Mobile Robot: Automatic Speech Recognition Application for Automation and STEM Education

Duy Thanh Tran
Vietnam National University Ho Chi Minh City
Dang Huy Truong
Vietnam National University Ho Chi Minh City
Hoanh Su Le
Vietnam National University Ho Chi Minh City
Jun-Ho Huh (72networks@kmou.ac.kr)
Korea Maritime and Ocean University https://orcid.org/0000-0001-6735-6456

Research Article

Keywords: Mobile Robot, Robot Software, ASR; Android, AIDL, Automation, AI, Emotion Care, Affective Computing, STEM.

Posted Date: September 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1287153/v1

Mobile Robot: Automatic Speech Recognition Application for Automation and STEM Education

Duy Thanh Tran[1,2], Dang Huy Truong[3], Hoanh Su Le[4,*], Jun-Ho Huh[5,]**

Abstract: Nowadays, Robots are widely applied in life as well as industrial production, medicine, rescue, learning and entertainment. There are many types of robots using different modern technologies that are divided into groups such as Robot Manipulator, Movement Robot, Biomimetic robot, each type is applied according to its own role. Robots are also applied in training, especially in STEM training. In this paper, we propose technical models to design Mobile Robot combined with Android OS Tablet with user interaction screen, voice control model and AIDL IPC interactive model for remote control. Mobile Robot's motion as well as giving a layered UML Model for motion handling towards open source. We research the Mobile Robot using RockChip AI Processor RK3399Pro, which focuses on understanding the structure and operating mechanism of the AI Processor that RockChip Robot uses, along with researching Android with IPC AIDL architecture and with Automatic Voice Recognition technology to control Robot to turn left, turn right, move forward, backward, jump or stop motion by voice with three languages: Korean, English and Vietnamese. From there, it will serve as a foundation for the research and development of a system to support students in learning by interacting with robots through voice in the form of gamification of the curriculum, especially for extensive research on training programs, created in the STEM field.

Keywords: Mobile Robot, Robot Software; ASR; Android; AIDL; Automation; AI; Emotion Care; Affective Computing; STEM.

[1] Ph.D. Candidate of Department of Data Informatics, (National) Korea Maritime and Ocean University, Busan 49112, Republic of Korea (South Korea); First Author Email: thanhtd@uel.edu.vn
[2] Full Lecturer (Assistant Professor) of Faculty of Information Systems, University of Economics and Law, Vietnam National University HCMC, Vietnam.
[3] GBS AI Lab, Vietnam.
Author e-mail: huytd17411@st.uel.edu.vn
[4] Dean (Tenure) of University of Economics and Law (UEL), Vietnam National University HCMC, Vietnam
*Co Corresponding Author: Hoanh Su Le e-mail: sulh@uel.edu.vn
[5] Associate Professor (Tenure) of Dept. of Data Science, National Korea Maritime and Ocean University, Republic of Korea.
**Corresponding Author Jun-Ho Huh e-mail: 72networks@kmou.ac.kr or 72networks@pukyong.ac.kr

1. INTRODUCTION

In the 4.0 technology era with the strong development of advanced technologies, including the development of Robots in various fields such as fire fighting, medical care, education, space exploration, ocean exploration. The appearance of Robot has effectively supported human production activities. Not only that, they also show their role in many other fields. Since its appearance, the robot has changed to better promote its role.

Meanwhile, manipulator is a type of robot [1] with multiple joints in series, each of which can be reciprocating or rotating, usually driven by a servo motor. Manipulators with clamps or handles with many degrees of freedom are suitable for applications such as painting, welding, workpiece picking, etc. Due to the development of technology and practical needs, today's manipulators have many applications new uses such as surgical support in medicine, assistance for the disabled, or in industry.

A movement robot [2] is a robotic system capable of performing tasks in many different locations with the ability to move by wheel, chain or foot depending on the terrain. For robots that move underwater or in the air, we need a propeller, propeller or jet engine to create motion for the robot. Robotic movement has many applications and requires solving many new problems. One of the common problems to be studied in movement robots is the ability to determine the direction of the robot.

In addition to movement robots running on wheels and chains, researchers are currently trying to apply the transport mechanisms of living organisms to robots to create new types of biomimetic robots with the ability to move automatically [3].

Around the world, many corporations in different countries have also developed many robots in the field of education, especially in the field of STEM teaching [4], [5], with a wide variety of categories. In the history of robots and their technology, the most important countries are the United States and Japan. The United States is still leading in the level of comprehensiveness of robot technology, while Japan has the highest number and diversity of robots in the world. However, the problem of combining the Android [6] operating system to let users interact directly as well as allowing programming support right on the Robot is still in the beginning stage for many reasons that may be related to copyright.

There are many types of educational robotics for STEM is reviewed in the paper [7] such as Do It Yourself (DIY) robots, Open hardware robots, Brick- based robots, pre-assembled robots, only for simple actions or specific purpose robots, Humanoid robots, Robots-based on tangible programming. However, the integration of the Android screen that allows users to interact directly is still limited. In the paper [8], the authors presented Android OS Mobile Technologies in the Robotics, the project for about Educational and STEM-Enhancing.

In this paper, we continue research the Mobile Robot model that combines interactive screens with Android operating system and integrate IPC AIDL architecture, Automatic Speech Recognition Application, Single TensorFlow to control movement Robot and sup-port education, entertainment and open-source application development. RK3399Pro AI Chip is a high performance, power processor will be used to design the Mobile Robot. Creating excitement in learning for learners, especially under-university students, they can manipulate directly on the Robot or separate the robot's face to learn depending on their needs. This model can add to the educational development ecosystem for STEM and it is possible to research and deploy application models for robotic business process automation and robotics application.

The United States and Republic of Korea (South Korea) are major powers in robotics [9], Vietnam is a developing country, and nowadays, robots are applied in production, business, education and entertainment a lot. Vietnam has had trade cooperation with the US and Korea, and the Korean Studies industry has developed very well in Vietnam. Therefore, the research team wants to focus on testing with three different languages including American and Korean to increase business and educational cooperation opportunities in Vietnam with the Mobile Robot products.

The paper structure includes the following sections: Introduction, approach, related studies, model design, system development, experiments and results, conclusions and development directions.

2. RESEARCH METHODOLOGY

In this article, we research the structure and function of the components as well as how to use and interact with the RockChip RK3399Pro AI Processor [10]. Research Artificial Intelligence Technology in Google cloud speech [11] in multilingual voice diagnostics and authentication to integrate into software to convert voice into control commands. We propose to interact with the Mobile Robot via Android Tablet using RockChip AI by integrate IPC AIDL architecture [12] into the system to receive interactive commands after voice conversion into the corresponding integer values to control Robot to move left, right, for-ward, backward, jump or stop moving the voice. In addition, we research and integrate more speech processing model using Sequential API in TensorFlow - Keras and then Freeze and Convert trained model into a Single TensorFlow .pb file to join with tensor-flow-android to recognize English speech even without an internet connection.

Finally, we design a software on Android platform and install on Mobile Robot to interact, this software includes features such as allowing three languages to be switched between English, Korean and Vietnamese. Corresponding to each language in which the user will use the selected language to talk and give commands to Mobile Robot, the control commands in this research are mentioned such as turn left, turn right, forward, backward, jump and stop moving. Users can also give voice commands as well as give commands by interacting directly on the face of Mobile Robot. In addition, we build the central software as an ecosystem in STEM training, focusing on math, foreign languages and Blockly drag-and-drop programming.

