Expertise
and insight
for the future

Jingxuan Wu

Household Garbage Pre-sorting Based on Voice Control

Metropolia University of Applied Sciences

Bachelor of Engineering

Information Technology

Bachelor's Thesis

28 April 2020

Author	Jingxuan Wu
Title	Household garbage pre-sorting based on voice control
Number of Pages	33 pages + 1 appendix
Degree	Bachelor of Engineering
Degree Programme	Information Technology
Professional Major	Smart Systems
Instructors	Sami Sainio, Senior Lecturer

Due to lack of knowledge of garbage classification, at least 33% of garbage is not well disposed of in an environmentally friendly manner. In order to make garbage sorting easier, the goal of this project was to investigate the status of garbage sorting, and design and create a garbage pre-sorting household product based on voice control. When people say the name of the item they want to throw away, the system would detect their speech and open the corresponding garbage bin.

In this project, an Arduino UNO was combined with a WEGASUN-M6 voice recognition module based on the Arduino programming platform. Some boxes and servo motors were used to simulate the garbage bins to enhance the sense of experience.

The voice recognition system has been successfully implemented with general garbage sorting mechanism. First, typical household wastes were divided into four categories based on their different disposal and utilization methods. The waste categories were input into the created system as a library. The system could recognize 150 different wastes which covers the majority of household garbage. The system succeeded in real-time pre-sorting voice control within a fixed distance. When receiving detected voice input, the cover of the garbage bins opens automatically. The recognition system is compatible with both English and Chinese (Mandarin).

This voice controlled garbage pre-sorting system not only improves people's awareness of garbage disposal, but also enhances the accuracy of garbage pre-sorting, which has a positive impact on the environment.

Keywords	household garbage pre-sorting, WEGASUN-M6, voice recognition, Arduino

Contents

List of Abbreviations

IWR	Intelligent Waste Recycling. A high-technology method to recycle wastes.
RVM	Reverse Vending Machine. A machine utilized to recycle bottles or cans. The machine gives refund to the end users.
SVM	Support Vector Machine. A class of generalized linear classifiers that perform binary classification of data according to supervised learning.
CNN	Convolutional Neural Network. A type of deep neural network most commonly used to analyze visual images.
MFCC	Mel-Frequency Cepstral Coefficients. An audio extraction method.
HMM	Hidden Markov Model. A statistical model widely used in various natural language processing applications.
LVCSR	Large Vocabulary Continuous Speech Recognition. A speech recognition system that recognizes continuous audio streams and automatically converts audio information into text.
GMM	Gaussian Mixture Model. A clustering algorithm that uses Gaussian distribution as a parameter model.
DNN	Deep Neural Network. A neural network with many hidden layers.

1 Introduction

According to a survey from Ormedia agency in the United States, the tap water people usually drink contains plastic particles that are invisible to the naked eyes. With a similar investigation of the tap water in 14 main countries around the world, plastic particles were found in 83% cases. The amount of plastic pollution is great. In other words, due to human behavior, rivers and oceans are flooded with plastic waste. [1.] According to the Global Waste Index 2019, a human being produces about 2,010,000,000 tons of garbage every year. Due to the lack of knowledge of garbage classification, at least 33% of solid waste is not well disposed of in an environmentally friendly way. [2.] Usually, when people do not know how to sort their garbage, they mix all the garbage together in one bag. However, it is simple for people to execute garbage pre-sorting if they realize what type of garbage they have at home. [3.] In order to improve people's awareness of the garbage classification and make contributions to environmental protection, the goal of this project was to design and create a garbage pre-sorting product based on voice control.

Unutilized and unrecycled garbage causes resource waste and environmental pollution every year [2]. Taking the condition in China as an example, according to statistics from the China Association of Urban Environmental Sanitation, urban domestic waste in the country exceeds 150 million tons per year and increases at an annual rate of 8-10%. The value of resource losses caused by garbage is about 25 billion to 30 billion Chinese Yuan each year. Rough estimated, one ton of waste paper can produce 850 kilograms of good quality paper, saving 300 kilograms of wood, which is equivalent to cutting down 17 big trees; one ton of waste plastics can be recycled to produce 300 kilograms of unleaded gasoline and diesel; one ton of scrap steel can make 750 kilograms of steel; and one ton of waste glass can make 20,000 glass bottles of 500 grams. [4.]

Going through a large amount of relevant literature shows that the existing common Intelligent Waste Recycling (IWR) methods such as laser processing, winnowing processing, and image processing are mainly used in the waste disposal industry [5]. This final year project provides a method to lighten the burden of final waste processing in recycling plants. This can be achieved by improving the pre-sorting level of household trash.

In this project, voice recognition was combined with garbage categories, to offer real-time garbage pre-sorting voice guidance when people dispose of their garbage at home. An Arduino UNO was used as the programming kit based on the Arduino platform. A WEGASUN-M6 voice module with a microphone and a speaker was used for voice detection and processing the input voice signals. Both English and Chinese (Mandarin) were implemented as the main languages for the recognition system. In order to enhance the sense of experience, some boxes were simulated as different garbage bins and some servos were placed onto the boxes to control their lids.

The voice controlled household garbage pre-sorting system reduces the workload for subsequent subdivision processing and improves people's awareness of garbage classification. This further has a positive impact on the environment. [3.] However, it still has its limitations: when people stand farther than the fixed recognition distance, the noise and the vague voice input would be detected together by the system. In this case, the error rate would be high.

2 Literature review

Everyone throws out much garbage every day. The composition of garbage is mainly about food, paper and cardboard. [2.] Figure 1 shows the global waste composition:

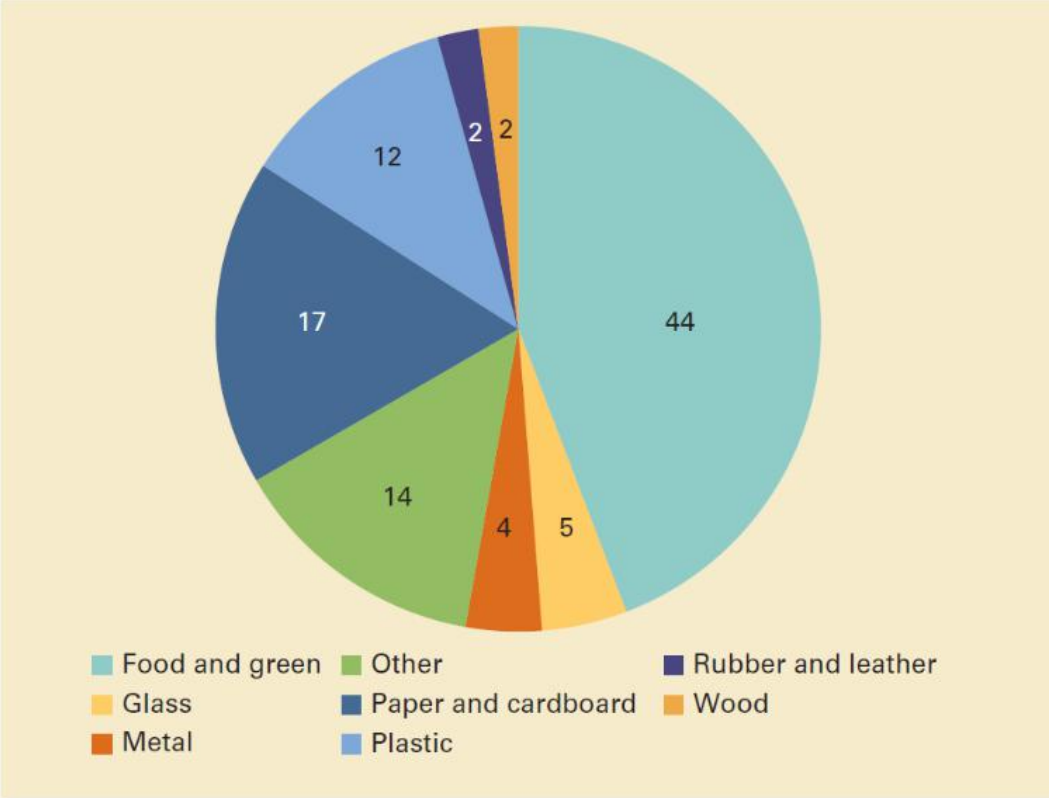

Figure 1. The percentage of global waste composition. Copied from What A Waste 2.0. A Global Snapshot of Solid Waste Management to 2050 [2].

Figure 1 illustrates that the recyclable waste contains most of the glass, metal, paper, cardboard, plastic, rubber, leather and wood, which account for 42% of the total wastes. However, according to figure 2, only 13.5% of waste is recycled. [2.]

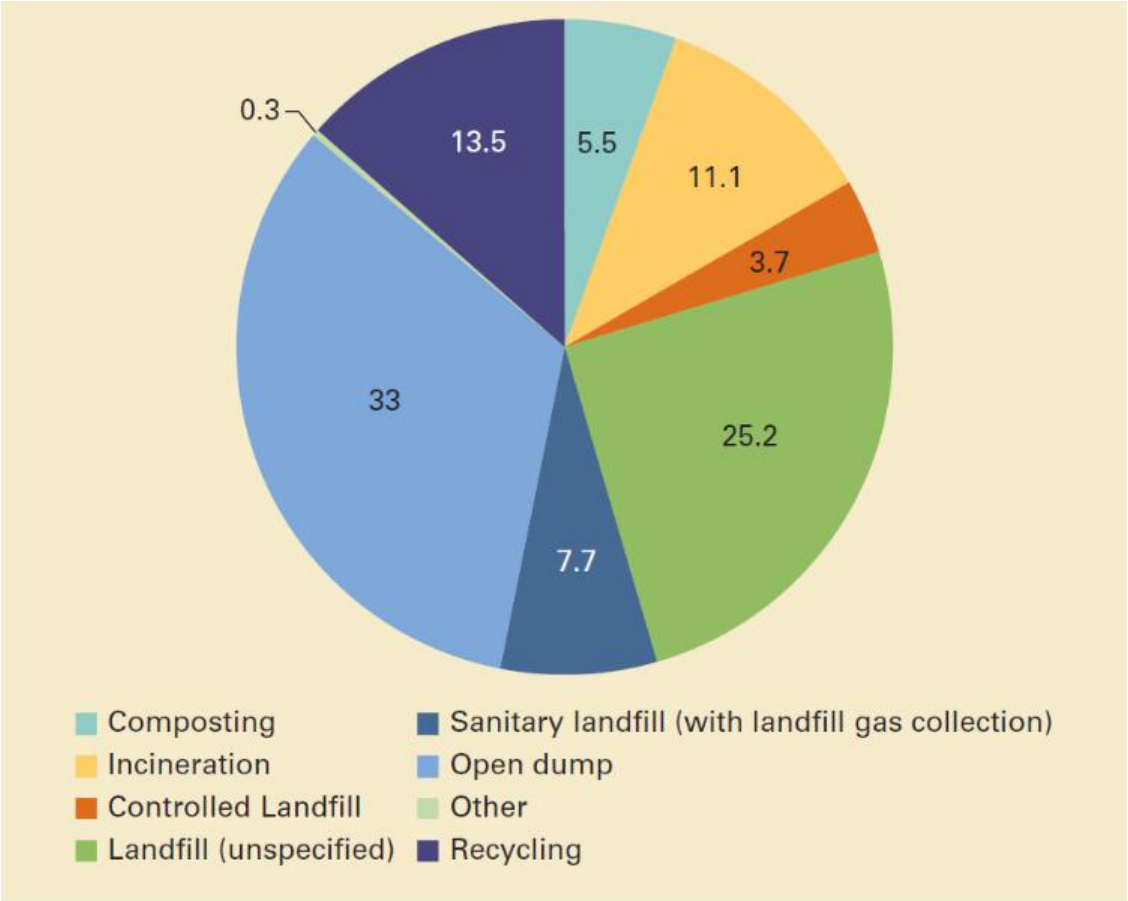

Figure 2. The percentage of global garbage disposal. Copied from What A Waste 2.0. A Global Snapshot of Solid Waste Management to 2050 [2].

