MINISTRY OF EDUCATION AND SCIENCE OF THE REPUBLIC OF KAZAKHSTAN
Non-profit join - stock corporation
ALMATY UNIVERSITY OF POWER ENGINEERING AND TELECOMMUNICATION named after G. Daukeev
Department of *Electronics and robotics*

«Allowed to defence»
The head of department of «Electronics and robotics»
Chigambayev T.O. c.t.s., associate professor
(Full name, academic degree, rank)

____«____» ________________2020 year
(sign)

DEGREE PROJECT

On the topic: *Development of a system of intelligent servicing visitors*
Done by: *Prmashov Yerkebulan* *PSa-16-3*
(Surname and initials of a student) (group)

Specialty *5B071600 Instrumentation Engineering*

Research supervisor *Balbaev G.K. Ph.D, Docent*
(Surname, academic degree, rank)

______«____» ________________2020 year
(sign)

Consultants
on economic part: *Tuzelbayev B.I. Ph.D, associate professor*
(Surname, academic degree, rank)

______«____» ________________2020 year
(sign)

on life and environmental safety part: *Begimbetova A.S. Ph.D, senior lecturer*
(Surname, academic degree, rank)
______«____» ________________2020 year
(sign)

Compliance supervisor: *Fazylova A.R.senior lecturer*
(Surname, academic degree, rank)

______«____» ________________2020 year
(sign)

Almaty, 2020

MINISTRY OF EDUCATION AND SCIENCE OF THE REPUBLIC OF KAZAKHSTAN
Non-profit join - stock corporation
ALMATY UNIVERSITY OF POWER ENGINEERING AND TELECOMMUNICATION named after G. Daukeev

Institute of *space engineering and telecommunications*
Department of *Electronics and robotics*
Specialty *5B071600 Instrumentation Engineering*

ASSIGNMENT
for execution of degree project

Student *Prmashov Yerkebulan Yerboluly*
(Full name)
Topic of the work *Development of a system of intelligent servicing visitors*

Approved by the order of the rector № 155 from « 23 » october 2020 y.

Deadline of the finished work « 8 » june 2020 y.

Initial data required parameters of the results and initial data:

1. *PCduino microcontroller*
2. *TP-link Router*
3. *Gboy Webcam*
4. *Asus laptop*

List of issues to be developed in a degree project or a summary:
1. *Consider the control of a prosthesis based on an image recognition system*
2. *The servicing of visitors without contacting, with voice contorol.*
3. *Programming part*
4. *Development of life safety measures*
5. *Economic justification of the project*

List of graphical material (with precise indication of mandatory drawings);
This degree project contains 42 figures and tables

Recommended basic literature:
1. *Lifeng Zhao and Jingfeng Liu "Introduction to pcDuino"*
2. *Stankevich L. A., Yurevich E. I. Artificial intelligence and artificial intelligence in robotics: textbook. Saint Petersburg: Polytechnic University Press, 2012.*
3. *CMU Sphinx. Open Basic concepts of speech. [Electronic resource]. - Access mode: http://cmusphinx.sourceforge.net/wiki/tutorialconcepts*

Consultants for work with indication of the relevant section

Section	Consultant	Date	Sign
Life safety	*Begimbetova A.S*	*16.05.2020*	
Economic part	*Tuzelbayev B.I*	*18.05.2020*	

SCHEDULE
Of degree project preparation

№	Title of section, list of issues to be developed	Deadline for submission to instructor	Note
1	*Theoretical part*	*29.02*	
2	*Engineering part*	*10.03*	
3	*Program part*	*01.05*	
4	*Life safety*	*08.05*	
5	*Economic part*	*27.05*	
6	*Conclusion*	*06.06*	

Date of issue of the assignment ____*« 20 »   01   2020 year*

The head of department _________*Chigambayev T.O.*________________________
(sign) (Surname and initials)

Supervisor: ____________________*Fazylova A.R.senior lecturer*_______________
(sign) (Surname and initials)

The assignment for execution is accepted by: _____________*Prmashov Y.*_______
(sign) (Surname and initials)

Аңдатпа

Бұл жұмыста дауыспен басқару жүйелерінде пайдаланылатын технологиялар туралы баяндалған. ARM-микрокомпьютер мен ESP8266 Wi-Fi платаның сипаттамасы берілген, сондай-ақ осы құрылғылар негізінде қонақтарға арналған, зияткерлік қызмет жүйесін дамыту сипатталған.

Сонымен қатар, осы жұмыста өмір қауіпсіздігі мәселелері қаралды, бизнес-жоспар мен жобаның өзін-өзі ақтау мерзімі есептелді.

Аннотация

В данной дипломной работе рассмотрены вопросы о технологиях, используемых в системах голосового управления. Описаны одноплатный ARM-микрокомпьютер и Wi-Fi плата на микросхеме ESP8266, и разработана система интеллектуального обслуживания посетителей общепита

Рассмотрены вопросы безопасности жизнедеятельности, составлен бизнес-план и рассчитан срок окупаемости разработанного проекта.

Summary

In this work concepts and technologies of voice recognition systems were defined, also this work includes descriptions of ARM-minicomputer and ESP 8266 Wi-Fi board. In this diploma has made intellectual voice control system using these devices and development of a system of intelligent servicing visitors.

Activity safety issues are considered, the business plan is made and the price of development of the project pays off.

Content

Introduction

The problem of creating means of verbal dialogue between a person and machines is one of the most pressing problems of cybernetics, computer science, and computer technology [1], therefore, the aim of this thesis is to develop a prototype voice control system. To achieve the goal, several tasks were set:

- selection of a speech recognition software package;
- selection of audio capture device;
- study of the principle of operation of ARM-minicomputers;
- Installation and configuration of the selected speech recognition program.

The first chapter discusses various open source audio recognition systems. Among them, the most suitable one was selected according to the criteria described in the work. The chapter also describes the various types of sound capture devices and their features. The second chapter describes the characteristics, shows the structural diagram of the PCDuino1 ARM minicomputer, which is the basis of the voice control system. The third chapter describes the installation, configuration and launch of the sphinx software package, which is an engine for recognizing voice commands.

1 Technological part

1.1 State of the problem of automatic speech recognition

Attempts to teach computers to communicate with people using a natural voice interface have been made since the first years of the history of computer technology. In the course of many years of research, it was found out that it is necessary to involve not only programmers, but also specialists in linguistics (linguistics), radio engineers, mathematicians, biologists, and even psychologists in solving the problem.

In fact, to create a speech recognition system you need to solve many problems.

First of all, you need to convert air vibrations into electrical signals using a microphone, filtering out noise and noise.

Further, somehow, the signal must be presented in digital form, available for processing using a computer (digitized). There are different possibilities: you can enter information about the amplitude of the audio signal into the computer, or you can analyze the spectral composition of the signal by extracting a set of fundamental frequencies from the signal. This information can be combined [2].

Linguistics specialists are necessary to study the principles of speech construction, highlighting those elements of speech that the program should recognize in the input stream - phonemes, morphemes, syllables, words, etc. Linguistics studies such aspects of speech, the account of which is simply necessary when creating speech recognition and synthesis systems.

It must be said that extracting linguistic constructions from the speech stream is not an easy task. Only a child who learns to speak or read spells words in syllables, separating syllables and words with pauses. In real life, speech is a continuous stream of sounds. In the process of forming continuous speech, sounds corresponding to the same letters can change when connected to other sounds.

Sounds of continuous speech contain a constantly changing spectrum of harmonic frequencies, as well as noise. The volume and pace of speech are also constantly changing. Moreover, the same phrase spoken by different people, or even by one person who is in different mental states, can have a different spectral-temporal color. This makes it very difficult to create universal recognition systems that "understand" the speech of different people.

In order to distinguish linguistic constructions from the digitized sound, various mathematical methods are used in combination with special computer equipment, such as, for example, hardware or software neural networks. Throughout the history of speech recognition systems, these methods have been constantly changing. At the same time, some methods died out due to their inefficiency, while others were developed and improved.

Biological research helps to study the mechanisms of education and speech recognition that are used by humans (and maybe animals). Trying to solve the problem of the speech interface, many researchers are trying to simulate the work of human organs, such as the speech tract, ear and neural systems of the brain.

And finally, when solving the problem of creating a speech interface of computer systems, psychologists are needed, since without taking into account the psychological characteristics of a person it is impossible to create a truly convenient speech interface.

Communication problems between humans and computers go beyond purely technical sciences. This happens because modern computers try to communicate with people using the same methods that people use to communicate with each other. And just as it is difficult for a person to communicate with patients suffering from mental disorders, it will also be difficult for him to communicate with a computer program that demonstrates inappropriate mental behavior.

1.2 Basic concepts of speech recognition

Speech is a complex phenomenon. People do not understand how it arises and is perceived. It is believed that speech consists of words, and each word is made of backgrounds. In reality, everything is different. Speech is a dynamic process without explicit parts. Using the audio editor, you can see the waveform when recording voice - Figure 1.1.

Figure 1.1 - The waveform when recording voice

All present day descriptions of speech are truly probabilistic, that is, there are no obvious boundaries among distinctive words. Speech recognition may not be absolutely true. This method is unusual for software builders who typically paintings with deterministic systems.

Speech is a continuous sound circulate where stable states are replaced by means of dynamic ones and vice versa. In this series of states, some may additionally define more or less similar units of speech - backgrounds. It is assumed that phrases are composed of heritage, however this is not absolutely true. The acoustic parameters of the sound wave of one history may additionally vary depending on diverse factors - from the context of the background, from the speaker’s speech style, etc., consequently, the backgrounds can also sound extraordinary from the “canonical” sound. Because transitions between backgrounds are more informative, such a idea as diphons is used. Diphons - a part of the background between two consecutive backgrounds. The idea of part backgrounds is also used. Often you can define three or greater parts of backgrounds.

The most commonly distinguished 3-part backgrounds. The first part of the heritage depends on the preceding heritage, the second one element is noticeably stable, and the third a part of the heritage is determined by using the following background, therefore in speech recognition 3 elements of the background are most customarily decided.

Backgrounds are once in a while viewed in context. Some backgrounds in the context can be taken into consideration as trifons or at the same time as pentaphones. For example, the heritage “u” inside the word “bad” follows after the history “b” and precedes the background “d”, and inside the word “ban” follows after the historical past “b” and precedes the background “n”. In these words, the history "u" sounds differently [3].

In calculations, it's far extra beneficial to determine the parts of the trifons than the whole. To do this, create a detector of the start of the trifon. A wide form of sound detectors are primarily based on some person small detectors. Typically around 4,000 small detectors are used to create detectors for the trifon. Such detectors are known as senons. Senons can define context much greater drastically and comprehensively than a simple previous-next. The precept of operation of senons may be represented as a selection tree.

Several consecutive backgrounds are considered as syllables. A characteristic feature of syllables is that their houses do now not depend on the rate of speech, in contrast to the heritage, however they rely upon the intonation of speech. Syllables, in turn, are words that restriction the variety of viable history combos. If there may be 40 background and the phrase on average consists of 7 history, then there need to be 407 one-of-a-kind words, however, even the most educated person rarely uses extra than 20 thousand phrases, which slightly simplifies the undertaking of speech recognition.

A simple way of recognizing speech is as follows: an audio sign is recorded, that is divided into fragments which can be separated via intervals of silence, then each fragment is recognized. To do this, all feasible combos of phrases are taken care of out, and the only that maximum suits the recorded fragment is selected. There are several crucial points on this process.

First is the precept of the attribute. Since the number of parameters is large, their optimization is required. Speech is divided into frames of approximately 10 ms long, and 39 numbers are allocated from every frame that describe speech. These numbers are referred to as function vectors. The precept of computing such vectors is a subject for a separate work, but if you describe it in words, then this is a spinoff of the spectrum.

Secondly - that is the precept of the version. The version describes a mathematical object, which is a hard and fast of parameters of a phrase, that is, it includes the most likely vector of the phrase attribute. The main issues of the model precept are how well the version corresponds to reality, whether the version can be improved, and how the adequacy of the version changes in distinctive conditions. The speech version is a hidden Markov model. This model describes tactics as a chain of states that succeed every different with some probability. The hidden Markov version is relevant not handiest in speech recognition, but also in different similar approaches.

Thirdly, it is a comparison method. Since the process of sorting thru all possible function vectors can take too long, many strategies are used to optimize the procedure.

In accordance with the structure of speech, three models are used for speech reputation.

The acoustic model incorporates acoustic properties (the most likely function vectors) and parameters for every syllable. Use context-unbiased and context-touchy models.

A phonetic dictionary suggests the correspondence of phrases to a historical past sequence. This method is not very effective, for the reason that phonetic version generally contains or 3 pronunciation variants, but this is sufficient to preserve recognition accuracy at a pleasant level.

The language version is used to optimize the phrase search manner. The language version contains feasible combos of phrases, and determines which word can comply with the previous one.

When growing speech popularity systems, the developer seeks to make it as correct and optimized (fast) as viable. For testing speech popularity systems, special take a look at recordings are commonly used. To evaluate speech reputation systems, use the alternatives below.

The frequency of erroneous words is determined by the formula (1.1):

$$WER = \frac{I+D+S}{N}, \tag{1.1}$$

where WER (Word Error Rate) is the frequency of erroneous words;
I – is the number of extra words;
D – is the number of missing words;
S – is the number of words moved;
N – is the number of words in the original text.

The accuracy is determined by the formula (1.2):

$$Accuracy = \frac{N-D-S}{N}. \tag{1.2}$$

Speed is defined as the ratio of recognition time to text reading time. For example, the text was read for 2 hours, and its recognition took 6 hours, therefore, the recognition speed will be 0.33 of real time [3].

1.3 Overview of Audio Signal Readers

The first recordings of sounds were made through direct cutting of the recording: via horn-like microphones, air vibration was transmitted to a gramophone needle, which produced slicing of those vibrations on the floor of a rotating wax cylinder. Today's microphones convert air vibrations into an electrical signal, and despite the fact that this principle underlies the operation of all microphones, their differences lie within the approaches that produce those transformations. The query of classifying microphones is not as easy as it might seem. They differ:

- at the precept of changing sound vibrations into electrical (mechanical and electrical traits);
- through the principle of the influence of sound vibrations at the diaphragm (mechanical-acoustic characteristics);
- in keeping with the dependence of the output stage on spatial orientation (directivity characteristics);
- at the principle of connecting to the audio path (switching traits).