3. RELATED WORKS

3.1. RockChip AI Processor RK3399Pro

Fuzhou RockChip Electronics Co., Ltd. at CES2018. RockChip released its first AI (artificial intelligence) processor RK3399Pro by adopting CPU + GPU [13] + NPU [14] hard-ware structure. The Chip integrates dual-core ARM Cortex-A72 MPCore processor and Quad-core ARM Cortex-A53 MPCore processor, both are high-performance, low-power and cached application processors. It has two CPU clusters, big cluster with dual-core Cortex-A72 is optimized for high performance and little cluster with quad-core Cor-tex-A53 is optimized for low power. Full implementation of the ARM architecture v8-A instruction set, ARM Neon Advanced SIMD (Single Instruction Multiple Data) support for accelerating media and signal processing. Based on Big.Little architecture [15], [16]. Equipped with one powerful Neural Network Process Unit (NPU), it supports mainstream platforms in the market, such as caffe, tensor flow, and so on. Fig. 1 shows RK3399Pro Develop Tool for AI.

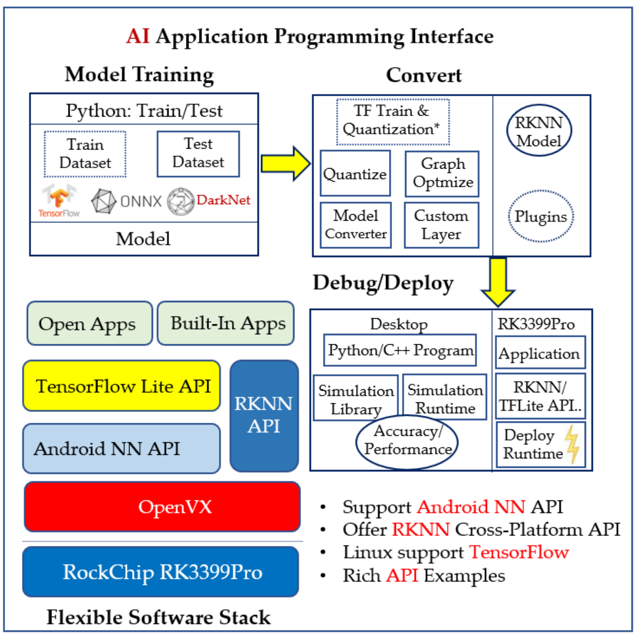

Fig. 1. RK3399Pro Develop Tool for AI.

RockChip RK3399Pro provides AI-related APIs for programmers to use such as RKNN API, TensorFlow Lite API, Android NN API.

Three important features of RockChip RK3399Pro AI solution, including High performance AI hardware, Superior platform compatibility and Easily development of turnkey solution.

High performance AI hardware: RK3399Pro adopted exclusive AI hardware design. Its NPU computing performance reaches 2.4 TOPs, and indexes of both high performance and low consumption keep ahead: the performance is 150% higher than other same type NPU processor; the power consumption is less than 1%, comparing with other solutions adopting GPU as AI computing unit.

Superior platform compatibility: RK3399Pro NPU supports 8 bit and 16 bit and is compatible with various AI software frameworks. Existing AI interfaces support OpenVX and TensorFlow Lite/AndroidNN API; AI software tools support the importing, mapping and optimizing of Caffe/TensorFlow model.

Easily development of turnkey solution: RockChip provides one-stop AI solution based on RK3399Pro, including hardware reference design and SDK. The solution can in-crease the AI products R&D speed of global developers and greatly reduce product launch time. It can significantly improve the speed of AI product development for global developers and greatly shorten time to market.

3.2. Automatic Speech Recognition (ASR)

Speech recognition, it is also known as automatic speech recognition (ASR) that develops methodologies and technologies that enable the recognition and translation of spoken language into text by computers. The speech recognition system is developed with major components that include acoustic front-end, acoustic model, language model, lexicon and decoder [17], [18], [19] as shown in the Fig. 2.

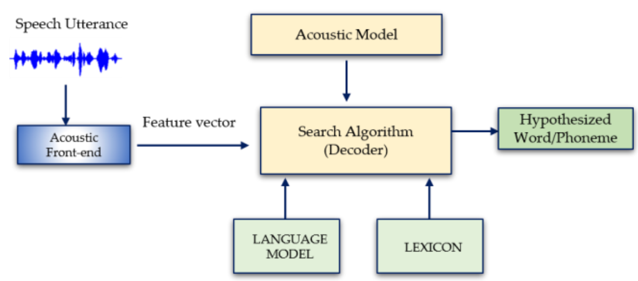

Fig. 2. Architecture of Speech Recognition System.

Acoustic front-end will convert the speech signal into appropriate features which provides useful information for recognition. The input audio is converted into a sequence of fixed-size acoustic feature vectors. The parameters of word /phone models are estimated from the acoustic vectors of training data. The decoder operates by searching through all possible word sequences to find the sequence of words that is most likely to generate. The function of the automatic speech recognition system can be described as the addition of several speech parameters from the audio speech signal to each word or sub-word unit. Speech parameters describe the word or particle according to their change over time, and together they form a pattern that characterizes the word or particle. During the training phase, the program will read all the words in the current application's vocabulary. Word patterns are stored and later when a word needs to be recognized it is compared with the stored samples and the word that gives the best match is selected.

The purpose of the ASR system [20], [21] is to obtain the most likely word order given by the speaker audio signal.

Fig. 3 shows google cloud speech to text platform. Google Cloud Speech-to-Text accurately convert speech into text using an API powered by Google's AI technologies. There are many functions are provided by this AI technology: Speech-to-Text client libraries to get started with Speech-to-Text in your language of choice, Cloud Speech REST API with v1 REST API Reference (Non-streaming JSON), Cloud Speech RPC API with v1 gRPC API Reference (Streaming and non-streaming Proto 3), Language support with the list of languages supported by Speech-to-Text such as Korean, Vietnamese, English language, Supported class tokens with the list of class tokens supported for speech adaptation by language, and Cloud Speech-to-Text On-Prem API with the Cloud Speech-to-Text On-Prem solution.

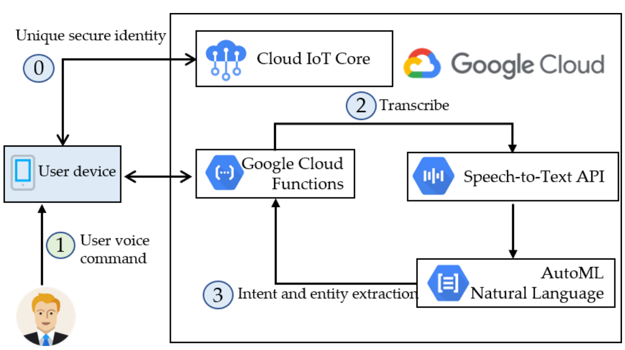

Fig. 3. Google Cloud Speech-to-Text Platform.

Google Cloud Platform is researched by many scientists to deploy software [22]. In this paper, the Speech-to-Text model in Google Cloud Platform will be applied to the system, use three languages Korean, Vietnamese and English to control the Robot.