In some areas with better garbage management, most of the garbage will be harmlessly treated by sanitary landfill, incineration, compost, etc. But the garbage in more places is often simply piled up or landfill, leading to the spread of odor and polluting the soil and groundwater. [3.]

With the development of technologies, in some areas, the garbage in the container could be transported underground in less than one minute through pipes by air flow in the pipes [6]. Nowadays, the rise of the "Internet plus" model has brought a new model for waste separation and recycling, which is expected to break the bottleneck that has long plagued garbage classification. Unless the stationary pneumatic refuses collection, some intelligent classification and collection systems for urban garbage based on the Internet of Things is proposed. [5.]

2.1 Intelligent garbage sorting methods

The existing smart appliance includes intelligent trash cans. One household automatic trash can with self-sealing is named Townew. It is implemented with infrared sensors so that it could detect the hand waving in 35 cm and open the lid automatically, waiting for the trash in. And when the trash bag is full, pressing the button on the can will seal the used bag and replace it with a new bag. This kind of product does not require to touch the can lid when opening the lid, which makes garbage throwing more convenient and sanitary. However, this product cannot help to sort the garbage with its technology. [7.]

As there are people who do not dispose of their garbage into the correct garbage bins, most of the existing IWR systems collect all types of wastes together and then separate different types of the waste into different containers to achieve waste sorting. In this section, three garbage sorting methods for the waste disposal industry would be introduced, including image processing, winnowing processing and laser processing. [5.]

2.1.1 Laser processing

The most widely used waste sorting device in Europe is named the Reverse Vending Machine (RVM), which is utilized for beverage containers. Normally, the refundable deposit for the containers would show up on the labels of the containers or outer packaging. The working process of the RVMs is that users put the used empty bottles or cans into its channel, and the scanner on the machine detects the materials, shapes, and the barcodes together with the refundable price of the item. The machine determines and calculates the number of bottles and cans in real-time. When placing the beverage containers, they are separated into different types of waste through the conveyor belt. After placing all the containers, the machine prints a receipt for the user. The value on the receipt from the RVM can be used as an offset coupon in the supermarket when shopping. At the same time, it can be used to exchange coins at checkout counters. The sorted bottles and cans are transported regularly and processed centrally. [8.]

The common barcode scanner inside RVMs is composed of a source of light, a scanning module, an optical lens, and an analog-to-digital conversion circuit. The system uses photoelectric elements to convert the detected light signal into an electrical signal. Then the light signal is converted into a digital signal by a converter through the analog-to-digital conversion circuit and transmitted to a computer for processing. [8.] The basic working principle of the scanners is as figure 3 below.

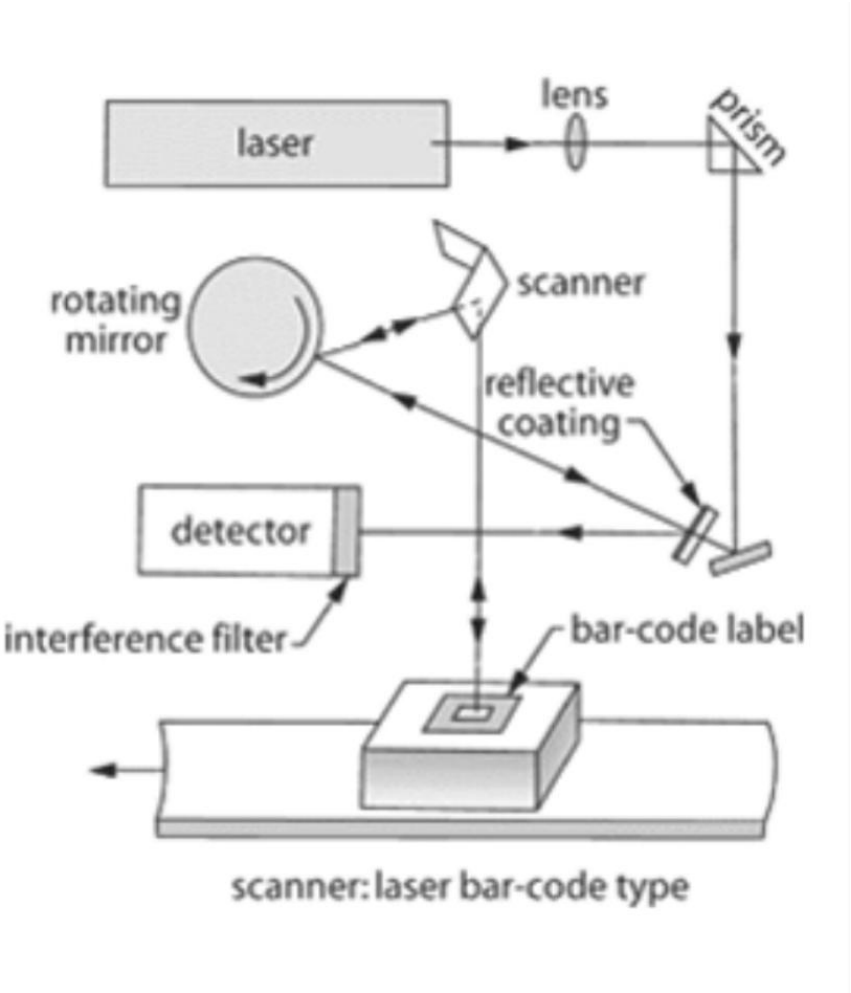

Figure 3. The working principle of a laser scanner for barcode detection. Copied from Optics 4 Kids [9].

As figure 3 demonstrates, the laser scanner emits a beam of light through a laser diode and irradiates a rotating prism or an optical lens. Then the reflected light passes through the reading window and hits the bar code surface. After reaching the bar-code label or doing an empty reflection, the light is returned to the scanner window. The light is collected and focused by a mirror, and a photoelectric converter will convert the light

to an electrical signal. The signal will be decoded by the scanner or the decoding software on the terminal. [9.]

Although the RVMs are useful to collect and recycle beverage containers, they also have some disadvantages. The first one is that the RVMs only can recycle beverage containers such as bottles, cans and glasses, which cannot cover all kinds of waste. Secondly, if the label on the containers is missing, the laser scanner cannot detect what waste it is and will refuse to recycle the waste. [8.] In this case, it would reduce the recycling rate. In order to eliminate the disadvantages above, another intelligent recycling method with wind has been considered [10].

2.1.2 Winnowing processing

The theoretical basis of winnowing selection is the gravity principle of substance. Winnowing technology has been widely used in sorting of materials. [10.] For example, in traditional agricultural societies, the winnowing technology was used to separate the bran and rice husk from the clean rice [11]. In addition to agricultural applications, coal preparation plants use the winnowing separation technology to achieve classification of different components in coal [10].

From the physics perspective, the principal of the winnowing system is that for an object with a mass of m, a volume of V, and a density of ρ, and the object falls freely at a place where H is high. Under the action of wind F in the horizontal direction (ignoring resistance in the horizontal direction), its motion state will change. [12.] The motion of the object is shown in figure 4.

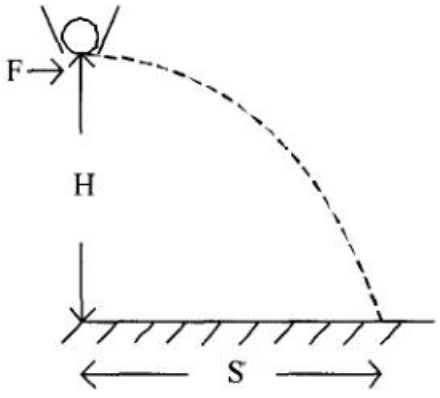

Figure 4. Object motion by wind power. Adapted from What Is 2D Projectile Motion [12].

Figure 4 above shows the motion of the object model by wind power, and the force of the object is shown in figure 5 below:

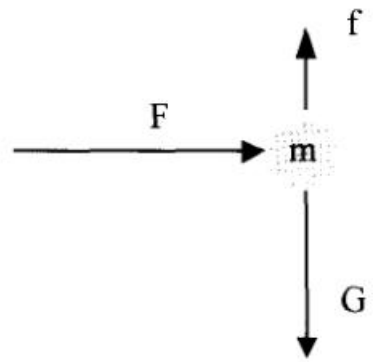

Figure 5. Force analysis of the object. Adapted from What Is 2D Projectile Motion [12].

If ay is set as the acceleration in vertical direction; ax is set as the acceleration in horizontal direction and in the horizontal direction, and the displacement of the object is S. According to the kinematics theorem:

In the vertical direction:

$$G - F = m * ay$$

$$H = 1/2 * ayt^2$$

In the horizontal direction:

$$F = m * ax$$

$$S = 1/2 * axt^2$$

Above all the equations, the displacement:

$$S = F * H / (\rho * V * g - f)$$

A conclusion can be drawn easily that for objects with different specific gravity, the displacement in the horizontal direction and the specific gravity is inversely proportional. [12.] As the theoretical background has been described, figure 6 shows the winnowing selection model.

