rotate by controlling the two back wheels to rotate at different speeds, while controlling the lower joint to make him nod so as to make it have a rich body expression.

Fig. 28 shows the electrical circuit diagrams for the control semiconductor chip.

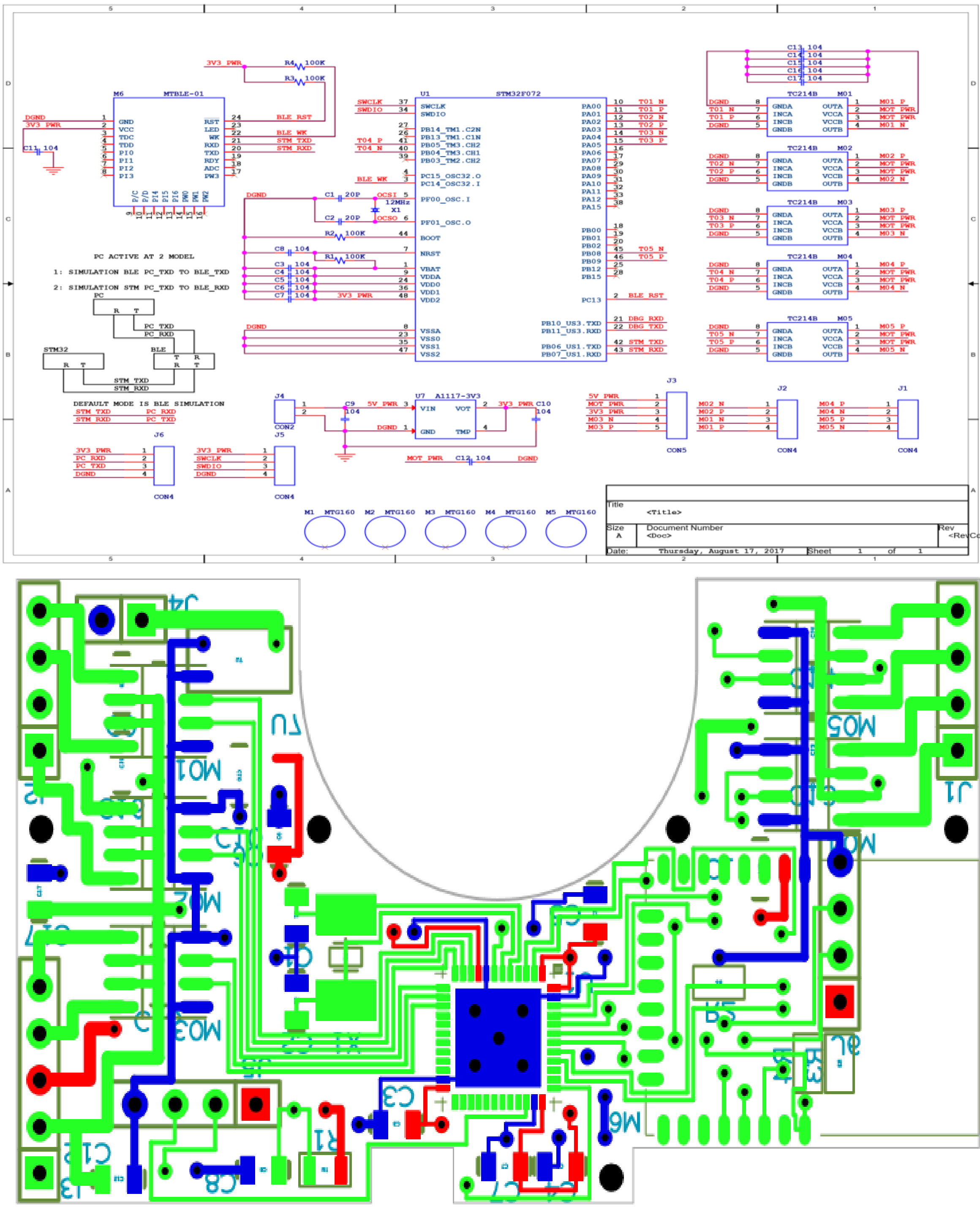


Figure 28: Electrical Circuit Diagrams for the Control Semiconductor Chip

3.4 Software Architecture Design

3.4.1 Overview

BaBi's software architecture is shown in Fig. 29 and Table II. BaBi's software is roughly divided into conversation module, motion design module, and control module. The interaction module mainly performs speech recognition, generation of the contents spoken by BaBi, calculation / generation of the emotions of BaBi. The control module mainly performs Bluetooth communication and camera control. The control module communicates with the control semiconductor chip described in Section 3.3 and controls the five servo motors.

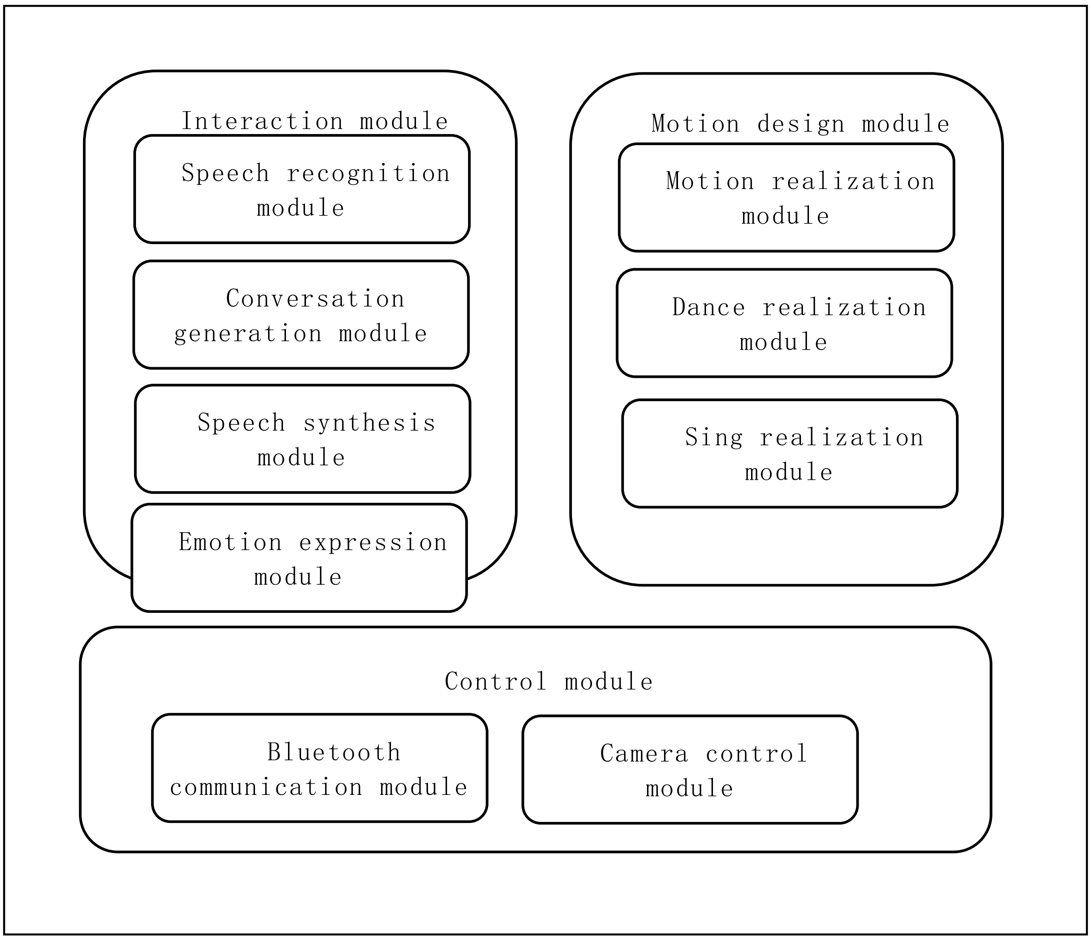


Figure 29: BaBi's Software Architecture

TABLE II： MAIN SOFTWARE MODULES IN BABI

Modules	Submodules
Interaction module	Speech recognition module Conversation generation module Speech synthesis module Emotion expression module
Motion design module	Motion realization module Dance realization module Sing realization module
Control module	Bluetooth communication module Camera control module

3.4.2 Interaction Module

The interaction module includes four submodules: speech recognition module, conversation generation module, speech synthesis module and emotion expression module. To realize human-robot interaction, we need to take people's voice and use a speech recognition to recognize the voice to obtain a text result.

(1) Speech recognition module

People are very comfortable to use speech as an input mode for various machines. Our goal for it can be achieved by automatic speech recognition (ASR) [66]. ASR system takes speech voice from the microphone of computer as an input and converts it into the text. Fig. 30 shows the basic structure of an ASR.

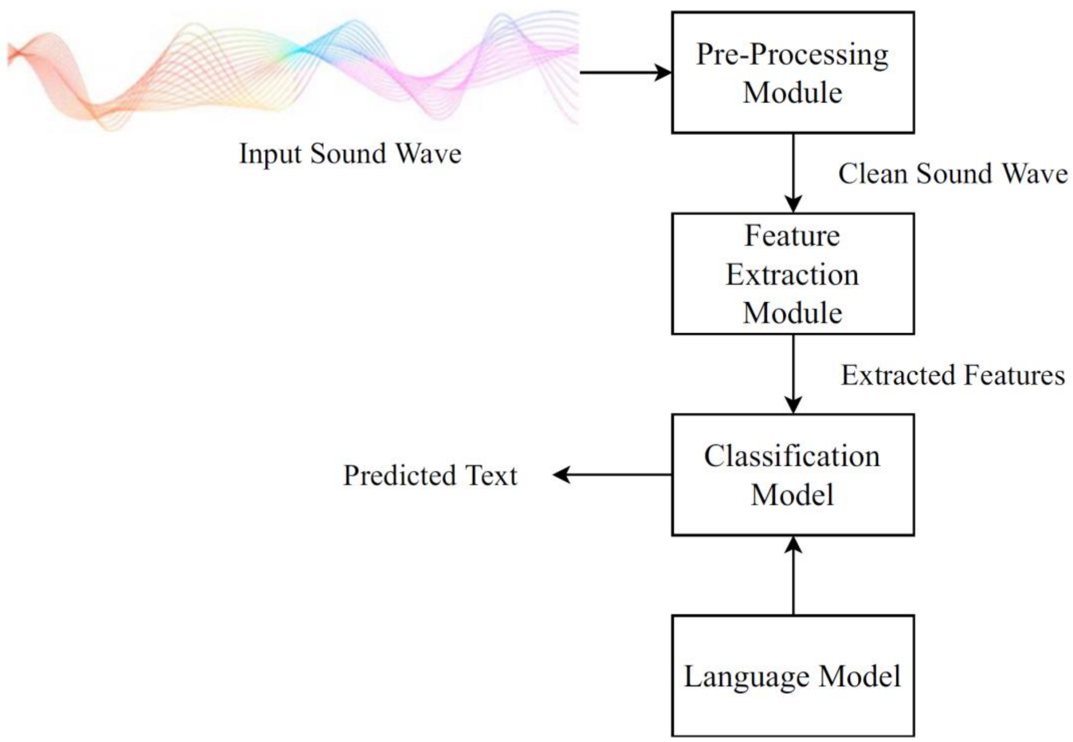


Figure 30: Basic Structure of an ASR [55]

As shown in Fig. 30, an ASR generally includes 4 modules: pre-processing module, feature extraction module, classification model, and language model. Usually, the input voice is captured by using a microphone. This implies that it may carries noise alongside the audio. The audio preprocessing is to reduce the noise ratio. Different filters and methods can be used to reduce the noise of a sound signal.

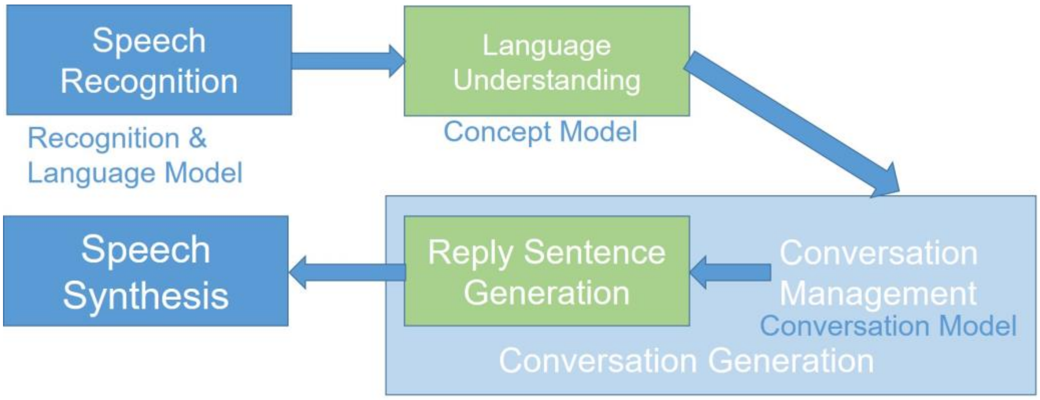


Figure 31: Structure of a Speech Conversation System

After pre-processing, ASR passes the speech signal to the feature extraction module. There are different methods for extracting features of speech signals. Features are obtained by using various methods to the input speech signal. The feature extraction methods that are the most commonly used are discrete wavelet transform (DWT), linear predictive coding (LPC) and Mel frequency cepstral coefficients (MFCCs).

The classification models take the extracted features to predict the text. It has different types of approaches to perform the speech recognition. The first approach is a generative approach. Gaussian mixture models (GMM) and HMM are the models based on the approach that are the most commonly used. The second approach is a discriminative approach. ANN and Support Vector Machines (SVM) are its examples that are the most commonly used.

The last ASR module is the language model. Language models applies trigrams, words or sentences. Using a language model can increases the efficiency significantly.

(2) Conversation generation module

Figure 32: Development of Speech Conversation System [67]

Interaction via language is a necessary skill for human beings. It is the most straightforward method that humans communicate with computers. If it is possible for a human to communicate with computer through natural language, he can access various information conveniently provided by computer systems. Conversation systems are necessary very much for various domains.

Fig. 31 shows a structure of a speech conversation system. In the conversation generation module, we need a conversation model to perform the reply sentence generation.

Fig. 32 shows the development of speech conversation system [67]. We can find the source of the conversation system in ELIZA and SHRDLU of the 1960s. SHRDLU talked based on the matching of the words, while ELIZA returned a reply after having understood a deep concept when limiting it to a world of a building block. Led by DARPA an ATIS project was carried out in the United States in the early 1990s. Targeting at the information guidance tasks of flights, the main research organizations in US participated in it and it was performed from data collection system evaluation based on a principle of the cooperation and competition. However, the boom of it was over in the late 1990s, and most of them did not lead to a practical use, and the researches become slightly being delayed.

It reached a boom of the new speech conversation system in around 2000. It was specialized in the service of the telephone (voice call) and a talking based on the typical grammar and talks flow described in hands. In other words, it just replaced the input of the telephone IVR (interactive voice response) system with a sound utterance from a numeric keypad, but it was introduced to a large scale in a call center and succeeded commercially.

The machine learning techniques have accomplished remarkable development and progress and have been attracting attention since around 2010. The speech conversation system also started introducing the machine learning.

With the success on machine learning, the model design for the conversation system has obtained a new research direction. The research of end-to-end conversation model form in deep learning. Conventional deep learning approaches include self-encoders, neural network models (NN), recurrent neural networks (RNN), convolutional neural networks (CNN), attention mechanisms, sequence-to-sequence (seq2seq), reinforcement learning (RL), memory networks, and generational adversarial networks (GAN) [68].

(3) Speech synthesis module

After obtaining the answer text from the conversation generation module, we need to use the speech synthesis module to generate a voice of the answer in order to BaBi can reply user by using the voice.