3.3. Android Interface Definition Language (AIDL) and IPC

The Android Interface Definition Language (AIDL) is similar to other IDLs [23],[24] you might have worked with. It allows you to define the programming interface that both the client and service agree upon in order to communicate with each other using Interprocess communication (IPC) [25] [26]. On Android, one process cannot normally access the memory of another process, they need to decompose their objects into primitives that the operating system can understand, and marshal the objects across that boundary for you. The code to do that marshalling is tedious to write, so Android handles it for you with AIDL. Using IPC AIDL allow clients from different applications to access your service with handling the multithreading in the services. Fig. 4 shows IPC AIDL architecture model.

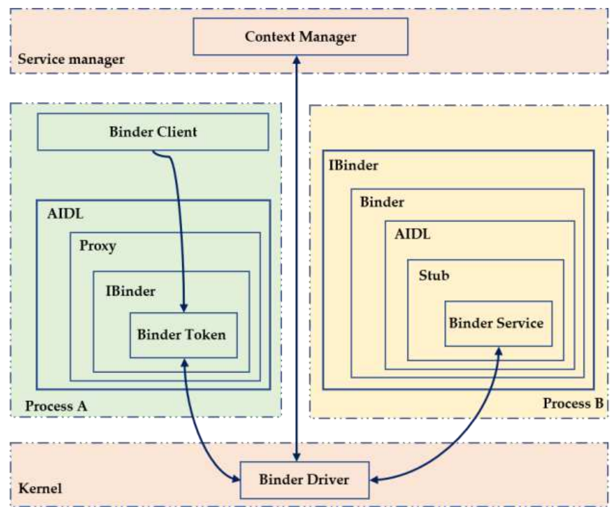

Fig. 4. IPC AIDL architecture model.

Before begin designing AIDL interface, be aware that calls to an AIDL interface are direct function calls. AIDL interface must be defined in an .aidl file using the Java programming language syntax, these are steps: (1) Create the .aidl file, this file defines the programming interface with method signatures. (2) Implement the interface, the Android SDK tools generate an interface in the Java programming language, based on the .aidl file. This interface has an inner abstract class named Stub that extends Binder and implements methods from the AIDL interface. It must be extended the Stub class and implement the methods. (3) Expose the interface to clients, Implement a Service and override onBind() to return implementation of the Stub class.

3.4. STEM education on Robotics

STEM is an idea-based curriculum that equips learners with knowledge and skills related to science, technology, engineering, and math – in an interdisciplinary and human-centered approach. Learning can be applied to solve problems in everyday life [27]. And now the trend of research and publication of STEM education technology is growing, the article [28] has listed 7 topic categories related to STEM publication. (1) K-12 teaching, teacher, and teacher education in STEM; (2) post-secondary teacher and teaching in STEM; (3)K-12 STEM learner, learning, and learning environment; (4) Post-secondary STEM learner, learning, and learning environments; (5) Policy, curriculum, evaluation, and assessment in STEM; (6) Culture and social and gender issues in STEM education; (7) His-tory, epistemology, and perspectives about STEM and STEM education. It is shown in the Fig. 5.

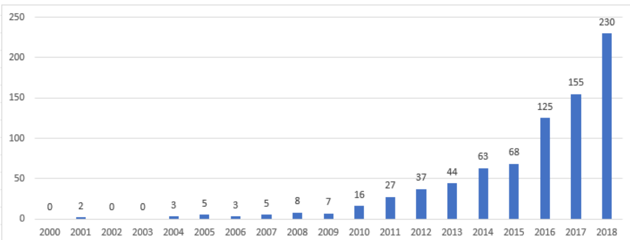

Fig. 5. The distribution of STEM education publications over the years (source [28])

In recent years, the trend of applying Robotics in the implementation of education to improve STEM education has been increasingly focused [29]. In developed countries like Vietnam, many STEM education associations have been built and gradually introduced into mass education in the country, in which the application of high technologies such as Robots in teaching is increasingly interested. Mobile Robot application with interactive interface will make it easier for learners to access, and at the same time can build software running on Android devices, which is already popular in the world and in Vietnam and other countries. In developing countries, this will promote the demand for applied STEM learning, making students more excited and creative in the learning process.

4. PROPOSAL AND IMPLEMENTATION OF RESEARCH MODEL

4.1. Computer and Tablet Environment

In this paper we experiment on Windows operating system (64 bit, 16GB RAM, Intel(R) Core (TM) i7-9700 CPU @ 3.00GHz), use .net frame work, Java language, Android Studio 4.2.1 tool, Android OS, IPC AIDL architecture to deploy software. Android Tablet with Rock-Chip AI with Android OS version 9.0, RAM 4GB.

4.2. Proposal and implementation of Research Model

In this research, we propose two models, the first model is an interactive master model that controls Mobile Robot using RockChip AI by Automatic Speech Recognition, and the second model is a detailed model that converts voice into code combined with IPC AIDL architecture to control the robot to turn left, turn right, move forward, move backward, dancing, or stop moving. Fig. 6 shows General Proposed Research of Model for a voice control processing model that combines Tensorflow Keras API and Google Cloud AI.

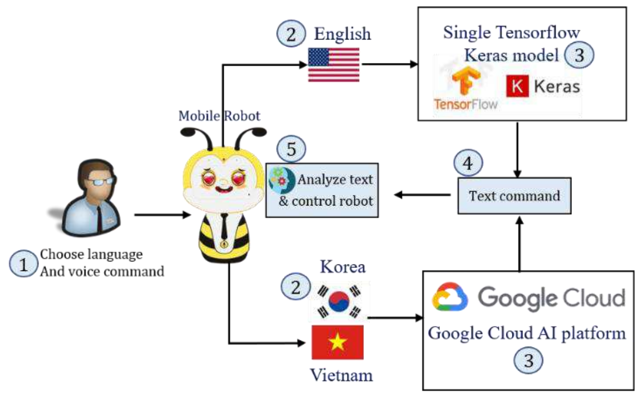

Fig. 6. General Proposed Research of Model for controlling Mobile Robot by voice

There are five main steps in the general model. In the **Step 1**, user selects the interactive language, which can be English, Korean or Vietnamese. And Interactive language recognition program will be detected in the **step 2**. After detect language, in the **Step 3** the program will apply models for speech processing corresponding to the selected language. If English is selected (including the environment with internet or without internet), the system will apply the Tensorflow Keras API model. If Korean and Vietnamese language are selected (The program is only applicable in the model with internet connection), the program will apply the Google Cloud AI platform model. The **step 3** will be details explained after general model. And in the **Step 4**, for each model applied in speech recognition, the program will receive text commands, the text examples are shown in the Table 1. And finally, in the **Step 5**, the program will receive the Text command from step 4 and conduct analysis, convert the text to integer number to call the commands in IPC AIDL architecture to control the movement of mobile Robot.

Table 1. Voice command for Mobile Robot movement

English Language	Korean Language	Vietnamese Language	Action of Mobile Robot
Left	왼쪽	Rẽ Trái	Turn left
Right	오른쪽	Rẽ phải	Turn Right
Up	앞으로	Tiến lên	Move forward
Down	뒤로	Lùi lại	Move backward
Stop	중지	Dừng lại	Stop movement

Fig. 7 shows the detailed architecture for IPC AIDL interaction model for Mobile Robot movement.