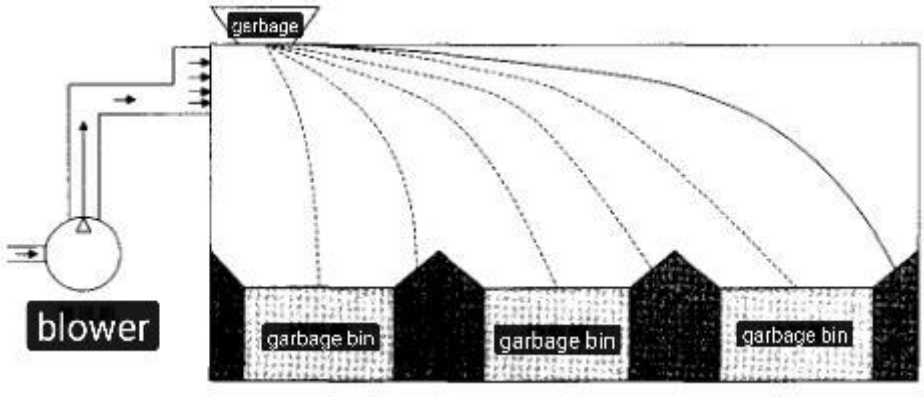

Figure 6. Garbage soring model. Adapted from Design of Intelligent Garbage Sorting Model Based on Winnowing [13].

Based on the physical principles, some garbage bins are put at the bottom of the box [12]. The garbage falls freely from the top, and the wind from the blower blows from the left to right. Due to the wind force, the garbage will deflect. According to the equation S = F*H / (ρ*V*g – f), the garbage which falls into the same garbage bin will have a smaller density difference. If the wind force is changed, the landing distance of the garbage will be changed as well. [13.] Therefore, the classification of garbage can be realized by adjusting the speed of the blower in reality [12].

The winnowing processing method can deal with most materials, but it still has some limitations: different materials have different densities and volumes, so that the volume of each piece of garbage should be detected and controlled when processing and the density should be the only independent parameter in this model. Otherwise, the sorting result would be inaccurate. [13.] Therefore, before using the winnowing system, the staff must spend much time on garbage collection and disposal, which requires a considerable amount of time and labor when processing. Another processing method named image

processing does not require pre-processing and can improve the efficiency and accuracy. [13.]

2.1.3 Image processing

Image processing is a visualized approach to sort garbage quickly and accurately in daily life. Most of the existing image recognition garbage sorting systems consist of cameras, information processors, trash can housings, several built-in trash cans, indicator controllers, several indicators and power supply modules. The most popular company in the visual recognition field is AlexNet, which won the championship in ImageNet Large-Scale Visual Recognition Challenge in 2012. It brought the convolutional neural network (CNN) to the society and made a great contribution to image classification. [14.]

This method is used when the garbage is collected together and processed on the conveyor belt. The basic working principle is that the camera is used to send the captured images of garbage to the information processor. [13.] In the meanwhile, the information processor is used to perform garbage recognition processing according to the images of garbage and the standard templates in its pre-stored standard template library [15]. After a while, the information processor generates a classification result, and sends the information based on the classification result. Then the conveyor belt sends the garbage to different containers and sorts out garbage successfully. [13.]

The existing studies on garbage classification are most related to Support Vector Machines (SVMs) and convolutional neural networks (CNNs). The researchers have received images of garbage on a clean white background through the camera on the pipeline and have used these two methods to sort the garbage into six different categories: paper, glass, plastic, metal, cardboard and trash. The basic SVM is a binary classification model, which defines a linear classifier with the largest interval in the feature space and the largest interval makes it different from the perception. [14.] Figure 7 illustrates the SVM sorting method.

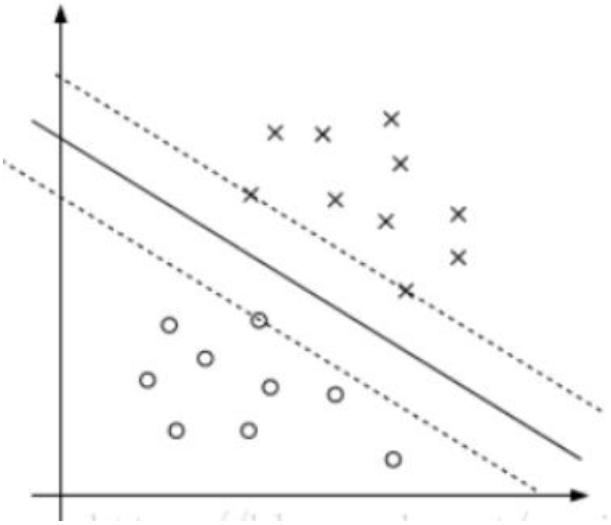

Figure 7. An ordinary SVM classification. Adapted from Classification of Trash for Recyclability Status [14].

According to figure 7 above, an ordinary SVM is used to solve the separation hyper-plane that can correctly divide a training data set and the hyper-plane has the largest geometric interval. This straight line is the most perfect of numerous straight lines that can be classified, because it is exactly in the middle of the two classes and the distance from the points of both classes is the same. The so-called Support Vector in the SVM is the points closest to the dividing line. If these points are removed, the straight line will most likely change their positions, so these points define the classifier. [14.]

The CNN has become a part of the most influential innovation in computer vision, and it has been widely used in image classification, localization, and other fields [14]. One of the advantages of the CNN compared to traditional image processing algorithms is that it avoids complex pre-processing of an image, especially manual participation in image pre-processing. Therefore, it can directly input the original image for a series of work. [16.] The following figure 8 shows the composition of CNN.

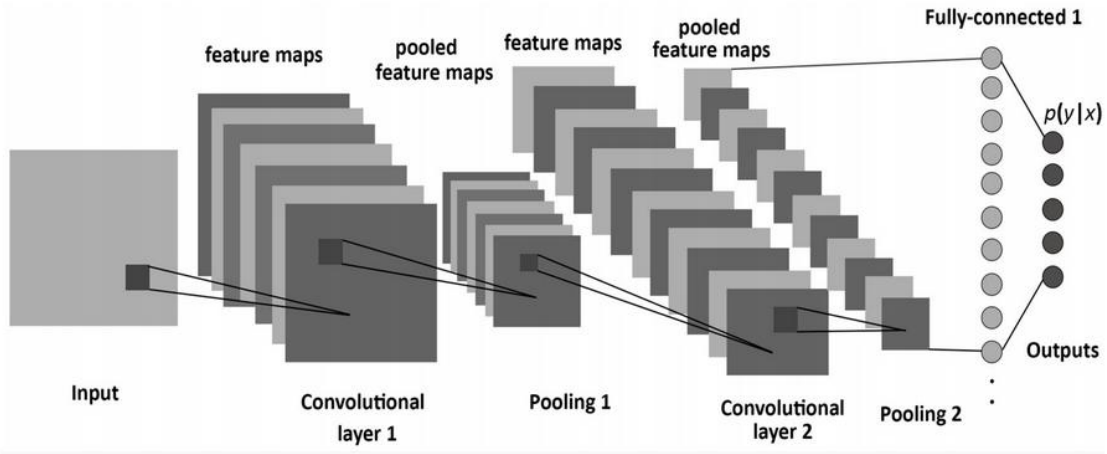

Figure 8. Typical CNN structure. Copied from Multilayer Hybrid Deep-Learning Method for Waste Classification and Recycling [16].

As shown in figure 8 above, a CNN generally includes convolutional layers, pooling layers and a fully connected layer [16]. Considering signal processing, the convolution operation in the convolution layer is a filter (convolution kernel) that performs frequency filtering on the signal. The CNN training is to find the best filter to make the filtered signal easier to classify. From the perspective of template matching, each kernel of the convolution can be considered as a feature template. [17.] The training is to obtain a suitable filter, so that there is a high activation of a specific mode to achieve the purpose of classification / detection. The convolution layers can set multiple filters to obtain different feature maps. [16.]

Pooling layer operations can also be called down-sampling or under-sampling with maximum pooling and average pooling. Pooling can reduce parameters to reduce the memory consumption and prevent over-fitting. In the meanwhile, it improves the receptive field to obtain more contextual information and its invariance enhances the model's robustness and generalization. [17.]

After capturing enough features to identify the image, the next step is classification. The input of fully connected layers is linked to each other. It plays the role of "classifier" in the entire convolutional neural network. The fully connected layer is also called feed-forward layer, which can be used to map the final output to a linearly separable space. [16.] Usually at the end of the convolutional network, the cuboid obtained at the end is

flattened into a long vector, and sent to the fully connected layer with an output layer for classification [17].

CNN and many network models derived from it can directly take image data as input. They do not only need to perform complex operations such as preprocessing and additional feature extraction of the image manually, but also the unique fine-grained feature extraction method could process images almost at human level. The image detection has the characteristics of no contact and high precision, which is extremely applicable for garbage sorting. [18.]

The image processing method for sorting garbage meets most of the requirements to detect and classify garbage. However, the equipment it uses costs a lot, such as the high precision cameras and its power supply modules. As a result of this, the image processing method is difficult to be widely used in people's daily life. [19.]

2.2 Voice recognition

Communicating with machines and letting the machines understand what people say is what people dream to achieve. Speech recognition technology is a process that converts spoken words into text. The technology uses speech as the input object and the machines receive speech signals, recognizing and understanding the content of speech automatically. Voice recognition technology is being used in instant messaging applications, search engines, in-vehicle systems, and home automation. [20.]

The existing recognition voice software includes Dragon Professional, Google Gboard, Otter, Amazon Transcribe, Microsoft Azure Speech to Text, Watson Speech to Text, etc. Among them, the Dragon Professional software is used for business-grade dictation, which can obtain 99% accuracy with typing 160 words every minute. The system processes the voice signal and its pattern to obtain the text result. It is widely used for individuals to record and manage documents. [21.]

2.2.1 Voice processing principles

Sound is a mechanical wave generated by the vibration of an object and voice is the external form of language. The common audio formats include CDA, WAV, MP3, FLAC, WMA, etc. Taking mp3 as an example, mp3 is an audio compression technology. It uses Moving Picture Experts Group Audio Layer 3 technology to discard data that is not important to human hearing in pulse code modulation data to obtain smaller size of files. When using a recognition system, an mp3 file should be uncompressed to a wav file because the audio needs to be transferred as a sound wave for the recognition system. A wav file is used to save audio information resources of the Windows platform. This technology is developed by Microsoft. [20.] Figure 9 below shows what a sound waveform looks like.

Figure 9. Sound waveform diagram. Copied from How Speech Recognition Works [20].