In addition, the microphone, combining the above traits in quite a few combinations, also can have a one-of-a-kind layout and purpose - lapel, hanging, handheld, on-camera, connected to a musical instrument, desktop, etc.

From the factor of view of the mechanical-electrical precept, there are not many varieties - simplest condenser and dynamic microphones are currently used. All others are poorly ideal for use in professional realities.

The tool of a dynamic microphone resembles a tool of a dynamic speaker (for this reason, the latter are often used as a microphone in intercoms, walkie-talkies, that is, in which compactness is most advantageous to sound quality). The diaphragm of the dynamic microphone is attached to the coil positioned within the gap across the magnet (Figure 1.2). The longitudinal vibrations of the adjoining air act on the diaphragm with the coil relative to the magnet with a regular magnetic field, which causes alternating voltage and contemporary at the ends of the coil, the amplitude and frequency of which are proportional to the strain degree and the frequency of the sound performing on the diaphragm.

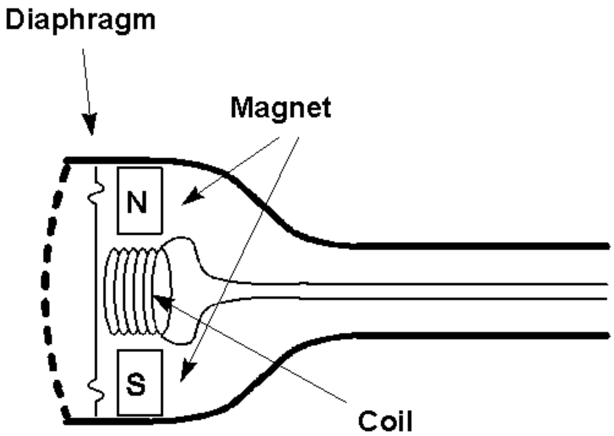


Figure 1.2 - Dynamic microphone device

In a condenser microphone, sound vibrations power the membrane, which is one of the plates of the condenser. (Figure 1.3) This capacitor is connected in collection with a DC supply. When sound acts on the membrane, it begins to transport forward, inflicting a change in capacitance, which turns the direct modern-day of the source into alternating. Due to a range of of functions of the use of a capacitor as a sound-electric powered converter, the condenser microphone is constantly equipped with a built-in preamplifier that matches the microphone output to the load input. Since the proposal to consist of a capacitor at the input of the low-frequency amplifier will reason an insufficient response from the digital engineer.

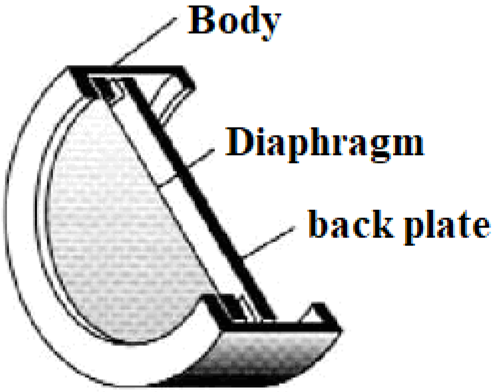


Figure 1.3 - Condenser microphone device

The big majority of mic preamps are primarily based on transistor circuits. However, there are a number of studio models with tube amplifiers. They are inaccurately called "tube microphones." Some sound engineers select tube amplifiers, even though the non-linear distortion coefficient of such devices is better than that of similar transistor ones.

Condenser microphones are divided into huge and small aperture microphones. Due to their length, layout and price, condenser microphones with a large membrane are in particular utilized in recording studios, whilst microphones with a small membrane have a correspondingly smaller size and are consequently more versatile.

There is also a condenser microphone primarily based electret microphone, in which the capacitor plates product of a special fabric are constantly charged and do not need a steady energy supply. This supply is still to be had in electret microphones, but simplest to electricity the microphone amplifier, which is simply as important in electret microphones as in traditional condenser microphones.

Most modern condenser microphones require a consistent voltage of 48 V for operation, which is supplied from a special strength source, or from a blending console that has the function of “phantom energy”. Some camcorders additionally have the ability to supply phantom power to be used in video recording of external condenser microphones.

According to the principle of the device, microphones are divided into strain receivers and strain gradient receivers. In pressure receiver microphones, sound vibrations act most effective on the the front aspect of the diaphragm, and within the stress gradient (difference) receiver, both at the front and rear sides. Differences within the design of acoustic receivers especially have an effect on their spatial traits.

According to the spatial characteristics of the microphones, at first, they're divided into non-directional and directional. Directivity is described as a exchange in microphone sensitivity when moving a sound supply of regular intensity relative to an axis perpendicular to the aircraft of the diaphragm. Naturally, the microphone is most sensitive alongside this axis. However, the conduct of the microphone as the supply deviates from this axis is different.

If the sensitivity changes very weakly, the microphone is non-directional, and its directivity is graphically depicted as a circle (Figure 1.4).

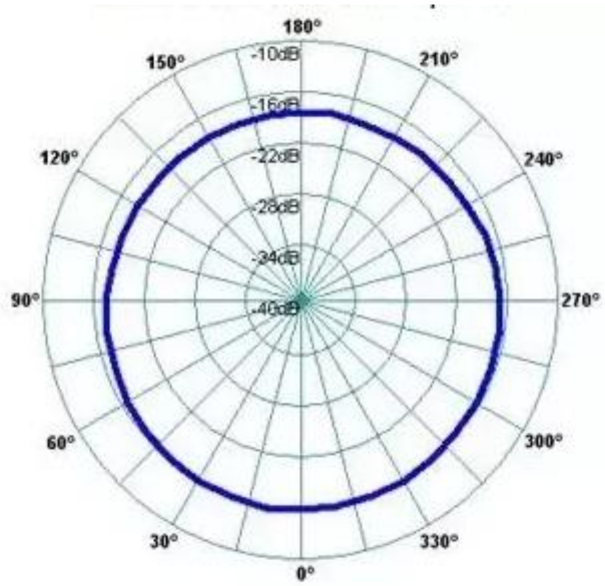


Figure 1.4 - Graphic depiction of directivity omnidirectional microphone

In the occasion of a pointy drop within the sensitivity of the microphone while deviating from the axis, this microphone is hypercardioid, or pointed. There are also bi-directional microphones, the graph of characteristics of that is the "eight".

It should be borne in mind that the directivity characteristics are strongly laid low with the ratio of the frequency of the sound wave or its length to the scale of the microphone. For sound waves of low frequency (long period), the path of the microphones is less, and for high frequencies (short waves) - more. It should be cited that the human ear has the same features.

According to the physical precept of connection, microphones are divided into conventional stressed and wireless (radio microphones). Wireless microphones are gadgets that include a microphone head, transmitter (transmitter) and receiver (receiver) in one housing. Lapel wi-fi microphones encompass parts: the microphone itself, installed on garments and a transmitter positioned on a belt or in a pocket and connected the usage of an audio cable.

Wireless microphones are created on the basis of general microphone heads (capsules), so their acoustic characteristics are almost equal to the fundamental stressed out counterparts.

Microphone parameters cover some of characteristics reflected, as a rule, in their technical documentation.

One of the important characteristics is the nominal frequency range. The nominal frequency variety is the range specified an audio signal produces an electrical signal on the output. The wider the frequency variety, the higher the microphone class.

It is also worth noting one of these microphone parameter as the unevenness of the frequency response, that is, the unevenness of the microphone's sensitivity from the sign frequency. The linear the frequency response of the microphone, that is, the smaller the difference among the maximum and minimum sensitivity of the microphone inside the nominal frequency variety, the higher the microphone.

Microphone sensitivity is the ratio of the output voltage to the sound pressure exerted on the diaphragm of the microphone, and is expressed in millivolts in keeping with pascal (mV / Pa). Since the sound strain on the microphone diaphragm may be very different relying at the frequency, the sensitivity dimension is standardized: it's far performed below the action of a right away sound wave at a frequency of one kHz.

As a rule, the sensitivity of condenser microphones is much higher than the sensitivity of dynamic ones.

Microphone sensitivity may be measured each in front of the microphone and behind. The ratio of these values is called the front / rear distinction in sensitivity. Obviously, this parameter strongly relies upon on the sort of directionality of the microphone.

You ought to also fit the output impedance of the microphone and the input impedance of the amplifier. These parameters, as a rule, also are indicated for a frequency of one kHz. These resistances ought to be of the same order, but the enter impedance of the amplifier must be slightly better than the output impedance of the microphone. This is essential to attain a excessive signal to noise ratio.

The sensitivity should be distinguished from the sensitivity level, depending on the nominal load resistance.

The standard sensitivity level is expressed in decibels (dB) and reflects the level of power developed by the microphone at rated load at a pressure of one pascal. Moreover, the lower the load resistance (and, therefore, the output impedance of the microphone), the higher the sensitivity level of the microphone.

Another parameter is the ultimate sound pressure. It is measured in the middle frequency range and shows at what level the coefficient of non-linear distortion exceeds 0.5%. For high-end microphones, the ultimate sound pressure can reach 140 dB.

Microphones also have their own noises. The noise level of the microphone is defined as the sound pressure level equivalent to the signal level at the microphone output in the absence of any effect (in conditions of silence). The lower the noise level of the microphone, the higher the signal-to-noise ratio and, accordingly, the quality of the output signal. For high-end microphones, the noise floor is 20 dB or less.

The dynamic range of the microphone is the difference between the ultimate sound pressure and the level of the microphone’s own noise, i.e. the level of intrinsic noise is a threshold value, because the useful signal will be lost in noise if its level is below the level of intrinsic noise. If the sound level exceeds the maximum sound pressure of the microphone, non-linear distortions will appear.

The ratio / signal noise is usually not indicated in the characteristics of the microphone, since this parameter is not standardized. You cannot directly compare this parameter for different microphones, because Each manufacturer makes such measurements under different conditions, but, basically, this value is calculated as the difference between the microphone’s own noise and the sound level of 94 dB.

Depending on the manufacturer, the technical data sheet of the microphone may contain many parameters. Large manufacturers can even attach various test schedules to their passport, such as frequency response, phase response, and others.

Depending on the recording conditions, a suitable microphone should be used. Microphones can be on legs, handheld, on-camera (for video shooting), fastened to clothes, etc. A suitable microphone design will allow you to flexibly position it depending on its radiation pattern.

The most common type is a hand-held microphone, as its design is the most versatile. The handheld microphone can be placed on a stand or held in the hand. A handheld microphone consists of a head where the condenser capsule itself is located, cylindrical handles and a cable connector. Typically, handheld microphones use an XLR connector, and connects to a preamplifier or recording device through a 6.3 mm jack. If this is a radio microphone, then in the handle there is a preamplifier, transmitter and an autonomous power source (battery or alkaline battery). Handheld microphones have a cardioid or supercardioid pattern.

As mentioned earlier, hand-held microphones have a similar shape, but in other types, the shapes of microphones can be very different. High-quality studio microphones are flat disk shaped. Such devices are mounted on a stand using springs, rubber bands or other elastic devices, which allows you to isolate the microphone on the vibrations of a building, floor or stand. Microphones designed for talking can be placed on flexible hoses and attached to headphones or clothing.

Unlike manual, "non-manual" microphones can have both directional and non-directional radiation patterns. Microphones with an omnidirectional characteristic are called microphones of the "common space of the hall".

A little later, compared to the types described above, table-type microphones with a hemisphere-shaped radiation pattern were invented. Such microphones are widely used on television and radio. Also this type of microphones is used in theaters, but they are already installed on the floor. Tabletop microphones are flat or slightly convex, rectangular or round. Thanks to this pattern, desktop microphones are also used at conferences, as can capture everyone’s speech at the table.

Microphones designed for attaching to clothing are called lapel microphones. Lapel microphones have a small sensitive head. Since the microphone is attached to clothes, it is difficult to direct it to a sound source, so lapel microphones have an omnidirectional radiation pattern. During human movements, the microphone can detect low-frequency vibrations, so filters are used in lapel microphones.

On-camera microphones used in video recording and reporting have a narrow directional sector, since in video recording it is required to record sound only from the frame area. Some camcorders have variable gain amplifiers that are directly proportional to the focal length of the lens. Those. with an increase in the focal length of the lens, the gain of the microphone amplifier increases, which allows you to increase the sound volume, while increasing distant objects.

Separately, it should be noted microphones with switchable focus. Such microphones are suitable in cases where it is impossible to say exactly under what conditions the microphone will be used. Switchable directivity is achieved due to the fact that several capsules are located in the microphone head, combining which in various combinations, directional, narrowly directed, non-directional and other directivity characteristics are achieved. Such microphones have a switch, which is located on the microphone body.

To expand the nominal frequency range, some microphones use two or more capsules that are used in different frequency ranges, similar to headphones with multiple reinforcing sound emitters, or with multi-band speaker systems. A feature of this technique is that it is necessary to coordinate the cutoff frequency and phase at the boundary frequencies. In the event that the signals from two capsules at the cutoff frequency are in different phases, a decrease in the frequency response of the microphone may form.

In microphones they realize not only switchable directivity, but also switchable frequency response. Such switches are commonly used in handheld microphones for performances. This allows you to quickly change the frequency response of the microphone: reduce low or high frequencies or equalize the response. In other situations, this is not entirely justified, because After recording, the levels of certain harmonics can be changed programmatically.

The microphone diaphragm is a very fragile part, so manufacturers design the microphone head in such a way as to prevent damage to the diaphragm and the capsule itself. In hand-held microphones, the capsule is protected from mechanical damage by a woven metal mesh. Underneath the mesh is a sound-transmitting fabric to protect against small particles and saliva. Microphones designed for outdoor use have a synthetic-fiber screen that protects the microphone from blowing and, therefore, from unwanted noise.

The microphone output signal, as a rule, does not exceed several tens of millivolts, therefore it is highly susceptible to external electromagnetic interference. If the connecting cables are long, the noise from interference will even exceed the microphone output signal level, in such cases a balanced connection is used. To do this, the microphone must have a balanced output, and the preamplifier with a balanced input. The balance cable must be shielded, which connects to the preamp enclosure.

All this information, it would seem, is obvious and well-known, but poor or incorrect screening, not to mention its absence, is the main source of problems, both in the studio and in sound reinforcement.

Almost all modern professional microphones have an XLR connector. The connectors used previously are not used now. The XLR connector has three pins: the first is the common wire, the second is the direct polarity (plus), the third is the reverse polarity (minus).

