

## Research Article

# An Intelligent Big Data Management System Using Haar Algorithm-Based Nao Agent Multisensory Communication

**Fatmah Abdulrahman Baothman** 

*Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21431, Saudi Arabia*

Correspondence should be addressed to Fatmah Abdulrahman Baothman; [fbaothman@kau.edu.sa](mailto:fbaothman@kau.edu.sa)

Received 4 March 2021; Accepted 4 May 2021; Published 14 July 2021

Academic Editor: Nawab Muhammad Faseeh Qureshi

Copyright © 2021 Fatmah Abdulrahman Baothman. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Artificial intelligence (AI) is progressively changing techniques of teaching and learning. In the past, the objective was to provide an intelligent tutoring system without intervention from a human teacher to enhance skills, control, knowledge construction, and intellectual engagement. This paper proposes a definition of AI focusing on enhancing the humanoid agent Nao's learning capabilities and interactions. The aim is to increase Nao intelligence using big data by activating multisensory perceptions such as visual and auditory stimuli modules and speech-related stimuli, as well as being in various movements. The method is to develop a toolkit by enabling Arabic speech recognition and implementing the Haar algorithm for robust image recognition to improve the capabilities of Nao during interactions with a child in a mixed reality system using big data. The experiment design and testing processes were conducted by implementing an AI principle design, namely, the three-constituent principle. Four experiments were conducted to boost Nao's intelligence level using 100 children, different environments (class, lab, home, and mixed reality Leap Motion Controller (LMC)). An objective function and an operational time cost function are developed to improve Nao's learning experience in different environments accomplishing the best results in 4.2 seconds for each number recognition. The experiments' results showed an increase in Nao's intelligence from 3 to 7 years old compared with a child's intelligence in learning simple mathematics with the best communication using a kappa ratio value of 90.8%, having a corpus that exceeded 390,000 segments, and scoring 93% of success rate when activating both auditory and vision modules for the agent Nao. The developed toolkit uses Arabic speech recognition and the Haar algorithm in a mixed reality system using big data enabling Nao to achieve a 94% success learning rate at a distance of 0.09 m; when using LMC in mixed reality, the hand sign gestures recorded the highest accuracy of 98.50% using Haar algorithm. The work shows that the current work enabled Nao to gradually achieve a higher learning success rate as the environment changes and multisensory perception increases. This paper also proposes a cutting-edge research work direction for fostering child-robots education in real time.

## 1. Introduction

Artificial intelligence (AI) was introduced half a century ago. Researchers initially wanted to build an electronic brain equipped with a natural form of intelligence. The concept of AI was heralded by Alan Turing in the 1950s, who proposed the Turing test to measure a form of natural language (symbolic) communication between humans and machines. In the 1960s, Lutfi Zadah proposed fuzzy logic with dominant knowledge representation and mobile robots [1]. Stanford University created the Automated Mathematician to explore new mathematical theories based on a heuristic algo-

gorithm. However, AI had become unpopular in the 1970s due to its inability to meet unrealistic expectations. The 1980s offered a promise for AI as sales of AI-based hardware and software for decision support applications exceeded \$400 million [2]. By the 1990s, AI had entered a new era by integrating intelligent agent (IA) applications into different fields, such as games (Deep Blue, which is a chess program developed at Carnegie Mellon that defeated the world champion Garry Kasparov in 1997), spacecraft control, security (credit card fraud detection, face recognition), and transportation (automated scheduling systems) [3–7]. The beginning of the 21st century witnessed significant advances in AI in

industrial business and government services with several initiatives, such as intelligent cities, intelligent economy, intelligent industry, and intelligent robots [3].

A unified definition of AI has not yet been offered; however, the concept of AI can be built from different definitions:

- (i) It is an interdisciplinary science because it interacts with cognitive science
- (ii) It uses creative techniques in modeling and mapping to improve average performance when solving complex problems
- (iii) It implements different processes to imitate intelligent human or animal behavior. Fourth, the developed system is either a virtual or a physical system with intelligent characteristics
- (iv) It attempts to duplicate human mental and sensory systems to model aspects of “humans” thoughts and behaviors
- (v) It passes the intelligence test if it interacts completely with other systems or creatures worldwide and in real time
- (vi) It follows a defined cycle of sense–plan–act

The present study proposes the definition of AI as follows: “AI is an interdisciplinary science suitable for implementation in any domain that uses heuristic techniques, modeling, and AI-based design principles to solve complex problems. Single or combined processes in perceiving, reasoning, learning, understanding, and collaborating can improve system behavior and decision-making. The goal of AI is to enable virtual and physical intelligent agents, including humans and/or systems that continuously upgrade their intelligence to attain superintelligence. Agents should be able to integrate with one another in fully learning, teaching, adapting themselves to dynamic environments, communicating logically, and functioning efficiently with one another or with other creatures in the world and real time through sense–plan–act–react cycles.” The three-constituent principle for an agent suggests that “designing an intelligent agent involves constituents, the definition of the ecological niche, the definition of the desired behaviors and tasks, and design of the agent [8, 9].” Therefore, an agent’s intelligence can grow in time using the “here and now” perspective during interactions in different dynamic environments. In the present study, the robot agent Nao’s design is not among the required tasks, but the other two constituents are related to the environment and involve interactions with a human agent. Therefore, this work defines the ecological niche using different environments (a classroom, a lab, and a home), focusing on a mixed-reality environment. Nao’s functions are present according to the desired behavior as teaching simple mathematics to a child. The objective is to improve Nao’s learning ability and increase its intelligence. Thus, the study shows that “the three-constituents principle, the definition of the ecological niche, and the definition of the desired behaviors and tasks [2]” are sufficient to increase Nao’s intelligence.

- (i) The “here and now” perspective: related to three-time frames and shows that the behavior of any ‘agent’s system matures over a certain period and is associated with three states
- (ii) State-oriented: describes the actual mechanism of the agent at any instance of time
- (iii) Learning and development: relates to learning and development from state-oriented action
- (iv) Evolutionary: explains the emergence of a higher level of cognition through a phylogenetic perspective by emphasizing the power of artificial evolution and performing more complex tasks

The Mixed Reality System TouchMe provides a third-person camera view of the system instead of human eyes [10]. The third-person camera view is considered more efficient for inexperienced users to interact with the robot [11]. Leutert et al. [12] reported using augmented spatial reality, a form of mixed reality, to relay information from the robot to the user’s workspace. They used a fixed-mobile projector. Socially aware interactive playgrounds [13] use various actuators to provide feedback to children. These actuators include projectors, speakers, and lights. These “interactive playgrounds can be placed at different locations, such as schools, streets, and gyms. Humans produce, interpret, and detect social signals (a communicative or informative signal conveyed directly or indirectly) [13].” Thus, their social signals can be used to enhance interactions with others. Various studies have been conducted on teaching humans to use robots in various environmental settings. RoboStage module implements learning among junior high school students through mixed reality systems [14–20]. Its creators compared the use of physical and virtual characters in a learning environment. RoboStage enables module interactions in robots to use voice and physical objects to achieve three stages of events: learning, situatedness, and blended. These events help students learn and practice activities, understand an environment, and execute an event. GENTORO uses a robot and a handheld projector to interact with children and perform a storytelling activity [21–27]. Its creators studied the effect of using a small handheld projector on the storytelling process. They also discussed the effects of using audio interactions instead of text and a wide-angle lens.

The agent matures into an adult by which the process in any state is affected by its previous state. The present study has focused on state-oriented and learning and development states to observe its outcome in association with the evolutionary state [28–30]. The proposed definition enhances research at the experimental design level using multisensory technologies to improve intelligence interaction and growth by applying the AI design principle [31–33]. Enhanced interaction between humans and robots improves learning, especially in the case of a child. Motion and speech sensor nodes are fused to this end. Contemporary children are familiar with handheld devices such as mobile phones, tablets, pads, and virtual reality cameras. Therefore, the toolkit developed in this study uses a mixed reality system featuring different