Except for a file header stored in the wav file, there are points of the sound wave. Sometimes it is necessary to remove the silence part at the end, which could improve the sound quality before starting to process and decrease the interference for the next steps. At the same time, to analyze the sound, it needs to be re-edited and cut into small segments with simple incisions, which is achieved by using window functions. [20.] Each segment is called a frame. After obtaining the pieces of segments, the waveform needs to be transformed because it has little ability to describe the information in the time domain. A common transformation method is to extract Mel-Frequency Cepstral Coefficients (MFCC) features, which could transform voice into a matrix named “observation sequence”. [22.]

Transforming the matrix to text format is the next step. Firstly, the system recognizes the frames as states. Many frames of speech correspond to one state, and every specific number of states is combined into a phoneme. [20.] The commonly used phoneme set for English is a set of 39 phonemes from Carnegie Mellon University [23].

A word consists of many phonemes; in other words, many phonemes make up the pronunciation of the word. Once the system finds which state is related to the frame of speech, the recognition result comes out. [20.]

The most appropriate method to obtain the state each frame phoneme corresponds to is to see which state corresponds to the frame that has the highest probability. The probability is from the acoustic model, which stores a large number of parameters. The method of obtaining these parameters is called “training”. It requires the use of a huge amount of voice data. [20.]

However, when processing audio data, each frame will obtain a status number, and finally the whole voice will obtain a mess of status numbers. The status numbers between two adjacent frames are basically different. For example, it is supposed that there are 100 frames of speech, and each frame corresponds to one state, and every three states are combined into one phoneme. Then the speech will probably be combined into 300 phonemes. However, there are not so many phonemes at all in this speech in reality, so that the obtained status numbers may not be combined in to phonemes at all. [20.]

In fact, each frame is short, so that the state of adjacent frames should be mostly the same. To avoid obtaining messy status numbers, a statistical model named Hidden Markov Model (HMM) is applied to limit the consequence for a pre-defined network. When performing the speech recognition, HMM first establishes a vocalization model for each recognition unit. The state transition probability matrix and output probability matrix will be obtained through long-term training. During recognition, the decision will be made based on the maximum probability in the state transition process. [24.]

2.2.2 Recognition working flow

The voice recognition system is essentially a pattern recognition system, which includes three basic units: feature extraction, pattern matching and reference pattern library [20]. Its basic structure is shown in figure 10.

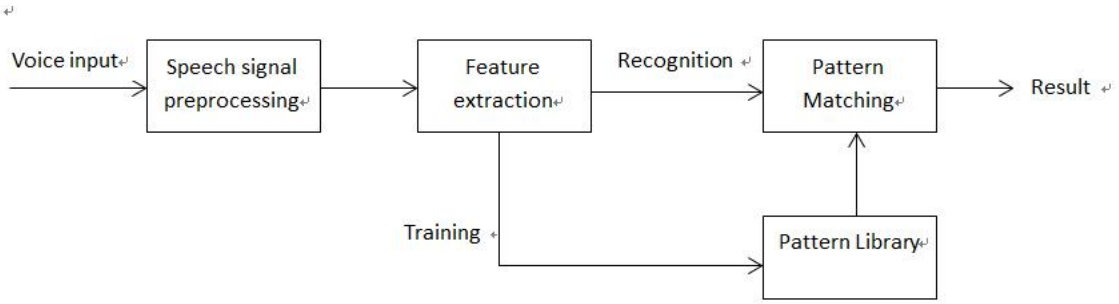

Figure 10. Basic structure of speech recognition. Adapted from How Speech Recognition Works [20].

As figure 10 above demonstrates, firstly, the voice needs to be input and the system will process the input speech. Then the system extracts the features of the speech, and builds the template needed for speech recognition on this basis. During the recognition process, the computer compares the existing voice template features with the features of the input voice signal according to the model of voice recognition. After that, according to the pattern library, it will use its unique strategy to find some optimal templates that match the input voice well. Then the recognition result could be given according to the template definition. Overall, the best recognition result is decided by the choice of features, the quality of the speech model, and the existing template accuracy. [20.]

The system of speech recognition construction process contains two sections: training and recognition. The network is usually offline when training is done. Signal processing and knowledge mining are performed on a large amount of pre-collected speech and language databases to obtain the "acoustic model" and "language model" required for the speech recognition system. The completion of recognition usually needs to be processed online and it recognizes the user's voice automatically in real-time. [25.] The recognition process generally contains two modules: "front-end" and "back-end": the front-end processing refers to processing the original speech before feature extraction to partially eliminate noise and the effects of different speakers. The processed signal can better reflect the essential characteristics of speech. [20.] The function of the "back-end" module is to utilize the trained "acoustic model" and "language model" to execute statistical pattern recognition (also known as decoding") on the feature vector of the user's speech. It is also used to obtain the text information it contains. [25.] In addition, the backend module has an "adaptive" feedback part, which could learn the

user's voice. Therefore, it has the function to "correct" the "acoustic model" and the "speech model" necessary to enhance the recognition accuracy. [20.]

2.2.3 Development history and applications

Voice recognition has been widely used in products in recent years. The most common application is the “voice assistant” on mobile phones, such as Siri, which can recognize people’s commands and make a corresponding reaction. Actually, in the early 1950s, speech research was made by Bell Labs. The earliest isolated word speech recognition system can only recognize ten English numbers through their features of vowels in the pronunciation. After ten years, Martin from RCA Labs proposed a time-regular mechanism, which decreased the impact of inconsistency on recognition score calculation. [26.] At the same period, Vintsyuk from the former Soviet Union proposed dynamic time warping, which effectively solved the similarity measure of two different length speech segments [27].

In the 1970s, three vital technologies, pattern recognition ideas, dynamic programming algorithms, and linear predictive coding, were introduced into speech recognition, which made the isolated word speech recognition system meet the requirement in practice [27]. Then the continuous speech recognition field was considered to be researched with a statistical model based on the Large Vocabulary Continuous Speech Recognition (LVCSR). The most successful system is Harpy, which was built by Lowerre’s team in Carnegie Mellon University in 1976. It could be used to recognize 1011 words effectively and accurately, and it was the first model that used FiniteState Network to reduce the amount of computation and achieve string matching. [26.]

The 1980s was a critical period for the HMM acoustic model and a language model based on N-Grams [28] was implemented into speech recognition, which developed the LVCSR system. The most representative system is the SPHINX system developed by Li Kaifu. [26.] It was the first non-specific continuous speech recognition system that developed based on the Gaussian Mixture Model (GMM), which modeled the observation probability of the speech states [29]. Except for the GMM, the HMM was used to model the timing of speech states. These two models were combined together to be the dominant framework for the speech recognition system during the rest of the

1980s. In the next ten years, the majority of the studies about speech recognition were based on GMM-HMM models. [27.]

In the 1990s, a series of discriminative training criteria for acoustic modes were proposed, such as the Maximum Mutual Information and Minimum Classification Error criteria [26]. This training improved the model performance on the basis of the HMM. At the same period, many speech recognition products appeared based on the existing technologies. For example, the Via-Voice system from IBM, the Whisper system from Microsoft and the DRAGON system are still being used. [30.]

In the 20 century, the speech recognition began to shift from standard reading conversations to more difficult daily communications in English [27]. Deep Neural Networks (DNNs) were used for acoustic modeling of speech and replaced the previous GMM model [31]. At the end of 2011, the researchers applied DNN technology in LVCSR, which greatly reduced the error rate of speech recognition [27]. This DNN-HMM system models every state and it is no longer necessary to assume or estimate the speech distribution data. The splicing of adjacent speech frames contains the time-series structure information of the speech, which significantly improves the classification probability of the state. At the same time**,** DNN also has an outstanding environment learning ability, which can improve the noise and the robustness of people’s accents. [31.]

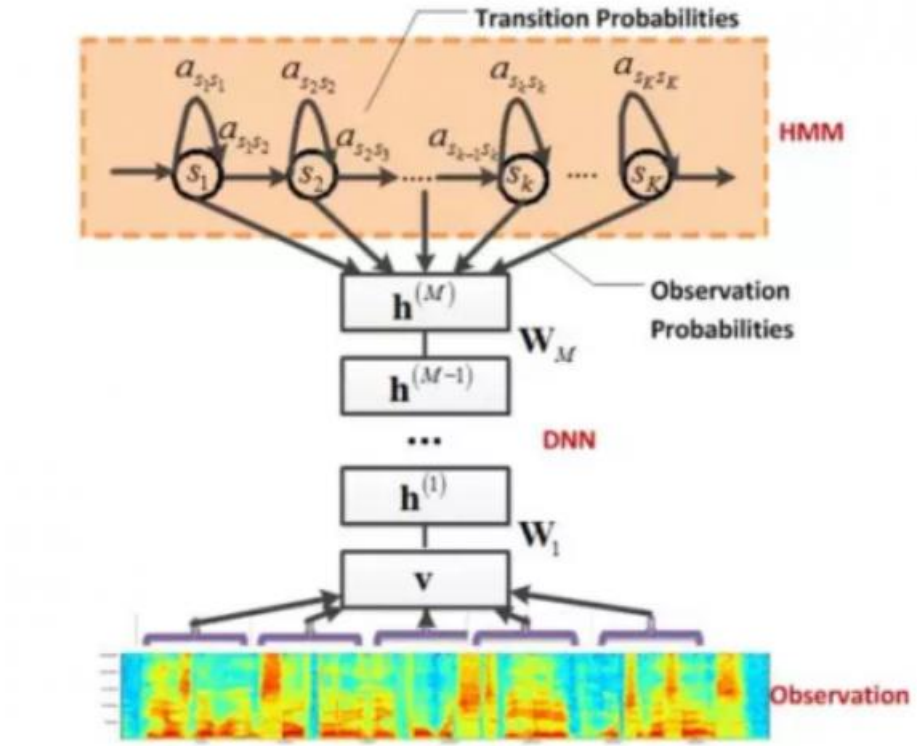

Figure 11. Working principle of DNN-HMM model. Copied from A Deep Learning Approach [32].

Figure 11 above shows that DNN is the state probability corresponding to the input string. Owing to the voice signal being continuous, there are no boundaries between each phoneme, syllable and word. In order to improve these conditions, the researchers proposed a new model, the CNN. [17.] It uses a large number of convolutions to directly model the entire speech signal. With the breakthroughs of recognition technology, some related products such as Siri and Google Assistant appeared before the 2010s. These products could only process some basic command until the creation of Amazon Echo. [30.]

From the perspective of technology and function of speech recognition and even natural language understanding, there is no substantial change compared to Amazon Echo to Siri and other intelligent voice recognition products. The core change of voice recognition technology is that it changed the near-file voice interaction to far-field voice interaction. [30.] Figure 12 shows the architecture of far-field voice interaction recognition systems.