For human speech an artificial process is performed in speech synthesis. A language text is converted into speech using a text-to-speech (TTS) system [69]. A database can store recorded speech and we can concatenate pieces of the speech to create synthesized speech. The largest output

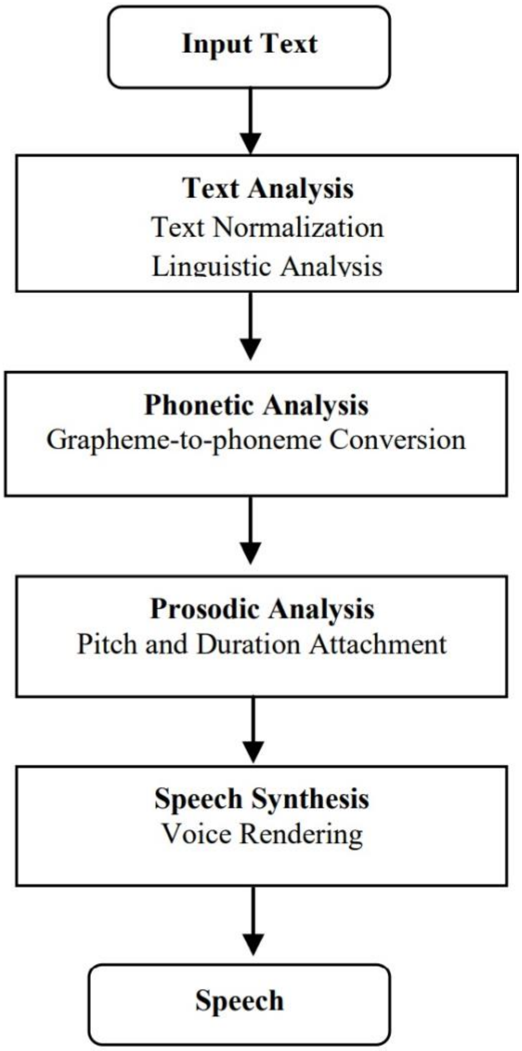


Figure 33: Block Diagram of Text-to-Speech Synthesis [69]

range can be provided by a system where phones or di phones are stored, and clarity may be lacked. In order to improve the accuracy, it is possible to store sentences or words. For the characteristics of human voice and vocal tract, the model can be incorporate by a synthesizer. Using the understood ability and the similarity with the voice of human, for a speech synthesizer its quality can be evaluated. Since the early 1990s, speech synthesizers have been used in many computers operating systems.

As shown in Fig. 33 the main phases of TTS system are text processing and speech generation.

Text processing:

A phonetic / linguistic representation is transcribed into from an input text after analyzing and normalizing the input text. The processing issues for low level including word / sentence segmentation are dealt with at the text processing components.

a) Detection of document structure. By using paragraph formatting and punctuation mark, it detects the structure of document.

b) Text normalization. Acronyms and abbreviation are handled by the text normalization. Matching the text is the goal for normalization. For example, September can be rendered from Sept. Good accuracy of output is resulted by proper normalization.

c) Linguistic analysis. In written text in order to avoid ambiguities syntactic analysis and morphological analysis are performed in linguistic analysis. Facilitating phrasing and accenting is handled in the syntactic analysis, while proper word pronunciation is dealt with in the morphological analysis.

Speech generation:

By using parameter, the speech is generated in the speech generation component.

a) Phonetic analysis. Within each word the phone level is focused on in it. About how to produce and what sound to produce, the information is tagged to each phone.

- Converting to phoneme from grapheme. On the input sentences for each word， it determines exact pronunciation.
- Homograph disambiguation: For input sentence whether the past tense or present tense is used is identified using dictionary.

b) Prosodic analysis. The importance of prosodic analysis lies in the fact that it provides us with a basis for marking prosodic effects in terms of utterance plans, for example, prosodic processing.

(4) Motion expression module

When BaBi speaks its eyes and mouth move like humans to emote and expresses feelings. BaBi performs rich express: facial expressions and motions.

The robot BaBi has three small wheels (two front wheels and one rear wheels) and two joints as shown in Fig.22. BaBi has five motors where two are for controlling the rotation of the front wheels, two is for opening and closing the two joints and one is for controlling the head camera to be lifted out from or inside the box.

We can ask the smart phone to send commands to the control semiconductor chip described in Section 3.3 to control the five motors to rotate, so as to make it have a rich body expression and motions.

The motion expression module calculates and generates the emotions of BaBi to express its emotions when making interaction with people.

3.4.3 Motion design module

The motion design module includes three submodules: motion realization module, dance realization module, sing realization module. BaBi uses the motion realization module to perform the motions such as open, close, and come on, while BaBi uses the dance realization module to perform a dance motion and uses the sing realization module to perform a sing motion.

The smart phone sends commands to the chip to control the five motors to rotate. Each command has six characters and each character has three types: 0, 2 and 1. 0 is to stop, 2 is to perform positive rotation and 1 is to perform reverse rotation.

Fig. 34 shows an example of a command. The first one controls the head camera. The second one controls the left front wheel. The third one controls the right front wheel. The fourth one controls the lower joint. The fifth one controls the upper joint. The last one shows the command end.

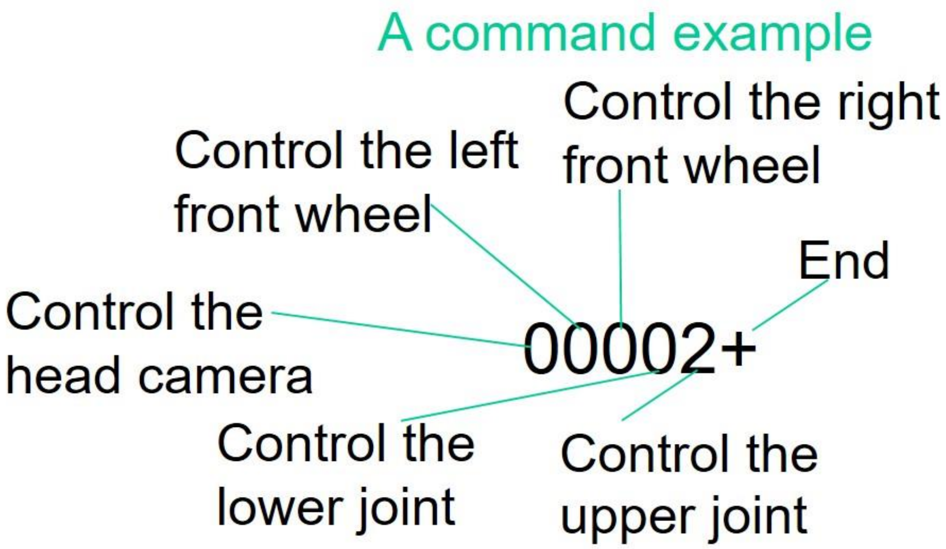


Figure 34: A Command Example

(1) Open motion

To make BaBi open the following steps are taken.

1. User speaks “Open”.

2. A voice “Open” is recorded by the smart phone.

3. The smart phone sends the voice to the Baidu speech recognition to obtain a result text.
4. The result text is searched into the command list.
5. If the “Open” command is found in the command list, the processes as shown in Fig. 35 are performed.
6. If the “Open” command is not found in the command list, it sends the text to the Turing Robot web to obtain an answer.

```
SendCommand("00002+");
Sleep(300);
SendCommand ("00000+");
Sleep(1000);
SendCommand ("00002+");
Sleep(300);
SendCommand ("00000+");
Sleep(1000);
SendCommand ("00002+");
Sleep(300);
SendCommand ("00000+");
Sleep(1000);
SendCommand ("00010+");
```

Figure 35: Open Processes

```
SendCommand(“00020+”);
Sleep(1000);
SendCommand(“00001+”);
Sleep(900);
```

Figure 36: Close Processes

(2) Close motion

To make BaBi close the following steps are taken.

1. User speaks “Close”.
2. A voice “Close” is recorded by the smart phone.
3. The smart phone sends the voice to the Baidu speech recognition to obtain a result text.
4. The result text is searched into the command list.
5. If the “Close” command is found in the command list, the processes as shown in Fig. 36 are performed.
6. If the “Close” command is not found in the command list, it sends the text to the Turing Robot web to obtain an answer.

(3) Coming on motion

To make BaBi come on the following steps are taken.

1. User speaks “Come on”.
2. A voice “Come on” is recorded by the smart phone.
3. The smart phone sends the voice to the Baidu speech recognition to obtain a result text.

```
SendCommand("01000+");
Sleep(900);
SendCommand("00100+");
Sleep(900);
SendCommand("00000+");
```

Figure 37: Coming on Processes

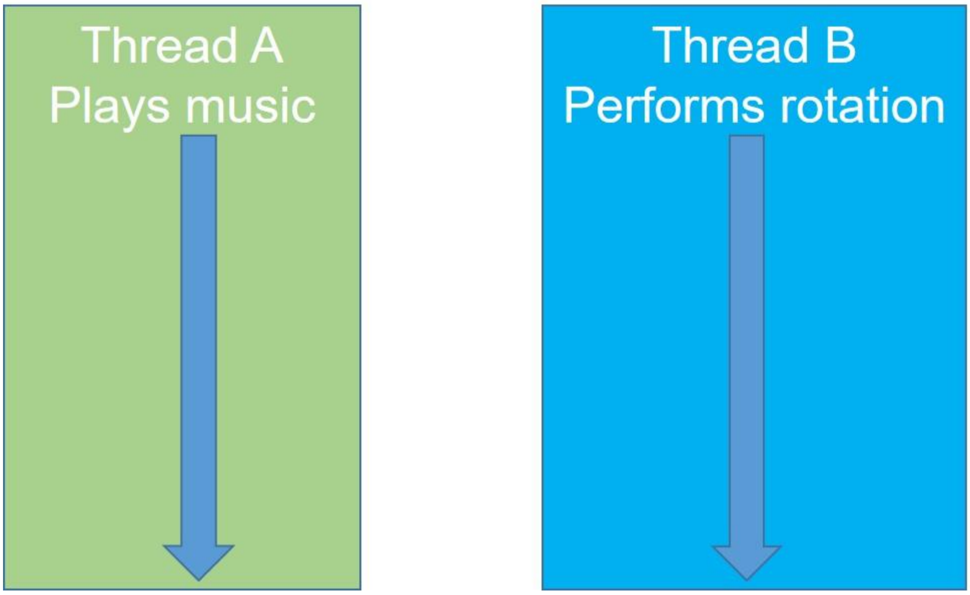


Figure 38: Dance Motion

```
SendCommand("01200+");
```

Figure 39: Rotation Processes

4. The result text is searched into the command list.
5. If the "Come on" command is found in the command list, the processes as shown in Fig. 37 are done.

6. If the “Come on” command is not found in the command list, it sends the text to the Turing Robot web to obtain an answer.

(4) Dance motion

To make BaBi dance the following steps are taken.

1. User speaks “Dance”.
2. A voice “Dance” is recorded by the smart phone.
3. The smart phone sends the voice to the Baidu speech recognition to obtain a result text.
4. The result text is searched into the command list.
5. If the “Dance” command is found in the command list, the processes as shown in Fig. 38 are done, where two threads are processed at the same time to perform the dance motion. Fig. 39 shows the rotation that is performed at the dance motion.
6. If the “Dance” command is not found in the command list, it sends the text to the Turing Robot web to obtain an answer.

(5) Sing motion

To make BaBi sing the following steps are taken.

1. User speaks “Sing”.
2. A voice “Sing” is recorded by the smart phone.

3. The smart phone sends the voice to the Baidu speech recognition to obtain a result text.
4. The result text is searched into the command list.
5. If the “Sing” command is found in the command list, the processes as shown in Fig. 40 are done, where two threads are processed at the same time to perform the sing motion.

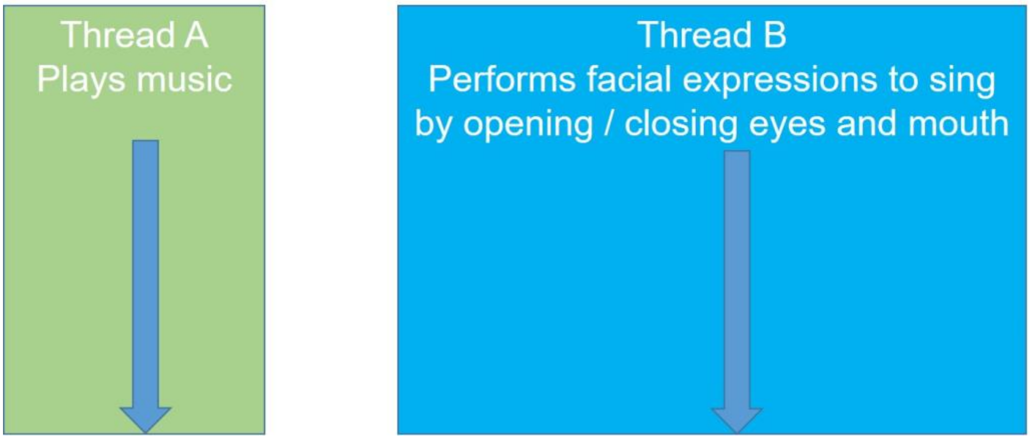


Figure 40: Sing Motion

6. If the “Sing” command is not found in the command list, it sends the text to the Turing Robot web to obtain an answer.

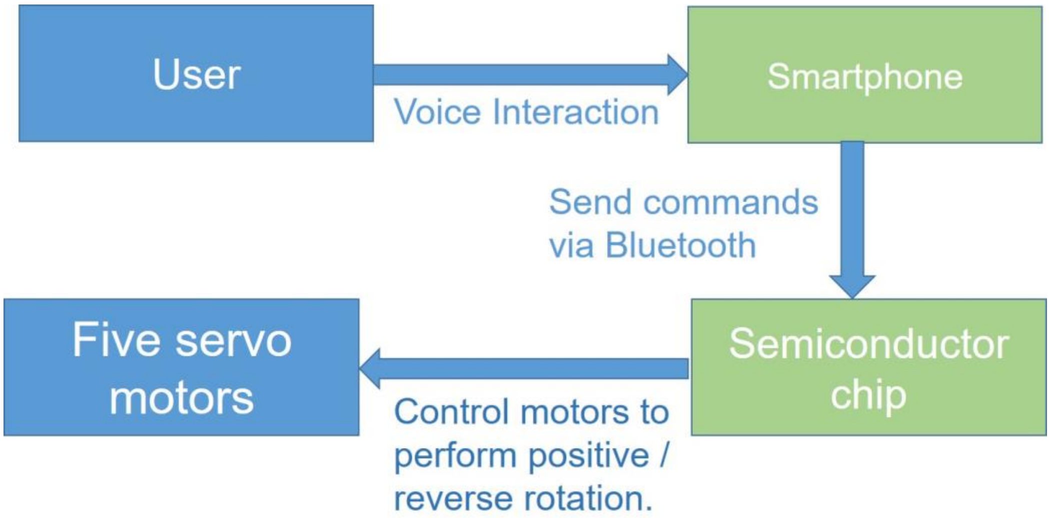


Figure 41: Control Module

3.4.4 Control module

The control module includes Bluetooth communication module and camera control module. The control module communicates with the control semiconductor chip described in Section 3.3 and controls the five servo motors.

Fig. 41 shows the structure of the control module. A user can speaks an instruction with voice interaction, and then the smartphone accepts the instruction and transforms it into commands to send the commands to the semiconductor chip via Bluetooth. After that the semiconductor chip controls the five motors to perform positive / reverse rotation to realize the robot motions.