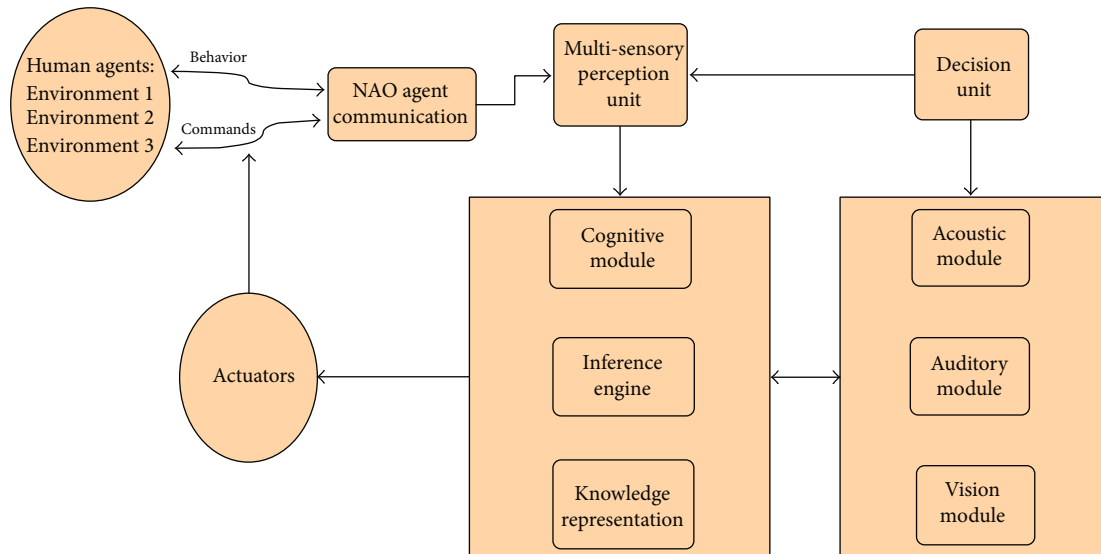


FIGURE 1: Nao agent multisensory perception communication concept design. Preliminary training datasets for Nao involved <https://github.com/IBM/watson-nao-robot>, Andreas Ess' webpage (ethz.ch), The NAL dataset (<http://inria.fr>), SpeechRecognition · PyPI, speech corpus [34], and Leap Motion Controller [35].

ways of interaction between a child and a robot agent. This study makes the following contributions.

- (i) Enhancing the humanoid robot Nao's learning capabilities with the objective to increase the robot's intelligence, using a multisensory perception of vision, hearing, speech, and gestures for HRI interactions
- (ii) Implementing Arabic Speech Agent for Nao using phonological knowledge and HMM to eventually activate child-robot communication [34]
- (iii) It developed a toolkit using Arabic speech recognition and the Haar algorithm for robust image recognition in a mixed reality system architecture using big data enabling Nao to achieve a 94% success learning rate featuring different environments, and for LMC, the highest accuracy of 98.50% using the Haar algorithm

The remainder of the study is organized mainly into Materials, Data, and Methods, which describe the architecture and experiment design, while the Discussion and Results section covers the intelligent big data management system using Haar algorithm-based Nao Agent Multisensory Communication in mixed reality and using LMC. Finally, the Conclusion and Future Work of the proposed study.

## 2. Materials, Data, and Methods

The experiment initiated at King Abdul-Aziz University with an Aldebaran representative was related to a three-year-old robot Nao, which could not speak Arabic or solve simple mathematics. The study analysis was initiated by selecting the suitable artificial intelligence principle design for the study. The experiment's goals and tasks were defined pre-

cisely to increase Nao's intelligence to at least seven years old. The Nao mathematics intelligence measurements were based on solving 100 children's exercises for basic addition, subtraction, and multiplication problems with human agents' help. Nao also reached the level of understanding simple sentences for Arabic language speech recognition. The experiment time scale was set for a total of two years. The study is aimed at involving the robot Nao in the learning-teaching process using interaction and multisensory Nao agent perceptions by exposing Nao to different environments (see Figure 1), enabling communication concept design. However, the present work focused more on the mixed reality environment.

The data collection involved two agents, the Nao robot and the user. For the user speech datasets (1) is collected, the recordings were for spoken Arabic numbers from 0 to 9. A total of five male and female speakers were asked to pronounce each number three times. All speakers were from Jeddah city, and the recordings were conducted in an ordinary quiet room. The speech data were captured at 16 KHz at an average speaking rate. For each person, the recording session lasted 60-90 minutes. The acoustic data were transmitted into a war sound recording file for later analysis. The Nao voice recognition consists of recording (ALProxy module ALAudioRecorder) and recognition modules tested by asking each person to repeat the number until the recognizer gets it correct. Python programming language is used. The environment Windows, Python IDLE (Python GUI), and NAOqi operating system using Choregraphe modules such as Almath and Python's Automatic Speech Recognition library were implemented. The computer is a static environment for data and processing and computational analysis.

For image collection, the hand dataset (2) is generated at KAU class, lab, home, and mixed reality; the author used The NAL dataset (<http://inria.fr>) for initial training and gathered the dataset by implementing the Leap Motion Controller

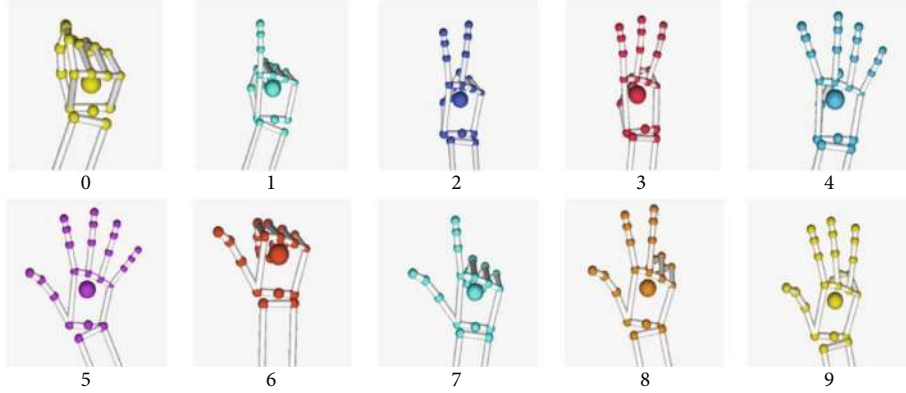


FIGURE 2: Visualization of gestures tracked by LMV [35].

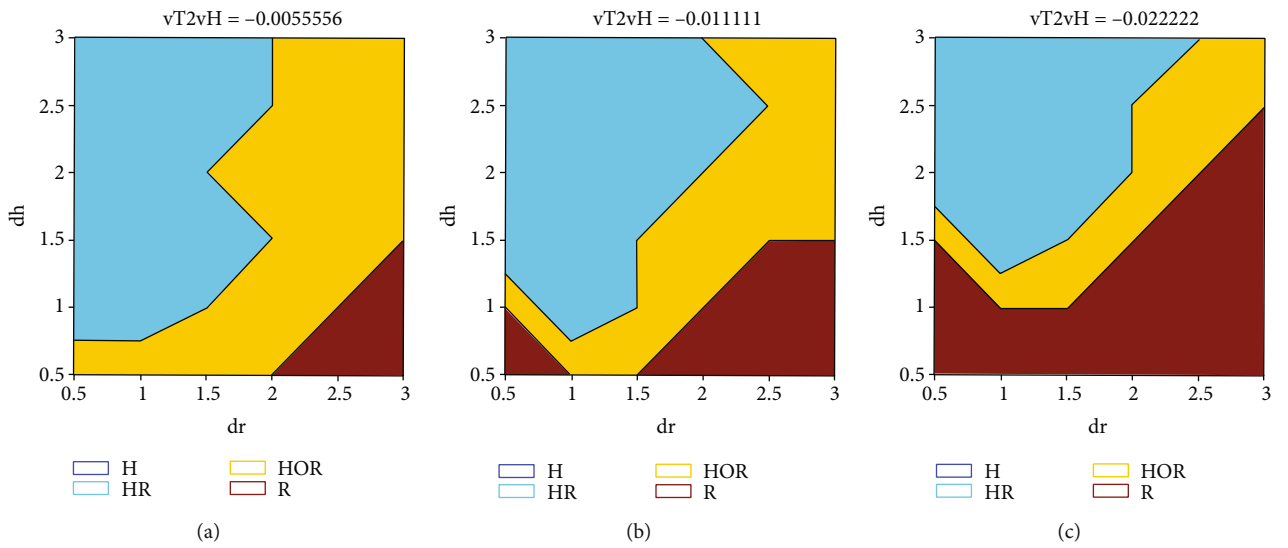


FIGURE 3: Time cost increases from graph (a) to graph (c) [36].

(LMC) Visualizer software for hand tracking for ten number gestures acquired from 34 subjects of three age groups (six and thirty) [35]. Figure 2 shows the gesture numbers in Leap Motion Visualizer (LMV) during hand tracking while LMV extracts the finger features based on path  $(x, y, z)$  and place  $(x, y, z)$  for thumb, index, middle, ring, and pinky.