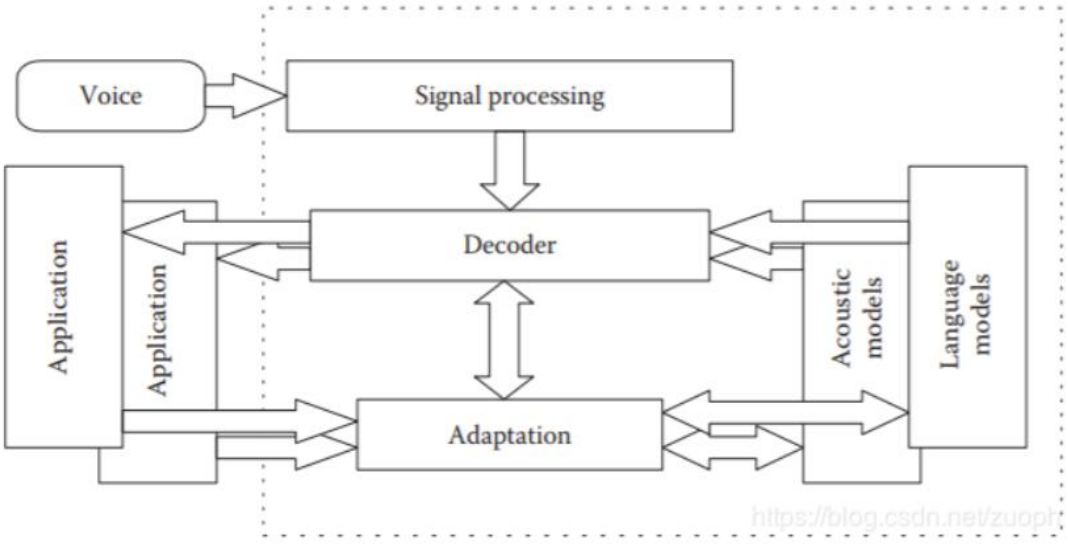

Figure 12. The architecture of far-field voice interaction system. Copied from Can You Hear Me Now? Far- field Voice [33].

Currently, commercial voice recognition systems are designed based on statistical principles [30]. As shown in figure10 above, the acoustic model is used to express variable factors such as acoustics, microphones, and environmental diversity. The language model expresses the combination of words and logical order defined by a linguistic perspective. The application interface with the decoder is used to better adapt the recognition result to other modules in the system. [33.]

In recent years, the development of voice recognition has mainly been based on two directions: one is based on the LVCSR system. For instance, the hospital uses a voice recognition system for electronic medical record entries, and the court clerks share the clerk's work through voice recognition. In these cases, voice recognition systems are mainly used for inputting, querying and recording a large amount of voice information based on a computer platform in a networked state. [34.] Another research direction is to develop the application of voice recognition in miniaturized portable products, such as smart driving assistants, voice assistants on mobile phones, smart toys, and smart home appliances. These application systems are mostly implemented with specific voice recognition chips and integrated circuits using specialized hardware systems. [35.]

3 Tools and implementation

After learning the theory of how the voice recognition works, it can be concluded that the existing IWR products are mainly used in the waste disposal industry. Considering home appliances, the voice controlled household garbage pre-sorting product belongs to miniaturized portable products [21]. This section of the thesis discusses what devices and models were used in the project and how garbage pre-sorting could be carried out based on voice control.

3.1 Tools

This project was developed based on the Arduino platform and M6SE-IDE. One Arduino UNO board, a WEGASUN-M6 voice recognition module with a speaker and four SG90 servos with simulated board garbage bins were used. Figure 13 shows the tools that were used.

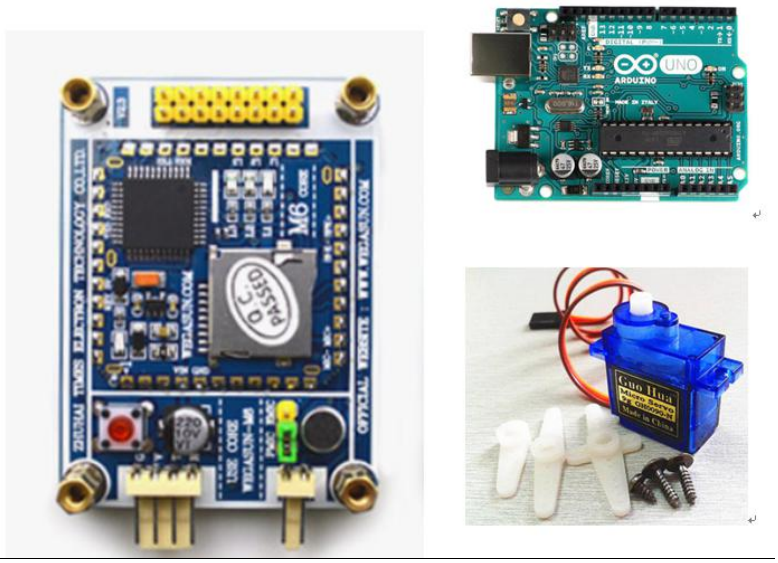

Figure 13. The tools that were used in this project.

The WEGASUN-M6 voice recognition module is a multifunctional module integrating speech recognition, speech synthesis, speech MP3 on demand, the RF function, and the infrared function. It has been implemented, for example, in smart home appliances

and educational robots. [36.] Figure 14 shows the circuit of the WEGASUN-M6 voice recognition module.

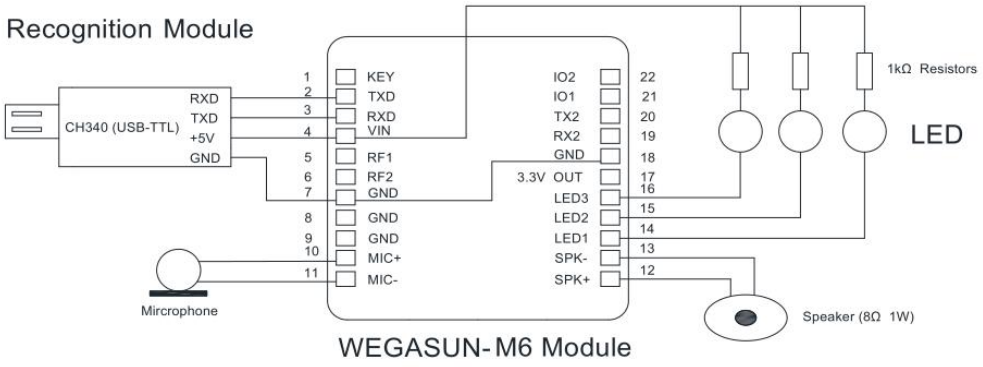

Figure 14. The circuit of the WEGASUN-M6 voice recognition module. Adapted from WEGASUN-M6 [36].

As figure 14 shows, the module contains a built-in speech recognition engine with a microphone [36]. The voice recognition module has its own software to set up and a voice processing package, which makes the implementation easier. In addition, when the system receives the voice input, the corresponding voice reply content can be set in advance, so this module is not required to be used in the network state. The module can be used with Arduino UNO through the TXD and RXD serial ports to control the servos on the garbage bins. [36.]

3.2 Implementation

After confirming the purpose of this project, the voice recognition would be the most important function. The implementation was divided into three parts: voice recognition, servo motion, and combination. At the beginning, Google Cloud Speech API was planned to implement the voice recognition system because it can recognize more than 80 languages worldwide. Google Cloud Speech API is based on DNN, and it can work in a noisy environment or reply in real-time. [37.]

After compiling the working environment, the Google Cloud Speech API was tested with some audio examples and the result shows in figure 15.

Name	↓ Requests	Errors (%)	Latency, median (ms)	Latency, 95% (ms)
Compute Engine API	100	2	210.362	830.123
Cloud Deployment Manager V2 API	22	0	174.763	428.169
Cloud Speech-to-Text API	8	50	262.144	1,677.722
Service Usage API	2	0	262.144	1,992.294
Cloud Resource Manager API	1	0	196.608	255.59

Figure 15. The testing result with Google Cloud Speech API.

As figure 15 shows, it took a long time to receive the reaction when processing an audio file. And the result was not as accurate as expected. One reason for the unexpected latency was the policy limitation for Google services in the author's country. Therefore, the Google Cloud Speech API was given up due to the inaccuracy and the latency.

In order to improve the recognition accuracy and efficiency, it was decided that WEGASUN-M6 would be used in this project. The first step was to wire all the devices correctly. After obtaining the WEGASUN-M6 voice recognition module, it was connected with an external speaker. At the beginning, its functions were tested on its software named M6SE-IDE.

After wiring and connecting correctly, the next step was to set the baud rate, volume, recognition distance, wake-up command and the maximum waiting time for recognition and voice recording. The waiting time was set as 6 seconds for both. The baud rate was set as 9600. In general, the volume range is 1-10, and in the project, it was set to 3. The recognition distance was set as "less than 3 meters". Then the recognition of words and their corresponding replies were set on the platform by using the "@WriteKeywords#" and the "@WriteFlashText#" commands. Most recognition tasks were operated on M6SE-IDE.