Chapter 4 Implementation and Discussion

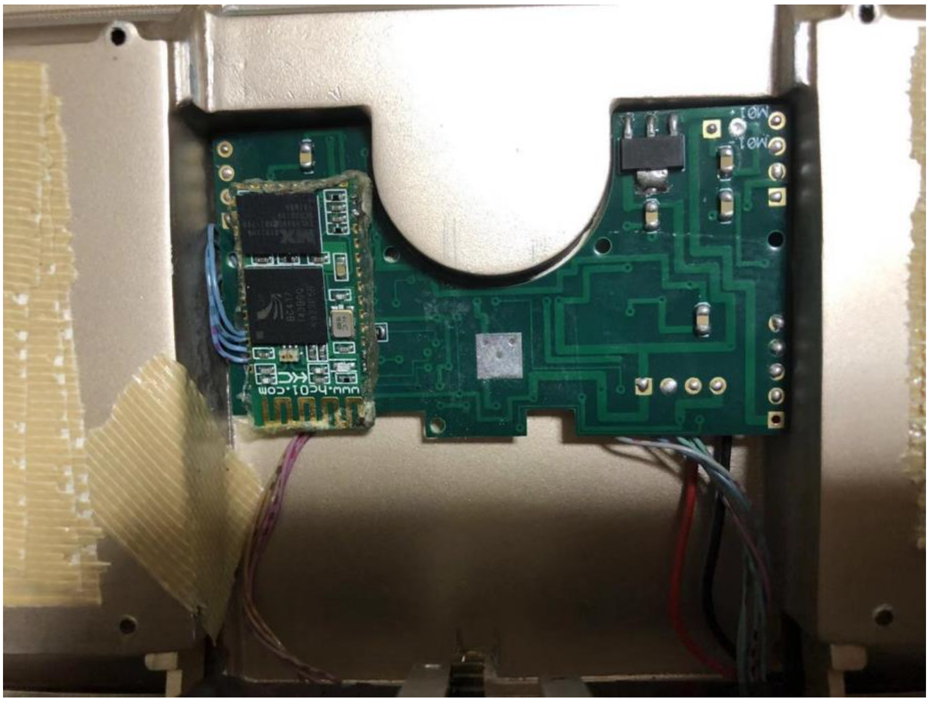


Figure 42: A Semiconductor Chip to Control the Motors to Rotate

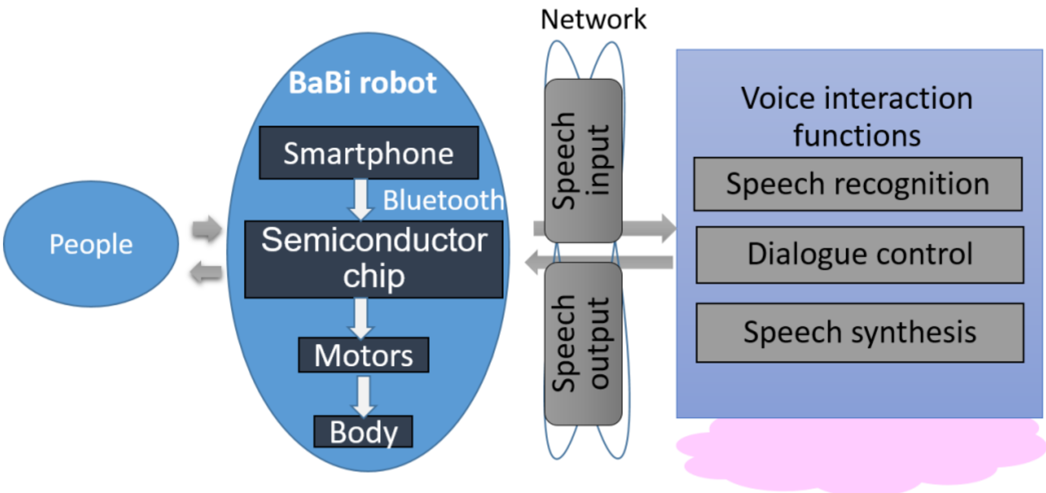


Figure 43: Implementation Architecture of BaBi Phone Robot

Chapter 4 Implementation and Discussion

4.1 Implementation Architecture

We have developed a semiconductor chip as shown in Fig. 42 that receives commands from the smartphone via Bluetooth and controls the motors to rotate to make BaBi's acts. We control its acts by voice interaction (speech) as shown in Fig.43.

BaBi has five motors where two are for controlling the rotation of the front wheels, two is for opening and closing the two joints and one is for controlling the head camera to be lifted out from or inside the box. The developed chip is driven at a voltage of 5V and powered by the smartphone and the motors are powered via the chip. We can make BaBi walk by send commands to rotate the back wheels and make BaBi rotate by controlling the two back wheels to rotate at different speeds, while controlling the lower joint to make him nod so as to make it have a rich body expression. BaBi's body is made of aluminum, so it's very light.

4.2 Functions

We prototyped a set of interactions and scenarios that users would be able to engage with. We show how our system can exist harmoniously with already existing practices.

(a)

(b)

(c)

Figure 44: BaBi's Eyes and Mouth Moving like Humans to Emote and Expresses Feelings

Figure 45: Interface Design

(1) Transformation from a phone to a robot

At the beginning, BaBi is a rectangular box (a phone), and it automatically opens to transform from the box to a movable robot when its owner calls to it, “open”. To do it, BaBi sends the people’s voice data to the Baidu speech recognition server and obtains a result text. When the text is understood as open instruction, BaBi gives an instruction to the chip by a Bluetooth message that instructs motors to work to open.

(2) Behavior, perception and reasoning

After opening BaBi will acts like a human and performs more varied express to extend current smart phone functionality to be far-field, context-driven and multimodal interactions. For example, BaBi moves, talks, emotes, sings, dances, and converses with people to make people happy, enhance people's lives, facilitate relationships, have fun with people, connect people with the outside world and assist and support people as an intelligent personal assistant partner.

We give instructions to ask BaBi to sing, dance, move, tell jokes, make conversation and do secretary works including scheduling of works, schedule reminders, sending emails, calling phones, booking, making reservations, searching information, etc. It can remember and recognize people's voices and understand its owner. BaBi is portable, transformable, movable and intelligent.

(3) Voice interaction

We make communication with BaBi and give instructions by voice interaction. BaBi works on a Windows Phone system. BaBi records people's voice from Microphone when people speak to BaBi, and then sends the voice data to a speech recognition server (Baidu speech recognition) via Web API and obtains a recognition result text. We apply a Voice Activity Detection (VAD) algorithm to detect the speech and the non-speech frames [70] and detect the stop of the people speech by the non-speech frames. When the stop time of the people speech is longer than a threshold, BaBi stops the voice record and sends the voice data to the Baidu speech recognition server to obtain a result text. According to the result text, BaBi recognizes it as an instruction if it includes the key words in the key word list, otherwise, considers it as a chat text and sends the text to the Turing Robot web server by web API [71] to obtain an answer. After that BaBi replies the obtained answer by synthesizing the answer text into a human sounding speech using the Microsoft Windows Phone Speech Synthesizer and playing the synthesized speech. When it is an instruction, BaBi replies a prepared answer that is also synthesized into a human sounding speech by the Microsoft Windows Phone Speech Synthesizer so as to play the speech. Then BaBi acts according the instruction such as singing, dancing and moving.

(4) Facial expressions

When speaking BaBi's eyes and mouth move like humans to emote and expresses feelings by displaying and changing continually the face pictures as shown in Fig. 44, where the face picture is changed from (a) to (b) and (c).

The interface design is shown in Fig. 45, where we show BaBi's answers on the left field and show some function buttons on the bottom-right. Because BaBi equips the smartphone, its face design can be easily changed according to user's preferences.

Chapter 5 Study on Three-Dimensional Robot Motion and Face Design by IEC and non-IEC fusion

5.1 The Basic Ideas of the Proposed Face and Motion Design Method

This study proposes a robot face and motion design method based on the fusion of IEC and Non-IEC [12, 13, 14]. This method allows users to perform optimization while switching IEC and Non-IEC search at any time. The proposed method performs IEC search when users make selections, while it does a non-IEC process when users do not make any selection, resulting in being a fusion of IEC and non-IEC. The user can add candidates to the case base when finding satisfying candidates and evaluates candidates' scores according to the case base. It applies a Different Evolution (DE) method to generate new candidates. Users may alternately push the buttons to generate robot motion without making any selection of candidates. The method will make selection with the scores when users do not make any selection of candidates. Adding more candidates to the case base can speed up generating the desired motion and promote the non-IEC progress and reduce the user workload.

In every generation it is possible for users to choose desirable candidate or grades some candidates, and directly edit chromosomes of candidates.

For user preference estimation in order to reduce the burden on users of evaluation, a Case-Based Reasoning (CBR) approach [15] is applied. Users add candidates with scores to the case base when finding satisfying candidates and system estimates user preference value of candidates using the case base.

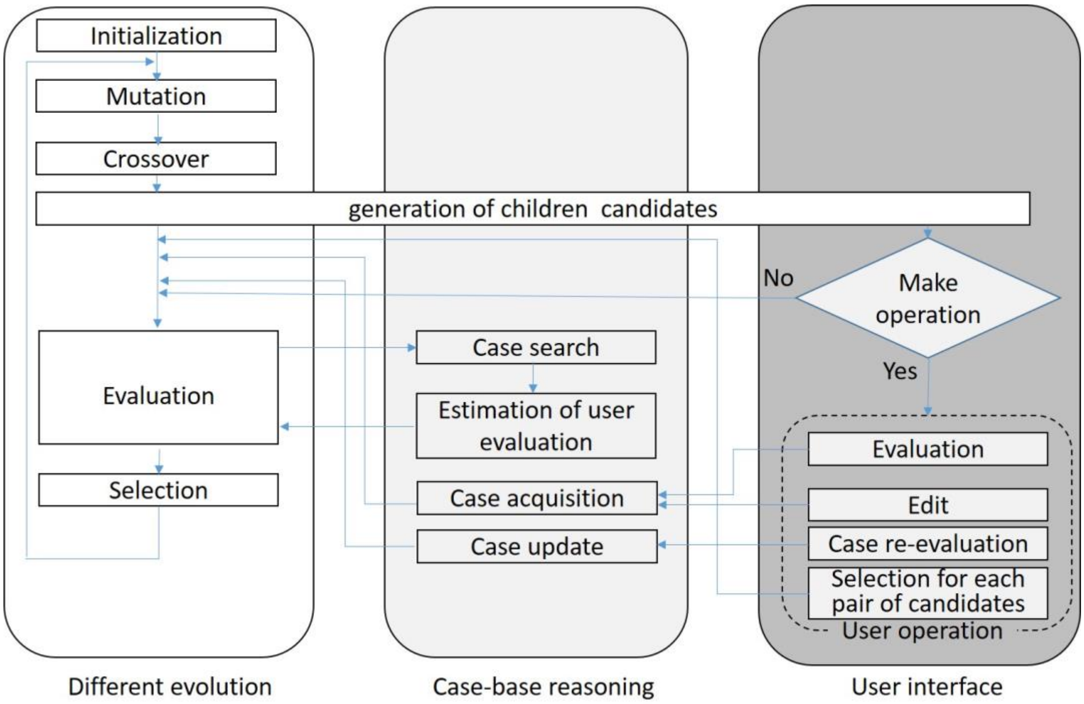


Figure 46: Algorithm of the Proposed Method Based on [11, 12]

5.2 Structure and Process Flow the Proposed Method

Fig. 46 shows the structure and process flow of the proposed method. It performs Non-IEC search and suggestion of solution candidates unless a user make operation. In Non-IEC search individuals are evaluated using a case base if there are stored cases in the case base.

A user can stop the search and make operation whenever finding a favorite individual or none of individuals suits the user taste. On the use operation, the user can directly edit each individual’s chromosome and score. The method stores the edited individuals into the case base as cases. The cases can also be re-evaluated by the user. On the use operation, the user can also manually select one for each pair of candidates.

The DE algorithm initializes the parameters of the parent candidates with random numbers. And then it generates children’s candidates for all parent candidates. The algorithm selects one for each pair of candidates (a parent and a child) and updates the selected one to parent and remove another. For each pair of candidates, when users make a selection with a selection interface, the DE algorithm takes the selected one by users, otherwise it will make the selection according to the scores, which are evaluated according to the case base or by users.

5.3 Chromosome Representations

We consider 10 parameters $\{p_1, p_2, \ldots, p_{10}\}$ as a chromosome to express pleasure, sadness, anger, etc. as shown in Fig. 47. The parameters are coded as integers. Parameter pairs of (p_1, p_2), (p_3, p_4), and (p_5, p_6) determine the shape of eyes, eyebrows, and mouth, respectively. The speed at which the eyes, eyebrows, and mouth change is constant. Parameters $p_7, p_8, \ldots, p_{10}$ are set as 10-level scores for the speed and angle of BaBi’s two body joints. Although DE algorithm processes variables as continuous values, we round them into integers so that the user can easily edit them when generating actions or displaying them on the GUI.

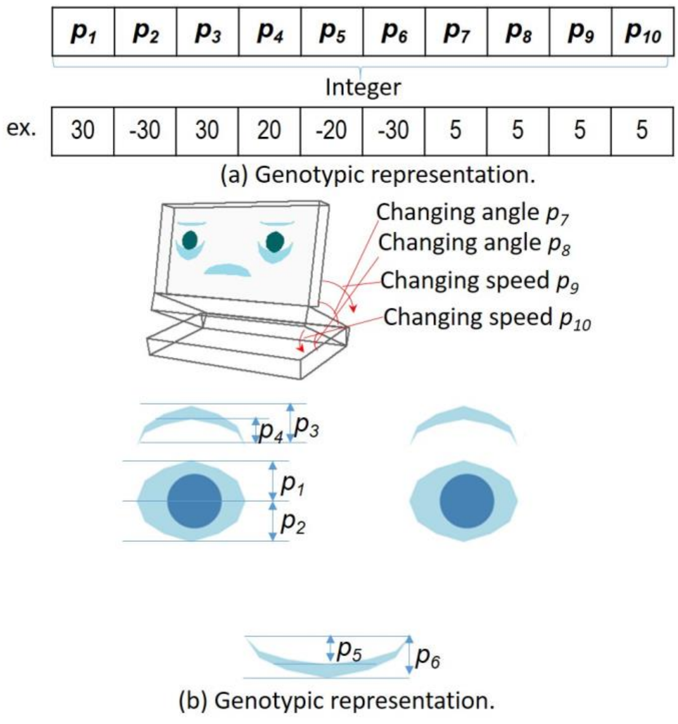


Figure 47: Chromosome Representation

5.4 Estimation of User Evaluation Value by Case-Based Reasoning

The system will make selection with the scores (fitness) evaluated according to the case base when users do not make any selection of candidates. As a fitness of an individual *C*, the proposed method employs an estimated user preference $\overline{P(C)}$ of the individual *C* that is calculated from the preference value *P(C')* of the nearest case *C'* and the Euclidean distance *d(C, C')* between *C* and *C'* by the following equation:

$$\overline{\mathrm{P(C)}} = \frac{\mathrm{dmax} - \mathrm{d(C,C')}}{\mathrm{dmax}} \mathrm{P(C')} \qquad \textbf{(1)}$$

where *dmax* is a maximum of the distances between individuals.

5.5 User Interface

Fig. 48 shows the interface of the proposed method. The interface shows 10 pairs of candidates on left part of the form, where 10 candidates are the parents, while other 10 candidates are the children. Each child is generated form its left parent candidate by DE algorithm.

Users can use mouse to select a candidate, and an enlarged image and 3D figure of the selected candidate are shown on the right part. User can rotate the 3D figure by pushing the left mouse button and moving the mouse simultaneously. We designed 32 lines and 122 vertexes for the 3D figure, and after rotation transformation, it was transformed into a planar graph by a perspective projection. We use this 3D figure to simulate the motion of the two joints of the robot.

The system initializes the parameters of the 10 parent candidates with random numbers. And then it generates 10 children's candidates for all parent candidates when users push button [Evolving]. Users can manually select one for each pair of candidates with the checkbox interface, then press button [Selection]. Users are asked to alternately push the buttons [Evolving] and [Selection] to generate new candidates. It allows users to add the selected candidate to the case base when finding the selected candidate is satisfying. When push the button [Selection], the system

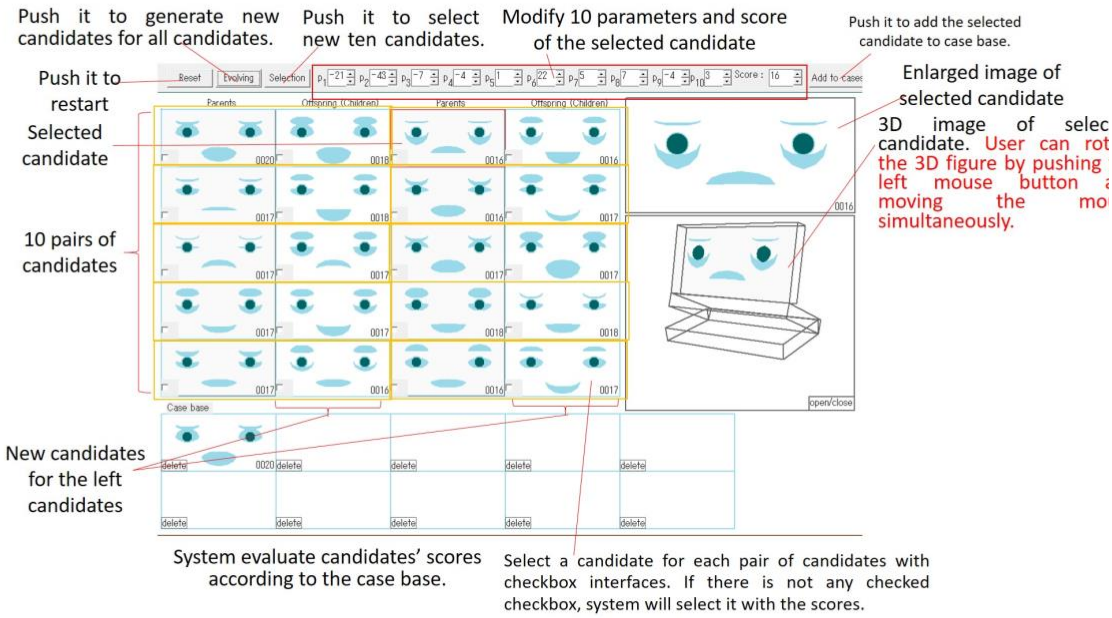


Figure 48: User Interface

selects one for each pair of candidates (a parent and a child) and updates the selected one to parent and remove another. For each pair of candidates, when users make a selection with the checkbox, the system takes the selected one by users, otherwise the system will make the selection according to the scores, which are evaluated by the system according to the case base or by users.