When using multiple microphones, their phase characteristics need to be the same. Even using microphones of the same manufacturer, one cannot be sure that their phase characteristics match. To check the common mode of the microphones, it is enough to sum their signals. If the volume increases, then we can say that the microphones have similar phase characteristics.

1.4 Selection of a speech capture device

When studying the manual for using a microphone, one of the main parameters is the concept of “resistance” or “impedance”. This is a parameter that shows the microphone output impedance value for AC. Resistance is measured in ohms (ohms). Depending on the magnitude of the resistance, high-impedance (10-50 kOhm) and low-impedance (50-600 Ohm) microphones are distinguished. The output impedance of the microphone should correspond to the input impedance of the amplifier, otherwise the ratio of the useful signal to noise may decrease. You should pay attention to this. It is worth noting that these resistances should not be exactly the same, the output impedance of the microphone should be within acceptable limits of the value indicated in the parameters of the microphone amplifier. If the amplifier has a low-impedance input, then the output impedance of the microphone should also be low-impedance. If you connect a high-impedance microphone to the low-impedance input of the amplifier, this will lead to a decrease in the signal level. On the other hand, when connecting a low-impedance microphone to the high-impedance input of an amplifier, the sound can become distorted, although, in a general sense, this is not a big problem.

The advantage of using a low impedance microphone is that such microphones have a lower noise level, even when using long cables.

When choosing a microphone, one should take into account not only the totality of its technical characteristics, but also the recording conditions, for this reason it is quite difficult to give concrete recommendations. However, general rules for choosing a microphone do exist.

An omnidirectional microphone can be used when recording voice and sounds in a very muffled room. It should also be used to transmit the general acoustic environment in multi-microphone recording.

A one-way microphone with a characteristic of the "cardioid" type is preferably used when recording in a room with a large number of sound reflections. It is also used in the event that extraneous noise penetrates the room where the recording is made. The microphone should be installed with the back to the sound source. Such a microphone is recommended for use with a wide front placement of artists. This microphone is used for unlikely amateur recording of multi-microphone recordings for a clear separation of sound groups, as well as when placing the sound source close to the microphone to reduce the low-frequency distortion inherent in this case, non-directional and two-way directional microphones.

A bi-directional microphone with a figure-eight diagram should be used when recording in a muffled room, when it is necessary to increase the relative level of reflected signals, as well as when recording individual sounds and voices to emphasize low frequencies in close proximity, sound sources at the microphone. Such a microphone is also used in the case when it is necessary to detach from directional noise sources. For this, the microphone is oriented with a zone of zero sensitivity to the noise source. A bi-directional microphone oriented in the horizontal direction is useful for attenuating sound waves reflected from the floor, ceiling and side walls of a room. This allows acoustic processing of only two walls: behind the source of sound waves and opposite it.

For the final choice you need to decide in which sound field the microphone will be used.

If the sound reaches the microphone without first reflecting from other surfaces (walls, ceiling, floor), then the concept of “direct sound field” is applied (Figure 1.5).

To achieve good recording results in a direct sound field, it is necessary to direct the low-sensitivity zones of the microphone towards the alleged sources of reflected sound. This will reduce distortion, as the reflected sound may have a different phase from the unreflected sound.

In the case when the sound, before reaching the microphone, reflects on any surface, the concept of “reflected (reverberation) sound field” is used describes the situation, (Figure 1.6).

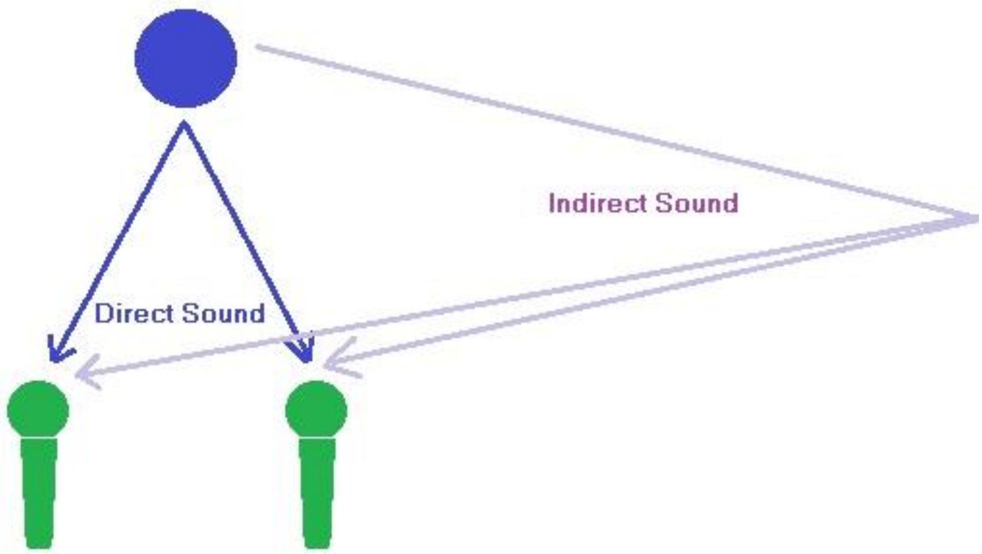


Figure 1.5 - Direct sound field and reflected sound

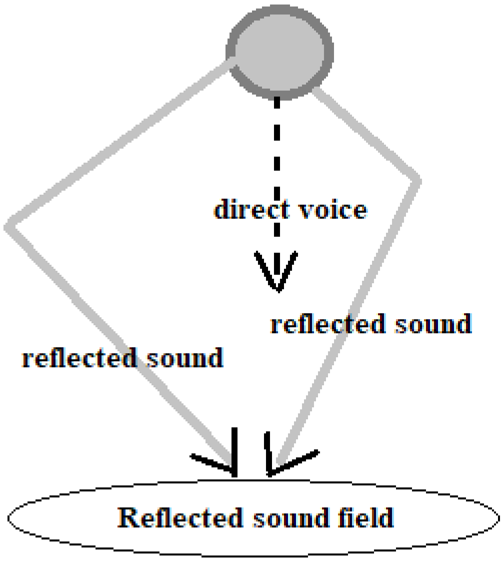


Figure 1.6 - Reflected sound field

In the case of a reflected sound field, to completely suppress unwanted noise, it is not enough to manipulate the low-sensitivity zone of the microphone, but it also reduces the level of distortion, although to a lesser extent. Therefore, unidirectional microphones should also be used in the reflected sound field, which provide more reliable protection against the feedback effect (due to the reflected sound of a different phase) and extraneous signals compared to omnidirectional ones. The smaller the directional sector of the microphone, the better it copes with the task. In this regard, microphones with a supercardioid orientation are most effective, followed by cardioid microphones. The microphone efficiency in the reflected sound field is expressed by a numerical measure, which is the directivity coefficient, which is inversely related to the directivity sector of the microfog. The coefficient values for microphones with different radiation patterns are presented in table 1.1:

Table 1.1 - Directivity patterns.

Microphone type	Directivity coefficient	Noise cutoff, dB
Omnidirectional	1	0 дБ
Cardioid	1.7	4.8 дБ
Supercardioid	1.9	5.7 дБ

The data in the table indicate that the directivity coefficient of a supercardioid microphone is 1.9 times higher than the corresponding indicator for an omnidirectional microphone. Under real conditions, this is expressed in the following: in a reflected sound field, a supercardioid microphone picks up 5.7 dB less reverberation noise.

To implement voice control, a high quality microphone is required. Since it is only necessary to capture speech of the operator, it is advisable to use a narrowly focused microphone, which will cut off extraneous sounds and voices. Also, to cut off extraneous wooks lying outside the range of a person's voice, filtering the sound signal is required.

Based on the foregoing, it is worth stopping your choice on the microphone built into the Gboy webcam (Figure 1.9). This is a high-quality narrowly focused microphone with a small internal resistance, which is a condenser microphone. The device has a built-in amplifier, ADC and a frequency filter, which allows to reduce the noise level. In addition, the frequency response of the model is optimized for recording human speech. The microphone characteristics of the Gboy webcam is presented in table 1.2.

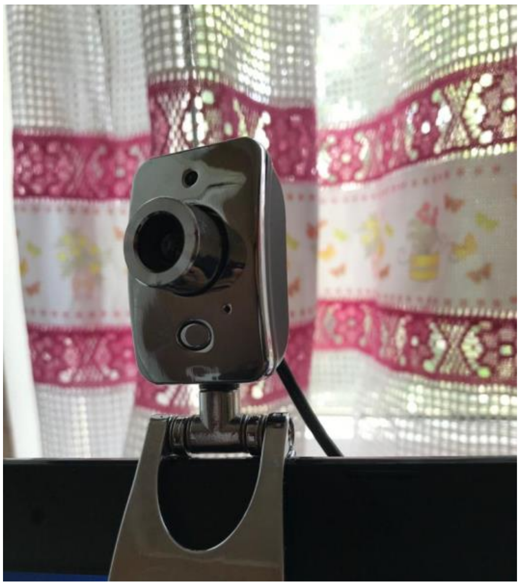


Figure 1.7 – GboyWebcam

Table 1.2 – Gboy WebCamera Microphone Features

Parameter	Value
Radiation pattern	narrowly targeted
Frequency response	211-6915 Hz
Microphone type	capacitor
Sensitivity	1.5 mV / Pa
Resistance	80 ohm
Max. sound pressure	85, 1.2 KHz at 4% THD
Signal to noise ratio	70 DB at 1.1 kHz
Weight	75 g

1.5 Search for an optimal audio recognition system based on open source speech recognition

Among the most common speech recognition software products, the following can be distinguished: CMU SPhinx, JulIus, SIMON software, IATROSLet's consider them in more detail.

CMU Sphinx - For brevity, it will be called Sphinx, it was written by the developers of speech recognition programs at Carnegie Mellon University. The Sphinx software package includes a series of speech recognizers (Sphinx 2-4), a program for adapting sound models (Sphinx train).

In 2000, Sphinx developers at Carnegie Mellon University approved open-source speech recognition system components, including Sphinx 2 and Sphinx 3 (later in 2001). The speech decoder included simple applications and acoustic models. Available resources included a language model, additional software for training an acoustic model, a pronunciation dictionary, and a linguistic compilation model that is in the public domain (cmudict).

Sphinx is a continuous speech recognizer that works with any speaker, uses the n-gram statistical language model and the Hidden Markov model. It was developed by Kei-Fu Li. Sphinx is capable of recognizing long speech, has a speaker-independent huge vocabulary of recognition, that is, those features that in 1986 caused great controversy in the speech recognition environment. Sphinx in historical development is notable for the fact that in its development eclipsed all previous versions in terms of performance.

Sphinx2 is the performance-oriented and fastest speech recognition program developed by Chudon Juan at Carnegie Mellon University and released with open source code based on BSD license for SourceForge by Kevin Lenso in Linux World in 2000. Sphinx2 is focused on real-time speech recognition and is ideal for creating a variety of mobile applications. Sphinx2 includes such functionality as an end pointer, partial hypothesis generation, connection of a dynamic language model, and so on. It uses a dialogue system and a language learning system. It can be used on PBX computers such as Asterisk. Sphinx2 code has been incorporated into numerous commercial products.

Unfortunately, Sphinx2 did not evolve for long (nothing more than planned maintenance). The current decoder for real-time recognition is integrated into PocketSphinx.

Sphinx3 became a half-non-stop acoustic version of speech reputation (used Gaussian combos of a wide variety with an man or woman model that took into account the burden of the vector over those Gausses). Sphinx3 adopted the giant non-stop model built on hidden Markov models and changed into at the start used for correct speech reputation, which was carried out within the "post-factum" mode (not actual time). Recent traits (algorithms and software program) have contributed to the truth that Sphinx3 could recognize in near real-time mode, even though the application become now not yet appropriate for first-rate use as an application. Sphinx3, after actively growing and reuniting with SphinxTrain, has furnished get admission to to numerous contemporary strategies and fashions, such as LDA / MLLT, MLLR and VTLN, which have stepped forward speech reputation accuracy.

Sphinx4 is the completely rewritten Sphinx speech engine, whose important intention is to provide a bendy framework for the implementation of developments in speech reputation. Sphinx4 is absolutely written inside the Java programming language. A extensive contribution to the development of Sphinx4 become made via

Sun Microsystems, and assisted inside the application understanding of the challenge. Individual venture developers are experts from MERL, MIT, CMU. Current development goals include:

- improvement of new acoustic fashions for training;
- implementation of a speech adaptation system (MLLR);
- improvement of configuration management;

ConfDesigner release - a graphical shell for PocketSphinx - this version of Sphinx may be incorporated into any other systems based totally on ARM microprocessors, including pcDuino, Raspberry Pi, Odroid. PocketSphinx is actively growing and integrating into diverse structures with fixed-point mathematics and into efficient fashions primarily based on a combined calculation version.

Julius is a high-performance continuous vocabulary popularity application with huge vocabulary recognition, a decoder of software for studies in the subject of linked speech and improvement. Julius is perfectly suited for real-time decoding on almost all current computers, with a dictionary of 60 thousand words, using the obligations of a phrase tiagram and a contextually impartial Hidden Markov version. The predominant function of the challenge is its full embeddability, it's also really worth noting that safe modulation can be independent of version structures and various types of Hidden Marklenov fashions, which maintains the general nation of trifons and the associated combination of fashions with many potions, phonemes and statements. Standard codecs are active due to the free modeling of tools. Basically, Julius is designed to work on Linux structures and different UNIX-like stations, but the gadget also works on Windows. Julius is an open-source gadget and is distributed with the BSD sort of license.

Julius - is being evolved as a part of loose software program for studies inside the discipline of reputation of the Japanese language, in view that 1997 and this paintings continued in the framework of the Continuous Speech Recognition Consortium (consortium for recognizing continuous speech) from 2000 to 2003 in Japan.

Starting with version 3.4, the grammar base of the analyzer's speech recognition gadget is referred to as Julian and is included into Julius. Julian is a modified shape of Julius, which makes use of its very own designed shape of the Finite-state gadget grammar as a language version. It can be used to construct voice navigation systems with a small dictionary or other conversational structures for spotting all varieties of dialogs.

To start the Julius speech recognizer, you need to choose a language version and an acoustic version for the favored language. Julius adapts the acoustic model of the HTK ASCII coding format, the pronunciation database of the HTK format, and 3-layer diagrams for building an ARPA language model (2 direct and 3 reversible education fashions wrapped in a speech corpus with an inverted phrase customer) .