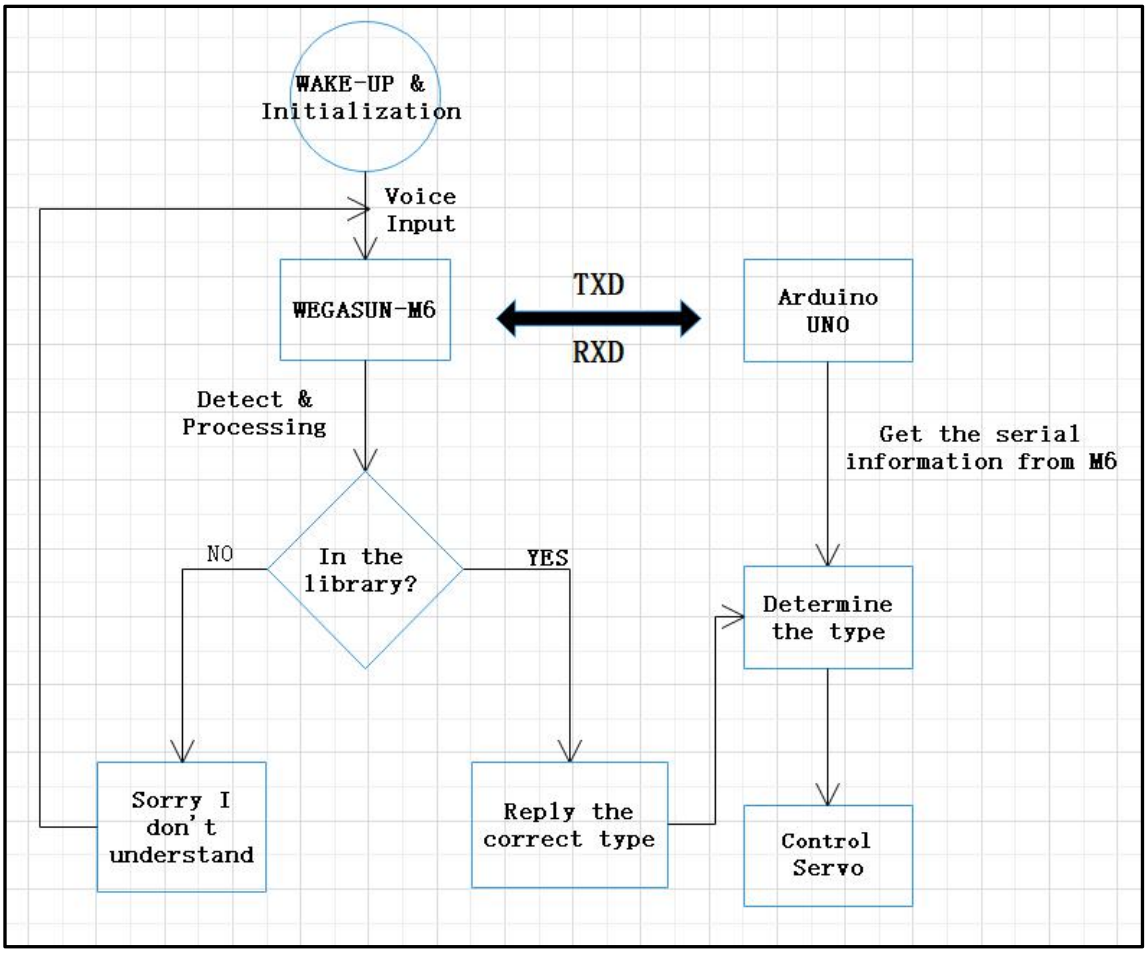

Figure 16. The workflow chart.

Figure 16 above illustrates the working principles for the devices combined together. After receiving the voice input, the sound wave would be compared to the existing library and converted into electrical signals, and then the recognition result would be sent to the Arduino UNO serial ports. The microcomputer would send a driving pulse signal, and under the action of the servo motor driver, Arduino UNO would control the opening and closing of the garbage bin covers according to the input voice signals through the WEGASUN-M6 voice recognition module. If the input command was not in the library, the system would reply “Sorry I don’t understand” and wait for another command.

In this project, all the garbage bins were made of cardboard in the spirit of environmental protection. Garbage was categorized into four main types based on disposal and utilization methods: hazardous waste, kitchen waste, recyclable waste and other waste [38].

Table 1. Four garbage categories with their disposal and utilization methods. Adapted from Garbage classification / Public Design [38].

Garbage Category	Disposal Methods	Utilization Methods
Hazardous Waste	Harmless disposal	
Kitchen Waste	Landfill, anaerobic fermentation, aerobic fertilizer production	Biogas, fertilizer
Other Waste	Incineration	Energy
Recyclable Waste	Reuse	Raw material

Table 1 shows the unique disposal and utilization methods of the four types of garbage after being collected and processed. In this voice controlled pre-sorting method, the recognition system is compatible with both English and Chinese (Mandarin).

Table 2. The implemented household garbage classifications.

Hazardous Waste	Kitchen Waste	Other Waste	Recyclable Waste
Lighter	Orange Peel	Basketball	Can
Band Plaster	Watermelon Peel	Football	Plastic Bottle
Alcohol	Banana Peel	Toilet Paper	Plastic Box
Palette	Apple Peel	Shell	Plastic Bag
Paint	Pea peel	Brick	Plastic Wrap
Expired Capsule Drug	Cucumber	Toilet Bowl	Plastic Comb
Thermometer	Cabbage	Clothes	Plastic Foam
Battery	Fish Bone	Sponge	Fresh-keeping Bag
Expired Pill	Strawberry	Bathtub	Glass Pot
Fluorescent Light	Blueberry	Peanut Shell	Glass Bowl
Medical Cotton Swab	Raspberry	Disposable Plate	Glass Plate
Gas Tank	Eggplant	Disposable Bowl	Glass Cup
Medical Gloves	Onion	Disposable Chopsticks	Glass Bottle
Pesticide	Potato	Disposable Knife	Glass Cup
Watercolor Pen	Tomato	Disposable Fork	Iron Pot
Pesticide Bottle	Biscuit	Disposable Spoon	Iron Wok
Mercury Thermometer	Cake	Disposable Cup	Iron Spoon
Perfume Bottle	Ketchup	Tire	Iron Knife
Disinfectant	Eggshell	Broken Vase	Iron Fork
Glow Stick	Corn	Broken Pot	Iron Nail

Expired Cosmetic	Mushroom	Broom	Tin Foil
Hair Spray	Shrimp Shell	Napkin	Induction Cooker
Paint Bucket	Pepper	Cigarette Butt	Rice Cooker
Medical Gauze	Chili	Dust	Kettle
Printer Cartridges	Bread	Mud	Broken Lock
Herbicide	Cereal	Wooden Comb	Key
Nail Polish	Nut	Pet Feces	Scissors
Hair Dye	Condiment	Hair	Shaver
LED Screen	Lettuce	Crayon	Metal Fish Hook
Led Light	Leftovers	Sanitary Napkin	Cable
Sphygmomanometer	Withered Flower	Stone	Headphone
Desiccant	Tea Leaves	Finger Nail	Wooden Spatula
	Coffee Grounds	Glue	Wooden Chopsticks
	Leaf	Lipstick	Paper Bag
	Rice	Mirror	Paper
		Facial Mask	Newspaper
		Cutting Board	Magazine
			Paper Box
			Carton
			Leaflets
			Toothbrush
			Old School Bag
			Old Toy
			Tire
			Curtain
			Pillow

The household garbage was implemented into the WEGASUN-M6 recognition library as Table 2 above shows. The listed garbage items included common products people have at home. Most of the household garbage seen in the table could be recycled and reused after being sorted out.

The programming code was written on the Arduino UNO based on the Arduino programming platform. The example code is in Listing 1 below:

```
void loop()
{
    input = Serial.read();//Read the value received by the serial port and
store it in a
    if (1 == input)//If the system detects “Turn on the light” instruction,
the light will be on.
    {
      digitalWrite(13, HIGH);// Turn on the light
```

```
}
if(3 == input)//Servo 1 is for controlling the lid of hazardous waste bin
{
    myservo1.write(135); // Lid opened
    delay(5); // Waiting for turning to the specific angle
    delay(3500);   //Waiting for the garbage
    myservo1.write(45); // Lid closed
    delay(5); // Waiting for turning to the specific angle
}
```

Listing 1. Programming code in Arduino for light and servos control

Listing 1 shows the example code of the light on Arduino UNO board control and servo control for the hazardous waste simulated trash bin. The “servo.h” library was directly used in this project. The code for the light on/off is similar and the code for controlling the four servos on simulated garbage bins is similar as well.

Table 3. The servo angle corresponding to the duration of the pulse width in one cycle (20ms). Adapted from Servo Motor: Types and Working Principle Explained [39].

Pulse width accounts for the entire period(20ms)	Rotation
0.5ms	0 degree
1ms	45 degrees
1.5ms	90 degrees
2ms	135 degrees
2.5ms	180 degrees

Table 3 shows the servo angle corresponding to the pulse width for the entire 20ms period. The control of the servos requires the MCU to generate a 20ms pulse signal to control the angle of the servos with a high level of 0.5ms to 2.5ms. [39.] In this project, the servo rotation angle for opening the lid was set as 135 degrees and for closing the lid, it was set as 45 degrees.

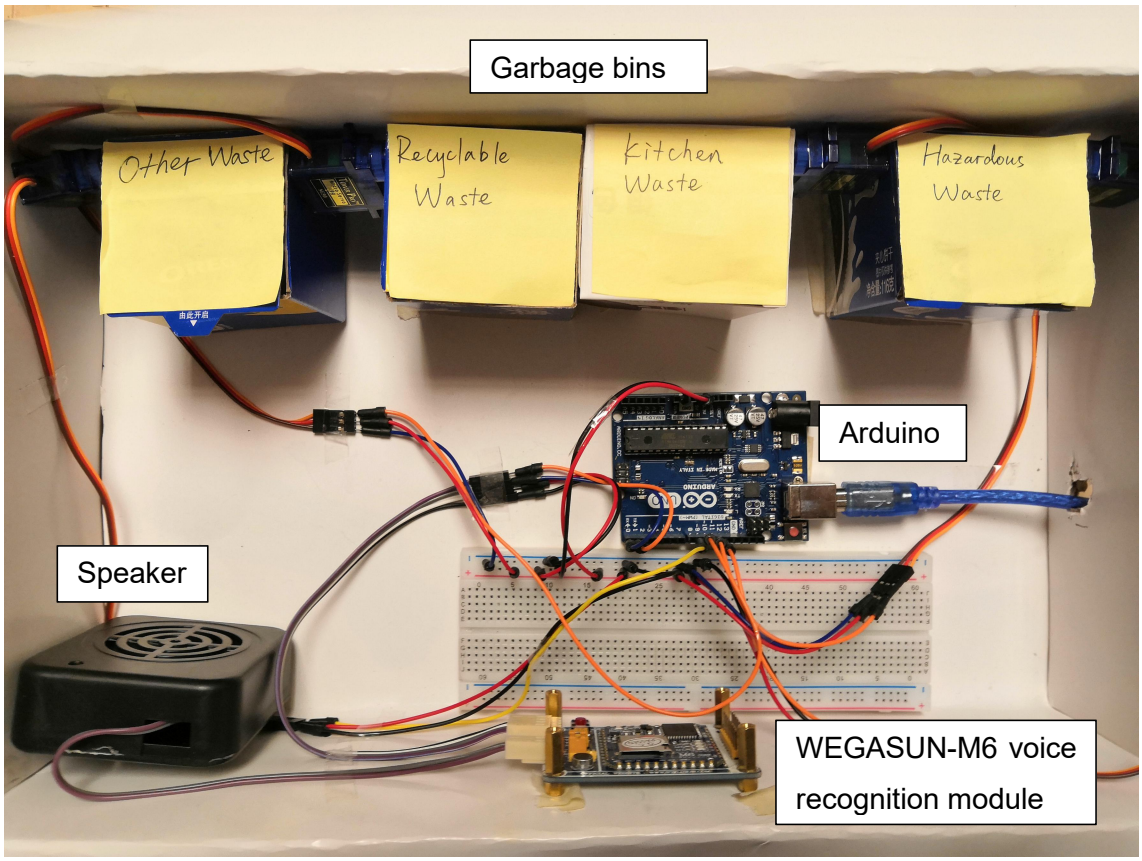

Figure 17. Demonstrating the setup of the project.