Uses can edit the parameters and score of the selected candidate by up/down interfaces on the top of the form, and confirm the modified result by the enlarged image and 3D figure of the selected candidate. By pushing button "Reset", the processes are restart.

5.6 Differential Evolution Algorithm

In this section, the canonical version of Differeintial Evolution is shown first, and then details of the paired comparison DE was described.

5.6.1 Canonical Differential Evolution

5.6.1.1 Overview

In DE, solution candidates are represented as D-dimensional parameter vectors such as equotion：

$$x_{i,g} = \left(x_{i,g,1}, x_{i,g,2}, \dots, x_{i,g,D}\right) \qquad (2)$$

where i denotes an index of a solution candidate ($i \in \{1, \dots, N_P\}$) and g indicates generation. The number of solution candidate N_P is called as a population size in many evolutionary computation algorithms inclulding DE. In canonical DE [16], each parameter $x_{i,g,d}$ ($d \in \{1, \dots, D\}$) has a real value rather than discrete values or symbols used in combinatorial optimization. Although discrete versions of DE have also been proposed such as [72] [73], this paper focuses only on continuous optimization because the target problem of designing facial emotion and body movement of the proposed robot involves continuous parameters only and DE shows its good performance mainly in the continuous optimization problems.

5.6.1.2 Initialization

At initialization, most of DE algorithms randomly select values of parameters in all solution candidates based on uniform probability distributions:

$$x_{i,1,d} = U(a,b) = \begin{cases} \dfrac{1}{b-a} \, for \; (a \leq x \leq b) \\ 0 \;\; for \; (x < a \; or \; x \; > b) \end{cases}, \quad (i \in \left\{1, \dots, N_p\right\}, d \in \{1, \dots, D\} (3)$$

where U denotes the uniform distribution.

It is also possible to employ a relativey good solution obtained by other techniques or human professionals. In that case, Gaussian distribution whose mean values are those of the good solution's variables is more suitable to sample initial solution candidates.

$$x_{i,1,d} = \mathcal{N}(\mu, \sigma^2)\,, \quad (i \in \{1, \dots, N_p\}, d \in \{1, \dots, D\}$$

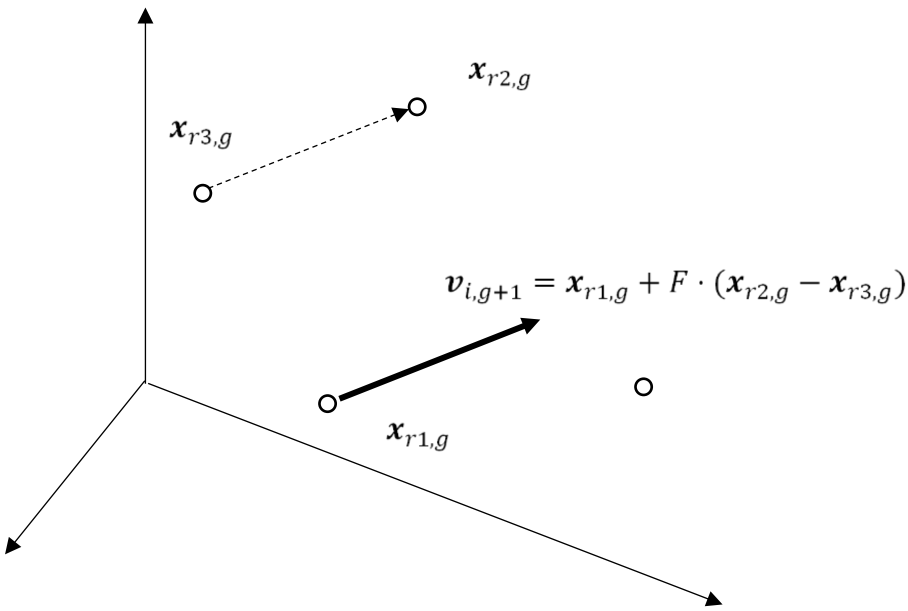


Figure 49: Calculating a Mutant Vector

where $\mathcal{N}$ denotes Gaussian distribution.

After initialization, DE iterates solution candidate generation and evaluation during the predetermined generations G_{limit}. DE generates new solution candidates (offsprings) by multation and crossover. The two operators are similar to those in GA, but DE applies them in the opposite order from GA, i.e., mutation is applied first, then crossover is applied. The two operators are applied to all solution candidates $\boldsymbol{x}_{i,g}$ separately, where $\boldsymbol{x}_{i,g}$ is called a target vector in canonical DE.

5.6.1.3 Mutation

Mutation is a operation that generate new vectors by adding the weighted difference between solution candidates to another solution candidate. Following is a basic strategy called 'rand/1/bin':

$$\boldsymbol{v}_{i,g+1} = \boldsymbol{x}_{r1,g} + F \cdot (\boldsymbol{x}_{r2,g} - \boldsymbol{x}_{r3,g}) \quad (4)$$

where $r1$, $r2$, and $r3$ are randomly chosen indices $r1, r2, r3 \in \{1, \dots, N_P\}, r1 \neq r2 \neq r3 \neq i$. F is a control parameter called *scale factor* that deeply changes DE convergence property ($F > 0$). In canonical DE, F is constant real value and usually set to a value between 0 and 2. Fig. 49 shows the process of calculating a mutant vector.

5.6.1.4 Crossover

After mutation, crossover operation is applied to the generated new vector $\boldsymbol{v}_{i,g+1}$ to create a trial vector $\boldsymbol{u}_{i,g+1}$ as follows Equotion:

$$u_{i,g+1,d} = \begin{cases} v_{i,g+1,d} & \text{IF } r_d^{01} \leq CR \text{ OR } d = r_i^D \\ x_{i,g,d} & otherwise \end{cases}, \ d = 1, \dots, D \quad (5)$$

where r_d^{01} is a uniform random real number between 0 and 1, CR is the second control parameter named *crossover rate*, which also changes DE algorithm characteristics. r_i^D is a random number in $\{1, \dots, D\}$ which is necessary to change at least one variable from the target vector $\boldsymbol{x}_{i,g}$.

Crossover creates the trial vector from the target vector and the one created by the mutation. In detail, it selects a vertex of a hyper-rectangle with $\boldsymbol{x}_{i,g}$ and $\boldsymbol{v}_{i,g+1}$ as vertices in the D-dimensional solution space as shown in Fig.50.

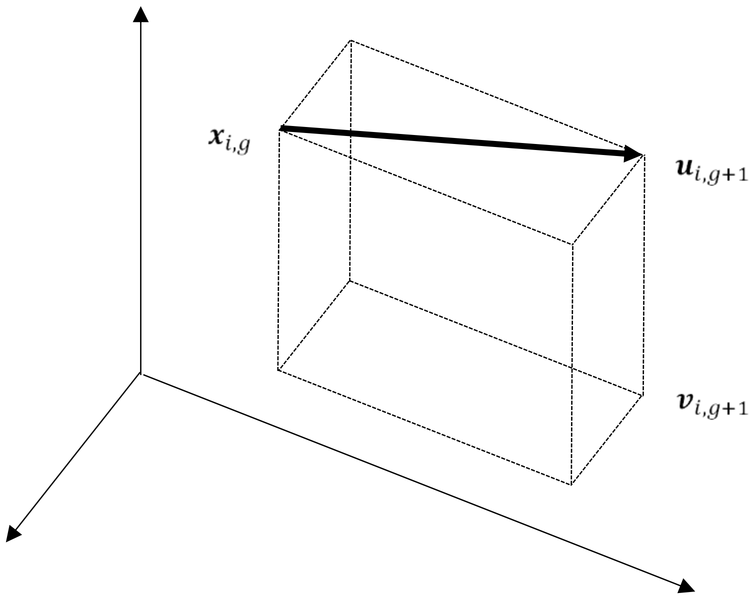


Figure 50: Schematic Diagram of the Crossover Process in DE (an example in 3-dimensional solution space)

5.6.1.5 Selection

The trial vector $\boldsymbol{u}_{i,g+1}$ created by the mutation and crossover processes is then compared with the target vector $\boldsymbol{x}_{i,g}$, i.e., objective function values $f(\boldsymbol{u}_{i,g+1})$ and $f(\boldsymbol{x}_{i,g})$ are compared. The better one is selected as $\boldsymbol{x}_{i,g+1}$ (greedy selection).

5.6.1.6 Control parameters

Canonical Version of DE includes mainly three control parameters as follows:

(1) Scale factor F: Scale factor amplitudes the difference vector calculated from $\boldsymbol{x}_{r2,g}$ and $\boldsymbol{x}_{r3,g}$. Larger F value promotes more global search. Recently, lots of self adaptation DE algorithms have been proposed that allow solution candidates having their own scale factor values F_i and changing their value according to the search progress. The representative self-adaptive algorithms are jDE [74], SADE [75], JADE [76], and SHADE [77].

(2) Crossover rate CR: Similar to the scale factor, crossover rate is also an important control parameters that significantly changes DE convergence properties. Low CR value is generally used in separable problems while high value such as 0.9 is used in non-separable problems. The self-adaptive DEs introduces parameters for each solution candidates CR_i, which are changed in the same or similar manner as F_i.

(3) Population size N_P: DE requires at least four solution candidates

when using a basic mutation strategy shown in eq.(4). In general, N_P should be set to $5D$ to $10D$. Recently, a self adaptation version of DE has been proposed named L-SHADE [78], which reduces N_p as the optimization progresses. This allows reducing the search cost because the last stage of the optimization performs local search and that requires less candidates than those required in global search at initial optimization stage.

5.6.2 Paired Comparison Differential Evolution Used in the Proposed Method

Similar to the canonical version of DE, the main steps of the paired comparison DE used in the proposed method is given below:

- Initialization
- Mutation
- Crossover
- Evaluation
- Selection
- Until (termination criteria are met)

(1) Initialization

First, as shown in Fig. 51 the system initializes the parameters of the 10 parent candidates with random numbers, where the DE algorithms randomly select values of parameters based on uniform probability distributions shown in Eq. (3).

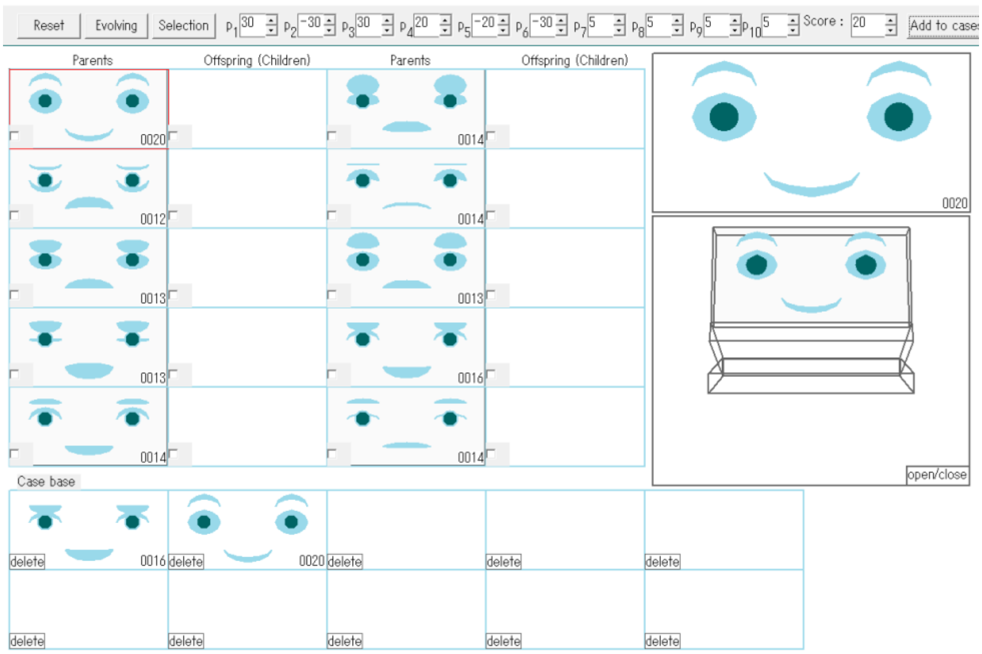


Figure 51: Initialization

It can select a candidate with mouse and uses the interface to modify parameters of the selected candidate, and then users can confirm the modified result by the enlarged image of the selected candidate.

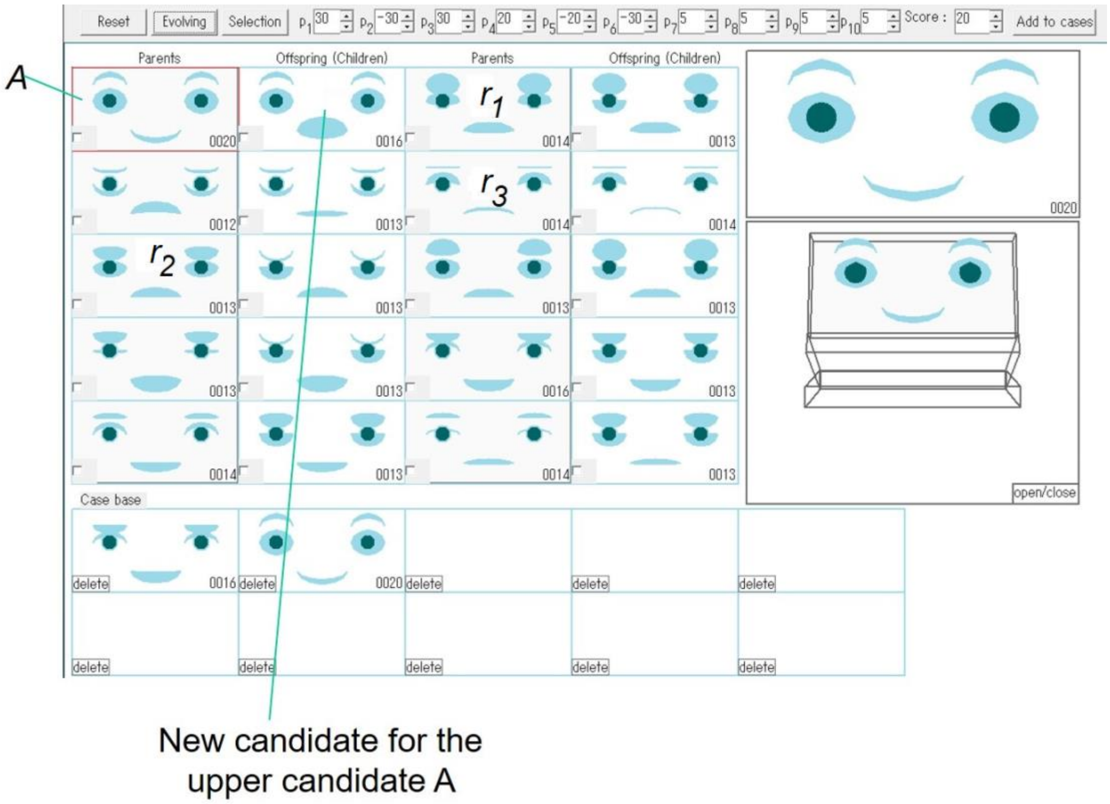


Figure 52: Mutation

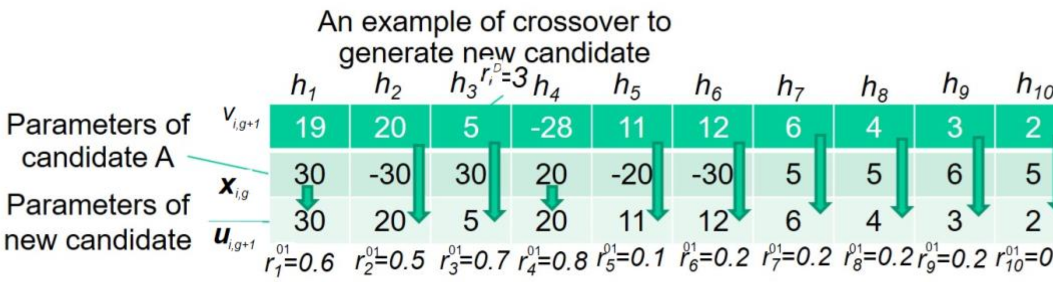


Figure 53: Crossover

(2) Mutation

And then the system generates a children candidate for each parent candidate when users push button [Evolving], using mutation and crossover methods.