NaoQi supports C++, Python, and JavaScript programming languages to be used on the robot. Several built-in modules include auditory, vision, and recognition. For example, the ALModule API has three methods to changeDatabase(), getParam(), and setParam() in C++. The Haar algorithm module was written in Python and applied to improve and stimulate Nao's vision. A snapshot model of a child showing fingers indicated a number. The Nao agent stores the number represented by a human agent in a database to improve its learning capabilities. Nao senses environment via sonar or pumper sensors and cameras; together, they support the Nao agent perceptions in addition to a trajectory algorithm. According to [36], "The human's sensitivities ( $d^h$ ) and the robot ( $d^r$ ) are ranged along  $X$  and  $Y$  axes human dominance with robot interaction reduces as the time cost increases from a graph 'a' ( $vT2vH = -0.0055$ ) to graph 'c'

( $vT2vH = -0.0222$ ), and consequently, the collaboration reduced." The following (Figure 2) represents the best performance level and shows that the decrease "from 92% in graph 'a' to 60% in graph 'c'" [36] in human dominance collaboration level as time cost and human response time increase, as indicated by the HR and HOR areas (see Figure 3).

In this work, an objective function is implemented as a collaborative model describing system performance within a specific environment; [19] used only four parameters (hits, false alarms, and missing target items) for a given process. To fit the experiments' objective function, the author added two more parameters to improve Nao's learning experience and robot-human interaction in different environments. There are two agents' interactions, a human and a robot, each could score a process, with a defined task, in four specified environments. Therefore, the author defined the objective functions (1) for both agents by six parameters, rather than four, to measure the system interaction performance  ${}^O SP$  as follows:

$${}^O SP = {}^O HAgInter + {}^O RobInter + {}^O HAgtask + {}^O RobAgtasks + {}^O RobEnv + {}^O Ts, \quad (1)$$

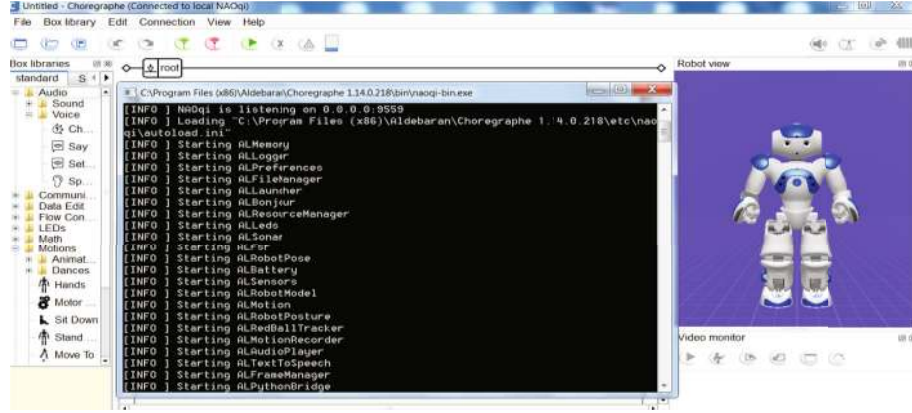


FIGURE 4: Nao agent initiation

where  $OHAgInter$  is the human information interaction by (0,1),  $ORobAgInter$  is the robot interaction by (0,1),  $^OHAgtask$  is the number of tasks by a human (1,2,n),  $^ORobAgtask$  is the number of tasks by a Nao robot (1,2,n),  $^ORobEnv$  is the robot environment (1 = class, 2 = lab, 3 = home, 4 = mixed reality), and  $^OT_s$  represents the time operational cost.

A loss function is a part of a cost function which is a type of objective function. In this work, the author calculated both objective and cost functions only. In function (2), the system operational time cost ( $^OT_s$ ) measures the activities' cost while detecting a true hit or a false miss for a target item. For example, the operation time for image processing by Nao to execute true hit in recognizing a number or counts false miss if wrong. The operational costs here are counted for two agents while they interact in different environments. It involves the cost of time spent within the operation, the decision time by the two agents (human/Nao robot) for identifying whether an item is a hit or not. The default cost value is chosen to be 4 seconds since two agents (Nao robot/human) are operating simultaneously. This work is calculated as follows:

$$O_{T_s} = t_{HS} \cdot C_t + [N \cdot P_{HMS} \cdot P_{H_s} + N \cdot (1 - P_S) \cdot P_{FHS}] \cdot O_c + t_{RS} \cdot C_t + [N \cdot P_{rMS} \cdot P_{r_s} + N \cdot (1 - P_S) \cdot P_{Frs}] \cdot O_c. \quad (2)$$

Following [36],  $t_{HS}$  and  $t_{RS}$  are the time needed to execute a task by a human and a Nao robot;  $O_c$  is the required operation cost for recognizing a solo item during the interactive mode,  $C_t$  represents the cost of a time unit for identifying an item,  $N$  represents the number of items for image detection, and  $P_{H_s}$  and  $P_{r_s}$  represent the probability of target item identified by a human and by Nao robot.  $P_{HMS}$  and  $P_{rMS}$  identify the system probability results for humans and a Nao robot, true or false.  $P_{FHS}$  and  $P_{Frs}$  are the false human and Nao robot probability for correctly rejected. The objective function was calculated for ten numbers of hand objects ( $N$ ). The three parameters  $N$ ,  $P_{r_s}$ , and  $P_{H_s}$  determine both Nao robot's and humans' environmental conditions with values set between 0.1 and 0.9 for each number. The system cost average value was estimated at 4.2 seconds for each

number. This high number is due to Nao multisensory and motor limitations.

The first experiment's main task was to enable Nao to interact with children and answer simple mathematical questions using hand gestures and speech. The children would interact with the physical Nao after activating its vision and speech modules to recognize the number of a human agent's fingers in a classroom environment. Figures 4 and 5 show the initiation of Nao agent speech and vision functions.

For the speech, Python code was first compiled using an Anaconda environment, and a file execution ("file01\_d.exe") was uploaded to the command NaoQi prompt. The voice module is tested by giving voice commands. Nao's response action is witnessed when getting orders from a human agent, and the command prompt indicates that Nao is ready. Next, a conversation and interaction command is loaded. For example, for the number recognition,

- (i) Use a simple pythonic OCR engine using opencv and numpy. <http://stackoverflow.com/questions/9413216/simple-digit-recognition-ocr-in-opencv-python>
- (ii) Run the program using Python example.py
- (iii) Press any key a few times for each processing

In the second experiment, the auditory module is activated, and a learning auditory guessing game was used. In this game, a child was asked to calculate the product of two numbers, and the robot reacted by making a clapping sound if the mathematics answer was correct or an alarm if it was not. Next, vision and auditory modules were activated and played interchangeably. This teaching game continued until the child learned from earlier errors, and the Nao recognition system continuously improved as it teaches more children and acquires more data. Thus, an agent would navigate and recall the learned action when an agent interacts with other human agents in the same lab environment [8]. A sample of an output file was generated (as shown in Figure 6).

To measure speaker independence, a kappa ratio measure was implemented to compare agreement of two classifiers having pairs of utterances, one recognized by human,

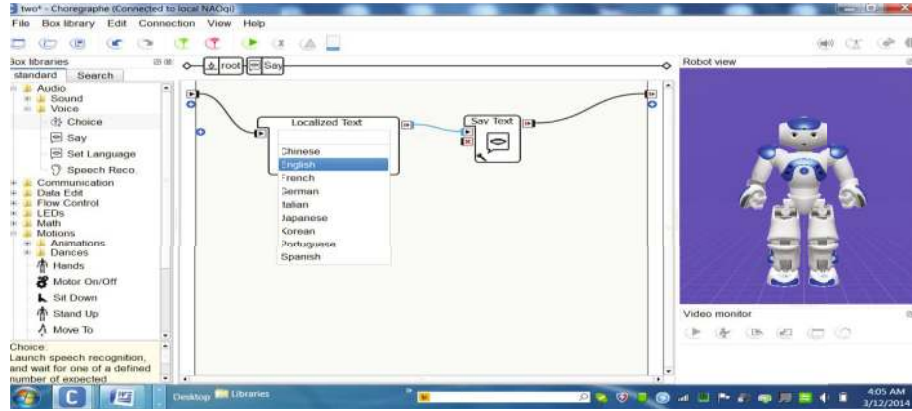


FIGURE 5: Nao agent speech initiation

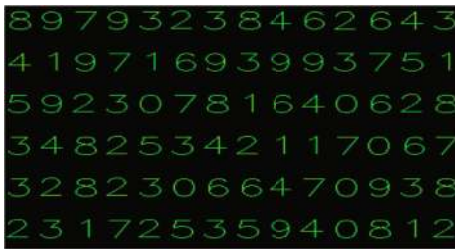


FIGURE 6: Results of accurately recognized numbers.



FIGURE 7: Nao participating in the sixth research event at KAU.

while the other by Nao agent, applying the same methodology of the linguistic speech agent of [34] by implementing an HMM syllabic recognizer having a corpus that exceeded 390,000 segment of total pairs. The probability of agreement is equal to 0.50833 measured by

$$\text{Kappa} = \frac{P(A) - P(E)}{(1 - P(E))}, \quad (3)$$

where  $P(A)$  is the actual recognizer and  $P(E)$  is the assumed random agreement. The rise of the learning curve for the HMM Nao syllabic recognizer showed a constant logistic growth of 0.030 per iteration with a kappa score of 32.5%, while the human agent scored 40.5% with 0.026 per iteration. For Nao agent, an HMM syllabic recognizer with a kappa

ratio 90.8% scored more than 93% of success rate when activating both auditory and vision modules for the agent Nao.