Although Julius is most effective to be had for the Japanese language version, the VoxForge challenge is working on an acoustic version for the English language the use of the Julius speech popularity engine.

RWTH ASR (RASR for short) is an open source speech recognition toolkit. The toolkit includes speech recognition technology to create computerized speech recognition systems. This technology is being evolved with the aid of the Natural Language Technology Center and the Model Recognition Group at the Rhine-Westphalian Technical University of Aachen.

RWTH ASR includes gear for growing acoustic fashions and decoders, as well as additives for adapting speaker speech, adaptive speaker speech schooling systems, uncontrolled studying structures, differential learning structures and lattice phrase-processing forms. This software program runs on Linux and Mac OS X. The challenge domestic page gives ready-to-use fashions for research with tasks, training structures and complete documentation.

The toolkit is published below the open source license, that's referred to as the "RWTH ASR License," that is derived from the QPL (Q Public License). This license represents unfastened use, which includes redistribution and change for non-commercial use.

Simon is a speech reputation gadget based at the Julius and HTK speech engines. Simon device is designed in such a way that it is pretty convenient for operating with diverse languages and numerous dialects. At the identical time, the speech reputation response is completely customizable and it is not appropriate for the exclusive reputation of unmarried voice requests and can not be configured for the needs of users.

To use the gadget easily, certain "scenarios" ought to be completed. Simon programs are configured for special obligations. Among the possible Simon scripts, for example, "Firefox" (launching and dealing with the "Firefox" browser) or "window manipulate bundle" (closing, moving, resizing windows) and so on. Scripts can effortlessly be created by users and distributed within the network via the Get Hot New Stuff gadget. To date, greater than 39 scripts had been written and 3 languages published in the opendesktop.Org repository.

Simon also supports simple universal and simple fashions much like GPL fashions from Voxforge, which customers use to understand English, German and Portuguese. At the equal time, there may be no want to conform or teach the model so as for it to begin operating correctly. A demonstration of Simon 0.3.0 may be found at the official website of the developers. At the same time, the user's speech includes technical terminology - that is the main function of the implementation in simon - and thereby it demonstrates how it's far feasible to use simon for customers and the way simon can be integrated via users to beautify their very own development.

IATROS is a redesigned ATROS speech recognition application that is suitable for recognizing both speech and handwriting recognition. IATROS is based totally on a modular structure and can be used each to build differentiated models, whose purpose is to perform a VetriBri search based totally on a hidden Markov version. IATROS can work each the usage of the Internet and offline (based totally on ALSA modules).

IATROS consists of 2 pre-processing modules (for speech and hand-written images) and a recognition engine module. Data preprocessing and module extraction capabilities are supplied by using module popularity vectors that use Hidden Markov fashions and language models which might be carried out through searching for guesses from the best speech recognition structures. All these modules are made within the programming language “C”.

SHoUt is a toolkit written by Marina Hubrechts on the Netherlands Institute of Image and Sound, that is designed to apprehend non-stop speech with a large records dictionary. SHoUt consists of an utility for training statistical fashions and for (non) speech detection, diarization and speech decoding.

VoxForge is a free speech engine and language acoustic version supplied in a public information warehouse. VoxForge turned into constructed as a speech transcription repository with a free GPL package deal to be used with speech engines which are open supply. Speech audio files may be assembled into acoustic models for use with open source speech reputation systems inclusive of Julius, Sphinx and HTK (however HTK has a distribution restriction).

HTK is a speech popularity toolkit the usage of a hidden Markov model. It is specially supposed for speech popularity, but it is also used for other different duties that use the hidden Markov model - speech synthesis, popularity of visible images (facial expressions, facial capabilities, etc.).

Thus, having examined the most common open supply speech reputation structures, it's far essential to be aware that the solution is primarily based on CMU Sphinx, specially PocketSphinx, is maximum strongly represented. It is maximum suitable for recognition responsibilities based totally on the ARM minicomputer.

2 Design part

2.1 single board ARM minicomputer pcDuino1

Currently, systems based on ARM minicomputers such as Raspberry Pi, BeagleBone, Odrioid, pcDuino and others are widely used. They differ in processor performance, RAM and ROM, the number of input / output ports, supported interfaces, price, and other indicators. The voice control system was implemented on the pcDduino V1 minicomputer manufactured by SparkFun.

Consider this platform in more detail and more carefully, because it is in it that the fulfillment of the task is concluded. The developers designated their product as PC + Arduino, i.e. Arduino-compatible minicomputer. Justified move, because SparkFun also produces many expansion cards for Arduino, and most of them can be used with pcDuino. The core of the board is an Allwinner A10 chip system containing one ARM Cortex-A8 core at 1 GHz and a Mali 400 graphics processor. Table 2.1 shows the main technical specifications of pcDuino V1 [4].

Table 2.1 - pcDuino V1 specifications

Parameter	Value
CPU	1 GHz ARM Cortex A8
GPU	Mali 400
RAM	1 GB
ROM	2 GB + expansion card slot
Video output	HDMI
OS	Lubuntu, Android
Communication	Ethernet 100 mbps, UART, I2C, SPI
Input / output ports	14 digital, 2 PWM, 2 6-bit and 4 12-bit analog inputs, 4 SPI, 2 I2C
USB ports	2

The pcDuino V1 minicomputer architecture and its interface are shown in Figure 2.1.

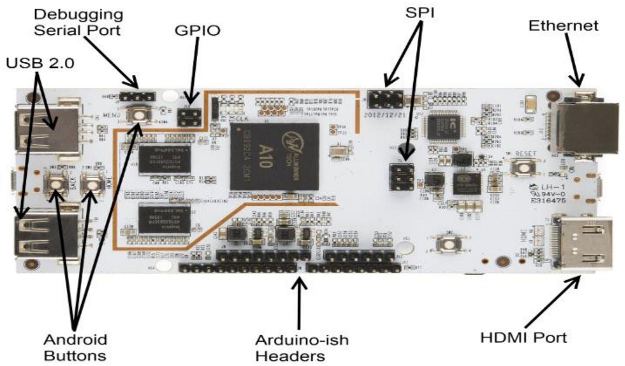


Figure 2.1 - Architecture of the pcDuino V1 minicomputer

2.2 Wi-Fi based on ESP8266 chip

The ESP8266 is designed for use in smart sockets, mesh networks, IP cameras, wireless sensors, wearable electronics, and so on. In a word, ESP8266 was born to become the brain of the upcoming “Internet of things”.

There are two options for using the chip: 1) in the form of a UART-WIFI bridge, when the module based on the ESP8266 is connected to an existing solution based on any other microcontroller and is controlled by AT commands, providing the solution with a Wi-Fi infrastructure; 2) implementing a new solution using the ESP8266 chip itself as a control microcontroller. The block diagram of the ESP8266 chip is shown in Figure 2.2.

As a result, a typical chip strapping consists of only a few elements. Less elements - lower cost of components, less cost of soldering, less placement area, less cost of printed circuit board. Which is perfectly confirmed by the current prices of modules based on the hero of our today's review. Manages all this integrated farming with an expanded version of the 32-bit Tensilica L106 Diamond series processor.

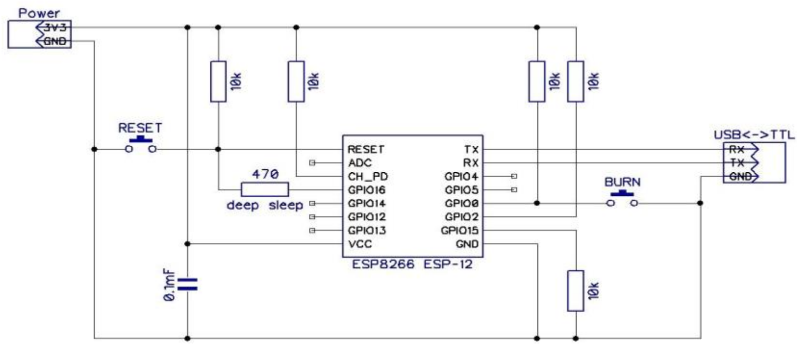


Figure 2.2 - Block diagram of the chip ESP8266

TENSILICA l-106 DIAMOND series is a 32-bit microcontroller with RISC architecture. The microcontroller operates at a frequency of 160 MHz, and uses 24- and 16-bit instructions.

The board (Figure 2.3) also has low power consumption. In transmission mode, it consumes 215 mA, in deep sleep mode - 60 μA, and in the mode of maintaining communication with the access point - 1 mA.

Figure 2.3 - Wi-Fi board based on ESP8266

2.3 Relay Modules

To implement the interface between the control objects and the microcontroller, two relay modules were used, the circuit and appearance of which are shown in Figures 2.4 and 2.5.

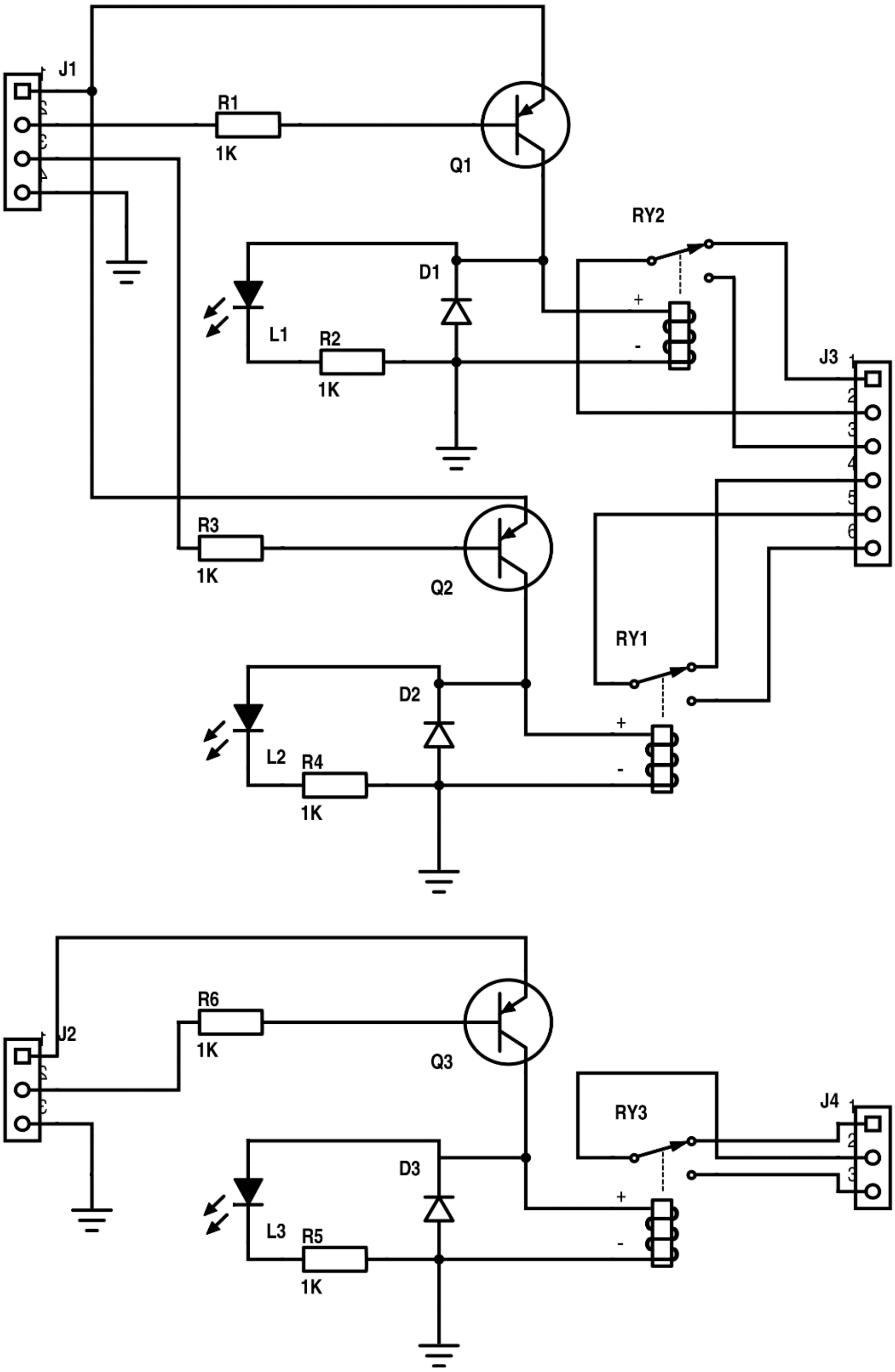


Figure 2.4 - Electrical diagram of relay modules

The relays are designed for a control voltage of 5 V, and can switch alternating voltage and current up to 250 V and 10A, respectively. In armature hold mode, the relay coil consumes 80 mA.

Figure 2.5 - 2-channel relay module

2.4 Wireless Router TP-LINK TL-WR702N

The TL-WR702N Wireless Router (Figure 2.6) is a multifunctional wireless network device that supports the operating modes of the access point, router, bridge, client, and relay, allowing users to more effectively work with various wireless applications.

Figure 2.6 - TP-LINK TL-WR702N Router

The router uses WPA-PSK / WPA2-PSK encryption modes that can effectively protect your wireless network. The pre-encryption function, which sets the initial wireless network name and password for users to protect the wireless network, increases its security. The TL-WR702N is equipped with a Micro USB port and can be powered by an external power adapter or by connecting to a computer’s USB port. Full characteristics of the device are given in table 2.2.

Table 2.2 - TP-LINK TL-WR702N Router Specifications

Parameter	Value
Wireless Data Standards	IEEE 802.11n, IEEE 802.11g, IEEE 802.11b
Dimensions (WxDxH)	57 x 57 x 18 mm
Antenna type	Built-in
Frequency range	2.4-2.4835 GHz
Wireless Security	64/128/152-bit WEP, WPA / WPA2, WPA-PSK / WPA2-PSK
Network Security (firewall)	Protection against DoS attacks, SPI firewall, filtering by IP address / MAC address / domain name, binding
Environmental parameters	Operating Temperature: 0 ℃ ~ 40 ℃ (32 ℉ ~ 104 ℉) Storage Temperature: -40 ℃ ~ 70 ℃ (-40 ℉ ~ 158 ℉)

2.5 Voice Control System Design

The basis of the device is the pcDuino1 ARM minicomputer. A USB digicam is attached to it through a USB camera with a integrated Gboy microphone, and Ethernet, the minicomputer is attached to the TP-LINK TL-WR702N router for communique with the ESP8266 board, to the terminals of which relay modules are connected. Then it is connected to display with HDMI. The block diagram of the voice control system is shown in Figure 2.7.