On the mutation processes that uses the mutation strategy shown in eq.(4), it selects 3 candidates with random: r_1, r_2, r_3, as shown in Fig. 52. The new vector $\boldsymbol{v}_{i,g+1}$: is generated according the strategy shown in eq.(4), where F is set as 0.5.

(3) Crossover

Fig. 53 shows an example of crossover to generate a new candidate according the strategy shown in Eq.(5).

(4) Evaluation and Selection

And then it generates 10 children's candidates for all parent candidates when users push button [Evolving]. Users can manually select one for each pair of candidates with the checkbox interface, then press button [Selection]. Users are asked to alternately push the buttons [Evolving] and [Selection] to generate new candidates. It allows users to add the selected candidate to the case base when finding the selected candidate is satisfying.

When push the button [Selection], the system selects one for each pair of candidates (a parent and a child) and updates the selected one to parent and remove another. For each pair of candidates, when users make a selection with the checkbox, the system takes the selected one by users, otherwise the system will make the selection according to the scores, which are evaluated by the system according to the case base or by users.

Chapter 6 Evaluation

We first verified the fundamental performance of the implemented robot, then demonstrated an example conversation, next conducted a questionnaire survey and evaluated it, and finally conducted an experiment with eight examinees to verify the effectiveness for the proposed three-dimensional robot motion design by IEC and non-IEC fusion.

6.1 Fundamental Specifications of the Implemented robot

We implemented the proposed robot BaBi using Windows smartphone (MADOSMA Q601) and the developed semiconductor chip as shown in

i) Front view

ii) Back view

iii) Side view

iv) Oblique view

Figure 54: The Implemented Prototype Robot (when closed)

i) Front view

ii) Back view

iii) Side view

iv) Oblique view

Figure 55: The Implemented Prototype Robot (when opened)

Figs. 54 and 55. The three wheels and five servomotors were equipped. The size of smartphone-like form is 83.2x160x36 [mm] while that of the robot-like form is 83.2x160x170 [mm]. Its weight is 200 [g].

We verified the basic performance of the implemented robot. We confirmed that we could set the rotation speed of the motors to control the BaBi's moving speed, and the fast speed for moving was about 0.2 [m/s]. The time necessary to transform was about 3 [s]. The sequential images of transforming process were shown in Fig. 56. We also confirmed that the robot could push an object whose size and weight were about the same as the robot itself as shown in Fig. 57.

Figure 56: Sequential Images of the Implemented Prototype Robot Being Transformed

Figure 57: BaBi Pushing an Object

6.2 Example Conversation to BaBi

We show some examples of Human-BaBi interactions in Chinese as follows:

Human: says, “打开。(Open.)”

BaBi: opens and replies, “你好，我叫波比。(Hello, I am BaBi.)”

Human: asks, “我能和你聊天吗！(Can I chat with you?)”

BaBi: says, “当然可以！(Of course, you can!)”

Human: asks, “你高兴吗？(Are you happy?)”

BaBi: says, “很高兴见到你。(I am very happy to meet you.)”

Human: says, “唱个歌吧！ (Sing please!)”

BaBi: says, “好的！(Okay!)” and then sings a song.

Human: says, “跳个舞吧！(Dance please!)”

BaBi: says, “好的！(Okay!)” and then moves and rotates to show a dance while playing a music.

Human: says, “红烧肉怎么做？(Please tell me how to cook the braised pork?)”

BaBi: tells the cook method of the braised pork.

Human: asks, “牛顿第一定律是什么？(What is Newton’s first law?)”

BaBi: describes Newton’s first law.

Human: says, “说个笑话。 (Tells joke please.)”

BaBi: tells a joke.

6.3 Questionnaire Survey for phone robot BaBi

We conducted an evaluation to (1) collect feedback about the system and (2) provide directions on the most promising scenarios to be investigated in future work. To achieve this, we ask participants to make interaction with BaBi to evaluate 14 scenarios. After each scenario, we asked how much the participants liked the presented scenario (a 5 item Likert scale was used). At the end of the survey, participants were free to write down comments.

Chapter 6 Evaluation

30 participants (11 female) aged 18 to 60 years (mean = 39) completed the survey. The results of the study are reported in Fig. 58. The figure shows a high tendency of positive results.

The results revealed that:

(1) **BaBi can brings more functions and services than an ordinary smart phone.** A high number (90% and 73%) of the participants very agree or agree with the scenario as shown in (1) (2) of Fig. 58.

(2) **BaBi is compact and portable, transformable, does not add more burdens.** The participants (90%, 83% and 73%) very agree or agree with the scenario as shown in (3) - (5) of Fig. 58.

(3) **BaBi has multimodal interactions and communication and more varied express to extend current smart phone functionality to be far-field and context-driven.** The participants (87%, 80% and 77%) very agree or agree with the scenario as shown in (6) - (9) of Fig. 58.

(4) **People operates phone through voice conversation, it can free hands and doesn't delay doing anything else.** The participants (80% and 73%) very agree or agree with the scenario as shown in (10) - (11) of Fig. 58.

(5) **BaBi is interesting and funny.** The participants (83%) very agree or agree with the scenario as shown in (12) of Fig. 58.

(6) **Users are satisfied with BaBi and willing to by BaBi.** The

participants (76% and 67%) very agree or agree with the scenario as shown in (13) (14) of Fig. 58.

In summary, 67% - 90% of the participants very agree or agree with these scenarios while 0% - 10% of them disagree or very disagree the scenarios. The results show that：

(1) BaBi is compact and portable, transformable, does not add more burdens while having multimodal interactions and communication and more varied express to extend current smart phone functionality to be far-field and context-driven.

(2) People operates phone through voice conversation, it can free hands and doesn't delay doing anything else.

(3) BaBi is interesting、funny and value added application.

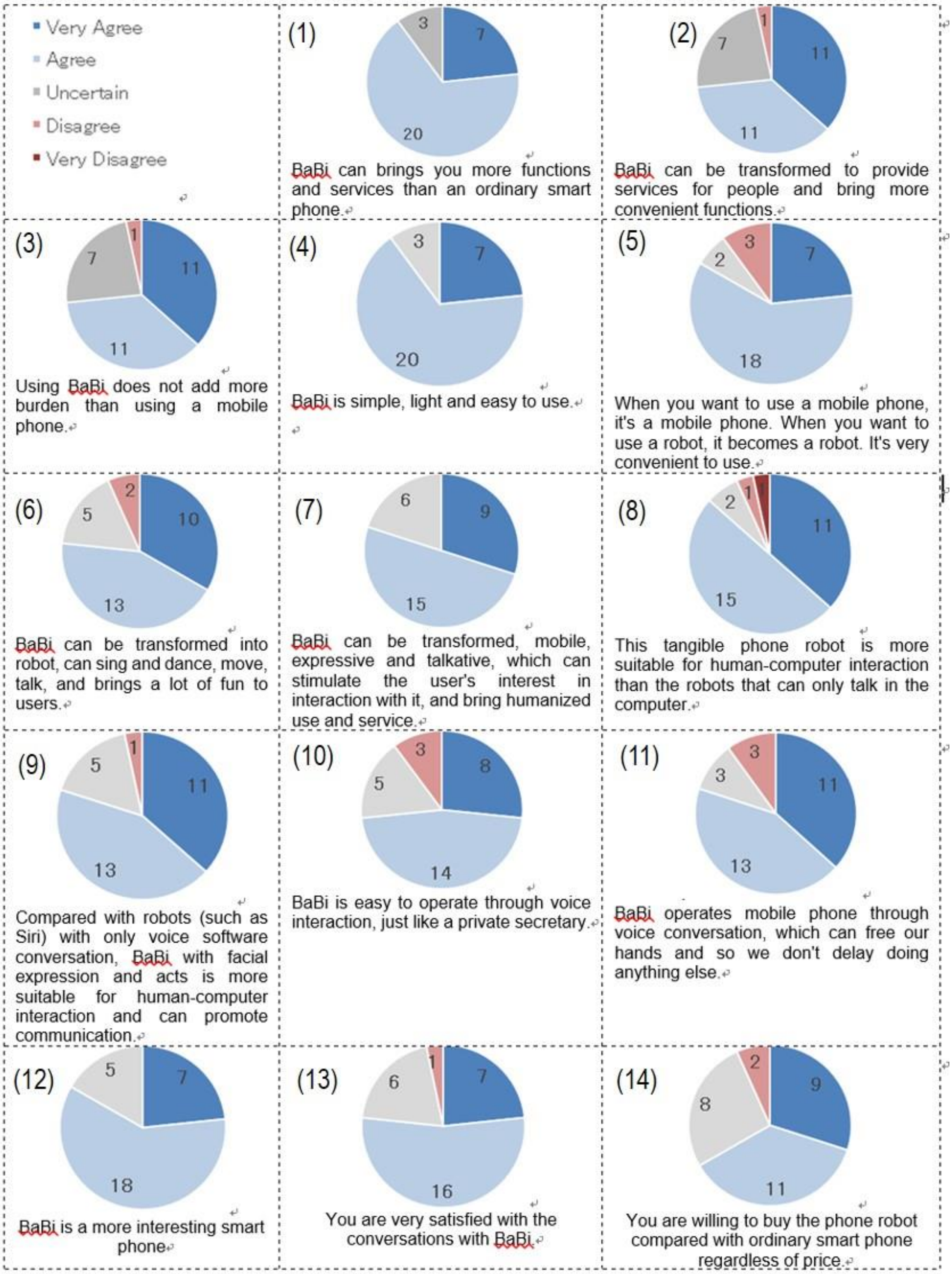


Figure 58: Results of Preliminary Questionnaire Survey

6.4 Evaluation of the study result on Three-Dimensional Robot Motion Design by IEC and non-IEC fusion

An experiment with eight examinees is conducted to verify the effectiveness for the proposed method. The following two methods are compared:

Selecting-all method: Requires each user to select one for each pair of parents (target vector) and its offspring (trial vector) in each generation. System evaluation of scores and the selection with scores for each pair of candidates are not performed. Users must manually select one for each pair of candidates, then press the [Selection]. It is not allowed to add any candidate to the case base.

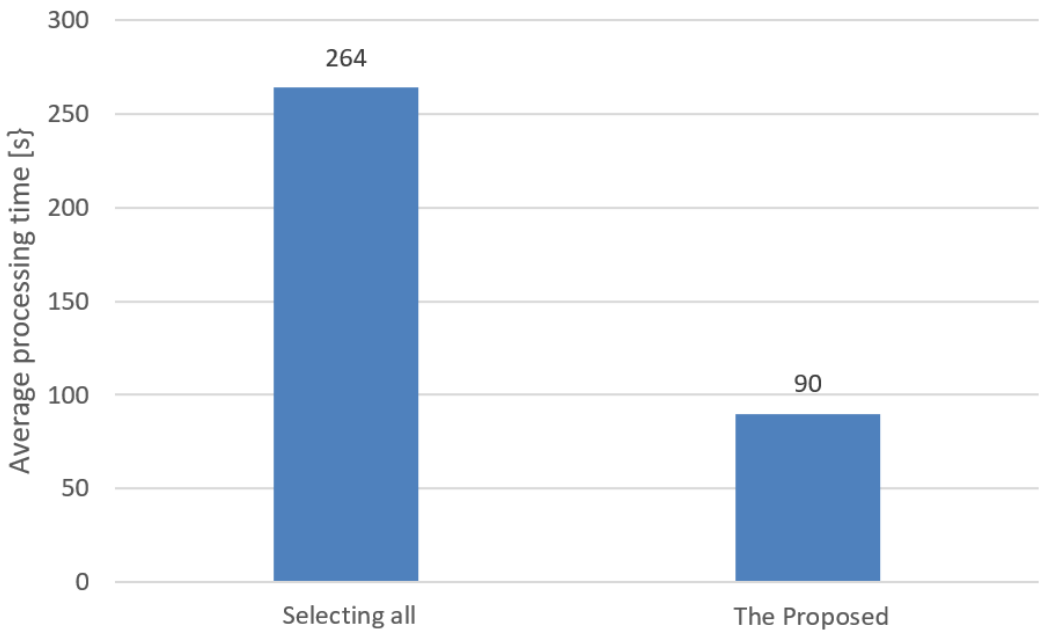


Figure 59: Processing Time

The proposed method: Allows users to add the selected candidate to the case base when finding the selected candidate is satisfying. Users may alternately push the buttons [Evolving] and [Selection] to generate face and body motion without making any selection with the checkbox

interface. System evaluation of scores and the selection with scores for each pair of candidates are performed. For each pair of candidates, users can select or not select.

In each method, “Reset” button were allowed to use at any time.

Parameters of DE were configured as follows: DE/rand/1/bin was employed as the fundamental DE strategy. The population size, crossover rate, and scale factor was set to 10, 0.5, and 0.5, respectively.

First, examinees were required to run this software and according to the above two methods, try to generate a “surprise” motion as a practice to familiar with the software operation. After it, we required examinees to test the above two methods in random order to generate three motions of “pleasure”, “sadness”, and “anger”. At each test, examinees were required to search until obtaining satisfactory solutions. At the end of the test, examinees were requested to complete a questionnaire involving the following questions on five-level rating system.

[Q1] Was the found motion satisfiable?

[Q2] Was the found motion in accordance with the desired motion image you had at the beginning of the search?

[Q3] Was the found motion better than what you expected?

[Q4] Did the method produce any unpredictable, good candidate?

[Q5] Do you think that this method is efficient and allows you to reduce the time to make a face motion?

[Q6] Do you think that this method helps you to think of something new?

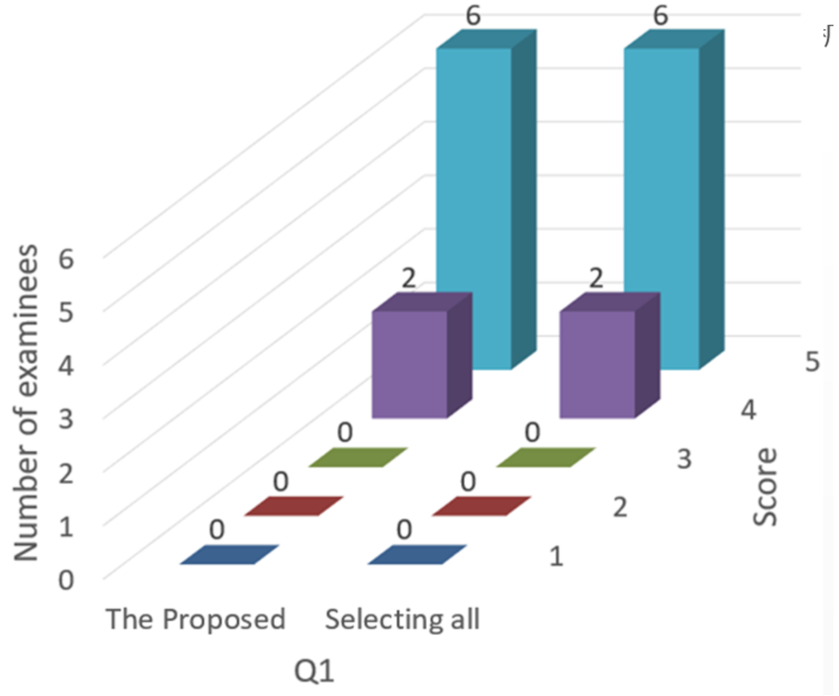


Figure 60: Questionnaires Result (Q1)