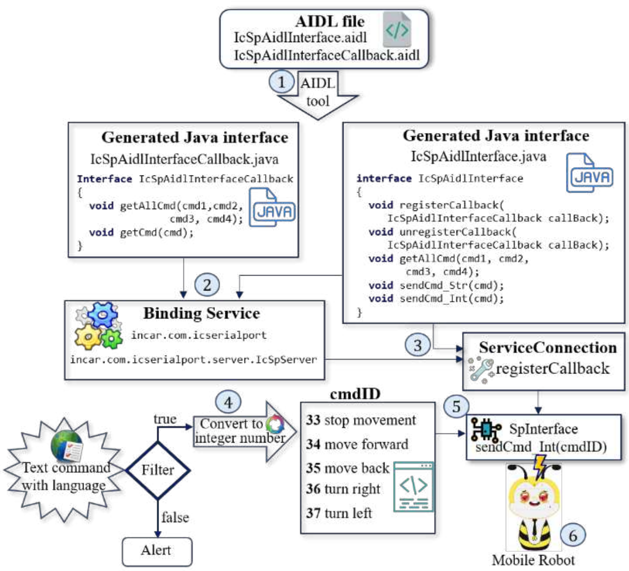

Fig. 7. IPC AIDL interaction model for Mobile Robot movement

In the Figure 7, model has 5 steps to process the Mobile Robot movement.

Step 1: Embed the .aidl formats into the project include IcSpAidlInterface.aidl and IcSpAidlInterfaceCallback.aidl through AIDL tool will generate 2 Java interfaces. These interfaces provide functions for us to interactively call Mobile Robot control commands. Such as sendCmd_Int(cmd) function to send the command to Robot.

Step 2: In this step, we will declare interface variables for binding, unbinding, and ServiceConnection references to support the registration and deregistration of callBack.

The Algorithm 1 demonstrates the bindService with action incar.com.icserialport.server.IcSpServer and package incar.com.icserialport and unbindService (in case of application shutdown).

Algorithm 1: bind and unbind service algorithms

```
private IcSpAidlInterface icsai = null;
private MyServiceConnection conn = null;
private void initAppActionCallback()
{
    conn = new MyServiceConnection();
}
public void bindService()
{
  Intent intent = new Intent();
  String action=
    "incar.com.icserialport.server.IcSpServer"
  intent.setAction(action);
  intent.setPackage("incar.com.icserialport");
  bindService(intent, conn, BIND_AUTO_CREATE);
}
public void unbindService(){
 if (icsai!=null) {
   try {
         icsai.unregisterCallback(callBack);
         if(conn != null)
           unbindService(conn);
    }catch (Exception e){ }
  }
}
```

Step 3: registerCallback for IcSpAidlInterfaceCallback by implementing the MyServiceConnection class for reference step 2 to use. This step after completing the service creation process will be used for step 2 to refer to bindService and unbindService. And onServiceConnected method creates IcSpAidlInterface object and from this object registers IcSpAidlInterfaceCallback to listen for interaction results. The details are shown in the Algorithm 2.

Algorithm 2: register and unregister callback algorithms

```
public class MyServiceConnection
             implements ServiceConnection {
 public void onServiceConnected(
      ComponentName name, IBinder service) {
    icsai = IcSpAidlInterface.Stub
               .asInterface(service);
    if (icsai!=null) {
       try{
          icsai.registerCallback(callBack);
        } catch (RemoteException e) {
              e.printStackTrace();
          }
      }
   }
 public void onServiceDisconnected
             (ComponentName name) {
    if (icsai!=null) {
       try {
          icsai.unregisterCallback(callBack);
       } catch (RemoteException e) {
              e.printStackTrace();
         }
      }
     icsai = null;
   }
}
```

Step 4: The program will filter the text command received from Google Cloud, if the command is in the group of defined commands, this text command will be converted to integer number to control Mobile Robot, and if the command is incorrect, the program will show the Alert warning. Table 2 shows Integer number for Mobile Robot movement.

Table 2. Integer number for Mobile Robot movement

Integer number	Functional meaning
33	stop movement
34	move forward
35	move back
36	turn right
37	turn left

Step 5: Invoke commands to control the robot to move using the codes analyzed in step 4. cmdID variable is the integer number to send control commands to Mobile Robot through the IPC AIDL architecture and it is shown in the Algorithm 3.

Algorithm 3: register and unregister callback

```
public void doSendCmd(int cmdID)  {
    try {
        if( icsai !=null) {
             icsai.sendCmd_Int(cmdID);
        }
    } catch (RemoteException e) {
        e.printStackTrace();
    }
}

```

Meanwhile, Fig. 8 is a summary of the UML model of the object class participating in the Mobile Robot control interaction.

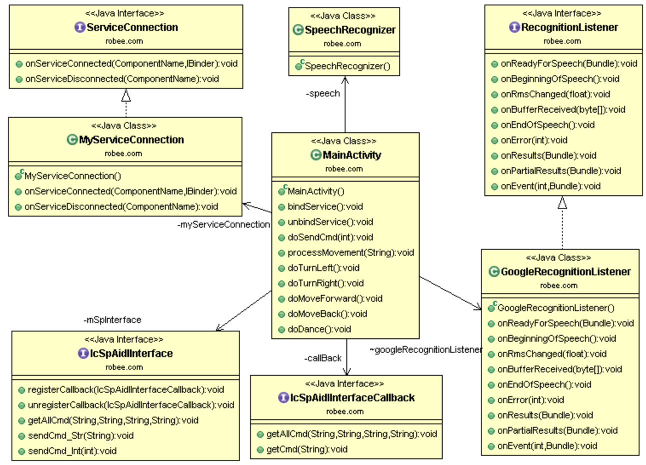

Fig. 8. Model class for Mobile Robot movement

IcSpAidlInterface is a java interface defined in the IPC AIDL architecture, which is automatically generated when the system uses the AIDL engine for processing. This interface provides several important methods by which we can interact with the circuit board system. These functions include registerCallback (used to register IcSpAidlInterface Callback), unregisterCallback (used to unregister IcSpAidlInterface Callback), getAllCmd (getAllCmd is a method that supports getting commands from the system), and sendCmd_Int (sendCmd_Int is a method to send control commands to Robot, the command is formatted as an integer number).

MyServiceConnection is a class that implements the ServiceConnection interface, providing two important functions, onServiceConnected and onServiceDisconnected to register and unregister the IcSpAidl callback. onServiceConnected is a function used for Stub to create an IcSpAidl object and register a callback to receive the system's return results. bindService task will refer to MyServiceConnection object when it is called. onServiceDisconnected is a function used for unregister a callback and unbindService task.

MainActivity is the central layer of the software system installed on Mobile Robot. This class is used to design and handle operations such as: Change the language to inter-act with Robot, transmit voice, receive text, command the Robot to perform some operations such as turning left, turning right, moving forward, move backwards, jump or stop moving. Meanwhile, the bindService, unbindService tasks are also performed in MainActivity.

Thus, we have presented the General Proposal and implementation of Research Model including the structure of Mobile Robot, interaction model, classes, system operation, IPC AIDL architecture applied.

Tensorflow Keras API is used to handle English language in case Mobile Robot has internet connection or no internet connection. The model is shown in Fig. 9 Implement Single TensorFlow .pb file in Mobile Robot for speech recognition. This model will process speech recognition through Tensorflow – Keras even when Robot is not connected to the network, in this case we do English speech processing.