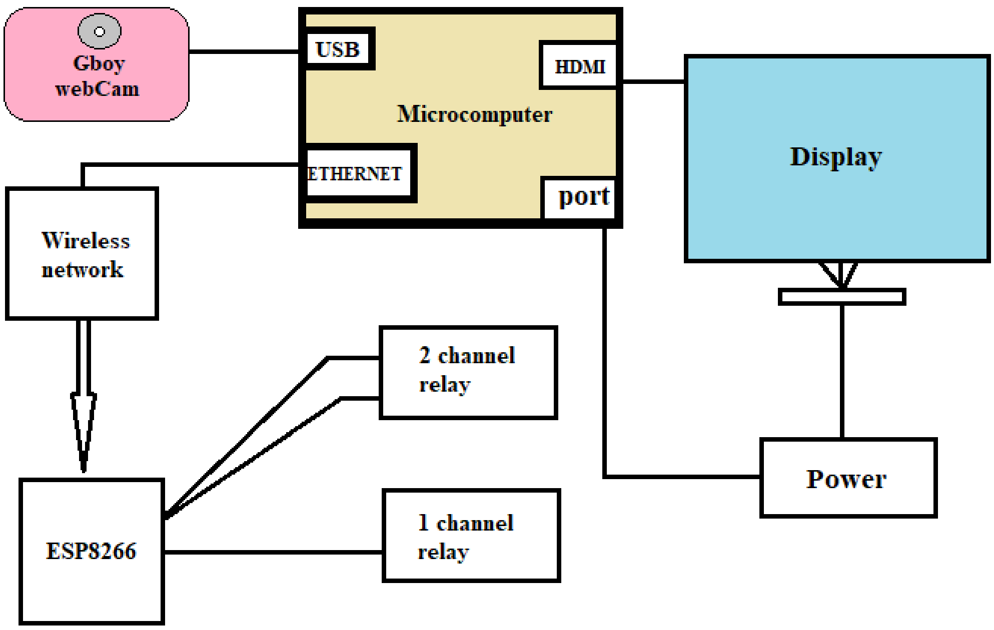


Figure 2.7 - Structural diagram of a voice control system

3 Software

3.1 Ubuntu operating system

Lubuntu working system turned into installed on pcDuino 1 minicomputer. The Lubuntu working machine is identical to the Ubuntu system, the main difference between Lubuntu is the low computer requirements due to using the LXDE computer environment, that is why all of the documentation for Ubuntu applies to Lubuntu. Ubuntu is undoubtedly the most popular distribution of the Linux operating gadget. Ubuntu is a serious contender within the world of laptop structures and Web servers. This device turned into introduced to life with the aid of Mark Shuttleworth. He based Canonical, which leads the improvement of Ubuntu Linux.

This machine is based on the idea of free, each economically and legally, software. This approach that you can deploy it for free, in addition to regulate and distribute with none royalties or royalties, which is completed by means of the discharge of Ubuntu beneath a special license that protects these rights.

Thus, you could installation Ubuntu on any laptop on which this system can work, and update it to the modern-day version as regularly as you want, without worrying about licenses, product activation, or unique keys that you have to enter. You can also distribute it - distribute it to your pals or family and help them set up the system on their computers.

In addition, Ubuntu is very easy to install, automatically spotting the gadgets you are using and installing the most appropriate drivers for them. From other running structures, the installation of which takes from one hour to 1/2 a day, Ubuntu is definitely advantageous, requiring installation on average much less than an hour.

For greater than a decade, Linux has provided the computing strength of thousands and thousands of Web servers, which isn't unexpected given the amount of work involved in growing graphical interfaces. Today, Linux computing device installations account for 1% of the full quantity of installations (and this range is developing rapidly) and Ubuntu takes the lion's proportion of them.

You can deploy the Lubuntu operating machine in pcDuino's NAND inner reminiscence, or you can create a bootable memory card. Because the amount of integrated NAND-reminiscence is most effective 2 GB, the selection was made in favor of a microSD card.

3.2 Creating a bootable memory card with Lubuntu

To prepare a bootable memory card, you must have a bootable image of the Lubuntu operating system and a microSD memory card. To create a boot map, Win32 Disk Image was installed on a personal computer running Windows. Using the Win32 Disk Image program, an image of the Lubuntu operating system was recorded on a memory card. The process of creating a bootable memory card is shown in Figure 3.1.

Figure 3.1 - The Win32 Disk Image program window

3.3 CMU sphinx software package and its installation

CMU sphinx software package - a leading tool in the field of speech recognition. The CMU sphinx software package includes several programs for various tasks and applications:

Pocketsphinx - a small library written in C;

Sphinxbase - an additional library required for the operation of Pocketsphinx;

Sphinx4 is a flexible speech recognition program written in Java;

Sphinxtrain is a program for adapting sound models.

On the pcDuino V1 minicomputer, pocketsphinx-0.8, sphinxbase-0.8, an acoustic model of the Russian language were installed, as well as a language model and a dictionary were created. The program requires an acoustic language model, a language model and a dictionary, where the acoustic model describes the sounds of words, and the language model describes the a priori probability of each speech fragment, for example, that the phrase "secret code" is more likely than the "secret cat". A dictionary is a collection of words that can be recognized. The dictionary consists of the following words: “turn on”, “turn off”, “relay”, “number”, “one”, “two” and “three”, because requires control of three relay modules.

Because CMU sphinx software package is not available in standard repositories, pocketsphinx-0.8 and sphinxbase-0.8 were manually installed. To do this, the archives of the pocketsphinx-0.8 and sphinxbase-0.8 packages were downloaded from the official site, unpacked using the "tar" command through the terminal of the Lubuntu operating system. Further, for each package the following commands were executed:

% ./configure

% make

% make install

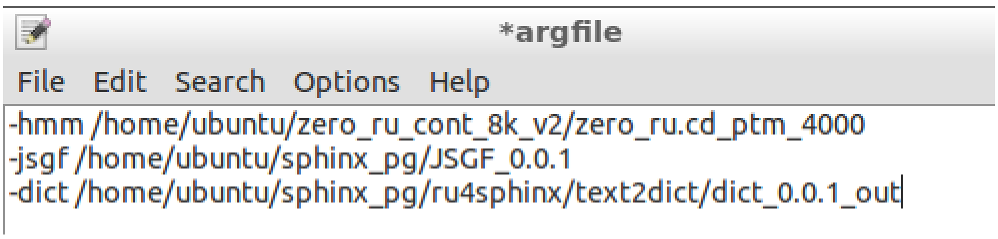


Figure 3.2 - The contents of the argfile argument file

Instead of a static language model, the JavaScript Grammar File (JSGF) was used, the contents of which are shown in Figure 3.3.

```
File   Edit   Search   Options   Help
#JSGF V1.0;

public <test> = ( <a> <b> <c> <d>);
<a> = ( включить | выключить );
<b> = ( реле );
<c> = ( номер );
<d> = ( один | два | три );
```

Figure 3.3 - The contents of the jsgf language model file

This file indicates that the command consists of four words. The first word can be "turn on" or "turn off", the second - "relay", the third - "number", and the fourth - "one", "two" or "three".

3.4 Configure and run pocketsphinx

The acoustic model of the Russian language was downloaded from the voxforge repository. Voxforge is a repository of acoustic models where you can download acoustic models of Russian, English, Bulgarian, French, Italian and many other languages for sphinx and julius.

Typing a command in the terminal

% pocketsphinx_continuous -argfile /home/ubuntu/pocketsphinx-0.8/argfile, we get the window shown in Figure 3.4. After initializing the program, the line "READY ..." appears, which indicates that the program has started work and is ready to recognize commands. When a signal appears on the microphone, the capture and processing of this signal begins, and within two seconds the result is displayed. The recognition result of the “enable relay number one” command is shown in Figure 3.5.

To manage devices, you need to highlight these commands. To do this, at startup, add the parameter "> / home / ubuntu / sphinx_pg / out", after which the recognition result will be written to a text file for further processing.

```
INFO: jsgf.c(581): Defined rule: <test.g00002>
INFO: jsgf.c(581): Defined rule: <test.a>
INFO: jsgf.c(581): Defined rule: <test.g00004>
INFO: jsgf.c(581): Defined rule: <test.b>
INFO: jsgf.c(581): Defined rule: <test.g00006>
INFO: jsgf.c(581): Defined rule: <test.c>
INFO: jsgf.c(581): Defined rule: <test.g00008>
INFO: jsgf.c(581): Defined rule: <test.d>
INFO: fsg_model.c(215): Computing transitive closure for null transitions
INFO: fsg_model.c(270): 36 null transitions added
INFO: fsg_model.c(421): Adding silence transitions for <sil> to FSG
INFO: fsg_model.c(441): Added 27 silence word transitions
INFO: fsg_search.c(366): Added 0 alternate word transitions
INFO: fsg_lextree.c(108): Allocated 2916 bytes (2 KiB) for left and right contex
t phones
INFO: fsg_lextree.c(253): 67 HMM nodes in lextree (40 leaves)
INFO: fsg_lextree.c(255): Allocated 7236 bytes (7 KiB) for all lextree nodes
INFO: fsg_lextree.c(258): Allocated 4320 bytes (4 KiB) for lextree leafnodes
INFO: continuous.c(371): pocketsphinx_continuous COMPILED ON: Apr 21 2014, AT: 1
0:26:31

Warning: Could not find Capture element
READY....
```

Figure 3.4 - Terminal window after starting pocketsphinx

```
000000002: включить реле номер три
READY....
Listening...
Recording is stopped, start recording with ad_start_rec
Stopped listening, please wait...
INFO: fsg_search.c(1032): 186 frames, 2055 HMMs (11/fr), 6008 senones (32/fr), 1
350 history entries (7/fr)

INFO: fsg_search.c(1417): Start node <sil>.0:2:32
INFO: fsg_search.c(1417): Start node <sil>.0:2:32
INFO: fsg_search.c(1417): Start node <sil>.0:2:32
INFO: fsg_search.c(1417): Start node <sil>.0:2:32
INFO: fsg_search.c(1456): End node <sil>.183:185:185 (-518)
INFO: fsg_search.c(1456): End node <sil>.183:185:185 (-518)
INFO: fsg_search.c(1456): End node <sil>.183:185:185 (-518)
INFO: fsg_search.c(1456): End node <sil>.183:185:185 (-518)
INFO: fsg_search.c(1456): End node <sil>.183:185:185 (-518)
INFO: fsg_search.c(1456): End node один.146:167:185 (-509)
INFO: fsg_search.c(1680): lattice start node <s>.0 end node </s>.186
INFO: ps_lattice.c(1365): Normalizer P(O) = alpha(</s>:186:186) = -172388
INFO: ps_lattice.c(1403): Joint P(O,S) = -172388 P(S|O) = 0
```

Figure 3.5 - the result of the recognition of the command

3.5 Processing recognition results

Following the steps in chapter 3.4, recognized speech commands are written to the text file "out". The contents of this file are shown in Figure 3.6.

```
READY....
Listening...
Stopped listening, please wait...
000000000:
READY....
Listening...
Stopped listening, please wait...
000000001:
READY....
Listening...
Stopped listening, please wait...
000000002:
READY....
Listening...
Stopped listening, please wait...
000000003: (null)
READY....
```

Figure 3.6 - the contents of the file out

From the figure it can be understood that each speech command has a nine-digit number, followed by the recognized command itself. To isolate and process the commands, a C program was written. The source code for the program is shown in Listing 3.1.

Listing 3.1 - Recognition processing program

```
#include <stdio.h>
#include <stdlib.h>
#include <iostream>
#include <fstream>

using namespace std;

char buffer[100];
char str1[]="Turn on relay number one";
char str2[]="Turn on relay number two";
char str3[]="Turn on relay number three";
char str4[]="Turn off relay number one";
char str5[]="Turn off relay number two";
char str6[]="Turn off relay number three";
int i,j,prev_n,b,lol,maximum,command;
int match[6];
double n;

void main(void)  {
  std::ifstream
sphinx_out("/home/ubuntu/sphinx_pg/out");

  while(1)
   {
  while(! sphinx_out.eof())
  {
  sphinx_out.getline(buffer,sizeof(buffer));
  if(buffer[0]=='0')
 {
    b=0;
    for(i=0;i<6;i++) match[i]=0;
    n++;
    for(i=0;i<=100;i++) {
         if(buffer[i]==' ' && lol==0) lol =i+1;
         printf("%c",buffer[i]);
         }
    for(i=0;i<=100;i++)
         {
         if(str1[i]==buffer[lol+i]) match[0]++;
         if(str2[i]==buffer[lol+i]) match[1]++;
         if(str3[i]==buffer[lol+i]) match[2]++;
```

```
        if(str4[i]==buffer[lol+i]) match[3]++;
        if(str5[i]==buffer[lol+i]) match[4]++;
        if(str6[i]==buffer[lol+i]) match[5]++;
        }

maximum = match[0];
command=0;

  for (i = 1; i <=5; i++)
  {
    if (match[i] > maximum)
    {
       maximum  = match[i];
       command = i;
    }
  }

if(match[command]<40 || match[command]==50) command=6;

switch(command)
{
    case 0:
        system("curl            --request            GET
'http://192.168.0.112/gpio?st=1&pin=5'");
        break;
    case 1:
        system("curl            --request            GET
'http://192.168.0.112/gpio?st=1&pin=4'");
        break;
    case 2:
        system("curl            --request            GET
'http://192.168.0.112/gpio?st=1&pin=2'");
        break;
    case 3:
        system("curl            --request            GET
'http://192.168.0.112/gpio?st=0&pin=5'");
        break;
    case 4:
        system("curl            --request            GET
'http://192.168.0.112/gpio?st=0&pin=4'");
        break;
    case 5:
        system("curl            --request            GET
'http://192.168.0.112/gpio?st=0&pin=2'");
        break;
```

```
}
     printf(" %d",command);
     printf("\n");

   }
   for(i=0; i<100; i++) buffer[i]='\0';
   }
   sphinx_out.clear();
   sphinx_out.seekg(0,std::ios::beg);
   for(i=1;i<=(n*4);i++)
sphinx_out.getline(buffer,sizeof(buffer));
   }
  }
```

The program extracts speech commands from a text file, and compares them with one of six possible, if they match, a GET request is sent to ESP8266, which in turn controls the relay modules. The block diagram of the program is shown in Figure 3.8.