===== HMM Nao Results  
Analysis =====  
Ref: NaospeechAgent.mlf  
Rec: recoutsNaoSpeechAgent.mlf

-----  
Overall Results

-----  
SENT: %Correct = 95.37 [H = 103, S = 5, N = 108]  
WORD: %Corr = **94.30**, Acc = **93.37** [H = 104, D = 0, S = 4, I = 1, N = 108]

=====

In the third experiment, the author increased the complexity by activating different learning behaviors using several modules within different physical environments, namely, classroom, home, and lab, for 50 hours of training per environment. The module, called ALBehaviorManager, consists of 18 built-in methods. Among these are `getInstalledBehaviors()`, `preloadBehavior()`, `runBehavior()`, `addDefaultBehavior()`, `getUserBehaviorNames()`, and `isBehaviorPresent()`. In the classroom environment, Nao interacted with students and answered their questions using a guessing game. Then, Nao was taken to the second environment at the university laboratory to solve basic math problems. Third, Nao was taken home and to a MOE theater (see Figures 7 and 8) to observe the family members' behaviors by interacting with them. The author implemented the three-constituent principle to focus on Nao's desired interactive behaviors within its morphological capability and limitations. The "here and now" perspective explains Nao's actions, such as answering math questions when asked. The learning and development behavior of Nao is described while answering new questions from previous responses, and evolution is concerned with how Nao's learning behavior evolved by answering unlearned questions or how new behavioral mechanisms emerge. In the "here and now" perspective, the mechanisms and principles are concerned with how behavior comes or how individual behavior results from an agent's interaction with the environments, as explained by Hafner and Möller [37]. The three-time-scale "here and now" led to Nao's instant action corresponding to specific learning-teaching situations and unexpected behavior that enable an emergent

unprogrammed action. Thus, Nao exhibited a defined behavioral relationship with each specified location.

The sensors and controllers developed their perceptual cues and representation models for each environment, and Nao linked specific tasks to each environment using the available module. The author agrees with Pfeifer and Bongard [9] regarding the concept of dynamical systems and attractor states that result from interactions among the input channels (vision, hearing), the ecological niche of different environments (mixed reality, physical, and virtual spaces), and the growth of knowledge from different experiments in understanding (signs, words, and numbers). Once the learning has been completed from a longer time scale, Nao can recognize a new, unprogrammed number and behavior, which indicates the notion of emergence, although this is not the study's focus. For example, when Nao was placed in the home environment observing Muslim prayers, a desired unprogrammed behavior emerged to allow Nao to simulate the human body performing the Muslim prayer actions, such as bending and hand movements. This shows that agent Nao's behaviors can emerge from interactions with the environment. In the final stage of the experiment, the learning was based on mixed reality system environments. Performing teaching-learning tasks over an extended period in different environments caused Nao to become more intelligent and develop a new task in addition to the desired tasks. When Nao is placed in a mixed reality environment, the robot recognized additional gestures, such as body bending and hand movements. The mixed reality environment is used in various fields, such as earth science, engineering, and medicine. They introduce extensive interaction along with a virtual environment to research. According to Schouten et al. [8], any game experience must meet three criteria: context-awareness, adaptation, and personalization. These criteria lead to enhanced interaction from the participant (child). Physical agents in the form of a humanoid robot Nao, with enhanced interaction using a mixed reality system, are making the learning process more interesting and effective than interaction with virtual agent character. For a sample of the time to task interaction and learning between Nao and human agents in different environments, see Table 1. As the number of tasks increases, the Nao agent requires much time to perform the tasks, especially if the interaction level is maximized. The NaoQi Library Python software is activated by calling the built-in modules for the interaction and multiperception recognition time measurements. Thus, the NaoVi module has the fastest reactions (see Table 2).

The interaction between humans and robot agents has been investigated for many years. The use of hardware-based devices and computer vision-related techniques for interaction has also been studied [38]. The choice between hardware- and software-based solutions leads to a tradeoff between accuracy and ease of interaction. In the case of computer vision-based solutions, marker-based and marker-less techniques are used. The use of gloves in marker-based computer vision techniques provides high accuracy but uses a virtual environment. The Haar classifiers technique has been implemented as part of a marker-less computer vision approach. The Haar Algorithm is developed by Viola and



FIGURE 8: Nao performing at the MOE.

Jones [39–44] and involves two main steps (features extractions and objects detections) and is known as “The Haar-training,” the normalization threshold in this experiment was set between  $[1, -1]$ , and the system process can be briefly described in the following steps:

- (1) Identify positively detected images patterns
- (2) Choose negative images patterns
- (3) Select from the positive images the training dataset
- (4) Select from the negative and positive images the testing dataset
- (5) Employ Haartraining to the selected training dataset
- (6) Calculate the performance for the testing dataset

The standard Haar algorithm using OpenCv for figure image processing is summarized below:

Haar uses layers of classifiers; each is trained to detect an object in a specified environment within an image. For each layer, a window is created to match the image and evaluate the information accuracy. If no image matching occurs, the classifier window is said to be “negative.” The window matching for the object is reinitiated with another classifier. If the result is “true positive,” the matching scores as a successful classification for the positive image. If the result is “false positive,” the matching indicates misclassification for negative values as a positive image. If the result returns “false negative,” the matching indicates misclassification for positive values as a negative image. The training for the classifier continues until the best score is reached before overtraining is reached. The window passes through all classification layers, with a positive score indicating successful detection [45]. The work shows that the Haar classifier's main disadvantage is that it highly scores false positives in real-time when the number of Haar classifier layers reached 25 if the object keeps moving. The best accuracies of the Haar layers classifier is 23 with 50 training images (see Table 3).

In the fourth environmental experiment, a camera focused on the working space featuring the robot and the child. Information from the robot agent was provided to the child using its motions and projections on a screen. The camera was mounted above the area to provide a top-down view [46]. According to Sugimoto et al. [11], such a view controls the robot from a 2D perspective easier. The presence of

TABLE 1: Time to task activity Nao-human agents in the four environments.

$^{\circ}$ RobEnv	$^{\circ}$ HAgInter	$^{\circ}$ RobAgInter	$^{\circ}$ HAgtask	$^{\circ}$ RobAgtasks	$^{\circ}$ Ts
1 = class	0	1	1	1	3.3 s
2 = lab	1	0.5	2	2	5.0 s
3 = home	0.5	1	3	3	6.9 s
4 = mixed reality	1	1	4	4	8.6 s

the top-down camera limits the area of operation in the environment. The focus area was fixed, and human-robot interaction only occurred with this specified locality. This limitation did not impair the system because its main objective was to enable the robot to teach the child, who is in its vicinity in any case. The recognition of the child's gesture intended for the robot was extracted using various known techniques using LMC that affected the research for two reasons: the hand's presence in front of the face and the background's lamination. The author used recognition and handheld device based for robot-human interaction to provide haptic feedback from the child to the system [35]. The robot taught the child how to perform mathematical tasks by projecting prerecorded audio or video or gestures. The projection required dimming the lights, whereas the camera focused on the robot, and the child needed proper illumination. Therefore, appropriate lighting or dimming or a handheld projector was necessary [13, 45].

### 3. Results and Discussion

The proposed system consists of five components: a projector, camera, child, robot, and server (see Figure 9).

- (i) Camera: it focuses on the area of coverage. Its position and focus are fixed. The camera's output is regularly passed to the server that uses Haar classifiers to identify the child using face recognition. In the absence of the camera, the robot's eyes are used to input the video, and the face recognition algorithm of the humanoid robot is used to detect the child's presence
- (ii) Child: it is the main component of the system, in which learning is the sole objective. To make learning fun for the children, they are given various choices from which to choose, including interaction through a handheld device, speech, fingers of the hand, or LMC
- (iii) Robot: it is used to detect the child's finger movements and recognize speech from him/her. The robot can perform face and speech recognition using its default module or pass the acquired data to the server for processing. A communication server runs on the robot to interact with the handheld device of the child
- (iv) Server: it is used to control the flow of operations. Instructions from the server can be given either to the robot or the projector. The server can also perform face recognition using a Haar classifier or other

TABLE 2: Multiperception recognition times in the Nao robot.