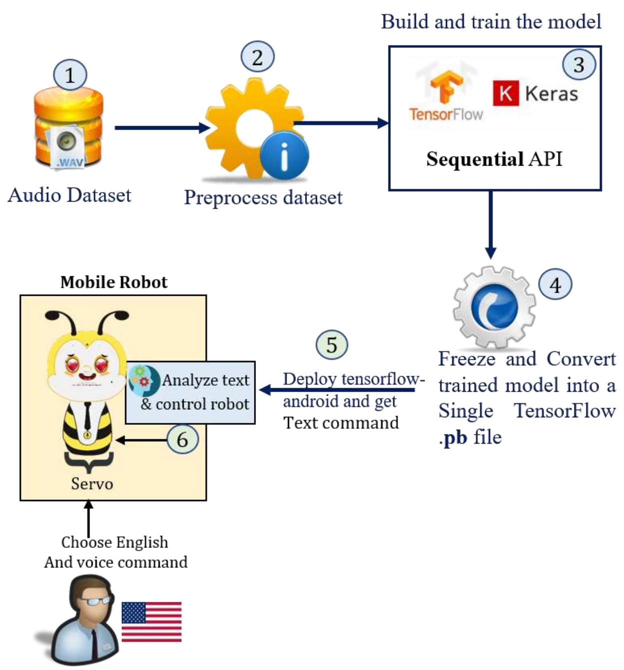

Fig. 9. Implement Single TensorFlow .pb file in Mobile Robot for speech recognition

Step 1: The Audio Dataset was reused from tensorflow.org, the original dataset consists of over 105,000 WAV audio files of people saying thirty different words. This data was collected by Google and released under a CC BY license [30]. Each utterance of a one-second file is stored in the '.wav' file format with 16 kHz sampling rate, dataset contains the words "up", "down", "left", "right", "stop". Google provided two version of Speech commands dataset to use [31], [32].

Step 2: Preprocess Dataset, includes: Import the Speech Commands dataset, Reading audio files and their labels. Convert the waveform into a spectrogram, which shows frequency changes over time and can be represented as a 2D image. This can be done by applying the short-time Fourier transform (STFT) [33] to convert the audio into the time-frequency domain.

Step 3: Build and train the model, the Sequential API of TensorFlow Kera is used. For the model, the simple convolutional neural network (CNN) [34] is used, since the audio files have been transformed into spectrogram images. The model also has the following additional preprocessing Normalization layer to normalize each pixel in the image based on its mean and standard deviation and A Resizing layer to down sample the input to enable the model to train faster.

Step 4: Freeze and Convert trained model into a Single TensorFlow .pb file by write_graph function. After having this package file, we will put it into the Mobile Android project to use in step 5.

Step 5: Use TensorFlowInferenceInterface class in tensorflow-android library to Load the TensorFlow model from Single TensorFlow .pb and then predict the speech recognition to text to get the text command.

Step 6: Deploy the application into Mobile Robot to experiment with the words "up", "down", "left", "right", "stop". These strings, when received, it will be analyzed text and converted to integers to control the Robot movement by AIDL architecture.

Fig. 10 shows the implementation Google cloud AI platform for speech recognition, when Mobile Robot has internet connection, it will support Korean language and Vietnamese language.

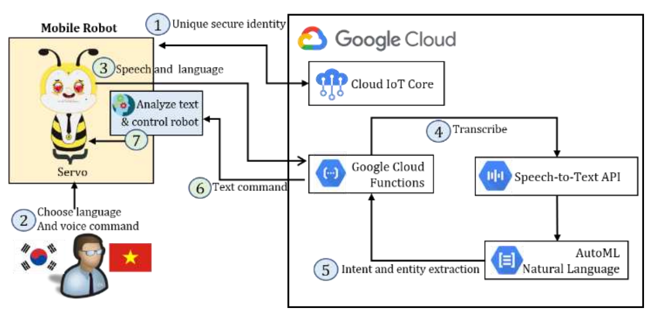

Fig. 10. Implement Google Cloud AI platform for speech recognition

The model consists of seven main implementation steps, described in detail below.

Step 1: Unique secure identity, a simple, secure, and flexible approach to identity and device management. Identity Platform is a customer identity and access management (CIAM) platform that helps organizations add identity and access management functionality to their applications, protect user accounts, and scale with confidence on Google Cloud.

Step 2: Choose language and voice command. In this step, the user will choose the interactive language designed on the software installed in Mobile Robot, select Korean or Vietnamese. Then will speak to Mobile Robot in the specified language. The voice command for Mobile Robot movement is shown in the Table 1, these voice commands will be passed to step 3 for processing.

Step 3: Speech and language. In this step, the program will receive the voice command and proceed to call the GoogleRecognitionListener to receive the text from the user's voice. The program will send voice to Google Cloud through Google Functions and receive text back after Speech-to-Text API and AutoML Natural Language processing. Result of text will be used for the integer transfer to process the Robot control, which will be detailed in other steps.

Step 4: Transcribe. In this step, the program will receive Voice from step 3 and call the Speech-to-Text API to convert voice to text, this API supports over 125 languages, Streaming speech recognition, customize speech recognition, Speech-to-Text On-Prem, Speech-to-Text can recognize distinct channels in multichannel situations and annotate the transcripts to preserve the order and API can handle noisy audio from many environments without requiring additional noise cancellation.

Step 5: Intent and entity extraction. After step 4 is done, the system will check the language using the AutoML Natural Language tool. This library allows to build and deploy custom machine learning models that analyze documents, categorizing them, identify entities within them, or assessing attitudes within them. AutoML Natural Language uses machine learning to analyze the structure and meaning of documents. We can train a custom machine learning model to classify documents, extract information, or understand the sentiment of authors.

Step 6: Text command. This step will receive the results returned from the Google Cloud platform after going through the processing processes in step 4 and step 5. The string returned from voice including language analysis is Korean or Vietnamese.

Step 7: Analyze text & control robot. Step 7 will receive the Text command from step 6 and conduct analysis, convert the text to integer number to call the commands in IPC AIDL architecture to control the movement of Mobile Robot. It is shown in the general model section. The GoogleRecognitionListener class is used to process the client-side command to receive voice input and return text results from Google Cloud. There are many functions supported in this class, like onBeginningOfSpeech (The function marks the beginning of voice processing), onEndOfSpeech (The function marks the voice processing as complete) and onResults (The result from Google Cloud analysis and it is stored in the Bundle object, we will extract data from this object).

Thus, we have presented Proposal and implementation of Research Model including interaction mobile robot model, classes, system operation, IPC AIDL architecture applied, combine with Single TensorFlow Keras API and Google Cloud AI platform. In the next section we will experiment the functions presented in the model.

5. EXPERIMENTAL AND RESULTS

We build a software system to experiment from the proposed model. The system includes the following Blueprint for Interactive Screen is shown in the Fig. 11.

Fig. 11. Blueprint for Interactive Screen

The above screen is the blueprint of the program, using Views such as: LinearLayout, TableLayout, ImageView, Edittext, TextView, Progressbar, Button. Fig. 12 shows an illustrative structure for BluePrint, applied to interactive interface design [35],[36].