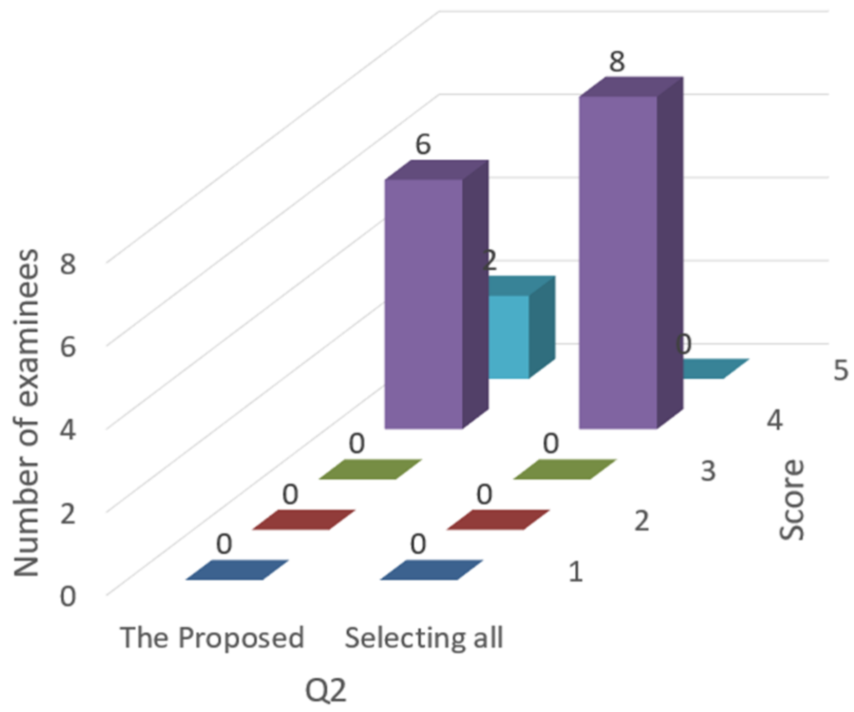


Figure 61: Questionnaires Result (Q2)

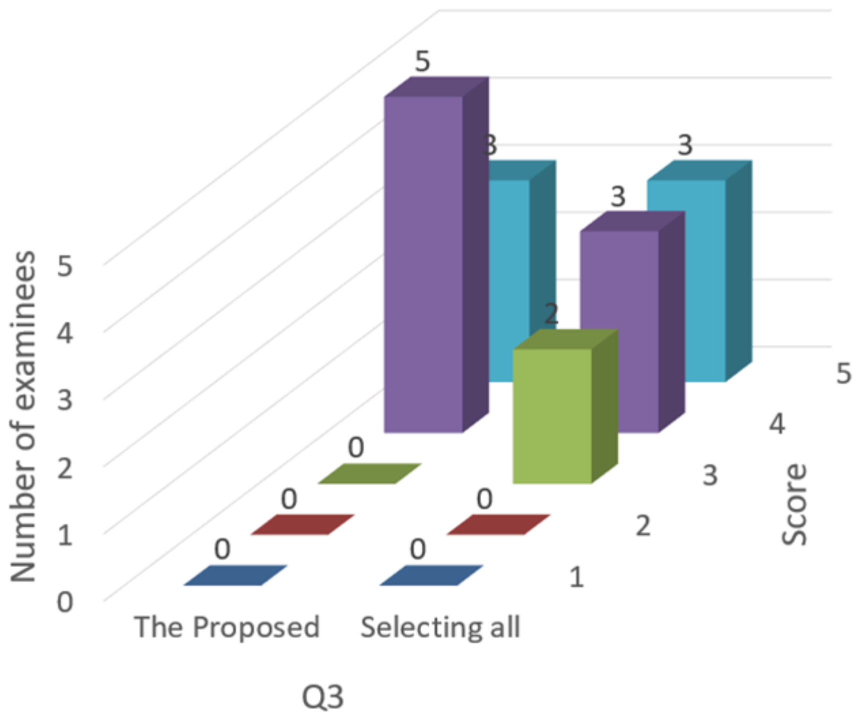


Figure 62: Questionnaires Result (Q3)

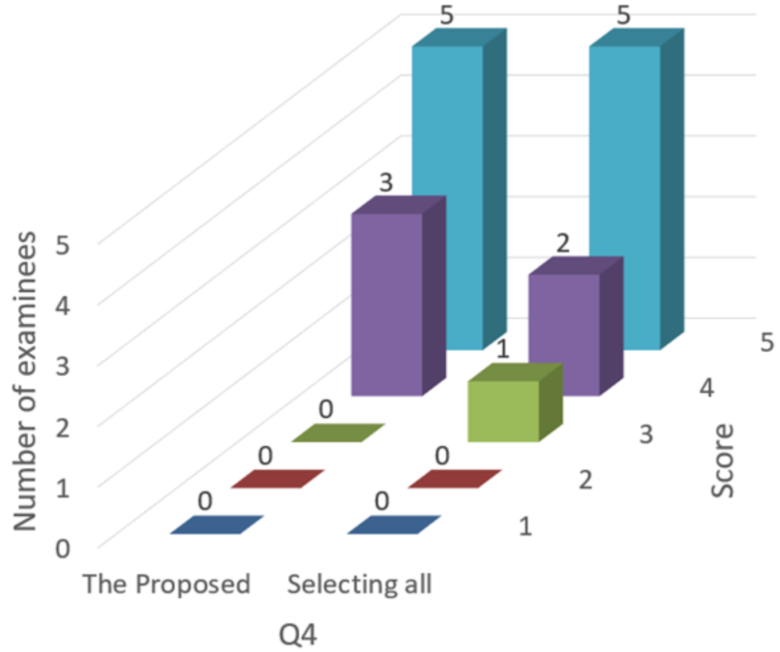


Figure 63: Questionnaires Result (Q4)

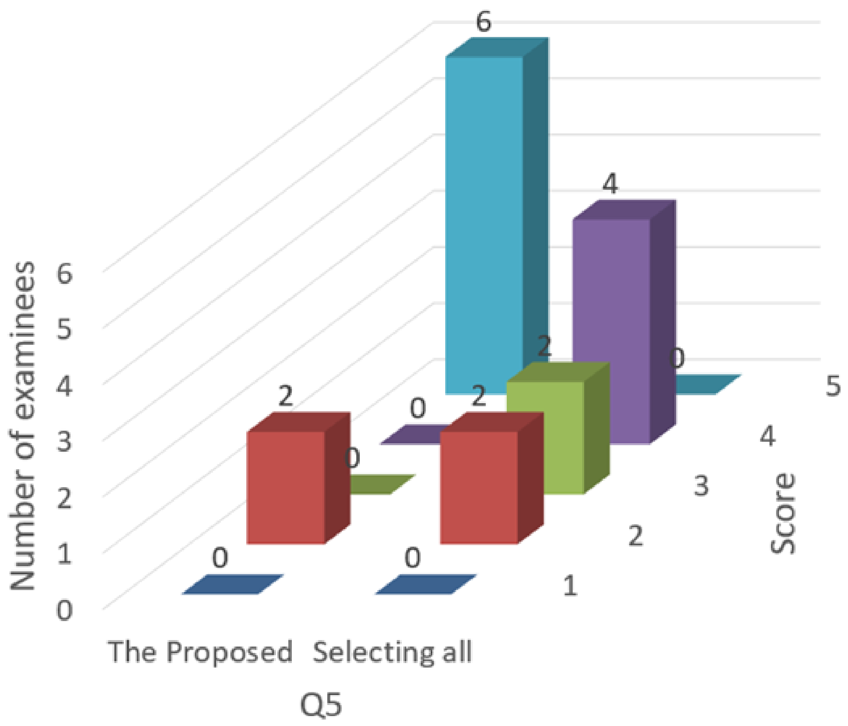


Figure 64: Questionnaires Result (Q5)

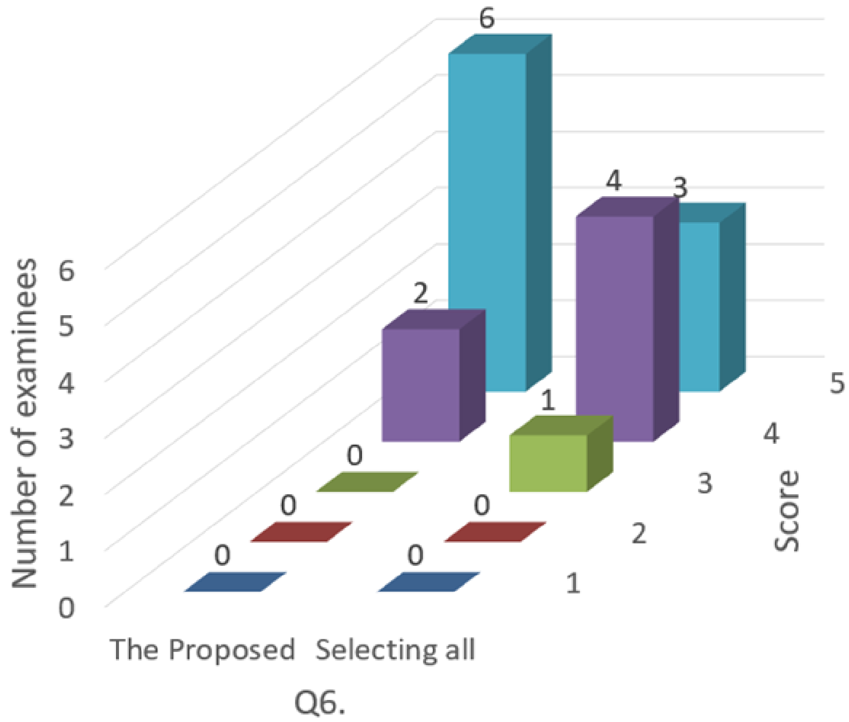


Figure 65: Questionnaires Result (Q6)

Fig.59 shows the average processing time for obtaining a motion. Fig.60-65 shows the questionnaires result.

In the selecting all method, the examinees were forced to select one for each pair of candidates at every generation, it took more process time (Fig.59). We also conducted a paired t-test for the time results and found a significant difference between the conditions ($t(7)=7{,}63$, $p<0.001$). Thus, the proposed method significantly increases the processing speed more than the selecting all method.

In the selecting all method, some examinees felt it as ineffective (Fig. 55). But most of the examinees thought they found the desired solutions (Fig. 60 and 61), and unpredictable design candidates (Fig. 63) by the Selecting all method.

By the proposed method more examinees have gotten the desired solutions, unpredictable, better design candidates, and have felt it is more effective and helps them to think of something new than the Selecting all (Fig.60-65).

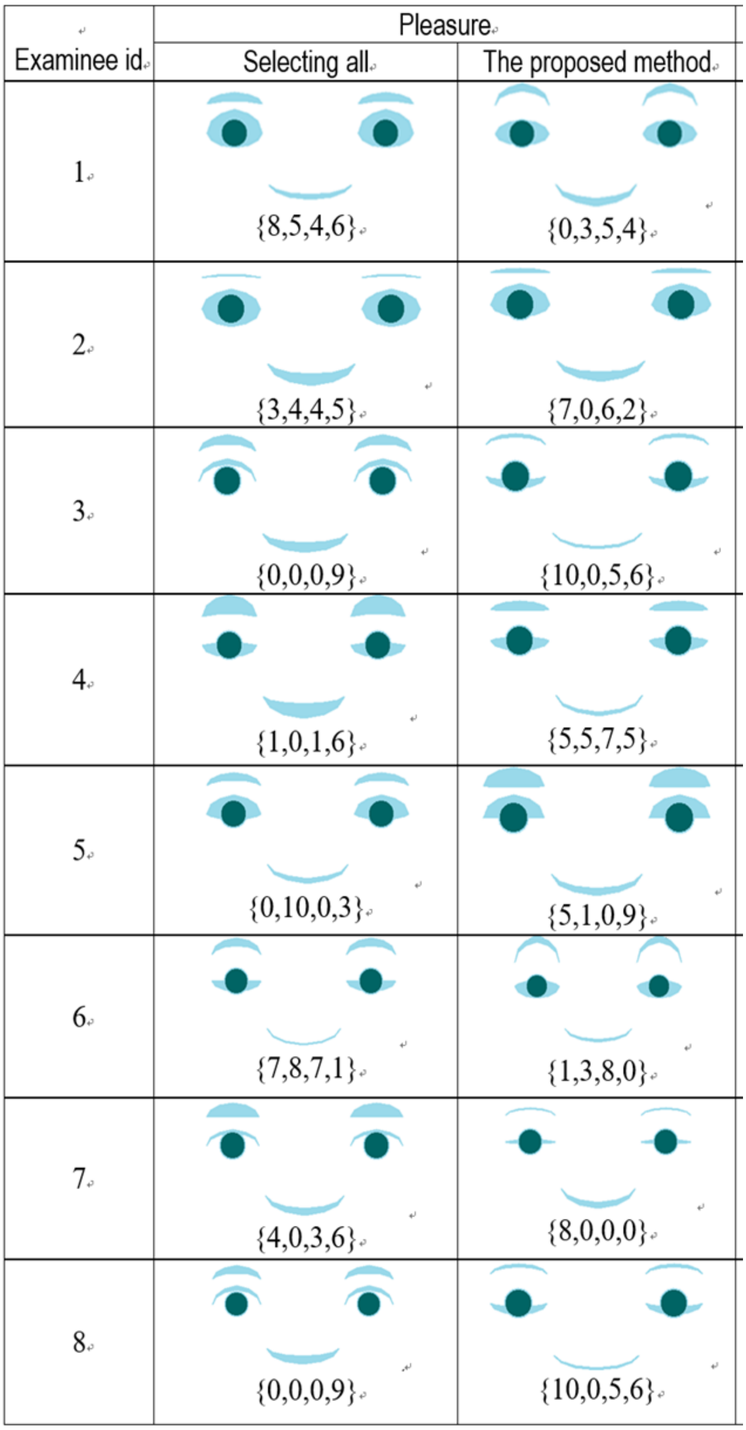
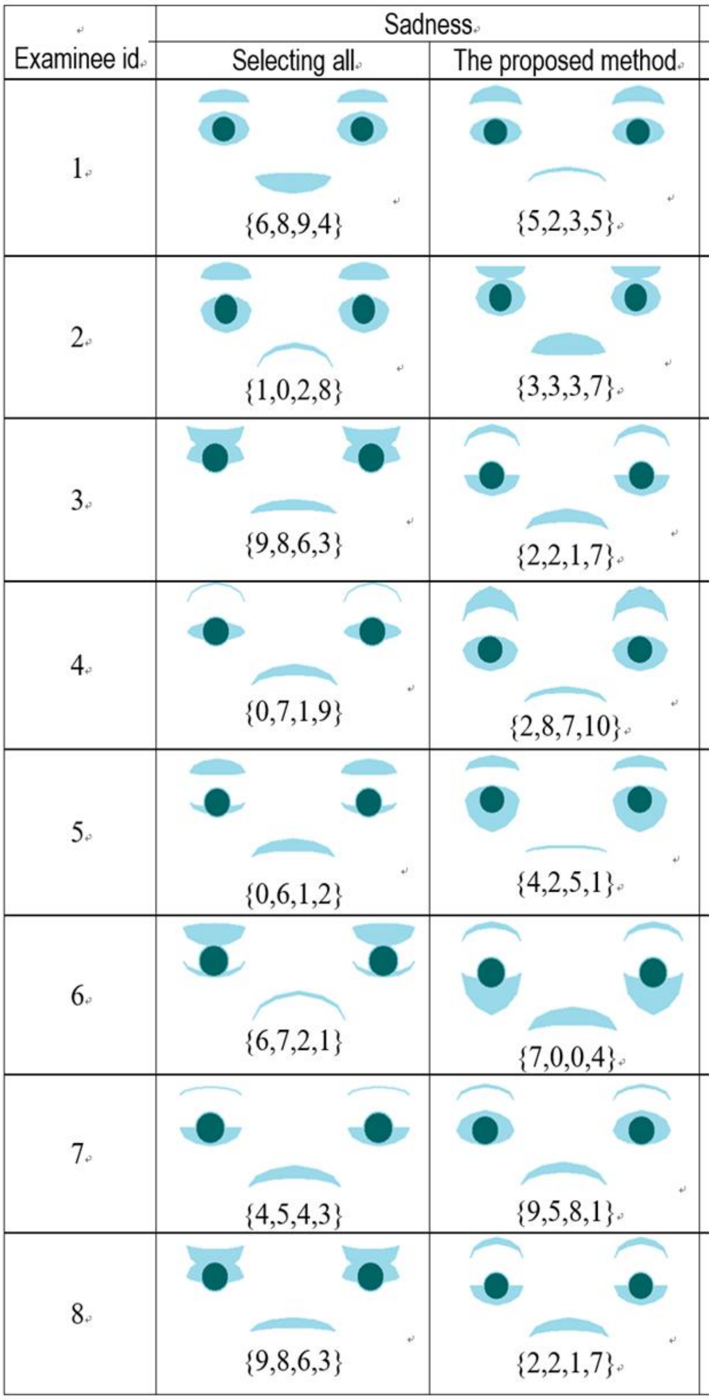


Figure 66: Generated Facial Expressions for Pleasure

(the number at the bottom of each box shows the parameters $\{P_7, P_8, P_9, P_{10}\}$, and in each box, left motion is obtained by the selecting all method while right that is obtained by the proposed method)