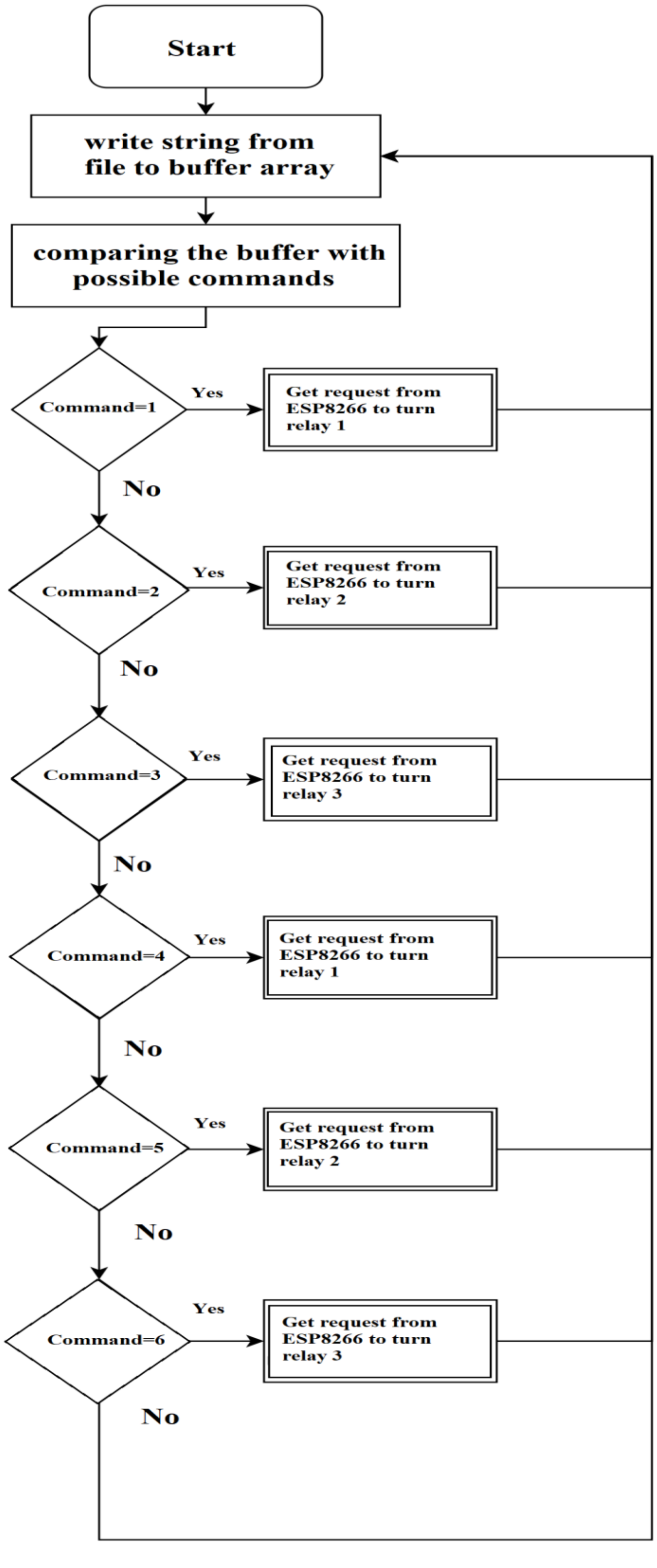
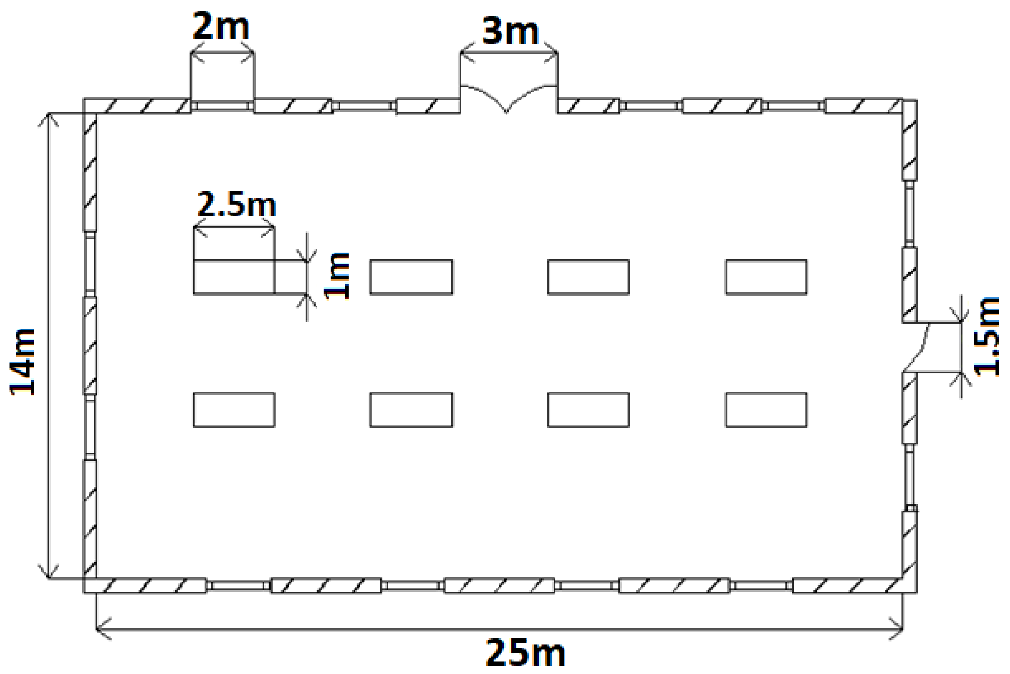


Figure 3.8 - Block diagram of a program for processing speech recognition

4 Life safety

4.1 We will analyze the premises and working conditions in it.

First we give a diagram of the workplace



	Window
	Door
	Work place
	Wall

Figure 4.1 - Plan assembly and assembly shop

For a more complete and detailed analysis of working conditions, the following factors must be considered:

- the number of people in the production workshop is 12 people, of which 8 are men and 4 are women;
- workshop area - 25x14x4 m;
- window size - 2x1.5 m;
- door size - 3 and 1.5 m;
- workplace size 2.5x1 m.

Sources of danger:

- danger of injury - parts of the workplace with sharp edges, cutting edges and cutting tools;
- electrical hazard - electric shock may result in severe personal injury;
- when soldering and varnishing, hazardous harmful substances are released. We set up fancoils of the type above the workstations, which we will choose later after calculating the ventilation. The most hazardous to health are heavy vapors of lead, rosin used in the assembly of equipment.

In the manufacture of equipment, the following technological operations are used:

- picking;
- installation of elements on the board (assembly);

All these operations are carried out in the assembly shop, with the following sanitary standards:

- area per worker not less than $4.5m^2$;
- the volume of the room per worker is not less than $15m^3$;
- the workshop uses artificial lighting;
- the severity category of works in workshop II (GOST 12.1.005-88), since the work is done while sitting, standing and accompanied by physical stress (energy consumption up to 120.00 kcal / hour).

Indoor weather conditions during the year are within the following limits:

- air temperature from 21 to 27 ° C;
- relative humidity from 40.0 to 60.0%;
- distance between workplaces not less than 3 m.

The location of jobs is shown in Figure 4.1. During assembly, Quick-702ESD soldering stations are used.

Table 4.1 - Specifications for the Quick-702ESD Soldering Station

Characteristic	Value
Length, mm	335
Width, mm	253
Height, mm	160
The coefficient of transition to the room of heat from the engine - η	0.8
Supply voltage, V	220
Power consumption, W	520
Soldering temperature, ° C	200 ... 480
Temperature condition of desoldering, ° C	150 ... 500
Temperature range of a desoldering pump, ° C	320 ... 480
Temperature Stability, ° C	± 1
Soldering iron voltage, V	24
Desoldering pump voltage, V	36
Weight, kg	13.3

Power is supplied from a three-phase AC network with a supply voltage of 220V / 380V and a frequency of f = 50Hz.

Workers work in the workshop from 9.00 to 18.00 in one shift. Since they work with small-sized parts, an average accuracy of 0.5–1 mm is required, which corresponds to the third category of visual work. And also for continuous operation during cloudy weather and the winter season (since the days are short), the shop should have artificial lighting.

When assembling the equipment, the workers are soldered, as a result of which heavy pairs of tin and rosin are released, as well as to maintain the microclimate. The workshop should have a ventilation system.

4.2.1 Protection against electromagnetic fields

The sources of electromagnetic radiation include: substations and overhead power lines, induction heating installations, radar devices, communications, television, etc.

The spectrum of electromagnetic fields is divided into frequency ranges:

- constant electrostatic fields due to the formation of electric charges;
- electromagnetic fields of industrial frequency 50 Hz (hertz);
- electromagnetic fields in the frequency range 10 - 30 kHz (kilohertz);
- electromagnetic fields in the frequency range 30 kHz - 300 GHz (gigahertz).

The impact of electromagnetic radiation on the human body leads to a violation of the nervous and cardiovascular systems, to changes in the composition of the blood. The degree of exposure depends on the frequency range, intensity, duration of radiation. Intense super-frequency radiation (above 300 MHz) cause pathology of various organs.

The safety criterion for a person in an electromagnetic field is the permissible electric field E in kilovolts per meter (kV / m) and magnetic field H in miles or microtesla (mT, μT) and amperes or kilo amps per meter (A / m , kA / m).

Electrostatic fields are characteristic of many production processes. The accumulation of electrostatic charges occurs on various surfaces, including clothing of workers, which creates a field of high tension, causing electric discharges. In explosive industries involving the use of combustible gases, flammable and combustible liquids, spark discharges of static electricity can cause an explosion and fire. Under certain conditions, static electricity can cause personal injury.

In accordance with the sanitary and epidemiological rules and regulations of SanPiN 2.2.4.1191-03 "Electromagnetic fields in the production environment (hereinafter - SanPiN 2.2.4.1191-03) and GOST 12.1.045-84" SSBT. Electrostatic fields. Permissible levels at workplaces and control requirements ”(hereinafter referred to as GOST 12.1.045-84) the maximum permissible level of electrostatic field strength (EFS) at workplaces of service personnel when exposed to 1 hour per shift is set to 60 kV / m. When exposed to more than one hour, the value is determined by the calculation method.

Electromagnetic fields of industrial frequency are part of the ultra-low frequency range of the radio frequency spectrum, the most common both in industrial conditions and in everyday life. The industrial frequency range is presented in Russia with a frequency of 50-60 Hz.

Hygienic regulation of electromagnetic fields of industrial frequency is carried out separately by electric magnetic zeros. In accordance with the requirements of these regulatory documents, the maximum permissible levels of electric fields for a full working day is 5 kV / m.

With tensions in the range of more than 5 to 20 kV / m inclusive, the permissible residence time is determined by the formula:

T = 50: E - 2, where

T is the permissible time spent in the electric field at the appropriate level of tension, h;

E - the intensity of the electric field in the controlled area, kV / m

Allowable time spent in an electric field can be realized once or fractionally during a work shift. In the rest of the working time, the electric field should not exceed 5 kV / m.

The maximum permissible levels of magnetic fields of industrial frequency are set depending on the length. If it is necessary for personnel to stay in areas with different magnetic field strengths, the total time for performing work in these areas should not exceed the maximum permissible time for the zone with maximum intensity.

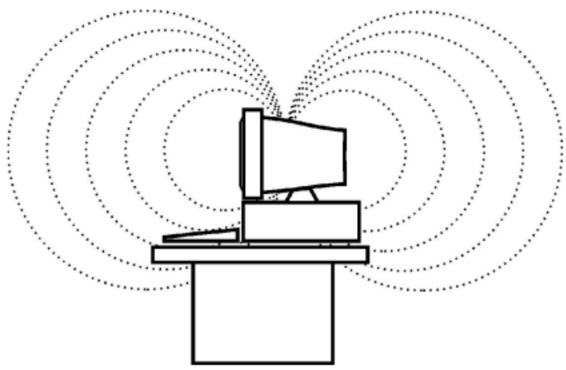


Figure 4.2 The magnetic field lines around the display

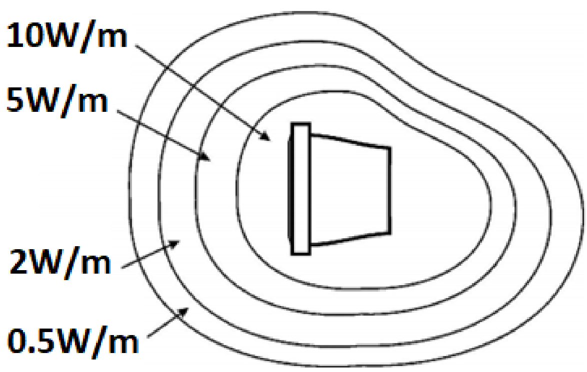


Figure 4.3 Spatial diagram of intensity distribution electric field around the display

4.2.2 Methods and means of protection against electromagnetic fields

Protection from exposure to electromagnetic fields of radio frequencies (EMR RF) is carried out by means of organizational and engineering, medical and preventive measures, as well as the use of personal protective equipment.

Organizational measures include: selection of rational equipment operating modes; limitation of the place and time of personnel in the area of exposure to EMR RF (protection by distance and time), etc.

Time protection provides for limiting the time a person spends in an electromagnetic field and is used when it is not possible to reduce the radiation intensity to acceptable values.

The value of the maximum permissible levels of electric and magnetic components depending on the duration of exposure are given in table. 4.1.

The values of the maximum permissible levels of energy flux density (MPLEFD) depending on the duration of exposure to EMR RF are given in table. 4.2.