Figure 17 shows the final design of the voice controlled garbage bins. One external speaker was connected because the sound of the internal speaker of the WEGASUN-M6 voice recognition module was too low.

4 Testing result

To start the system, the wake-up command was set as “Hello, Lisa”. After receiving its reply “Hi”, the system has been initialized and started to work. The recognition system could detect the type of garbage when a person said “Throw XXX” (where XXX stands for an item to be thrown away), and it would reply what type their garbage is. At the same time, the cover of the garbage bin would open for six seconds waiting for the garbage, and then it would close.

In order to test and verify the working ability of the voice controlled garbage pre-sorting system under normal working conditions, ten samples of garbage (banana peel, milk carton, crab shell, lighter, newspaper, medical mask, lipstick, hair, gas tank, and fish bone) were tested with different distances to determine the recognition accuracy. The voice input used the same mobile phone with the same volume and the test was carried out in the same quiet room, so that the distance was the only parameter in this verification.

Table 4. Voice recognition testing with different distances.

Testing number	Distance (m)	Testing count (times)	Error rate (%)	Correct rate (%)
1	0.1	100	0	100
2	0.5	100	0	100
3	1.0	100	0	100
4	1.5	100	3	97
5	2.0	100	5	95
6	2.5	100	9	91

As Table 4 illustrates, the smart garbage sorting system could detect the voice input accurately when the distance was between 0.1 meter and 1.0 meter. If the distance was greater than 1.0 meter, the recognition would not be totally accurate. In general, the most appropriate placement distance is about 1.0 meter to avoid the spread of bacteria from the garbage in real conditions.

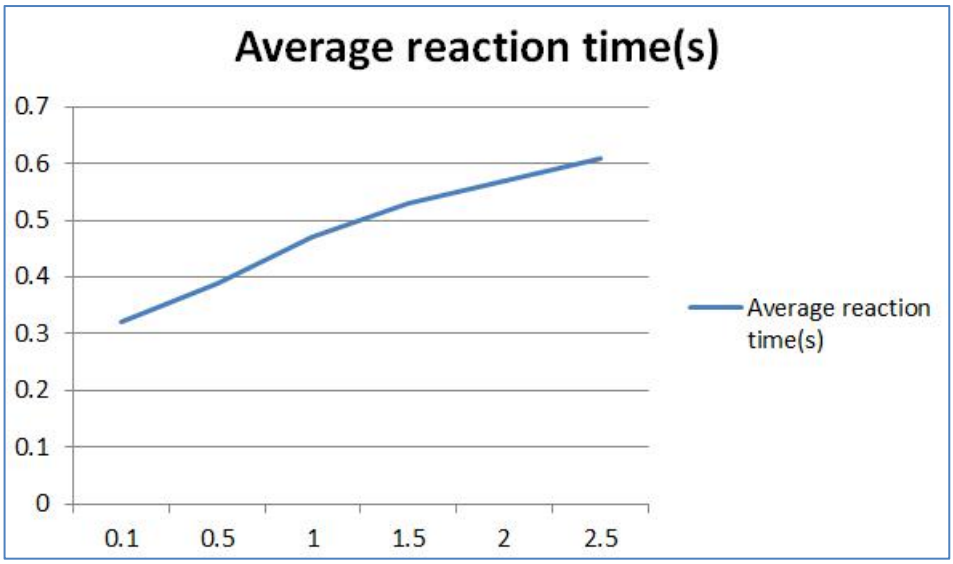

Figure 18. Line charts of average reaction time in different distances.

As figure 18 above suggests, the reaction time was counted as well with the 100 different garbage samples. The line chart shows that the average reaction time of the system was slightly longer as the distance increased. These results may have been caused by the noise when the recognition system received the voice input.

In order to observe the differences between the processing of this system and human knowledge about garbage classification, a survey was conducted with 50 people. The survey included 100 different kinds of garbage (see Appendix 1) and the requirement was to sort them into the correct types. The pie charts show the accuracy of garbage sorting of 50 people and the voice recognition system:

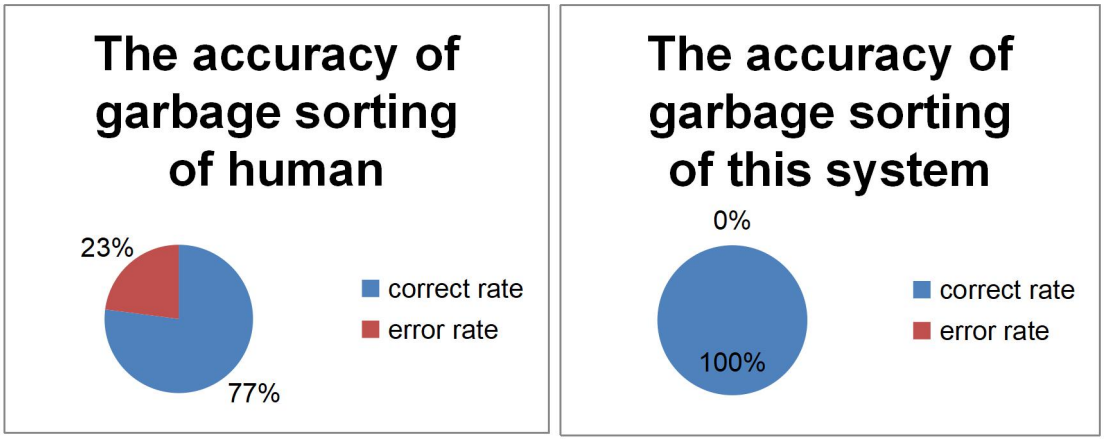

Figure 19. Pie charts of accuracy of garbage sorting.

Figure 19 compares the average correct rate of 50 people in the investigation with the rate of the voice recognition system. Looking at the information in more detail, the pie charts indicate that people who participated in the survey were able to sort out only 77% of the garbage correctly. With the same garbage, the voice recognition system s accurately sorted out all the garbage when the testing distance was 0.1 meter.

5 Conclusion

In this final year project, the idea of the combination of voice recognition and ordinary garbage bins was planned and implemented successfully in both English and Chinese (Mandarin). An Arduino UNO board and a WEGASUN-M6 voice recognition module were used based on the Arduino platform and M6SE-IDE software. The household voice recognition garbage pre-sorting system could detect and recognize garbage type when saying the name of the garbage item to the system at a fixed distance. According to the tests, the input garbage type could be recognized 100% correctly when the test person was standing in front of the garbage bins at a distance of less than one meter.

Overall, this thesis introduces the basic working principles of voice recognition technology and its theoretical basis, which could be useful to ICT students who want to work on the integration of the Internet of Things and voice recognition. In addition, the theory of garbage classification could provide inspiration and reference for the development of related products for garbage classification.

Although the method of implementing the voice recognition system changed from Google Cloud Speech API to the WEGASUN-M6, the pre-sorting garbage processing system based on voice control still met the initial goal set for it. However, some functions of the pre-sorting system were not achieved in this project. For example, background noise may affect the accuracy of voice input. Therefore, there are high requirements for the recognition environment in which the system operates. In addition, due to the diversity of garbage types, it would become difficult to cover all kinds of garbage if this project was implemented in real life. Therefore, only four main garbage types were used in this project. Another disadvantage is that the input data could not be stored in the system, so it could not be combined with a garbage processing plant for better garbage sorting.

For further improvement, a database for the voice recognition system could be implemented. When people speak to the system, the data could be stored in the database. The data in the database could be analyzed and sent to a waste disposal plant to help the plant know how many employees need to be assigned, what size

garbage bins are needed and how often the employees should collect and deal with the garbage. In addition, the system could be processed with noise reduction technologies through filtering the voice input. It could make the recognition more accurate.

Another function that could be implemented is that infrared sensors could be added on the body of the garbage bin, so that it would be possible to detect whether the user is still in the room to connect or disconnect the battery of the voice controlled garbage pre-sorting system, which could reduce power consumption.

The voice controlled household garbage pre-sorting product makes it possible to avoid all kinds of garbage ending up in the same bin because of people's ignorance, which has a positive impact on the environment. In addition, the product allows people to open the trash lid without touching it when holding things in their hands. From a broader perspective, the product improves the efficiency of garbage sorting and has a positive impact on the environment.

References

1 Ayer David, Merino Valeria. Green Cities. The Invisible Plastic Particles in Our Drinking Water [online]. 23 March 2018. URL: https://www.earthday.org/the-invisible-plastic-particles-in-our-drinking-water/. Accessed 19 April 2020.

2 The World Bank. What A Waste 2.0. A Global Snapshot of Solid Waste Management to 2050 [online]. URL: https://datatopics.worldbank.org/what-a-waste/trends_in_solid_waste_management.html. Accessed 19 April 2020.

3 Xie Echo. Can China Sort Its Household Waste Recycling Problem By 2020? [online]. 11 April 2018. URL: https://www.scmp.com/news/china/politics/article/3015647/can-china-sort-its-household-waste-recycling-problem-2020. Accessed 12 April 2020.

4 Liu Zhen, Yu Xiaodan, Li Shanshan, Liu Jingjing. If the Garbage Is Not Sorted, How Many “Treasures” Are Thrown Every Day (Chinese). CCTV [online]. 05 July 2019. URL: http://news.cctv.com/2019/07/05/ARTIqlaFECryclkilidnHvKi190705.shtml. Accessed 19 April 2020.

5 Xue Yanyan, Wen Zongguo, Bressers Hans, Ai Ning. Can Intelligent Collection Integrate Informal Sectors for Urban Resource Recycling in China [online]. 15 October 2018. URL: https://www.alliedvision.com/en/blog/detail/news/embedded-vision-what-is-the-real-price-of-image-processing.html. Accessed 8 April 2020.

6 Kaliampakos Dimitrios, Benardos Andreas. 2013. Underground Solutions for Urban Waste Management: Status and Perspectives. In: National Technical University of Athens Lab. Of Mining and Environmental Technology. pp. 11-12. URL: https://www.iswa.org/index.php?eID=tx_iswaknowledgebase_download&documentUid=3157. Accessed 10, March, 2020.

7 Ricker Thomas. The Automatic Trash Can Is Here. The Verge [online]. 31 December 2019. URL: https://ifworlddesignguide.com/entry/155085-garbage-classification. Accessed 19 April 2020.