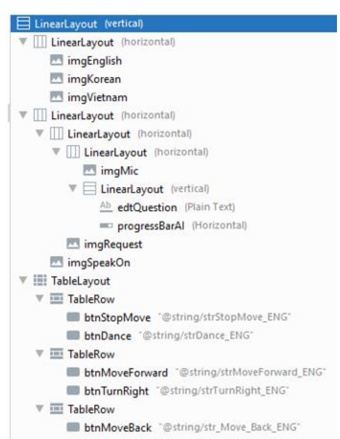

Fig. 12. Structure of Interactive Screen

The interactive screen structure consists of two parts: the interface part and the part that handles user interaction. The interface is designed in XML, the interactive part is written in java. When running the software, the Android operating system in the Mobile Robot's face will proceed to compile the XML layout into java code, from which we can access the elements in the graphic user interface similar to the Java object-oriented model.

The screen of the program will support switching between many languages including English, Korean and Vietnamese. Fig. 13 shows the Interactive Screen in English language.

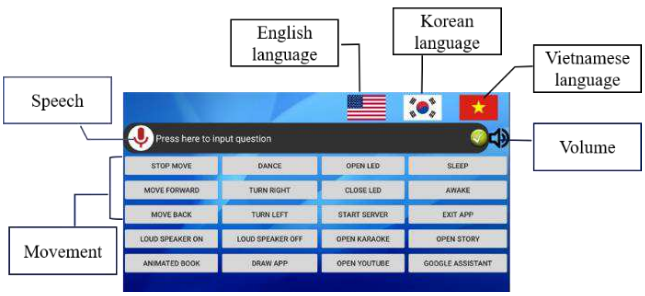

Fig. 13. Interactive Screen in English language

The main functions of the test software include: Change language, command Mobile Robot with this voice, change volume, movement functions such as turn left, turn right, go forward, go backward, jump or stop moving, exit the application. The software will be compiled and packaged into **.apk** format and deployed to Mobile Robot. It is shown in the Fig. 14.

Fig. 14. Deploy application on Mobile Robot

When choosing to switch the interface to Korean, the program will automatically change the interfaces on the software to Korean, and now the system will only listen to Korean when the user commands for Mobile Robot. Fig. 15. Shows the screen with the interface using the Korean language.

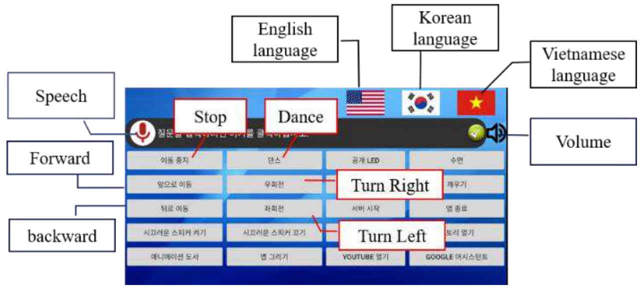

Fig. 15. Interactive Screen in Korean language

And the last one, Fig. 16 shows the Interactive Screen in Vietnamese language when user switches the interface to Vietnamese, the program will automatically change the interfaces on the software to Vietnamese, and now the system will only listen to Vietnamese when the user commands for Mobile Robot.

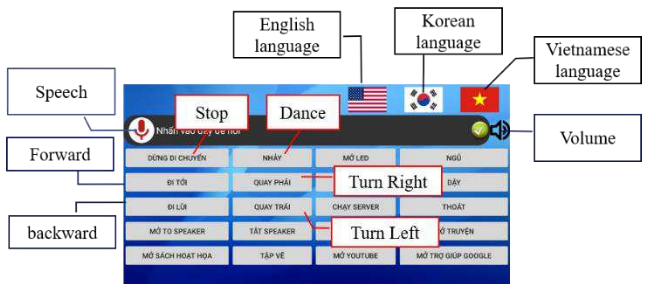

Fig. 16. Interactive Screen in Vietnamese language

Android Broadcast receiver technology will be applied in the experiment, it will listen to the internet connection state, when the connection is drop then Korean and Vietnamese language will be disabled automatically and enabled again when the internet connection is on. It is shown in the Fig. 17.

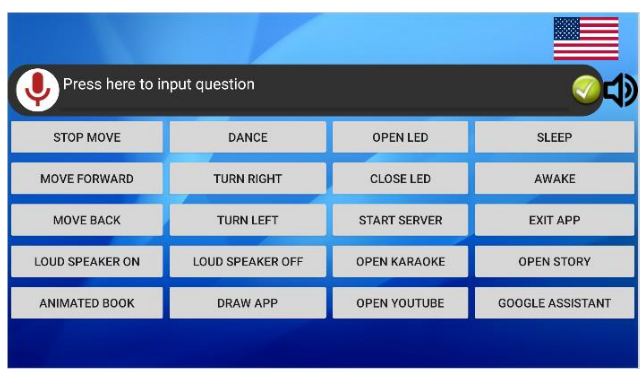

Fig. 17. Android Broadcast receiver disabled Korean and Vietnamese language

Fig. 18 shows the synchronize command strings with integer conversion while experimenting.

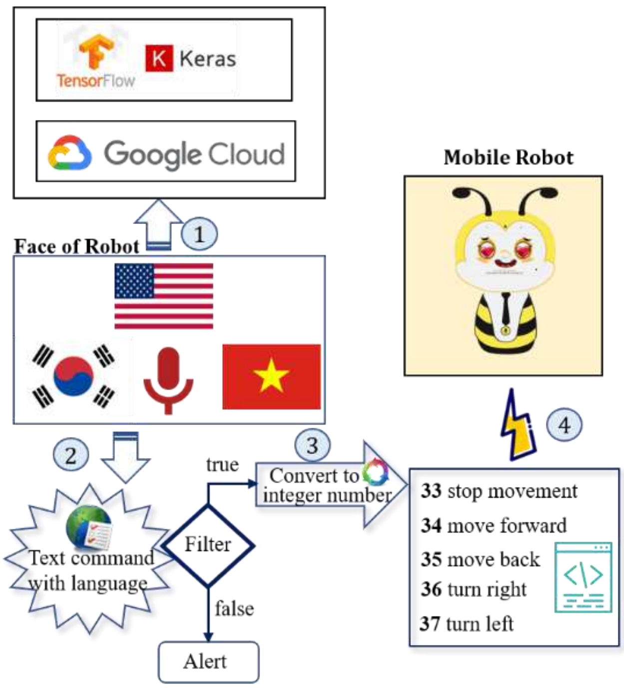

Fig. 18. Synchronize command strings with integer conversion

This model is to simplify the analysis process, the languages English, Korean, Vietnamese after receiving the results returned from Google's machine learning system or Single TensorFlow Keras model. If the command is in the group of defined commands, this text command will be converted to integer number to control Mobile Robot by IPC AIDL, and if the command is incorrect, the program will show the Alert warning.

The Algorithm 4 illustrates send the command (through the parameter cmdID which is an integer value).

Algorithm 4: doSendCmd method

```
public void doSendCmd(int cmdID)  {
    try {
        if( icsai != null) {
             icsai.sendCmd_Int(cmdID);
        }
    } catch (RemoteException e) {
        e.printStackTrace();
    }
}
```

From this function, we can call Mobile Robot commands according to IPC AIDL architecture simply as the example is shown in the Table 3.