Table 4.2.1. Maximum permissible levels of electric and magnetic components in the frequency range 30 kHz - 300 MHz, depending on the duration of exposure

The duration of exposure, T, h	EMPL. W/m			MMPL. A/m	
	0,03-31 MHz	3-30 MHz	30-300 MHz	0,03-3 MHz	30-50 MHz
8.0 and more	50	30	10	5,0	0,30
7,5	52	31	10	5,0	0,31
7,0	53	32	11	5,3	0,32
6,5	55	33	11	5,5	0,33
6,0	58	34	12	5,8	0,34
5,5	60	36	12	6,0	0,36
5,0	63	37	13	6,3	0,38
4,5	67	39	13	6,7	0,40
4,0	71	42	14	7,1	0,42
3,5	76	45	15	7,6	0,45
3,0	82	48	16	8,2	0,49
2,5	89	52	18	8,9	0,54
2,0	100	59	20	10,0	0,60
1,5	115	68	23	11,5	0,69
1,0	141	84	28	14,2	0,85
0,5	200	118	40	20,0	1,20
0,25	283	168	57	28,3	1,70
0,125	400	236	80	40,0	2,40
0,08 and less	500	296	80	50,0	3,00

Engineering measures include: rational placement of equipment; the use of means restricting the supply of electromagnetic energy to the workplaces of personnel (power absorbers, shielding, the use of minimum generator power); designation and enclosure of areas with increased RF EMR level.

Treatment and preventive measures are carried out in order to prevent, early diagnose and treat employee health problems associated with exposure to EMR RF, and include preliminary (upon admission to work) and periodic medical examinations.

Personal protective equipment includes safety glasses, shields, helmets, protective clothing (overalls, bathrobes, etc.).

Depending on the exposure conditions, the nature and location of the sources of EMR RF, various means and methods of protection against radiation can be applied: protection by time; distance protection; radiation source shielding; radiation reduction directly in the radiation source; shielding jobs; individual protection means; allocation of radiation zones.

Table 4.2.2. Maximum permissible levels of energy flux density in the frequency range 300 MHz - 300 GHz depending on the duration of exposure

The duration of exposure, T, h	MPLEFD. μW / sq.cm
8.0 and more	25
7,5	27
7,0	29
6,5	31
6,0	33
5,5	36
5,0	40
4,5	44
4,0	50
3,5	57
3,0	67
2,5	80
2,0	100
1,5	133
1,0	200
0,5	400
0,25	800
0,20 and less	1000

When simultaneously irradiated from several sources of RF EMR, for which the same maximum permissible levels are established, the following conditions must be met:

sum from i = 1 to n ((E1) (2) xT1) <Ee

(sum from i = 1 to n (E1) (2)) (1/2) = Esumm <Empl

amount from i = 1 to n ((M1) (2) xT1) <EE npd

(sum from i = 1 to n (M1) (2)) (1/2) = Msumm <Mmpl

amount from i = 1 to n (ППЭ1T1) <ЭЭпэпд

amount from i = 1 to n (PES1) = EFDsumm <EFDmpl

where Ei is the electric field created by the source of electromagnetic radiation under the i-th number;

Mi is the magnetic field strength created by the EMR source under the i-th number:

EFDi is the energy flux density created by the EMR source under the i-th number:

Ti - the exposure time of the i-th source:

n is the number of sources of electromagnetic radiation.

In the frequency range 30 - 60 kHz (LF) of the remote control, the electric and magnetic field, depending on the exposure time, are determined by the formulas (4.1) and (4.2)

$$\text{EMPL. W/m}=\sqrt{\frac{20000}{T}} \tag{4.1}$$

$$\text{MMPL. A/m}=\sqrt{\frac{200}{T}} \tag{4.2}$$

Protection by distance is applied if it is impossible to reduce the intensity of exposure to other measures, including reducing the time spent by a person in the danger zone. In this case, resort to an increase in the distance between the emitter and maintenance personnel.

Reducing the radiation power directly in the radiation source is achieved through the use of special devices. In order to prevent radiation into the working room, power absorbers (the equivalent of the antenna and the load of EMR RF sources) are used instead of open emitters as a generator load, while the radiation intensity is attenuated to 60 dB or more. The industry produces antenna equivalents calculated for absorption of 5.10.30, 50, 100 and 250 W with wavelengths of 3.1-3.5 and 6-1000 cm.

Protection by distance is also associated with a decrease in tension when moving away from the source. The space of live parts in which the field strength is more than 5 kV / m is called the influence zone. In some cases, combined protection by time and distance is possible. In particular, it is allowed to work on the ground in the zone of influence of overhead lines with a voltage of 400 ... 500 kV without time limits within 20 m from the axis of the support of any type and no more than 90 minutes when working in flight; in the zone of influence of overhead lines with a voltage of 750 kV - no more than 180 minutes within 30 m from the axis of the intermediate support and no more than 10 minutes when working in the span or near the anchor support.

One of the practical ways to reduce the effect of the field on the personnel serving the outdoor switchgear is to reduce the field strength using grounded cables, which are suspended in the working area under current-carrying wires. For example, the use of grounded cables suspended at a height of 2.5 m above the ground under the phases of 750 kV outdoor switchgear busbars reduces the potential in the working area at a height of 1.8 m, i.e. at the level of human growth, from 30 to 13 kV.

4.3 Collective protective equipment and their classification

Collective protective equipment (VHC) is the means used to prevent or reduce the exposure of workers to harmful and dangerous production factors, as well as to protect them from pollution.

Collective protective equipment should be located on production equipment or at the workplace in such a way that it is constantly possible to monitor its operation, as well as the safety of maintenance and repair.

Types of collective protective equipment

1) Means of normalizing the air environment of industrial premises and workplaces: devices for maintaining the standardized value of barometric pressure; ventilation and air purification; air conditioning; localization of harmful factors; heating; automatic control and alarm; deodorization of air.

2) Means of normalizing the lighting of industrial premises and workplaces: light sources; lighting; light openings; light protection devices; light filters.

3) Means of protection against an increased level of ionizing radiation: protective devices; warning devices; sealing devices; protective coatings; devices for capturing and purifying air and liquids; protective equipment during transportation and temporary storage of radioactive substances; safety signs; containers for radioactive waste.

4) Means of protection from high levels of infrared radiation: protective devices; sealing devices; heat insulating devices; ventilation devices; automatic control and signaling devices; remote control devices; safety signs.

5) Means of protection against increased or decreased levels of ultraviolet radiation: protective devices; devices for air ventilation; automatic control and signaling devices; remote control devices; safety signs.

6) Means of protection against high levels of electromagnetic radiation: protective devices; protective coatings; sealing devices; automatic control and signaling devices; remote control devices; safety signs.

7) Means of protection against increased tension of magnetic and electric fields: protective devices; protective grounding devices; insulating devices and coatings; safety signs.

8) Means of protection against increased levels of laser radiation: protective devices; safety devices; automatic control and signaling devices; remote control devices; safety signs.

9) Noise protection equipment: protective devices; soundproofing, sound-absorbing devices; silencers; automatic control and signaling devices; remote control devices.

10) Means of protection from high levels of vibration: protective devices; vibration isolating, vibration damping and vibration absorbing devices; automatic control and signaling devices; remote control devices.

11) Means of protection against an increased level of ultrasound: protective devices; soundproofing, sound-absorbing devices; automatic control and signaling devices; remote control devices.

12) Means of protection against an increased level of infrasonic vibrations: protective devices; safety signs.

13) Means of protection against electric shock: protective devices; automatic control and signaling devices; insulating devices and coatings; protective grounding and grounding devices; automatic shutdown devices; potential equalization and voltage reduction devices; remote control devices; safety devices; lightning rods and arresters; safety signs.

14) Means of protection against high levels of static electricity: grounding devices; neutralizers; moisturizing devices; anti-electrostatic substances; screening devices.

15) Means of protection against low or high temperatures of the surfaces of equipment, materials and workpieces: protective devices; automatic control and signaling devices; thermal insulation devices; remote control devices.

16) Means of protection against increased or decreased air temperatures, temperature extremes: protective devices; automatic control and signaling devices; thermal insulation devices; remote control devices; devices for heating and cooling.

17) Means of protection against mechanical factors: protective devices; automatic control and signaling devices; safety devices; remote control devices; braking devices; safety signs.

18) Means of protection against chemical factors: protective devices; automatic control and signaling devices; sealing devices; devices for ventilation and air purification; devices for removing toxic substances; remote control devices; safety signs.

19) Means of protection against biological factors: equipment and preparations for disinfection, disinsection, sterilization, disinfestation; protective devices; sealing devices; devices for ventilation and air purification; safety signs.

20) Means of protection against falling from a height: fences; protective nets; safety signs.

4.3.1 By types of industrial lighting, they distinguish

1) natural light
2) artificial lighting
3) combined lighting

Calculation of artificial lighting in the workshop

The calculation task is to determine the required power of an electric lighting installation to create a given illumination in a production room.

To calculate the total uniform illumination with a horizontal working surface, the main method is the luminous flux (utilization factor), The luminous flux of the lamp (lm) for incandescent lamps is calculated by the formula (4.3):

$$\Phi_l = \frac{E_n \cdot S \cdot z \cdot k}{N \cdot n \cdot \eta}, \tag{4.3}$$

where E_n is the normalized minimum illumination, 300 lux (for the category of visual work III b for artificial lighting, with a general lighting system);

S - the area of the illuminated room, 350 m^2(25x14);

z - minimum illumination coefficient equal to the ratio Esp / Emin, the value of which for incandescent lamps is 1.15;

k is the safety factor of 1.5 (for assembly and assembly shops with gas discharge lamps);

N is the number of fixtures in the room;

n is the number of lamps in the lamp

η is the coefficient of use of the luminous flux of lamps, depending on the efficiency and curve floor ρfloor and walls ρc, the suspension height of the fixtures and room indicator i.

The room indicator i is determined by the formula (4.4):

$$i = \frac{A \cdot B}{h \cdot (A + B)}, \tag{4.4}$$

where A is the length of the room, 25 m;

B - the width of the room, 14 m;

h - lamp height above the working surface, m.

We calculate the height of the suspension above the working surface according to the formula (4.5):

$$h = H - h_c - h_p, \tag{4.5}$$

where h_c – distance from the lamp to the ceiling, 1 m;

h_p – distance of the working surface above the floor, 1 m;

H – room height, 4 m.

$$h = 4 - 1 - 1 = 2m,$$

$$i = \frac{25 \cdot 14}{2(25 + 14)} = 4{,}487.$$

According to GOST 2239-79, we select the nearest standard lamp and determine the electrical power of the entire lighting system. From the calculated one to minus 10% and plus 20% is allowed, otherwise a different arrangement.

Reflections from the ceiling ρ_{ceil}, ρ_{fl}, floor and ρ_w, walls, take the minimum value of 30, 10 and 10, respectively. We choose the LSP01 2x150 lamp, for which the value of η with i = 4.487 (rounded to 4.5) is 54.

Table 4.2 - Technical specifications, LSP01 2x150 lamp

Lamp type	Number of lamps	Power, W	Length, mm	Width, mm	Height, mm	Weight, kg
LSP01	2	300	1536	674	184	15,5

In each lamp LSP01 2x150 there are 2 lamps of the LHB150 type

Table 4.3 - Technical characteristics of the lamp type LHB150

Lamp type	Power, W	Lamp voltage, V	Lamp current, A	Luminous flux, lm	Lamp length, mm	Diameterof a lamp, mm
LHB150	150	90	1,90	8000	1500	40

We calculate from the formula (4.3) the number of fixtures necessary for lighting formula (4.6):

$$\Phi_l = \frac{E_n \cdot S \cdot z \cdot k}{N \cdot n \cdot \eta}, \tag{4.6}$$

$$N = \frac{300 \cdot 330 \cdot 1{,}15 \cdot 1{,}5}{800 \cdot 2 \cdot 0{,}54} = 20{,}94 \approx 21 pcs.$$

Take 21 pcs and check the validity of our actions according to the formula (4.3). In this case, the actual illumination will be:

$$N = \frac{800 \cdot 2 \cdot 0{,}54 \cdot 21}{350 \cdot 1{,}15 \cdot 1{,}5} = 300{,}52.$$

The deviation of the luminous flux of lamps from value does not exceed 10%:

$$\frac{300{,}52 - 300}{300{,}52} = 0{,}17\%.$$

We will calculate the distance between the fixtures and their layout.

$$L_{AB} = \lambda_{AB} \cdot h,$$

where

$$\lambda_A = 1{,}6;\ \lambda_B = 1{,}8.$$

$$L_A = \lambda_A \cdot h = 1{,}6 \cdot 2 = 3{,}2\ m.$$
$$L_B = \lambda_B \cdot h = 1{,}8 \cdot 2 = 3{,}6\ m.$$

The distance from the extreme lamps to the wall:

$$l_A = \frac{L - 6 \cdot L_A}{2} = \frac{25 - 6 \cdot 3{,}2}{2} = 2{,}9m.$$
$$l_B = \frac{L - 6 \cdot L_B}{2} = \frac{14 - 2 \cdot 3{,}6}{2} = 3{,}5m.$$

The location of the fixtures is shown in Figure 4.4.

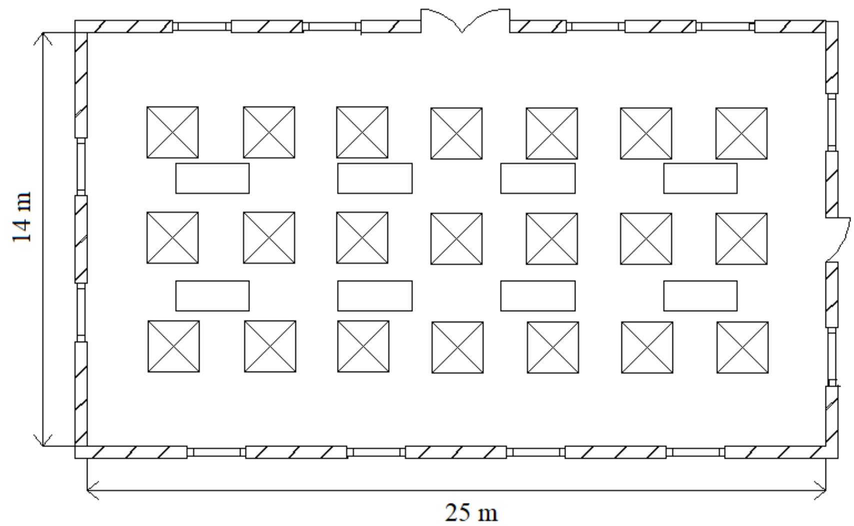


Figure 4.4 - Layout of lighting lamps

Conclusion of life safety

As can be seen from the calculations, 10 luminaires of the LHB 2x150 type are sufficient to provide illumination of 300 lx of a lab room with an area of 350 m^2. Thus, in the evening, artificial lighting will provide the necessary illumination.