Library	Hear	See	Speak	Move
NaoQi/actuators	0.857 s	0.231 s	0.252 s	0.752 s

means. It can also process the detection of the child's fingers using the convex hull approach and can perform speech recognition

- (v) Projector: it is used to display appropriate learning material for the child, which is done only when the child must be taught through already recorded videos. The process can be affected by light in the coverage area; therefore, a suitable projector must be used

The proposed toolkit description is shown in Figure 10. The process starts with Nao's eye as input using the camera for image processing while the audio is input via Nao's ear. Both inputs will end up with the learning management system. The fixed camera and/or Nao's eye are used to acquire the video of the child. The initial image processing is carried out to check whether the child is in front of the robot. The image acquisition and image processing are described in the flowchart (see Figure 11).

The steps involved in image processing are as follows.

- (i) Face identification: the human face in front of the robot is detected using the ALFaceDetection module in Nao. The detected face is written to the ALMemory periodically. Once the Nao robot detects the child's face, it welcomes the child and sends a request to the camera or itself to start acquiring the image
- (ii) Restrictive face recognition: the Haar classifiers regulate face recognition on the control server to enable the system to work with only specific children. The data acquired are passed to the control server to recognize already known faces
- (iii) Acquisition: vision is implemented through Nao's eyes in the form of an image, and a sequence of images periodically is considered equivalent to a video
- (iv) Obtaining image from Nao: the specifications of the image are entered using ALProxy with the ALVideoDevice module. The generic video module (GVM) is used to provide the necessary image format and specifications. The image is obtained using



Input: read image file  
Output: number of fingers detected in already detected hand  
Step 1: Change the image into grayscale for feature extraction using both cv2.cvtColor () and cv2.inRange ()functions  
Locate the hand(s) of the grayscale image and maps onto the colored image and resize using resizing using cv2.getPerspective functions  
Match features of the hand(s)/finger(s) segments with the rectangular box by  
Detecting pixel coordinates for hand(s) previously saved using cv2.threshold.function  
Detecting fingers pixel coordinates inside the hands' coordinates and draw the rectangles  
(1) All the number of fingers detected and compared in hands using function cv2.matchTemplate () and calculated using Haar learning algorithm  
If (p1, q1) and (p2, q2) are coordinates of the hand and (r1, s1) and (r2, s2) are coordinates of the number of fingers (fingers are present inside the hands)  
Then,  
    p1 < (r1 and r2 both) < p2  
    q1 < (s1 and s2 both) < q2  
Otherwise,  
: the number of fingers is Discarded (lie outside the hand).  
Step 2: Detection for the number of fingers.  
For the remaining fingers, checked existence, add to count for detecting another finger.  
If (there is a finger within the defined hand coordinates  
Then,  
Add one.  
Loop: Search for more fingers  
For each finger found, find the region of interest of hand which makes the  
Square, which represents the finger in the hand  
Condition: The number of fingers in one hand should not be greater than 5  
the rectangle is extended on both sides  
Step last: End.

ALGORITHM 1

TABLE 3: Best accuracies of the Haar classifier with 50 training images.

Detected (true positive)	False-positive static objects	False-positive moving objects	Negative (not detected)	False negative	Percentage
47	1	2	1	1	94%

- the getImageRemote method and is converted from a pixel image into a PIL image
- (v) Conversion: the obtained image is converted into a grayscale or HSV scale. The author used grayscale, where no morphological effects are observed. The HSV scale should be used in case of morphological effects
  - (vi) Children wearing gloves for LMC are processed using HS. Colored gloves are extracted from the image, thereby separating the fingers of interest from other acquired image/video parts
  - (vii) Morphological effects: certain morphological effects, such as erosion, dilation, and gradient, are applied to the image/video if the acquired image is different from the user's requirements. This process aims to improve the value of the acquired information
  - (viii) Blurring: a blur using a Gaussian filter is applied to the gray image to delete the image's Gaussian noise
  - (ix) Thresholding: the blurred image is then processed in a threshold. This process converts grayscale into a binary image based on the threshold value
  - (x) Finding contours: contours refer to the outline of the given image. The hierarchy or relationship between contours is obtained, and they are compressed to save space
  - (xi) Contour areas: the area of each contour, is obtained. The contour with the maximum area is identified and passed to the next stage
  - (xii) Convex hull and moments: the convex hull is used to find the approximate curve by considering convexity defects. The moments are used to find the center of the given contour. A red circle shown in Figure 12 is drawn at the center of the contour
  - (xiii) Polygonal curve: the Douglas-Peucker algorithm is used to draw a polygonal curve on the given image. The convex hull is again applied to the output of the polygonal curve

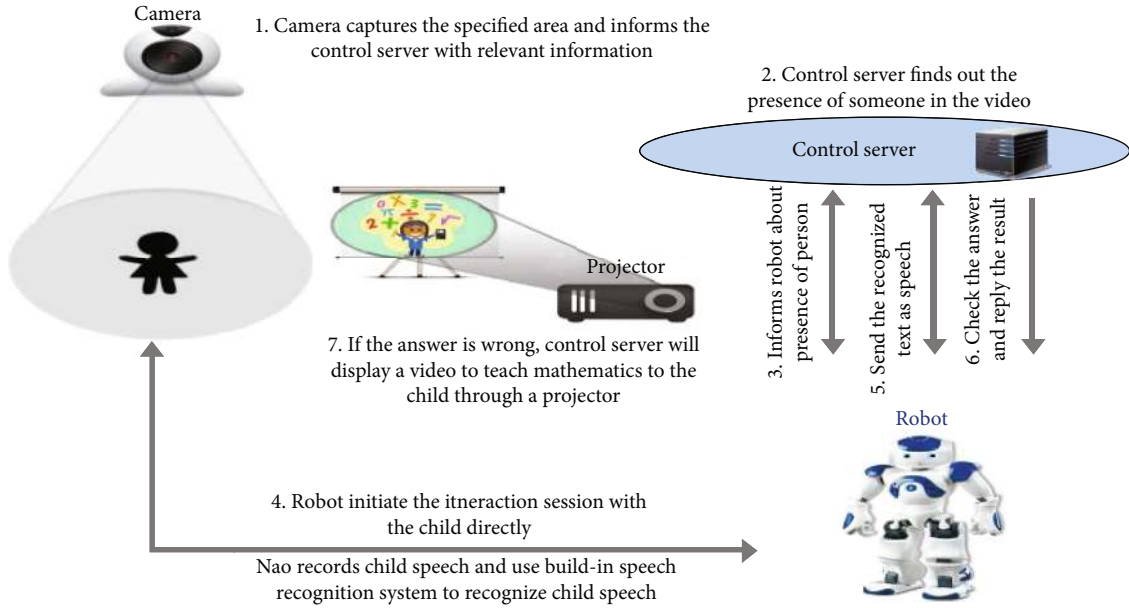


FIGURE 9: System architecture.

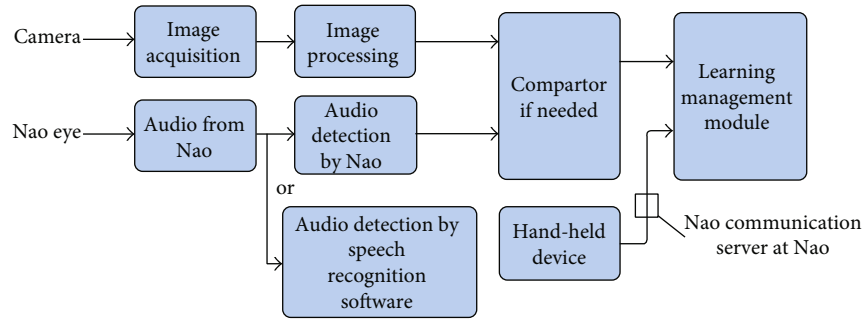


FIGURE 10: Toolkit description.

- (xiv) Convexity defect: any deviation from the convex hull is found with representations as starting and ending points. The appropriate lines and circles are drawn using the result of the convexity defec-tion operation
- (xv) Calculating the number of fingers: the points obtained from the convexity defect are used to find the angles between them to decide on the fingers' number

The fingers' number held up by the child must be identified and matched with the trained datasets, the number that the child is showing to the robot Nao due to a simple operation. Various factors hinder the identification of the fingers:

- (i) The effect of the cloth (see Figure 12(c))
- (ii) The effect of the position of the face (when it is away from the robot) and the fingers (see Figure 12(i))
- (iii) The consequence of wearing an ill-fitting glove by the child (see Figure 12(h) and see Figure 11(j))

- (iv) The use of colored gloves affecting identification by Nao (see Figure 12(j))

As shown in Figure 12, the system detected the fingers correctly, except for when hindered by the face or other objects. Thus, a glove with an appropriate color was used.