8 Tomra. Reverse Vending 101: A Beginner's Guide [online]. 15 August 2017. URL: https://newsroom.tomra.com/reverse-vending-101-a-beginners-guide/. Accessed 13 March 2020.

9 Optics 4 Kids. Scanner [online]. URL: https://www.optics4kids.org/what-is-optics/terms/terms/s/scanner. Accessed 13 March 2020. Accessed 13 March 2020.

10 Campbell QP. Le Roux M, Morgan L. Winnowing as a dry separation method for fine coal. In: North-West University, South Africa: School of Chemical and Engineering [online]. 2018. URL: https://tu-freiberg.de/sites/default/files/media/institut-fuer-energieverfahrenstechnik-143/publikationen/36-freiberg_coal_conference_2018_poster_winnowing.pdf. Accessed 13 March 2020.

11 Darlina Md. Naim, Mardiana Idayu Ahmad. 01 January 2015. Natural Resources And Technology (Penerbit USM). Electronic book. Malaysia: Penerbit Universiti Sains Malaysia (Universiti Sains Malaysia Press). URL: https://books.google.com.tw/books?id=MhaWCwAAQBAJ&pg=PT31&lpg=PT31&dq=traditional+winnowing+technology+has+been+used+in&source=bl&ots=dabg6X0_Hy&sig=ACfU3U0qQ_tXqjDJ6CXUMxdont1DX6TYsw&hl=en&sa=X&ved=2ahUKEwjKnu2F-p3oAhXGc94KHfNyBXAQ6AEwDXoECAEQAQ#v=onepage&q=traditional%20winnowing%20technology%20has%20been%20used%20in&f=false. Accessed 13 March 2020.

12 Khan Academy, What Is 2D Projectile Motion [online]. URL: https://www.khanacademy.org/science/physics/two-dimensional-motion/two-dimensional-projectile-mot/a/what-is-2d-projectile-motion. Accessed 13 March 2020.

13 Yuan Ye-lin, He Song, Liu Zhong-xiao, Feng Yang, Zhang Xiang, Jiang Shi-ming. Design of Intelligent Garbage Sorting Model Based on Winnowing [online]. 23 October 2012. URL: https://www.doc88.com/p-2962078633319.html. Accessed 13 March 2020.

14 Yang Mindy, Thung Gary. Classification of Trash for Recyclability Status [online]. URL: http://cs229.stanford.edu/proj2016/report/ThungYang-ClassificationOfTrashForRecyclabilityStatus-report.pdf. Accessed 13 March 2020.

15 Chu Yinghao, Huang Chen, Xie Xiaodan, Tan Bohai, Kamal Shyam, Xiong Xiaogang. Multilayer Hybrid Deep-Learning Method for Waste Classification and Recycling [online]. 01 November 2018. URL: https://www.hindawi.com/journals/cin/2018/5060857/. Accessed 13 March 2020.

16 Özkaya Umut, Seyfi Levent. Fine-Tuning Models Comparisons on Garbage Classification for Recyclability [online]. 15 August 2017. URL: https://arxiv.org/pdf/1908.04393.pdf. Accessed 13 March 2020.

17 Saha Sumit. A Comprehensive Guide to Convolutional Neural Networks-the ELI5 way [online]. 16 December 2018. URL: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. Accessed 17 March 2020.

18 Arunava. Convolutional Neural Networks. An Introduction to Convolutional Neural Networks [online]. 26 December 2018. URL: https://towardsdatascience.com/convolutional-neural-network-17fb77e76c05. Accessed 17 March 2020.

19 Zalewski Paul Maria. Embedded Vision: What Is the Real Price of Image Processing [online]. 20 March 2018. URL: https://www.alliedvision.com/en/blog/detail/news/embedded-vision-what-is-the-real-price-of-image-processing.html. Accessed 8 April 2020.

20 Grabianowski ED. How Speech Recognition Works [online]. URL: https://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speech-recognition1.htm. Accessed 20 March 2020.

21 Fearn Nicholas, Tuener Brian. Best Speech to Text Software in 2020: Free, Paid and Online Voice Recognition Apps and Services [online]. 05 February 2019. URL: https://ifworlddesignguide.com/entry/155085-garbage-classification. Accessed 19 April 2020.

22 Hui Jonathan. Speech Recognition-Feature Extraction MFCC&PLP [online]. 29 August 2019. URL: https://medium.com/@jonathan_hui/speech-recognition-feature-extraction-mfcc-plp-5455f5a69dd9. Accessed 22 March 2020.

23 Lenzo Kevin. The CMU Pronouncing Dictionary [online]. URL: http://www.speech.cs.cmu.edu/cgi-bin/cmudict. Accessed 23 March 2020.

24 Tomer Amit. Introduction to Hidden Markov Models [online]. 8 June 2019. URL: https://towardsdatascience.com/introduction-to-hidden-markov-models-cd2c93e6b781. Accessed 23 March 2020.

25 Vyas Krunal. Everything You Need to Know about Voice Recognition Technology [online]. 06 March 2018. URL: https://dzone.com/articles/everything-you-need-to-know-about-voice-recognitio. Accessed 20 March 2020.

26 Donges Niklas. A Brief History of ASR: Automatic Speech Recognition [online]. 31 March 2019. URL: https://towardsdatascience.com/a-brief-history-of-asr-automatic-speech-recognition-95de6c014187. Accessed 22 March 2020.

27 Juang B. H., Rabiner Lawrence R. Automatic Speech Recognition – Abrief History of the Technology Development [online]. 10 August 2004. URL: https://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf. Accessed 23 March 2020.

28 Kapadia Shashank. Language Models: N-Gram, ASstep into Statistical Language Modeling [online]. URL: https://towardsdatascience.com/introduction-to-language-models-n-gram-e323081503d9. Accessed 8 April 2020.

29 Oscar Contreras Carrasco. Gaussian Mixture Models Explained, From intuition to implementation [online]. 3 June 2019. URL: https://towardsdatascience.com/gaussian-mixture-models-explained-6986aaf5a95. Accessed 23 March 2020.

30 Boyd Clark. Speech Recognition Technology: The Past, Present, and Future [online]. 10 January 2018. URL: https://medium.com/swlh/the-past-present-and-future-of-speech-recognition-technology-cf13c179aaf. Accessed 23 March 2020.

31 Hinton Geoffrey, Deng Li, Yu Dong, Dahl George, Mohamed Abdel-rahman, Jaitly Navdeep, Senior Andrew, Vanhoucke Vincent, Nguyen Patrick, Sainath Tara, and Kingsbury Brian. Deep Neural Networks for Acoustic Modeling in Speech Recognition [online]. November 2012. URL: https://www.cs.toronto.edu/~hinton/absps/DNN-2012-proof.pdf. Accessed 23 March 2020.

32 Yu Dong, Deng Li. Automatic Speech Recognition. A Deep Learning Approach. [online] 12 November 2014. URL: https://link.springer.com/content/pdf/10.1007%2F978-1-4471-5779-3.pdf. Accessed 22 March 2020.

33 Lu Jerry. Can You Hear Me Now? Far- field Voice [online]. 02 August 2017. URL: https://www.cs.toronto.edu/~hinton/absps/DNN-2012-proof.pdf. Accessed 23 March 2020.

34 Saon George, Chien Jen-Tzung. Large-Vocabulary Continuous Speech Recognition: A Look at Some Recent Advances [online]. 18 October 2012. URL: https://ieeexplore.ieee.org/document/6296522. Accessed 8 April 2020.

35 Liu Hong, Qian Yanmin, Liu Jia. 2011. English Speech Recognition System on Chip. In: Tsinghua Science and Technology 16(1); 95-99. 10 June 2010. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6077942. Accessed 8 April 2020.

36 WEGASUN-M6. Voice Recognition Module Instruction [online]. URL: http://www.ankux.com/catalog/voice-recognition-module-wegasun-m6-imported-101071006301109721.html. Accessed 27 March 2020.

37 Google Cloud. Speech-to-Text [online]. URL: https://cloud.google.com/speech-to-text. Accessed 19 April 2020.

38 Li Hao. World Design Guide. Garbage classification / Public Design [online]. URL: https://ifworlddesignguide.com/entry/155085-garbage-classification. Accessed 8 April 2020.

39 Goel Askhat, Servo Motor: Types and Working Principle Explained. Engineering. 06 November 2018 [online]. URL: https://engineering.eckovation.com/servo-motor-types-working-principle-explained/. Accessed 20 April 2020.

100 multiple choices about garbage classification with 50 random people

A. Hazardous waste
B. Kitchen waste
C. Other waste
D. Recyclable waste

Hazardous waste	Kitchen waste	Other waste	Recyclable waste

1. Band plaster
2. Eggplant
3. Onion
4. Expired pill
5. Battery
6. Perfume bottle
7. Basketball
8. Chilli
9. Newspaper
10. Soap
11. Carton
12. Paper cup
13. Plastic comb
14. Wooden comb
15. Toilet paper
16. Banana peel
17. Watercolour pen
18. Hair spray
19. Sponge
20. Milk carton
21. Old clip
22. Cans
23. Glow stick
24. Old toy
25. Broom
26. Brick
27. Eggshell
28. Gas tank
29. Pesticide bottle
30. Medical gauze
31. Fish bone
32. Shrimp shell
33. Peanut shell
34. Doll
35. School Bag
36. Toothpaste package
37. Crayon
38. Scissors
39. Pen
40. Lighter
41. Key
42. Plastic bag
43. Lipstick
44. Glue
45. Shampoo
46. Ink
47. Flashlight
48. Cable
49. Bread
50. Light bulb
51. Plate
52. Book
53. Hanger
54. Curtain
55. Pillow
56. Knife
57. Suitcase
58. Medical gloves
59. Medical masks
60. Cigarette butt
61. Leaf
62. Glass
63. Paint bucket
64. Dust
65. Pet feces
66. Disposable chopsticks
67. Mercury thermometer
68. Umbrella
69. Carpet
70. Withered plant
71. Staples
72. Hair
73. Empty alcohol bottle
74. Pen refill
75. Iron pot
76. Ketchup
77. Makeup brush
78. Broom
79. Crab shell
80. Bathtub
81. Orange peel
82. Flower pot
83. Corn
84. Cutting board
85. Pear
86. Mirror
87. Football
88. Brush
89. Palette
90. Syringe
91. Shell
92. Fish tank
93. Tire
94. Walnut
95. Ceramic bowl
96. Iron bowl
97. Stone
98. Button
99. Advertising flyer
100. Helmet