5 Feasibility study for the project

5.1.1 Justification for the need for development

Within the framework of this graduation project, a prototype of a voice control system was developed, which can be used in various fields: automation of residential and industrial premises, automation of elevators, household appliances, i.e. wherever there is a human-machine interface.

5.1.2 Goal and objectives

The purpose of this chapter is to calculate the economic efficiency of the commercial production of voice control systems. To achieve the goal, you must complete the following tasks:
- calculation of investment costs;
- calculation of the cost of commercial products;
- calculation of income from sales.

5.1.3 Market analysis

Our products may be of interest to both individuals and legal entities. Among individuals, special needs for such products may be shown by people with disabilities, as controlling lighting, heating, and other devices and processes in a home is difficult. Also, products will interest ordinary people who are interested in new products. Among legal entities, a voice control system may be of interest to construction companies, for installation in residential buildings and cottages, companies engaged in automation of enterprises.

5.2 Financial plan

5.2.1 Calculation of investment costs

We calculate the volume of capital investments necessary for the organization of production. The list of necessary equipment, quantity and total cost are presented in table 5.1.

Table 5.1 - The cost of the necessary equipment and software for the implementation of the project for the production of voice control systems

Name	Quantity, pcs.	Sum, tenge
High-Z S-1400T milling machine	2	1000202
WIN PCNC Professional (CD) Software Suite + I / O Card	1	67246
T-slot table for S-1400T	2	87604
Assembly table	2	34188
Vacuum table 400x1000mm	2	42240
Vacuum pump for a vacuum table 0,70 KW	2	57350
Personal Computer	2	184000
Total		1472830

The total investment in creating a network includes investment in hardware and software, investment in installation and investment in transportation costs, and is determined by the formula (5.1).

$$\sum K = K_t + K_i + K_f, \qquad (5.1)$$

where $\sum K$ – total capital invectment;

K_t – capital investment in equipment;

K_i – capital investment in installation work; (20% of the cost of equipment) K_f – capital investment in transportation costs; (5% of the cost of)

$K_i = 1472830 \cdot 0{,}2 = 294566$ tenge,

$K_f = 1472830 \cdot 0{,}05 = 73642$ tenge,

$\sum K = 1472830 + 294566 + 73642 = 1841038$ tenge.

5.2.2 Calculation of the cost of commercial products

The cost of commercial products includes the following cost items:
- costs of materials and components;
- salaries of employees;
- social tax;
- electricity for industrial needs;
- depreciation deductions;
- overhead costs.

The cost of commercial products are determined by the formula (5.2):

$$C = C_{mat} + PF + SN + D + E + O, \qquad (5.2)$$

where C_{mat} - costs of materials and components;

PF - payroll fund;

SN - deductions for social needs;

D - depreciation;

E - electricity for industrial needs;

O - overhead.

The list and cost of materials and components necessary for the production of products are given in table 1.2.

Table 1.2 - Costs of materials and components

Name	Quantity, pcs.	Sum, tenge
Fiberglass foil 200x300mm	20	23800
Solder paste	10	15600
ABS plastic	20	18950
Chips Allwinner A10	50	100000
Total		158350

The initial number of employees for organizing the production of the device is presented in table 1.3. Here are the average monthly salaries of staff.

Table 1.3 - Salaries of company employees

Performers	Number of persons	Monthly salary, tenge
Supervisor	1	150000
Programmer	2	120000
Engineer	2	80000
Total	5	450000

The payroll is determined by the formula (5.3):

$$Payroll = S_{main} + S_{add}, \tag{5.3}$$

where S_{main} – main salary;
S_{add} – additional salary;

The main salary for the year will be:

$$S_{main} = 450000 \cdot 12 = 5400000\ tenge.$$

Additional salary is 10% of the main salary:

$$S_{add} = S_{main} \cdot 10\% = 5400000 \cdot 0.01 = 540000\ tenge.$$

Then payroll is equal to:

$$Payroll = 5400000 + 540000 = 5940000\ tenge$$

The deductions for social taxes directly depend on the payroll fund. To date, current deduction rates are characterized by the following data: social tax is 9.5% of the payroll tax according to the tax code of the Republic of Kazakhstan, less deductions to the pension fund in the amount of 9.5% of the payroll fund in accordance with the Law on Pension Provision in the Republic of Kazakhstan.

Deductions for social needs are determined by the formula (5.4):

$$S_{need} = 0.095 \cdot \big(Payroll - (Payroll \cdot 0.1)\big), \tag{5.4}$$

$$S_{need} = 0.095 \cdot \big(5940000 - (5940000 \cdot 0.1)\big) = 507\ 870 tenge.$$

Electricity costs are calculated by the formula (5.5):

$$E = W \cdot T \cdot S, \tag{5.5}$$

where W – is the power consumption;
T - the number of hours of equipment operation (8 hours, 245 working days);
S is the cost of kilowatt hours of electricity.

$$W = 1.2 \cdot 2 + 1 \cdot 2 + 0.8 \cdot 2 = 6\ kW.$$
$$T = 8 \cdot 245 = 1960\ h.$$
$$S = 17.8\ tenge/kW \cdot h$$
$$E = 6 \cdot 1960 \cdot 17.8 = 209\ 328\ tenge.$$

Overhead costs make up 9.5% of the total amount and are determined by the formula (5.6):

$$S = 158350 + 5940000 + 507\ 870 + 209\ 328 = 6\ 815\ 548\ tenge. \quad (5.6)$$
$$O = 0.095 \cdot 7\ 002\ 658 = 647\ 477 tenge.$$

We calculate the cost of commercial products according to the formula (5.1):

$S = 158350 + 5940000 + 507\ 870 + 209\ 328 + 647\ 477 =$
$7\ 463\ 025\ tenge.$

Consider the cost structure of commodity products, which are presented in table 5.4.

Table 5.4 - Cost structure of commercial products, tenge

The name of indicators	Plan	weight, %
Materials and components	158 350	1,9
PHOT	5 940 000	73,1
Social Security Contributions	507 870	8,6
Electricity Costs	209 328	2,2
Overhead	647 477	9,1
Total cost	7 463 025	100
Released Products, Units	50	
Unit Cost	149 260	

5.2.3 Calculation of income from the implementation of voice control systems

Income from product sales (1.8):

$$I = N + C, \quad (5.8)$$

where N - the number of implemented voice control systems (50 units);
C - the cost of the product without VAT, i.e. 83.3% of the market value.

The market value of products is calculated by the formula (5.9):

$$C = U + P + VAT, \quad (5.9)$$

where U is the unit cost of production;
P - profit (15% of cost);
VAT - value added tax (12% of U + P).

Calculate the market value of the products:

C= 162491 + 162491 · 0.15 = 186864 tg;
VAT = 186864 · 0.12 = 22424 tg;
U = 186864 + 22424 = 209288 tg.

Thus, the income from the sale of 50 units of products will be:

D = 50 · 186864 = 9343200 tenge.

According to the plan, the number of manufactured and sold products will be 50 units.

5.2.4 Calculation of economic efficiency

As a result of the calculations carried out in section 1.2, the commercial cost of the project was obtained, equal to 1472830 tenge.

The annual economic effect will be calculated by the formula (5.10):

$$E_e = \frac{I-C}{K_t} = \frac{N}{K_t}, \tag{5.10}$$

where C - is the cost of production;
I - income from core activities;
N - net income;
K_t - one-time capital investments.

The normative coefficient of economic efficiency of capital investments is taken equal to = 0.2. The equipment of fixed assets is being introduced new, so capital costs will be equal to the commercial cost of the production of voice control systems.

$$I - C = 9343200 - 7671877 = 1671323 \text{ тенге}$$

$$E_e = \frac{1671323}{1472830} = 1.13$$

The payback period is calculated by the formula (5.11):

$$T = 1/E_e. \tag{5.11}$$

$$T = \frac{1}{1.13} = 0.76\, years.$$

Based on the above calculations, it can be said that the implementation of the developed system will allow to obtain a significant economic effect. The annual economic effect amounted to 1671323 tenge, with a payback period of 0.76 years, which indicates the feasibility of production of this system.

The discount factor is calculated by the formula (5.12):

$$D_f = 1/(1+E)^t, \tag{5.12}$$

where D_f is the discount factor;
E is the discount rate;

t calculation step number (years).
We calculate the discount factor for three years:

$$D_1 = \frac{1}{(1+0{,}2)^1} = 0.83;$$
$$D_2 = \frac{1}{(1+0{,}2)^2} = 0.69;$$
$$D_3 = \frac{1}{(1+0{,}2)^3} = 0.58;$$

PV – current cash flow value throughout the economic life of the project (5.13):

$$\mathrm{PV} = \sum_{t=1}^{n} P_t/(1+E)^t, \qquad (5.13):$$

where E – is the discount rate;
n – the number of periods of the project;
P_t– net cash flow in period t.

The net present value of the project (NPV) is calculated using the formula (5.14):

$$NPV = \sum_{t=1}^{n} \frac{P_t}{(1+E)^t} - I_0, \qquad (5.14)$$

where I_0 – is the amount of initial costs, the amount of investment at the beginning of the project.

PI return on investment index; calculated by the formula (5.15):

$$PI = \frac{\sum_{t=1}^{n} \frac{P_t}{(1+E)^t}}{I_0}. \qquad (5.15)$$

If PI> 1, then the project should be accepted.

DPP Discounted payback period of the project (5.16):

$$DPP = Ca\frac{1}{P_t}, \qquad (5.16)$$

where Ca – is the investment;
P_t – net cash flow in period t.

Using formulas (5.12) - (5.15), we calculate the performance of the enterprise for three years, and write them in table 5.5.

For the implementation of the project 1.488 million tenge will be required. one-time investments. The unit cost of production will be 149 260. Income from the sale of 50 units of products produced during the year will amount to 9.343 million tenge.

Table 5.5 - Enterprise performance

Indicators	1 y.	2 y.	3 y.
Investment costs, million tenge	1,473		
Net cash flow, million tenge	1,219	1,219	1,219
The discount rate,%	20	20	20
Discount coefficient	0,833	0,694	0,579
Net present value (PV)	1,016	0,846	0,705
million tenge	1,094		
Net present value (NPV), million tenge	1,743		
Profitability Index (PI)	-0,457	0,389	1,094

5.3 Conclusion

In this chapter, calculations have been done to determine the cost-effectiveness of the project. According to the plan, the project will pay off in 1 year and 3 months.

Table 5.6 - Summary table of the economic efficiency of the project

№	Indicators	
1	Capital investments	1478300 tenge
2	Income	9343200 tenge
3	Cost of production	8124562 tenge
4	Profit	1671323 tenge
5	Coefficient econ. effectiveness	0.76
6	Payback period	1,2 years
7	Discount payback period	1,5 years

According to calculations, for the implementation of the project 1.488 million tenge will be required. one-time investments. The unit cost of production will be 149 260. Income from the sale of 50 units of products produced during the year will amount to 9.343 million tenge. The summary data is given in table 5.6. Based on the calculation results, we can consider the project profitable and effective.

Conclusion

This work consists of five main parts. The first and second parts are theoretical. In this work concepts and technologies of voice recognition systems were defined, also this work includes descriptions of ARM-minicomputer and ESP 8266 Wi-Fi board. In this diploma has made intellectual voice control system using these devices and development of a system of intelligent servicing visitors.

In the first part showed the main topic is the use of sound, the benefits of a market economy, and the simplification of people's lives.

In the second part of the thesis we analyzed modern models of multiservice in cousines. According to our theoretical data, we conducted a study on the quality of service in multiservice networks and gave a basic understanding of self-access traffic.

In the third part before you start modeling traffic I explained that Ubuntu is undoubtedly the most popular distribution of the Linux operating gadget. Ubuntu is a serious contender within the world of laptop structures and Web servers. This device turned into introduced to life with the aid of Mark Shuttleworth. He based Canonical, which leads the improvement of Ubuntu Linux. That’s why we calculated and made programming part by this operations.

The section "Life safety" analyzed the working conditions and microclimate. The calculation of artificial lighting and air conditioning is presented.

In the economic section, a business plan of the project is developed and to calculate the payback period of 1 year.

Bibliography

1. A. Pererva, V. Ivanova, The Way of the Analyst. The practical guide of an IT specialist, - Moscow, Publisher: Peter, 2019, 282 pp.
2. A. Tyugashev: programming languages. Tutorial. The third generation standard, - Moscow, Publisher: Peter, 2018. 336 pp.
3. Alex Elline: C ++. From lamer to programmer (including C ++ 11), - Moscow, Publisher: Peter, 2019, 480 pp.
4. Alexander Zhadaev: PHP for beginners, - Moscow, Publisher: Peter, 2020, 288 pp.
5. Baranova E. N., Fundamentals of informatics and information protection, - Moscow, Publisher: RIOR, Infra-M, 2013, 192 pp.
6. Baranova EN, Babash AN, Information Security and Information Protection, - Moscow, Publisher: RIOR, Infra-M, 2013, 256 pp.
7. Brian Hogan: HTML5 and CSS3. Web development according to the standards of a new generation, - Moscow, Publisher: Peter, 2014, 320 pp.
8. Brett McLaughlin: PHP and MySQL. An exhaustive guide, - Moscow, Publisher: Peter, 2014, 544 pp.
9. Bruce Tate: Seven languages in seven weeks. A practical guide to the study of programming languages, - Moscow, Publisher: DMK-Press, 2014, 384 pp.
10. Vadim Dunaev: HTML, scripts and styles, - Moscow, Publisher: BHV, 2015, 824 pp.
11. Veselov, Gorbachev, Piskunova: Internet marketing and sales. How to make a site sell, - Moscow, Publisher: Feniks, 2014, 222 pp.
12. Wigers, Beatty: Development of software requirements, - Moscow, Publisher: BHV, 2014, 736 pp.
13. Vitalii Potopakhin: The Art of Algorithmization, - Moscow, Publisher: DMK-Press, 2014, 320 pp.
14. Gashkov S. B., E. A. Primenko, M. A. Cherepnev Cryptographic methods of information protection, Moscow, Publisher: Akademiya, 2010, 304 pp.
15. Georgy Isaev: Designing information systems. Textbook, - Moscow, Publisher: Omega-L, 2015, 424 pp.