Table 3. Experimenting with movement actions

Action	Statement
stop movement	doSendCmd(33);
move forward	doSendCmd(34);
move back	doSendCmd(35);
turn right	doSendCmd(36);
turn left	doSendCmd(37);

Thus, we tested the system according to the proposed model, took advantage of Google's artificial intelligence system and Single TensorFlow Keras model to apply Mobile Robot voice control for multi-language selection. With the IPC AIDL architecture applied in Android, the system can control the board through the generated java interfaces when using the AIDL tool.

According to Pete Warden (from Google Brain) in the paper "Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition" as mentioned before. Training the default convolution model from the TensorFlow tutorial based on Convolutional Neural Networks for Small-footprint Keyword Spotting using the V1 training data gave a Top-One score of 85.4%. A model trained on V2 data, but evaluated against the V1 test set gives 89.7% Top-One.

6. CONCLUSION AND FUTURE WORK

In this research, we have implemented voice controlling model and IPC AIDL Interaction model for Mobile Robot, Model Layer for motion by Mobile Robot and Synchronize command sequence with integer conversion.

Successfully researched and integrated Google Cloud Artificial Intelligence Technology together with IPC AIDL architecture to recognize voice and pass control commands to Mobile Robot. We have successfully built and tested an Android application installed on Robot Face Tablet with the following features: Choice of three interactive languages English, Korean and Vietnamese, along with Mobile Robot voice control feature with operations to turn left, turn right, forward, backward or stop moving. We recorded video for experiment mobile robot with auto speech recognition [37].

In addition, the article also proposes a model to integrate trained model by Tensor-Flow-Keras into Mobile Robot by Freeze and Convert trained model into a Single Tensor-Flow .pb file to recognize English speech even when Mobile

Robot has no Internet connection. This is also a potential application direction for Mobile Robot, which can add data for different languages to conduct training and use multiple languages in an environment without network connection.

In the future works, we will apply model to build the STEM hub education all in one Mobile Robot. Depending on the educational institution's lesson design, each STEM com-ponent can be developed into modules that can be installed and invoked from the STEM hub system, it is shown in the Fig. 19. In addition to the STEM system, Mobile Robot will be designed with English learning modules (Vietnam market and developing countries have a lot of demand for learning English). At the same time, the system will provide modules for Entertainment. We recorded video the future works for STEM hub and Mobile Robot face emotions [38].

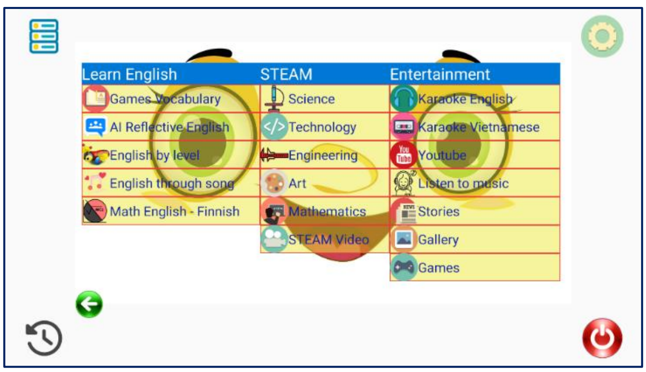

Fig. 19. STEM hub education all in one Mobile Robot

And the last one, we will apply the proposed model related to Automatic Speech Recognition to provide a learner interaction module with Mobile Robot through researches on Emotion Analysis in Human–Robot Interaction, Plutchik's wheel of emotion [39] and the future of service - the power of emotion in human-robot interaction [40].

We propose an artificial intelligence system for Mobile Robot to appear twelve emotions (contempt, sad-ness, distraction, joy, interest, rage, optimism, admiration, love, surprise, apprehension, disapproval), these will increase the excitement for learners, as well as create conversations with Robots for entertainment. The Fig. 20 shows the twelve emotions Mobile Robot Interaction, that we will apply in the future works.

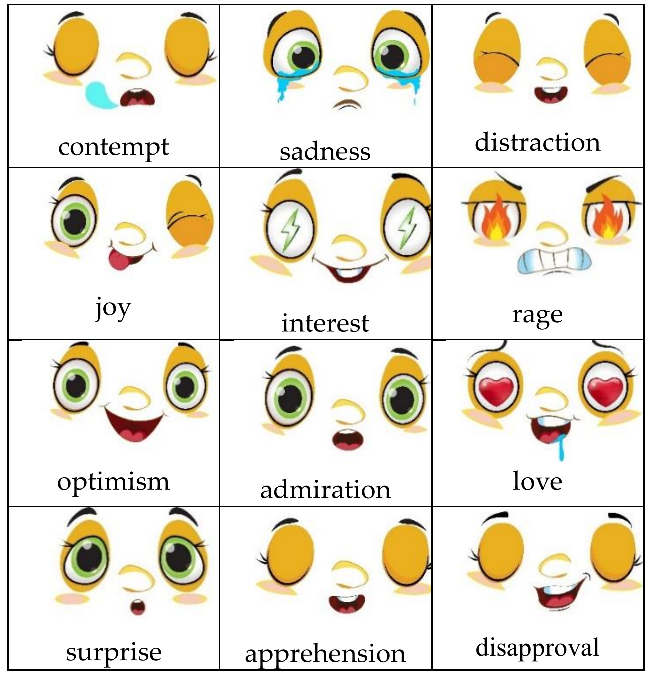

Fig. 20. Emotion Mobile Robot Interaction

In addition, Mobile Robot can develop integration with IoT systems, using Edge computing technology to control devices in the home.

Declarations

Video abstract 1: Video demo Movement control by Automatic Speech Recognition [37]

https://youtu.be/bJ8pKxXy0dU

Video abstract 2: Video Future works for Mobile Robot [38]

https://youtu.be/WnaFSKuzp3s

Conflict of interest: Authors don’t have any financial relationship with an organization that sponsored the research and didn’t receive any compensation or consultancy work. There aren’t any potential conflicts of interests that are directly or indirectly related to the research.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Kah, P., Shrestha, M., Hiltunen, E. et al. “Robotic arc welding sensors and programming in industrial applications,” Int J Mech Mater Eng, 10, 13 (2015). https://doi.org/10.1186/s40712-015-0042-y

[2] Cruzr Robot https://www.ubtrobot.com/products/cruzr?ls=en (access on Jan-21-2021)

[3] Human Robot https://www.asme.org/topics-resources/content/10-humanoid-robots-of-2020 (access on Jan-21-2021)

[4] S. B. Nite., M. Margaret., R. M. Capraro., J. Morgan., C. A. Peterson., "Science, technology, engineering and mathematics (STEM) education: A longitudinal examination of secondary school intervention," 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, 2014, pp. 1-7, doi: 10.1109/FIE.2014.7044214

[5] P. P. Merino., E. S. Ruiz., G. C. Fernandez., M. C. Gil., "Robotic Educational Tool to engage students on Engineering," 2016 IEEE Frontiers in Education Conference (FIE), 2016, pp. 1-4, doi: 10.1109/FIE.2016.7757417.

[6] David Griffiths., Dawn Griffiths., (2015), "Head First Android Development," O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

[7] Fachantidis, N., Dimitriou, A.G., Pliasa, S., Dagdilelis, V., Pnevmatikos, D., Perlantidis, P., Papadimitriou, A.P., (2017). "Android OS Mobile Technologies Meets Robotics for Expandable, Exchangeable, Reconfigurable, Educational, STEM-Enhancing, Socializing Robot," Interactive Mobile Communication, Technologies and Learning. Springer, 487-497.