Initially, the sound is detected in Nao using its Sound-Detected module. The ALSpeechRecognition module recognizes the voice of the child in Nao. A specified vocabulary dataset (1) list containing one, two, three, ...,  $n$ , start, and the end is given to the ALS speech recognition module to recognize. Based on the recognizer's confidence level, its efficiency is identified, leading to its acceptance or rejection. As an alternative, speech recognition can be carried out on the server. Streaming audio from Nao is not possible at present; therefore, the stored audio received for a few seconds is transferred to the server for processing. While the server was processing the received audio, Nao continued to record further audio. The storage format used was WAV with four channels at 16kHz. These channels were used to acquire audio signals from Nao's four receivers, namely, left, right, front, and rear. File transfer was

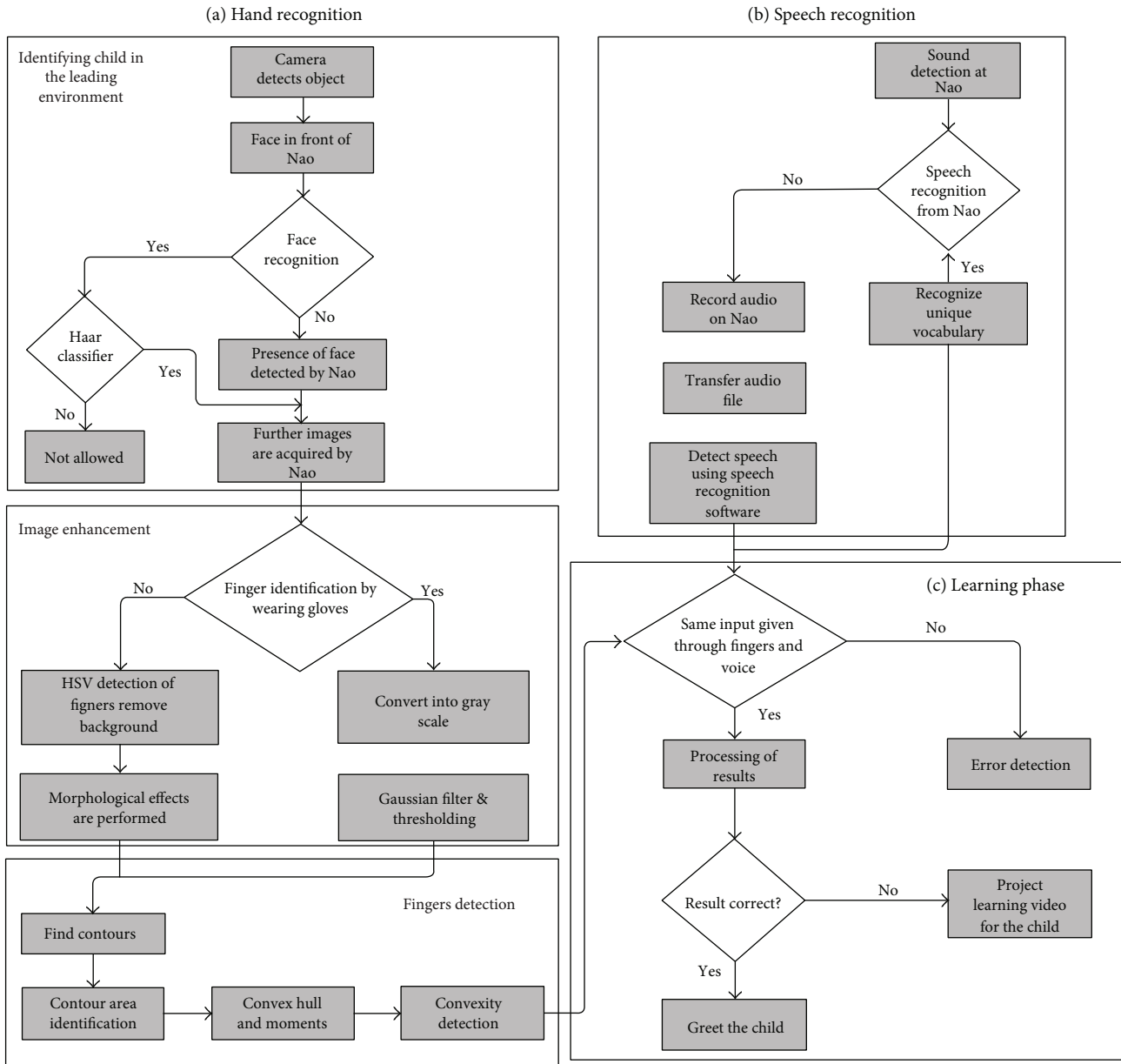


FIGURE 11: Flowchart of components of the toolkit.

performed using the ALFileManager module. Recognition at the server was carried out using speech recognition software, such as DragonFly. The fixed camera is used to follow the child’s location, while Nao initiates simple mathematics teaching for the child. The child’s learning is tested based on the output of the learning process. The outcome is displayed on the projected screen by the robot. Thus, both the robot and the projector screen respond to the child. The child also had a mobile device to interact with the robot, and its process of operations is as follows.

The camera recognizes the location of the child.

- (i) The child initiates communication with the robot using the mobile application available on his/her mobile device

- (ii) The robot teaches a lesson based on tracking audio, video, and movements. Given that Nao has only three fingers, using them to indicate numbers is not feasible
- (iii) Nao generates a question for the child to answer
- (iv) he response of the child is stored and tested using speech recognition HMM tools to check the answer
- (v) On the basis of speech recognition results, an appropriate response is displayed on the projector by the Nao

The system retains ambiguities, as shown in Figure 12, despite the image processing. Changes to the background and luminosity influence the quality of image processing,

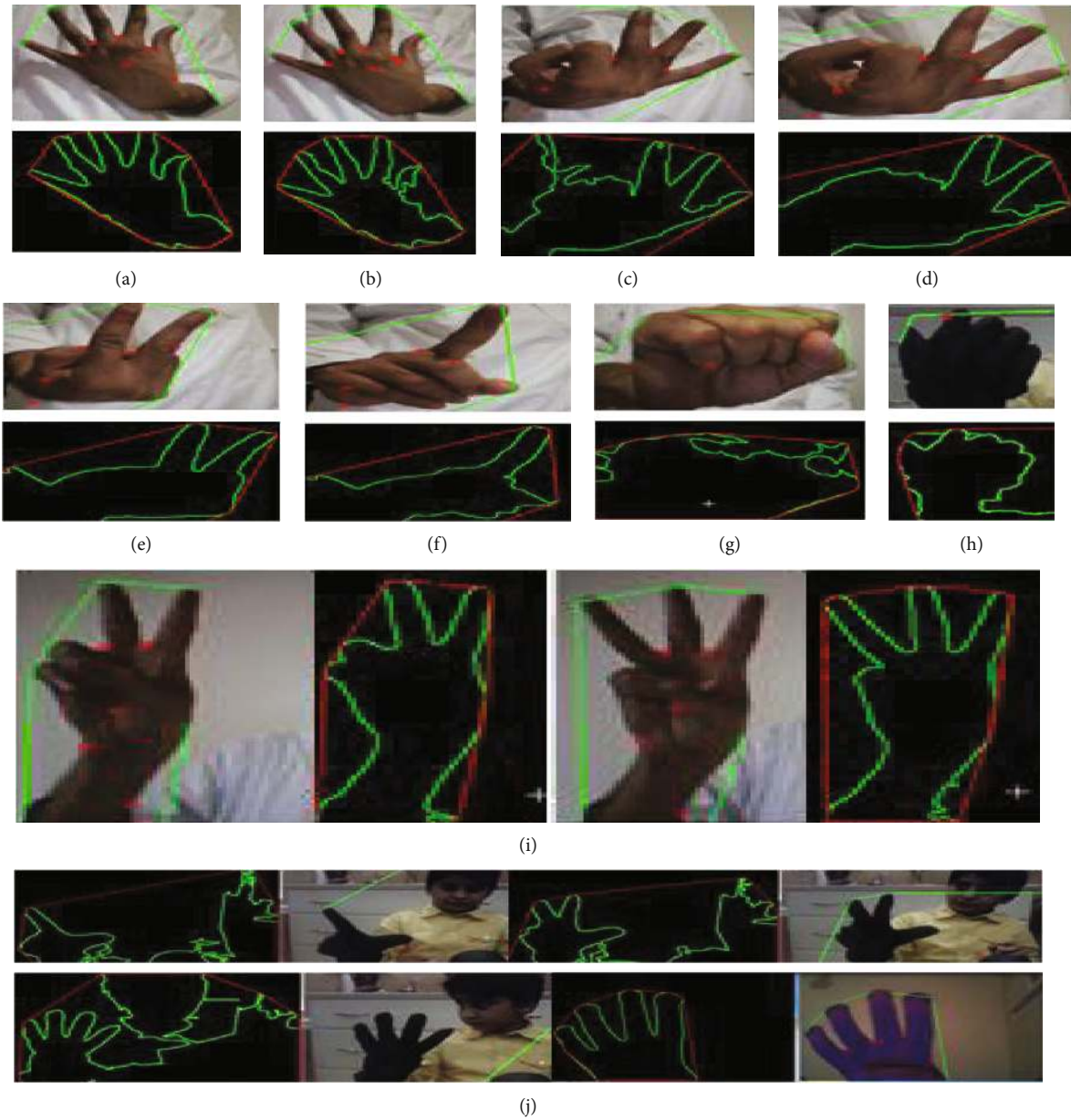


FIGURE 12: Recognition of number of open fingers.

thereby affecting finger recognition. Therefore, a handheld (Android) device-based application was used that communicates with the communication server in Nao. The communication server waits for a connection at its socket [47]. The android application’s user interface contains numbers from zero to nine and a few mathematical operations (see Figure 13). The request is passed to the communication server in Nao when the child clicks on the operands, operation, and result. The server checks the result. If it is correct, Nao greets the child with a clapping sound and hand movement action. If not, a video on the specific operation is projected for the child from the control server. Options are provided for the child or a parent to change the robot’s voice volume and speed.