Examinee id	Sadness	
	Selecting all	The proposed method
1	{6,8,9,4}	{5,2,3,5}
2	{1,0,2,8}	{3,3,3,7}
3	{9,8,6,3}	{2,2,1,7}
4	{0,7,1,9}	{2,8,7,10}
5	{0,6,1,2}	{4,2,5,1}
6	{6,7,2,1}	{7,0,0,4}
7	{4,5,4,3}	{9,5,8,1}
8	{9,8,6,3}	{2,2,1,7}

Figure 67: Generated Facial Expressions for Sadness

(the number at the bottom of each box shows the parameters {P_7, P_8, P_9, P_{10}}, and in each box, left motion is obtained by the selecting all method while right that is obtained by the proposed method)

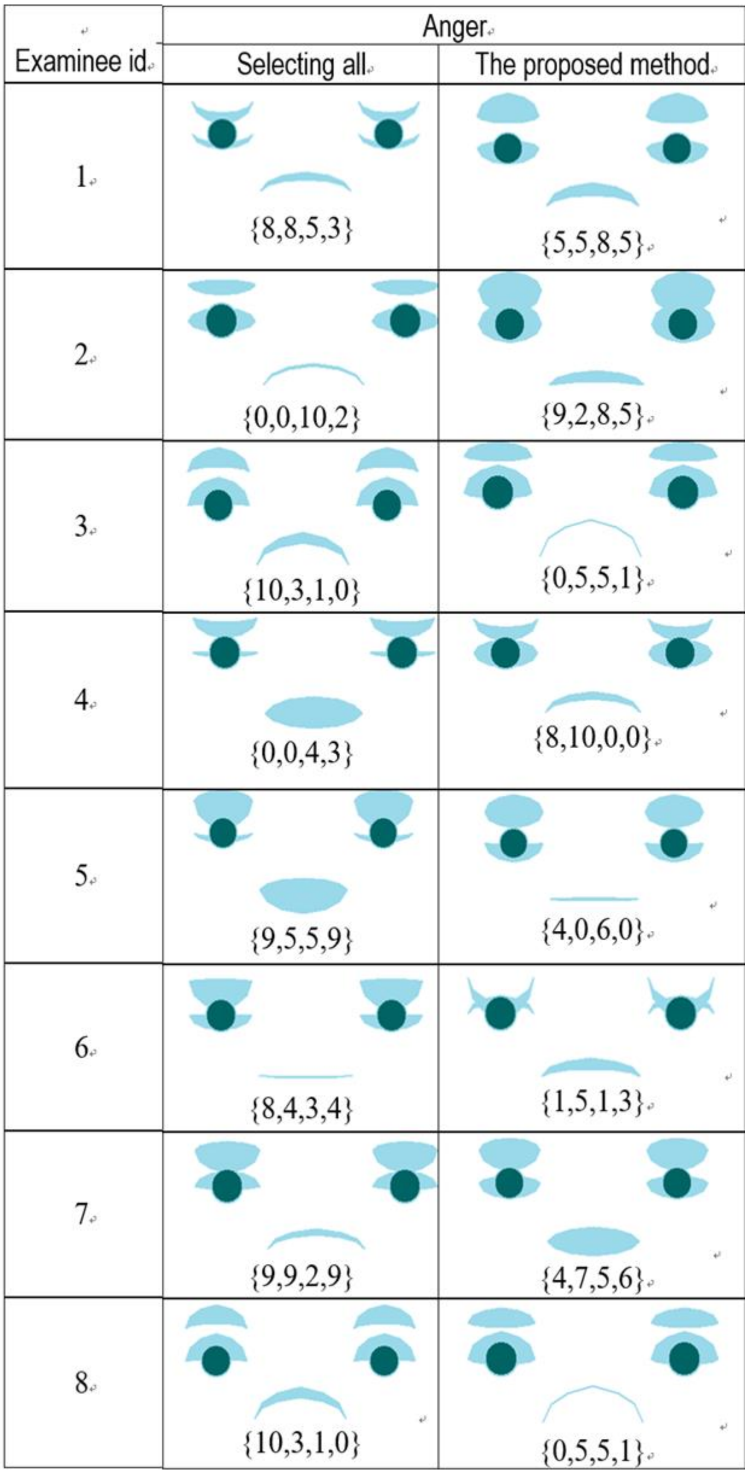


Figure 68: Generated Facial Expressions for Anger

(the number at the bottom of each box shows the parameters $\{P_7, P_8, P_9, P_{10}\}$, and in each box, left motion is obtained by the selecting all method while right that is obtained by the proposed method)

Fig. 66-68 shows the generated motions by the eight examinees, where the number at the bottom of each box shows the parameters {P_7, P_8, P_9, P_{10}}, and in each box, left motion is obtained by the selecting all method while right that is obtained by the proposed method. From the results we can see that different users expect different expressions.

On the other hand, obvious differences and characteristics among “pleasure”, “sadness”, and “anger” cannot be observed for the body motion parameters {P_7, P_8, P_9, P_{10}}. We consider that maybe it is because the examinees did not pay much attention to the body motions and they only cared the facial expressions when performing experimental evaluation. In particular, the user interface we implemented focused on the facial expression and it was hard for examinees to compare body motions between candidates. It is also necessary to give the examinees more instructions and hints. Moreover, we need to improve the expression model for the body motion. Fig. 69 shows an example of a generated motion of anger.

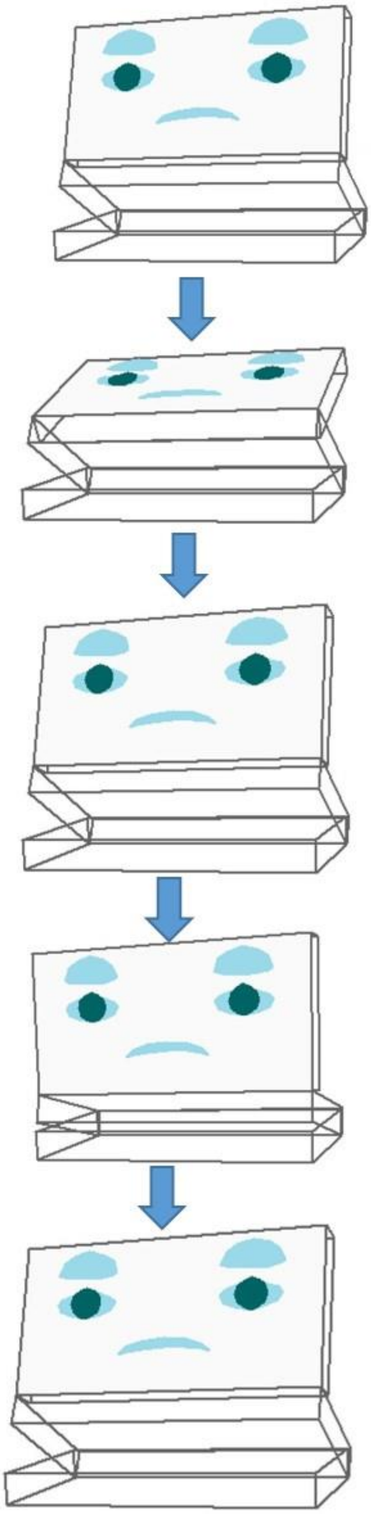
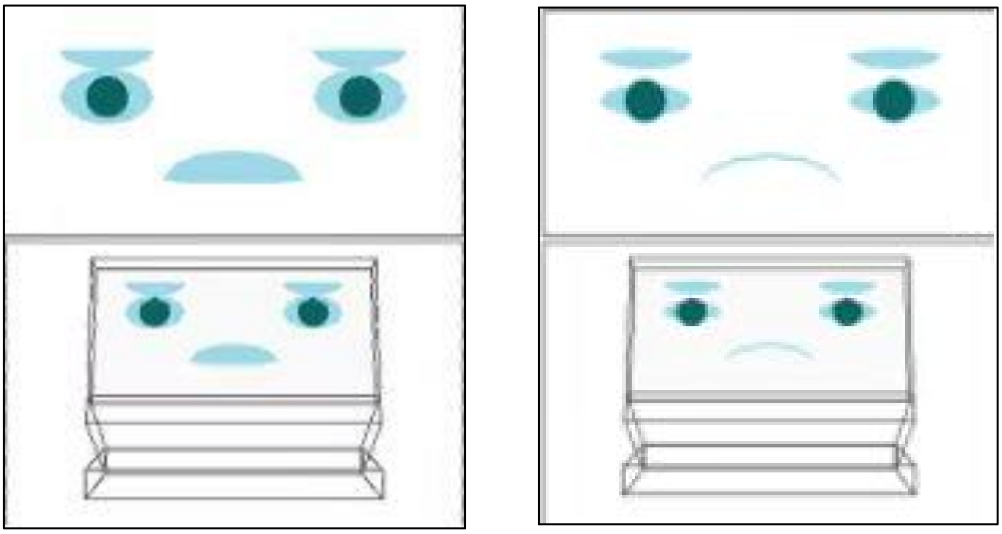


Figure 69: Example Generated Motion



(a) Suitable for BaBi. (b) Not suitable for BaBi.

Figure 70: Exmples of Facial Expressions

From the motions results we also found designing a CG face alone and designing a BaBi’s face are different. For instance, it's not so good when looking at the face alone, but it suits the BaBi body well when displaying it on the BaBi as shown in Fig. 70(a), where it seems too sad when only showing the face, but it looks good when displaying it on the BaBi body. On the other hand, it's a good design when looking at the face alone, but it is not so good when displaying on BaBi as shown in Fig. 70(b), where it seems good when only showing the facial expression. but it does not look so sad when displaying on the BaBi body because the mouth is so thin that it stands out less than the frame and body of the BaBi.

Chapter 7 Conclusion

After studying and sorting out the achievements of some smart phone robots at present, we found that many researchers and teams continue to involve in the field, and there are quite a lot of academic and application achievements. On this basis, we study and explore some new function and design application space of expanding smartphone into service robot, and this thesis provides a method to find and optimize the robot action and facial expression in designing 3D phone robot when users interact with it.

We have researched and implemented BaBi smartphone robot and used it in different scenarios. The preliminary study shows 14 scenarios we envisaged, and the preliminary feedback of users has been collected and statistics. Generally speaking, participants have a good understanding of these situations. Further evaluation of research is carried out according to different dimensions, such as attraction, function and availability. Moreover, users can carefully select and set their favorite actions and facial expressions to let the BaBi smartphone robot convey emotion.

Therefore, we propose an intelligent transformable smartphone robot as an intelligent assistant partner: BaBi. BaBi opens and automatically transforms from the moonlight treasure box (in the form of a smartphone) into a service robot. It can perform many functions similar to human beings, such as moving, talking, facial expression, singing, dancing and supporting people’s work, so as to make people happy, help and support people's daily activities. BaBi is a portable, transformable and moveable intelligent partner. We have studied and designed it to have multi-mode interaction and communication functions. Through voice interaction, people do not need to hold it with hands when interacting with BaBi, and their hands are liberated without delaying people to do other things. BaBi phone robot can also and provide more kinds of express. Users can set their favorite expressions and actions according to their preferences. We study and complete the expansion of the current smartphone functions to remote and contextual semantic interaction and operation. We have evaluated the availability of new features developed and the results show a high trend of positive results. In conclusion, 67% - 90% of participants strongly agree and agree with these scenarios.

This thesis also studies and proposed an approach for three-dimensional robot motion design by IEC and non-IEC fusion, the purpose to reduce the heavy burden of users in setting the motion and expression of 3D robot. In this study, via the intelligent transformable smartphone robot (BaBi), The experimental results show that using our proposed method, users can design the motion mode of the robot with less workload. In the experiment, 8 Examinees designed and got the happy, angry and sad facial expressions which their favorites with less workload and time.

As a comparison, Examinees should do the test with two methods, Select-all method and proposed method. In the selecting-all method, some examinees felt it as ineffective, average time spent is 264 seconds. But in the proposed method, the examinees spent more less average time 90 seconds only. A log file will be generated when examinee finished a test, so we can get all details about all examinees tests.

We also conducted a paired t-test for the time results and found a significant difference between the conditions ($t(7)=7,63$, $p<0.001$). Thus, the proposed method significantly increases the processing speed more than the selecting-all method. By the proposed method more examinees have gotten the desired solutions, unpredictable, better design candidates, and have felt it is more effective and helps them to think of something new than the Selecting-all method.

Finally, for making more contribution, in the future, we plan to improve body movements and conduct experiments with more subjects to obtain better movements, so as to realize the practical application of the new functions developed by BaBi. And make more in-depth phone robot application and practice of our new research methods for expression and motion design of three-dimensional robots.

References

[1] J. Wallén, "The history of the industrial robot," in Linköping University Electronic Press, 2008.

[2] James Auger, "Living with robots: as speculative design approach," J. Hum-Robot Interact.3, 1, pp.20-42, Feb. 2014.

[3] Oros, Nicolas, and Jeffrey L. Krichmar, "Smartphone based robotics: Powerful, flexible and inexpensive robots for hobbyists, educators, students and researchers," IEEE Robotics & Automation Magazine 1, 3. 2013.

[4] Wang, Yu, et al., "Samba: A smartphone-based robot system for energy-efficient aquatic environment monitoring," 14th International Conference on Information Processing in Sensor Networks, 2015.

[5] Jibawi, Lojain, et al., "Smartphone-based Home Robotics," ArXiv preprint arXiv:1803.02122, 2018.

[6] Berengueres, Jose, "Using smartphones for low-cost robotics," ArXiv preprint arXiv:1807.06003, 2018.

[7] Gautam Bose, Marisa Lu, Lucas Ochoa, Dan Lockton, "Emoto: From Phone to Emotive Robotic AI Sidekick," 13th International Conference on Tangible and Embedded Interaction, 2019.

[8] Ji-Dong Yim, Christopher D. Shaw, "CALLY: The cell-phone robot with affective expressions," 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2009.

REFERENCES

[9] H. Takagi, and M. Ohsaki, "Interactive Evolutionary Computation-Based Hearing Aid Fitting," IEEE Trans. on Evolutionary Computation vol.11, no.3, pp.414-427, Jun. 2007.

[10] J. Woo, J. Botzheim and N. Kubota, "Facial and gestural expression generation for robot partners," International Symposium on Micro-Nano Mechatronics and Human Science (MHS), pp.1-6, Nagoya, Japan, 2014.

[11] M. Hailemariam, B. Goertzel, T. Yohannes, "Evolving 3D Facial Expressions Using Interactive Genetic Algorithms," Proceedings of the Advances of Science and Technology (ICAST 2018), Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol.274, pp.492-502, Springer, Cham, 2018.

[12] S. Ono, and S. Nakayama, "Fusion of interactive and non-interactive evolutionary computation for two-dimensional barcode decoration," IEEE Congress on Evolutionary Computation, pp.1-8, Barcelona, Spain, 2010.

[13] S. Ono, H. Maeda, K. Sakimoto and S. Nakayama, "User-System Cooperative Evolutionary Computation for both Quantitative and Qualitative Objective Optimization in Image Processing Filter Design," Applied Soft Computing, vol.15, no.2, pp.203-218, 2014. https://doi.org/10.1016/j.asoc.2013.10.019.

[14] M. Kamizono, K. Shimomura, M. Tajiri, and S. Ono, "Two-Dimensional Barcode Decoration Using Module-wise Non-

systematic Coding and Cooperative Evolution by User and System," in Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '16), Association for Computing Machinery, pp.245-252, New York, NY, USA, 2016.
DOI: https://doi.org/10.1145/2908812.2908953

[15] C. K. Riesbeck and R. C. Schank, "Inside Case-Based Reasoning," Hillsdale, New Jersey, Lawrence Erlbaum, 1989.

[16] R. Storn, and K. Price, "Differential Evolution-A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," Journal of Global Optimization, vol.11, pp.341-359, 1997.

[17] ROS Robots. (n.d.) "Robot Video Montages," Retrieved August 2, 2021. From: https://robots.ros.org/

[18] J. Woo and N. Kubota, "Conversation System Based on Computational Intelligence for Robot Partner Using Smart Phone," in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, pp.2927-2932, Manchester, UK, 2013.

[19] J. Woo, J. Botzheim, and N. Kubota, "Conversation System for Natural Communication with Robot Partner, " in Proceedings of the 10th France-Japan and 8th Europe-Asia Congress on Mechatronics, Tokyo, Japan, 2014.

[20] E. Groff, "Laban Movement Analysis: Charting the Ineffable Domain of human Movement," J. Phys. Educ. Recreate. Dance., vol.66, no.2, pp. 27-30, Feb. 1995.