[8] Kaleci, Devkan., Korkmaz, Özge., (2018). "STEM education research: Content analysis," Universal Journal of Educational Research. Vol. 6. 2404-2412. 10.13189/ujer.2018.061102.

[9] Science Robotics Special Edition Booklet, 24 July 2019 Vol. 4, No. 32, aax2352, www.ScienceRobotics.org

[10] AI processor Rockchip RK3399Pro Datasheet, Revision 1.0, Nov. 2018, Fuzhou Rockchip Electronics Co., Ltd, https://rockchip.fr/RK3399Pro datasheet V1.0.pdf (access on Jan-21-2021)

[11] Google Cloud AI Platform

https://cloud.google.com/speech-to-text

(access on Jan-21-2021)

[12] Android Interface Definition Language (AIDL) https://developer.android.com/guide/components/aidl

[13] D. Steinkraus., I. Buck., P. Y. Simard., "Using GPUs for machine learning algorithms," Eighth International Conference on Document Analysis and Recognition (ICDAR'05), Seoul, Korea (South), 2005, pp. 1115-1120 Vol. 2, doi: 10.1109/ICDAR.2005.251

[14] H. Amrouch., G. Zervakis., S. Salamin., H. Kattan., I. Anagnostopoulos., J. Henkel., "NPU Thermal Management," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, IEEE, Vol. 39, No. 11, pp. 3842-3855, Nov. 2020, doi: 10.1109/TCAD.2020.3012753

[15] ARM, big. LITTLE Processing technologies, https://www.arm.com/why-arm/technologies/big-little

[16] K. Yu., D. Han., C. Youn., S. Hwang., J. Lee., "Power-aware task scheduling for big. LITTLE mobile processor," 2013 International SoC Design Conference (ISOCC), Busan, Korea (South), 2013, pp. 208-212, doi: 10.1109/ISOCC.2013.6864009

[17] S, Karpagavalli., Chandra, Evania., (2016). "A Review on Automatic Speech Recognition Architecture and Approaches," International Journal of Signal Processing, Image Processing and Pattern Recognition. 9. 393-404. 10.14257/ijsip.2016.9.4.34.

[18] V. T. Pham et al., "Independent Language Modeling Architecture for End-To-End ASR," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 7059-7063, doi: 10.1109/ICASSP40776.2020.9054116

[19] E. Palogiannidi., I. Gkinis., G. Mastrapas., P. Mizera., T. Stafylakis., "End-to-End Architectures for ASR-Free Spoken Language Understanding," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 7974-7978, doi: 10.1109/ICASSP40776.2020.9054314

[20] M. G. Tulics., G. Szaszák., K. Mészáros., K. Vicsi., "Using ASR Posterior Probability and Acoustic Features for Voice Disorder Classification," 2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Mariehamn, Finland, 2020, pp. 000155-000160, doi: 10.1109/CogInfoCom50765.2020.9237866

[21] C-W. Huang., Y-N. Chen., "Learning Asr-Robust Contextualized Embeddings for Spoken Language Understanding," ICASSP 2020, 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 8009-8013,

doi: 10.1109/ICASSP40776.2020.9054689

[22] A. Gupta., P. Goswami., N. Chaudhary., R. Bansal., "Deploying an Application using Google Cloud Platform," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 236-239, doi: 10.1109/ICIMIA48430.2020.9074911

[23] K. Shannon., R. Snodgrass., "Mapping the Interface Description Language type model into C," in IEEE Transactions on Software Engineering, IEEE, Vol. 15, No. 11, pp. 1333-1346, Nov. 1989, doi: 10.1109/32.41327

[24] L. Lamport., "On Interprocess Communication", SpringerVerlag, 1986.

[25] M. Kashyian., S. L. Mirtaheri., E. M. Khaneghah., "Portable Inter Process Communication Programming," 2008 The Second International Conference on Advanced Engineering Computing and Applications in Sciences, Valencia, Spain, 2008, pp. 181-186, doi: 10.1109/ADVCOMP.2008.38

[26] J. Choi., H. Gill., S. Ou., Y. Song., J. Lee., "Design of Voice to Text Conversion and Management Program Based on Google Cloud Speech API," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2018, pp. 1452-1453, doi: 10.1109/CSCI46756.2018.00286

[27] Sapounidis. T., Alimisis Dimitris., (2020). "Educational robotics for STEM: A review of technologies and some educational considerations," Science and Mathematics Education for 21st Century Citizens: Challenges and Ways Forward; Nova Science Publishers: Hauppauge, NY, USA, 167-190.

[28] Li Yeping., Wang Ke., Xiao Yu., Froyd, Jeffrey. (2020). "Research and trends in STEM education: a systematic review of journal publications," International Journal of STEM Education, Vol.7, No.1, 1-16. 10.1186/s40594-020-00207-6.

[29] A. Eguchi., L. Uribe., "Robotics to promote STEM learning: Educational robotics unit for 4th grade science," 2017 IEEE Integrated STEM Education Conference (ISEC), 2017, pp. 186-194, doi: 10.1109/ISECon.2017.7910240.

[30] Pete Warden., "Speech commands: A dataset for limited vocabulary speech recognition" arXiv, vol. 2018, no.1804.03209.

[31] Speech commands dataset version 1 (2017). Available: http://download.tensorflow.org/data/speech_commands_v0.01.tar.gz (access on Jan-21-2021)

[32] Speech commands dataset version 2 (2018). Available: http://download.tensorflow.org/data/speech_commands_v0.02.tar.gz (access on Jan-21-2021)

[33] D. Griffin., Jae Lim., "Signal estimation from modified short-time Fourier transform," IEEE Transactions on Acoustics, Speech, and Signal Processing, IEEE, Vol. 32, No. 2, pp. 236-243, April 1984, doi: 10.1109/TASSP.1984.1164317.

[34] Ghosh Anirudha., Sufian A., Sultana Farhana., Chakrabarti Amlan., De Debashis., (2020). Fundamental Concepts of Convolutional Neural Network. 10.1007/978-3-030-32644-9_36.

[35] Dave MacLean., Satya Komatineni., Grant Allen (2015), "Pro Android 5," Apress Media, LLC 233 Spring Street, 6th Floor, New York, NY 1001.

[36] Dave Smith (2015), "Android Recipes," 4th Edition. Apress Media, LLC 233 Spring Street, 6th Floor, New York, NY 1001.

[37] Duy Thanh Tran et al., Video abstract 1: Video demo Movement control by Automatic Speech Recognition https://youtu.be/bJ8pKxXy0dU

[38] Duy Thanh Tran et al., Video abstract 2: Video Future works for Mobile Robot

https://youtu.be/WnaFSKuzp3s

[39] Szabóová, M.; Sarnovský, M.; Maslej Krešňáková, V.; Machová, K. "Emotion Analysis in Human–Robot Interaction," Electronics 2020, 9, 1761. https://doi.org/10.3390/electronics9111761

[40] Stephanie Hui-Wen Chuah, Joanne Yu, "The future of service: The power of emotion in human-robot interaction," Journal of Retailing and Consumer Services, Vol. 61, 2021.

https://doi.org/10.1016/j.jretconser.2021.102551