The mixed reality system uses different entities to enhance learning. The performance of the system when using



FIGURE 13: Initial screen of the handheld application.

TABLE 4: Performance comparison.

Criteria	Measure
Communication distance between Nao and human	0.77 m to 1.12 m
Best communication distance between Nao and human	0.9 m

the Haar classifier and Nao's facial recognition algorithm was compared. Both correctly identified the face in front of the robot. Nao's facial recognition drawback was that it expected the face to be close and needed some time to recognize the face. According to documentation for the ALVisionRecognition-Nao Software 1.14.5 [48], the recognition process works between half and twice the distance used for learning purposes, achieving a 95% success rate, while the LMC using Haar classification scored 98.50%. The closeness of the human face to Nao restricted the child's free movements indicated in Table 4.

The use of different image processing techniques to recognize the number of fingers was significantly affected by the position of the face, the background, and the area's luminosity [49–53]. Therefore, a handheld device-based interaction system was provided for the child to interact with the robot. The device was connected via WiFi to the robot to reduce the restriction of a child's mobility. The projector in this mixed reality system also improved the child's learning capability due to excitement in dealing with Nao. The basic limitations of the study are

- (i) Children should be familiar with Nao using Arabic number language to communicate in gestures
- (ii) The fast movement speed affect LMC and recorded as errors
- (iii) Nao and the participants' distance should not exceed 0.09 m due to Nao's hardware audio-visual limitations
- (iv) The data sets size and quality could be improved for robust real-time recognition
- (v) Setting up the proper illumination level is necessary
- (vi) Streaming multisensory perception cues from Nao is not possible at present
- (vii) Nao has only three fingers, not five, to communicate the number visually with the child

#### 4. Conclusions and Future Work

In this work, the author developed a toolkit and evaluated the results in a mixed reality environment to enhanced learning by children and increased the robot Nao's intelligence level from a 3- to 7-year-old child. The teacher, in this case, is the robot Nao who interacted with the child through various means and environments. Four experiments were conducted to test interaction in four different environments (class, lab, home, and mixed reality using Leap Motion Controller).

The author showed that implementing an AI principle design, namely, the three-constituent principle, helped grow the robot's intelligence using different environments. The developed toolkit, using Arabic speech recognition and the Haar algorithm for robust image recognition in a mixed reality system architecture implementing big data, enabled Nao to gradually achieve a higher learning success rate ranged from 90.8%, 93%, 94%, to 98.50% as environment changes and multisensory perception increases. The highest learning level was achieved using LMC hand sign gestures with the Haar algorithm featuring a mixed reality environment. Activating a multisensory perception of vision, hearing, speech, and gestures for Human-Robot Interactions (HRI) in real-time increases children's learning math experience and makes it more enjoyable. The implementation of Arabic Speech Agent for Nao using phonological knowledge activated for HRI communication. The study shows that Nao's robot intelligence could be increased by learning similar to human intelligence and teaching simple mathematics to children. A cutting-edge research work direction for fostering Child-Robots education could be achieved using an active warehouse multidata system. Improving the Haar algorithm to operate in real-time with multihuman agent interaction and a single robot is one step toward future learning. An ultimate digital or physical robot teacher that operates in real time could be set as a goal for the years to come.

#### Data Availability

Preliminary training datasets for Nao involved: <https://github.com/IBM/watson-nao-robot>, Andreas Ess' webpage ([ethz.ch](http://ethz.ch)), The NAL dataset ([inria.fr](http://inria.fr)), SpeechRecognition · PyPI, Baothman speech corpus [49], Leap Motion Controller [19].

#### Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

#### Acknowledgments

The author thanks the Science and Technology Unit at King Abdulaziz University. This project was funded by the National Plan for Science, Technology, and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, the Kingdom of Saudi Arabia (award number 03-INF188-08).

#### References

- [1] D. Poole, "The independent choice logic for modelling multiple agents under uncertainty," *Artificial Intelligence*, vol. 94, no. 1-2, pp. 7–56, 1997.
- [2] R. Solomono, "The time scale of artificial intelligence: reflections on social effects," *Human Systems Management*, vol. 5, no. 2, pp. 149–153, 1985.
- [3] H. S. Nwana, "Software agents: an overview," *Knowledge Engineering Review*, vol. 11, no. 3, pp. 205–244, 1996.