[21] M. M. Gross, E. A. Crane, and B. L. Fredrickson, "Effort-Shape and

kinematic assessment of bodily expression of emotion during gait," Hum. Mov. Sci., vol.31, no.1, pp.202-221, 2012.

[22] C. D. Kidd and C. Breazeal, "Effect of a robot on user perceptions,".

[23] Radford, A., Metz, L., Chintala, S., "Unsupervised representation learning with deep convolutional generative adversarial networks," ArXiv preprint arXiv:1511.06434, 2015.

[24] Risi,S., Lehman,J., D'Ambrosio, D.B., Hall, R., Stanley, K.O.: "Petalz, Search-based procedural content generation for the casual gamer," IEEE Transactions on Computational Intelligence and AI in Games 8(3), pp.244-255, Sep. 2016.

[25] Secretan, J., Beato, N., D Ambrosio, D.B., Rodriguez, A., Campbell, A., Stanley, K.O.,"Pic breeder: evolving pictures collaboratively online," in proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.1759-1768. ACM, 2008.

[26] Takagi, H.: "Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation," Proceedings of the IEEE 89(9), pp.1275-1296, 2001.

[27] Forrest, S. Genetic algorithms: "Principles of adaptation applied to computation," Science 261: pp.872-878, 1993.

[28] Bonabeau, E. "Agent-based modeling: methods and techniques for simulating human systems," Proceedings of the Nat. Acad. SCI. USA pp.7280-7287, 2002.

[29] Naoki Ohshima, Yasuke Ohyama, Yuki Odahara, P. Ravindra S. De Silva, and Michio Okada, "Talking-Ally: The Influence of Robot

Utterance Generation Mechanism on Hearer Behaviors," International Journal of Social Robotics, 2014.

[30] Cynthia Breazeal, and Brian Scassellati, "A context-dependent attention system for a social robot," IJCAI '99 Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, pp.1146-1153, 1999.

[31] Kazuyoshi Wada, Takanori Shibate, Tomoko Saito, and Kazuo Tanie, "Effects of robot-assisted activity for elderly people and nurses at a day service center," Proceedings of the IEEE, Vol.92, No.11, Nov. 2004.

[32] Shintaro Watanabe, "Implementation and application of robot information processing functions-Using communication robots for elderly care support," New Breeze Autumn, pp.10-13, 2015.

[33] Kazuko Obatashi, "Elderly support by robot and sensors with cloud robotics," Universal Accessibility Evaluation Organization, 2016.

[34] Yoshihiro Fujita, "Personal Robot R100," Journal of the Robotics Society of Japan, Vol.18, No.2, pp.198-199, 2000.

[35] Yamato Iwamura, Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro and Norihiro Hahita, "Chat and Appearance Effects for Robots which Assist Elderly's Shopping," Journal of the Robotics Society of Japan, Vol.31, No.1, pp.60-70, 2013.

[36] Takayuki Kanda, "How a Communication Robot can Contribute to Education," Journal of Japanese Society for Artificial Intelligence, Vol.23, No.2, 2008.

References

[37] Testsuo Ono, Michita Imai, Hiroshi Ishiguro and Ryohei Nakatsu, "Embodied Communication Emergent from Mutual Physical Expression between Human and Robots," Transactions of Information Processing Society of Japan, Vol.42, No.6, 2001.

[38] Kazuko Obayashi, and Shigeru Masuyama, "The effects of Introducing Communicative Robots with Infra-red Radiation Monitoring System on Night Shift Duties of Nursing Facility Care Worker," Journal of the Robotics Society of Japan, Vol.31, No.1, pp.60-70, 2013.

[39] M. Gurman, "IOS 9.2 to bring Arabic support to Siri following UAE store openings," 9 to 5 Mac. Nov. 6, 2015.

[40] Chris Lau, "Why Cortana Assistant Can Help Microsoft in the Smartphone Market," The Street, Mar. 18, 2014.

[41] Morteza Husainy Yar, Vahid Rahmati, and Hamid Reza Dalili Oskouei, "A Survey on Evolutionary Computation: Methods and Their Applications in Engineering," Modern Applied Science, Vol. 10, No. 11, pp.1913-1852, 2019.

[42] Shigenobu Kobayashi, "Invitation to Evolutionary Computation," The Japanese Society for Evolutionary Computation, Vol. 1, No. 1, 2010.

[43] Bäck, T., & Schwefel, H. P. "An Overview of Evolutionary Algorithms for Parameter Optimization. Evolutionary Computation," 1(1), pp.1-23, 1993.

[44] Das, S., Abraham, A., & Konar, A. "Particle swarm optimization

and differential evolution algorithms: Technical analysis, applications and hybridization perspectives," Studies in Computational Intelligence, 116, pp.1-38, 2008.

[45] Wierstra, D., Schaul, T., Peters, J., & Schmidhuber, J. "Natural Evolution Strategies," IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), pp.3381-3387, 2008.

[46] Hansen, N. "The CMA evolution strategy: a comparing review, Towards a new evolutionary computation," Advances on estimation of distribution algorithms, Springer, pp.1769-1776, 2006.

[47] Kumar, M., Husian, M., Upreti, N., & Gupta, D. "Genetic Algorithm: Review and Application," International Journal of Information Technology and Knowledge Management, 2(2), pp.451-454, 2010.

[48] Whitley, D. "A genetic algorithm tutorial. Statistics and Computing," 4(2), pp.65-85, 1994.

[49] Fogel, D. B. "Introduction to Evolutionary Computation. Modern Heuristic Optimization Techniques: mas.ccsenet.org Modern Applied Science," Vol. 10, No. 11; 137 Theory and Applications to Power Systems, pp.1-23, 2019.

[50] Kennedy, J., & Eberhart, R. (b). "Particle swarm optimization. IEEE International Conference on Particle Swarm Optimization," 1995.

[51] Mahor, A., Prasad, V., & Rangnekar, S. "Economic dispatch using particle swarm optimization: A review. Elsevier," 13(8), 2134-2141.

2009.

[52] Kennedy, J., & Eberhart, R. (a). "Particle swarm optimization. Neural Networks," Proceedings IEEE International Conference on, 4, pp.1942-1948, 1995.

[53] Poli, R., Kennedy, J., & Blackwell, T. "Particle swarm optimization. Swarm Intelligence," 1(1), 33-57, 2007.

[54] Yan Pei and Hideyuki Takagi, "Research progress survey on interactive evolutionary computation," Journal of Ambient Intelligence and Humanized Computing, May 2018.

[55] Hideyuki Takagi, "Interactive Evolutionary Computation: Fusion of the Capabilities of EC Optimization and Human Evaluation," Proceedings of the IEEE. 89 (9), pp.1275-1296, Sep. 9, 2001.

[56] S. Bergen, B.J. Ross, "Automatic and interactive evolution of vector graphics images with genetic algorithms," The Visual Computer, vol.28, pp.35-45, Jan. 2012.

[57] D. Buchsbaum, P. Funes, J. Budynek, H. Koppermann and E. Bonabeau, "Designing collective behavior in a group of humans using a real-time polling system and interactive evolution," Proceedings of the IEEE Swarm Intelligence Symposium, pp.15-21, Pasadena, CA, USA, 2005.

[58] P.Bontrager, W. Lin, J. Togelius,and S. Risi, "Deep Interactive Evolution," Computational Intelligence in Music, Sound, Art and Design (EvoMUSART), Lecture Notes in Computer Science, vol.10783, Springer, Cham, pp267-282, 2018

References

[59] J. Inthiam, E. Hayashi, W. Jitviriya and A. Mowshowitz, "Development of an emotional expression platform based on LMA-shape and interactive evolution computation," Proceedings of 4th International Conference on Control, Automation and Robotics (ICCAR), pp.11-16, Auckland, 2018.

[60] J. Woo, J. Botzheim and N. Kubota, "Facial and gestural expression generation for robot partners," International Symposium on Micro-Nano Mechatronics and Human Science (MHS), pp.1-6, Nagoya, Japan, 2014.

[61] Y. Suga, Y. Ikuma, D. Nagao, S. Sugano and T. Ogata, "Interactive evolution of human-robot communication in real world," Proceedings of 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.1438-1443, Edmonton, AB, Canada, 2005.

[62] M. Virčíková, P. Sinčák, "Discovering Art in Robotic Motion: From Imitation to Innovation via Interactive Evolution," Proceedings of the Ubiquitous Computing and Multimedia Applications (UCMA), Communications in Computer and Information Science, vol.150. Springer, pp.183-190, Berlin, Heidelberg, 2011.

[63] M. A. Lewis, A. H. Fagg, and A. Solidum, "Genetic programming approach to the construction of a neural network for control of a walking robot," in IEEE Int. Conf. on Robotics and Automation, vol.3, (Nice, France), pp.2618-2623, May 1992.

[64] F. Gruau and K. Quatramarau, "Cellular encoding for interactive

evolutionary robotics," Cognitive Science Research Paper 425, University of Sussex, Jul. 1996.

[65] G. Dozier, "Evolving robot behavior via interactive evolutionary computation: From real-world to simulation," in 16th ACM Symposium on Applied Computing (SAC), (Las Vegas, USA), pp.340-344, Mar. 2001.

[66] Mishaim Malik, Muhammad Kamran Malik, Khawar Mehmood and Imran Makhdoom, "Automatic speech recognition: a survey," Multimedia Tools and Applications, 80(3), pp.1-47, Mar. 2021.

[67] Tatsuya Kawahara, "A Brief History of Spoken Dialogue Systems- Evolution and Recent Technical Trend," The Japanese Society for Artificial Intelligence, Vol.28, No.1, 2013.

[68] Yifan Fan, Xudong Luo, Pingping Lin, "A Survey of Response Generation of Dialogue Systems," International Journal of Computer and Information Engineering, Vol.14, No.12, 2020.

[69] Suhas R. Mache, Manasi R. Baheti, C. Namrata Mahender, "Review on Text-To-Speech Synthesizer," International Journal of Advanced Research in Computer and Communication Engineering, Vol.4, Issue 8, Aug. 2015.

[70] Javier Ramırez et al, "Efficient voice activity detection algorithms using long-term speech information, " In: Speech communication, 42.3, pp.271-287, 2004.

[71] Turing Robot. (n.d.) Turing Robot. Retrieved August 2, 2021. From http://www.turingapi.com

References

[72] V.Ho-Huu, T.Nguyen-Thoi, M.H.Nguyen-Thoi, L.Le-Anh, "An improved constrained differential evolution using discrete variables (D-ICDE) for layout optimization of truss structures," Expert Systems with Applications, Volume 42, pp.7057-7069, Issue 20, 15 Nov.15, 2015.

[73] M.Fatih Tasgetiren, P.N.Suganthan, Quan-KePan, "An ensemble of discrete differential evolution algorithms for solving the generalized traveling salesman problem," Applied Mathematics and Computation, Volume 215, pp.3356-3368, Issue 9, Jan.1, 2010.

[74] J. Brest, A. Zamuda, B. Boskovic, M. S. Maucec and V. Zumer, "Dynamic optimization using Self-Adaptive Differential Evolution," IEEE Congress on Evolutionary Computation, pp.415-422, 2009. DOI: 10.1109/CEC.2009.4982976.

[75] V. L. Huang, A. K. Qin and P. N. Suganthan, "Self-adaptive Differential Evolution Algorithm for Constrained Real-Parameter Optimization," IEEE International Conference on Evolutionary Computation, pp.17-24, 2006. DOI: 10.1109/CEC.2006.1688285.

[76] Jingqiao Zhang and A. C. Sanderson, "JADE: Self-adaptive differential evolution with fast and reliable convergence performance," IEEE Congress on Evolutionary Computation, pp.2251-2258, 2007. DOI: 10.1109/CEC.2007.4424751.

[77] Tanabe R, Fukunaga A. "Success-History Based Parameter Adaptation for Differential Evolution," IEEE Congress on Evolutionary Computation (CEC), pp.71-78, 2013.

References

[78] Awad NH, Ali MZ, Suganthan PN, Reynolds RG. "An ensemble sinusoidal parameter adaptation incorporated with L-SHADE for solving CEC2014 benchmark problems, " Proceedings of the IEEE Congress on Evolutionary Computation, pp.2958-2965, 2016.

[79] Jiansheng Liu, Zhu Bilan, "A prototyping of BoBi secretary robot, "Proceedings of the Future Technologies Conference (FTC), pp.1-5, 2017.

[80] Jiansheng Liu, Fujia Sun, Satoshi Ono, Mutsumi Watanabe and Bilan Zhu, "An Intelligent Transformable Phone Robot: BaBi," IEEE International Conference on Advanced Robotics and Mechatronics (ICARM), pp.368-373. 2020.

[81] Jiansheng Liu, Satoshi Ono, and Bilan Zhu, "Three-Dimensional Robot Motion Design by Combining Interactive and Non-Interactive Evolutionary Computation for an Intelligent Transformable Phone Robot: BaBi," Proceedings of the IEEE Congress on Evolutionary Computation (CEC), pp.5-6, 2021.

Figures

Figures

Figures

Figures

Tables

Acknowledgements

Time flies, time goes by, and in a twinkling of an eye, the three-year doctoral student's life is coming to an end. Looking back, thoughts abound. After several years of hard study and work, I have had the joy of success and the depression of failure. There are too many people and things worth remembering.

First, I would like to thank my mentor Professor Satoshi Ono and Professor Bilan Zhu for their teach and admonish with patience and care. In the process of my study, from the topic selection, experimental design and implementation, to the writing and modification of the paper, every link condenses the sweat and efforts of my mentor. My mentor's international vision, cutting-edge and quintessential academic attainments, rigorous and diligent academic style, calm, optimistic, open-minded and personal style of life make me understand, not only how to do valuable research, but also how to learn to write a good paper. which have a profound impact on my future work and life. Hereby now， I take this opportunity to express my most sincere thanks to my mentors!

I would like to thank Vice-President professor Mutsumi Watanabe and professor Takayasu Fuchida， They have rich experiences in the research field of robots and human-computer interaction algorithms. During the research process of my doctoral thesis, they pay much serious attention to and review each progress report of my presentation, and patiently put forward guidance. Many problems are solved with their specific guidance and help. Their exquisite academic experience and research philosophy, broad scientific research vision and rigorous work style have set a benchmark for my future research work!

This experiment was supported by the State Key Laboratory of intelligent technology and system of Tsinghua University and the laboratory of Mechanical Research Institute of Shanghai University of technology. In the process of experimental design and experimental implementation verification, we have gotten the support and help of Professor Liu Huaping and Sun Fujia. I would like to express my sincere thanks to them!

Thank you very much to Xie Huilong、Liu Jian、Huang Jing and others colleagues of Beijing Shangjiao NewReal-WinTop Intelligent Technology Research Institute, and thanks to engineer of Shinji Yamada and Tian Shuan from Zerone design Consulting Institute for their guidance and help in the topic selection of the shape design and the experimental of my research!

On the occasion of graduation, I would like to thank all teachers and classmates in Kagoshima University. Classmates are Dai Lulin、Zhang Yi、 Chen Bo and other doctoral students in the laboratory for their selfless help in studies. Special thanks to elder martial brother Zhang Yi and elder sister Dai Lulin for their great help in the process of experiment and my research sorting! Hereby, I would like to express my sincere thanks to them and all the teachers and students of Kagoshima University who have cared and helped me!

ACKNOWLEDGEMENTS

Thank my classmates Zhang Yuyan, Zheng Lei and Li Guoqing, Dr. Yuan Xiaohu of Tsinghua University and Dr. Hu Bin of the Institute of automation of the Chinese Academy of Sciences, Thank my teachers Professor Liu Changkui and teacher Han Ji of Donghua University for their encouragement and support during my doctoral career!

As soon as I finish my studies, I would like to thank my family for their understanding and support for my studies over the years. Because of my overseas study, I don't have enough time to accompany them, care for them and often make them worry about me. My son and daughter are still very obedient, study very hard, and their grades are excellent. They didn't disappoint me. They and my wife all supported my doctoral study without complaints and blames. Thank my wife for her understanding and support of my studies! It is her silent and selfless dedication that makes me have the confidence and perseverance to complete my studies. Without them, I can't imagine how I would spend this challenging and difficult learning career. I sincerely wish my relatives health and happiness!

Finally, I would like to pay high tribute to the teachers, friends and relatives who have made important contributions to the research and experiment of this paper and the one who completed the questionnaire test, software program test and prototype test, Thank you very much!

This paper is dedicated to all my relatives, teachers, students and friends who care, help, support and encourage me! I will continue to strive to create more achievement and progresses.