- [4] W. Brenner, R. Zarnekow, and H. Wittig, "Application areas for intelligent software agents," in *Intelligent Software Agents*, Springer, Berlin, Heidelberg, 1998.
- [5] W. Shen and D. H. Norrie, "Agent-based systems for intelligent manufacturing: a state-of-the-art survey," *Knowledge and Information Systems*, vol. 1, no. 2, pp. 129–156, 1999.
- [6] J. Tweedale, N. Ichalkaranje, I. Agents, and T. Applications, "Intelligent agents and their applications," in *Knowledge-Based Intelligent Information and Engineering Systems. KES 2006. Lecture Notes in Computer Science*, vol. 4252 Springer, Berlin, Heidelberg.
- [7] V. Kumar, A. Dixit, R. G. Javalgi, and M. Dass, "Research framework, strategies, and applications of intelligent agent technologies (IATs) in marketing," *Journal of the Academy of Marketing Science*, vol. 44, no. 1, pp. 24–45, 2016.
- [8] B. A. Schouten, R. Tieben, A. van de Ven, and D. W. Schouten, "Human behavior analysis in ambient gaming and playful interaction," in *Computer Analysis of Human Behavior*, A. Salah and T. Gevers, Eds., pp. 387–403, Springer, London, 2011.
- [9] R. Pfeifer, F. Iida, and J. Bongard, "New robotics: design principles for intelligent systems," *Artificial Life*, vol. 11, no. 1-2, pp. 99–120, 2005.
- [10] S. Hashimoto, A. Ishida, M. Inami, and T. Igarashi, "TouchMe: an augmented reality-based remote robot manipulation," *the 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011, 2011*, 2011 [http://www.ic-at.org/ICAT2011\\_Proceedings/pdf/061-Hashimoto.pdf](http://www.ic-at.org/ICAT2011_Proceedings/pdf/061-Hashimoto.pdf).
- [11] M. Sugimoto, G. Kagotani, H. Nii, N. Shiroma, M. Inami, and F. Matsuno, "Time follower's vision," *ACM SIGGRAPH Emerging technologies*, p. , 200429, 2004, <http://dl.acm.org/citation.cfm?id=1186185>.
- [12] F. Leutert, C. Herrmannand, and K. Schilling, "A spatial augmented reality system for intuitive display of robotic data," *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, pp. , 2013179-180, 2013, <http://dl.acm.org/citation.cfm?id=2447626>.
- [13] A. Moreno, A. van Delden, R. Poppe, and D. Reidsma, "Socially aware interactive playgrounds," *IEEE Pervasive Computing*, vol. 12, no. 3, pp. 40–47, 2013.
- [14] I. F. Siddiqui, N. M. F. Qureshi, B. S. Chowdhry, and M. A. Uqaili, "Edge-node-aware adaptive data processing framework for smart grid," *Wireless Personal Communications*, vol. 106, no. 1, pp. 179–189, 2019.
- [15] N. M. F. Qureshi, I. F. Siddiqui, M. A. Unar et al., "An aggregate mapreduce data block placement strategy for wireless IoT edge nodes in smart grid," *Wireless Personal Communications*, vol. 106, no. 4, pp. 2225–2236, 2019.
- [16] J. Koo and N. M. F. Qureshi, "Fine-grained data processing framework for heterogeneous IoT devices in sub-aquatic edge computing environment," *Wireless Personal Communications*, vol. 116, no. 2, pp. 1407–1422, 2021.
- [17] I. F. Siddiqui, N. M. F. Qureshi, B. S. Chowdhry, and M. A. Uqaili, "Pseudo-cache-based IoT small files management framework in HDFS cluster," *Wireless Personal Communications*, vol. 113, no. 3, pp. 1495–1522, 2020.
- [18] A. K. Podder, A. Al Bukhari, S. Islam et al., "IoT based smart agrotech system for verification of urban farming parameters," *Microprocessors and Microsystems*, vol. 82, article 104025, 2021.
- [19] M. K. A. GHANI, M. A. Mohammed, M. S. Ibrahim, S. A. Mostafa, and D. A. IBRAHIM, "Implementing an efficient expert system for services center management by fuzzy logic controller," *Journal of Theoretical & Applied Information Technology*, vol. 95, no. 13, 2017.
- [20] M. A. Mohammed, B. Al-Khateeb, and D. A. Ibrahim, "Case based reasoning shell framework as decision support tool," *Indian Journal of Science and Technology*, vol. 9, no. 42, pp. 1–8, 2016.
- [21] S. A. Mostafa, A. Mustapha, S. S. Gunasekaran et al., "An agent architecture for autonomous UAV flight control in object classification and recognition missions," *Soft Computing*, 2021.
- [22] A. Lakhan, Q.-U.-A. Mastoi, M. Elhoseny, M. S. Memon, and M. A. Mohammed, "Deep neural network-based application partitioning and scheduling for hospitals and medical enterprises using IoT assisted mobile fog cloud," *Enterprise Information Systems*, pp. 1–23, 2021.
- [23] K. H. Abdulkareem, M. A. Mohammed, A. Salim et al., "Realizing an effective COVID-19 diagnosis system based on machine learning and IOT in smart hospital environment," *IEEE Internet of Things Journal*, 2021.
- [24] M. A. Mohammed, S. S. Gunasekaran, S. A. Mostafa, A. Mustafa, and M. K. A. Ghani, "Implementing an agent-based multi-natural language anti-spam model," in *2018 International Symposium on Agent, Multi-Agent Systems and Robotics (ISAMSR)*, pp. 1–5, Putrajaya, Malaysia, August 2018.
- [25] K. H. Abdulkareem, M. A. Mohammed, S. S. Gunasekaran et al., "A review of fog computing and machine learning: concepts, applications, challenges, and open issues," *IEEE Access*, vol. 7, pp. 153123–153140, 2019.
- [26] S. A. Mostafa, A. Mustapha, A. A. Hazeem, S. H. Khaleefah, and M. A. Mohammed, "An agent-based inference engine for efficient and reliable automated car failure diagnosis assistance," *IEEE Access*, vol. 6, pp. 8322–8331, 2018.
- [27] A. A. Mutlag, M. Khanapi Abd Ghani, M. A. Mohammed et al., "MAFC: multi-agent fog computing model for healthcare critical tasks management," *Sensors*, vol. 20, no. 7, article 1853, 2020.
- [28] V. Lahoura, H. Singh, A. Aggarwal et al., "Cloud computing-based framework for breast cancer diagnosis using extreme learning machine," *Diagnostics*, vol. 11, no. 2, p. 241, 2021.
- [29] S. N. Makhadmeh, M. A. Al-Betar, Z. A. A. Alyasseri et al., "Smart home battery for the multi-objective power scheduling problem in a smart home using grey wolf optimizer," *Electronics*, vol. 10, no. 4, p. 447, 2021.
- [30] S. A. Mostafa, S. S. Gunasekaran, A. Mustapha, M. A. Mohammed, and W. M. Abdullallah, "Modelling an adjustable autonomous multi-agent Internet of Things system for elderly smart home," in *Advances in Neuroergonomics and Cognitive Engineering. AHFE 2019. Advances in Intelligent Systems and Computing*, vol. 953, H. Ayaz, Ed., Springer, Cham, 2020.
- [31] N. Bostrom, "Are we living in a computer simulation?," *The Philosophical Quarterly*, vol. 53, no. 211, pp. 243–255, 2003.
- [32] G. Fisher, "User modeling in human-computer interaction," *User Modeling and User-Adapted Interaction*, vol. 11, no. 1/2, pp. 65–86, 2001.
- [33] R. Pfeifer and J. Bongard, *How the Body Shapes the Way We Think—A New View of Intelligence: A Bradford Book*, The MIT Press, Cambridge, Massachusetts, 2006.
- [34] F. Baotrhan, *Phonology Based Automatic Speech Recognition for Arabic*, PhD thesis, The University of Huddersfield, 2003.

- [35] A. A. Almarzuqi, *Gesture-Based Assistive Robotics Children Education through Enhanced Interaction*, MSc. Thesis, King Abdulaziz University, 2017.
- [36] D. Yashpe, *Influence of Human Reaction Time in Human-Robot Collaborative Target Recognition Systems*, M.Sc thesis, University of the Negev, 2009.
- [37] V. V. Hafner and R. Möller, "Learning of visual navigation strategies," *Proceedings of the European Workshop on Learning Robots*, vol. 1, pp. 47–56, 2001.
- [38] A. B. Raij, K. Johnsenand, and D. S. Lind, "Comparing interpersonal interactions with a virtual human to those with a real human," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 3, pp. 443–457, 2007.
- [39] A. Kumar, M. S. Hashmi, A. Q. Ansari, and S. Arzykulov, *Haar Algorithm for The Analysis of Fractional Order Calculus Based Computation Problems Associated With Electromagnetic Waves*, Research Square Co.,Inc, 2021.
- [40] P. Taunk, G. Jayasri, J. P. Priya, and N. S. Kumar, "Face detection using Viola Jones with Haar cascade," *Test Engineering and Management*, vol. 83, 2020.
- [41] Y. Feng, Q. Jia, and W. Wei, "A control architecture of robot-assisted intervention for children with autism spectrum disorders," *Journal of Robotics*, vol. 2018, Article ID 3246708, 12 pages, 2018.
- [42] J. Huang, Y. Shang, and H. Chen, "Improved Viola-Jones face detection algorithm based on HoloLens," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, 2019.
- [43] F. S. Daniş, T. Meriçli, Ç. Meriçli, and H. L. Akin, "Robot detection with a cascade of boosted classifiers based on Haar-like features," in *RoboCup 2010: Robot Soccer World Cup XIV. RoboCup 2010. Lecture Notes in Computer Science*, vol. 6556, J. Ruiz-del-Solar, E. Chown, and P. G. Plöger, Eds., pp. 409–417, Springer, Berlin, Heidelberg, 2011.
- [44] R. Ogla, A. Ogla, A. Abdul Hussien, and M. Mahmood, "Face detection by using OpenCV's Viola-Jones Algorithm based on coding eyes," *Iraqi Journal of Science*, vol. 58, pp. 735–745, 2017.
- [45] *Microvision Display Technology Solutions*. n.d.August 2014, <http://www.microvision.com/solutions/>.
- [46] C. Guo, J. E. Young, and E. Sharlin, "Touch and toys: new techniques for interaction with a remote group of robots," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*, pp. 491–500, New York, NY, USA, 2009.
- [47] *Nao Communication Server*. n.d. [northernstars-wiki.wikidot.com/projects:naocom](http://northernstars-wiki.wikidot.com/projects:naocom)<http://northernstars-wiki.wikidot.com/projects:naocom>.
- [48] *AL Vision Recognition-Nao Software 1.14.5 documentation*. n.d.March 2015, <http://doc.aldebaran.com/1-14/naoqi/vision/alvisionrecognition.html>.
- [49] M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa et al., "Voice pathology detection and classification using convolutional neural network model," *Applied Sciences*, vol. 10, no. 11, article 3723, 2020.
- [50] S. A. Kashinath, S. A. Mostafa, A. Mustapha et al., "Review of data fusion methods for real-time and multi-sensor traffic flow analysis," *IEEE Access*, vol. 9, pp. 51258–51276, 2021.
- [51] X. Zhou, Y. Ma, Q. Zhang, M. A. Mohammed, and R. Damaševičius, "A reversible watermarking system for medical color images: balancing capacity, imperceptibility, and robustness," *Electronics*, vol. 10, no. 9, article 1024, 2021.
- [52] O. A. Mahdi, Y. R. B. Al-Mayouf, A. B. Ghazi, A. W. A. Wahab, and M. Y. I. B. Idris, "An energy-aware and load-balancing routing scheme for wireless sensor networks," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 3, pp. 1312–1319, 2018.
- [53] M. S. P. Subathra, M. A. Mohammed, M. S. Maashi, B. Garcia-Zapirain, N. J. Sairamy, and S. T. George, "Detection of focal and non-focal electroencephalogram signals using fast Walsh-Hadamard transform and artificial neural network," *Sensors*, vol. 20, no. 17, article 4952, 2020.