STATE COMMITTEE FOR COMMUNICATIONS, INFORMATION AND TELECOMMUNICATION

TECHNOLOGIES OF THE REPUBLIC OF UZBEKISTAN

TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES

To Protect
Supervisor
____
«____»_________2014 y.

GRADUATION QUALIFYING WORK OF BACHELOR

THEME: **Software of voice recognition in LIS**

Graduate	__________ (signature)	**Abdukhakimov S.A.** (Surname)
Supervisor	__________ (signature)	**Rakhmatullayev M.A.** (Surname)
Reviewer	__________ (signature)	________________ (Surname)
Consultant on SLP and ST	__________ (signature)	________________ (Surname)

Tashkent 2014.

The aim of this paper is to improve the efficiency and quality of recognition in speech recognition systems (SRS) with dynamically extensible dictionary teams in ILS.

Bu ishning maqsadi Axborot kutubxona tizimida dinamik kengayib boruvchi lug’atlar orqali ovoz buyruqlarini qayta ishlash tizimining ovoz buyruqlarini tanib olish sifati va samaradorligini oshirish

Целью этой работы является повышение эффективности и качества распознавания в системах распознавания речи (СРР) с динамически расширяемым словарем команд в ИБС.

State Committee for Communications, Information and Telecommunication Technologies of the Republic of Uzbekistan

Tashkent University of Information Technologies

"Faculty of professional training", "Information - library systems"
5320200 "Informatization and library science"

CONFIRM

Chief of Department____________

<<_____>>___________2014 y

Task For Final Qualifying Work

Student **Abduhakimov Sarvar Abduqaxxor o'g'li**

(surname, name, middle name)

1. Theme: Software of voice recognition in LIS.
2. Confirmed by University order №____ from «___» ________ 2014
3. Submission term finished work _______________________________
4. Source data to work: scientific and technical literature, Internet sites, Distance Education service in libraries, programming language.
5. Content of estimated explanatory records (questions list of elaboration): Introduction, Analysis of existing methods speech recognition in information systems, Models and algorithms for speech recognition in LIS, Software implementation of voice processing queries, Security of life activity, Conclusion
6. List of graphic materials: Presentation slides of Microsoft PowerPoint program
7. Date of task issue ____

Supervisor _______________
(signature)

Task received _______________
(signature)

8. Consultants on separate parts of final qualifying work

Units	Name of instructor	Signature date	
		Task issued	Task received
Chapter I. Analysis of existing methods speech recognition in information systems.	Rakhmatullaev M.A.	05.02.2014	05.03.2014
Chapter II. Models and algorithms for speech recognition in LIS	Rakhmatullaev M.A.	06.03.2014	06.04.2014
Chapter III. Software implementation of voice processing queries	Rakhmatullaev M.A.	06.04.2014	06.05.2014
Chapter IV. Safety of vital activity	Agzamova M.	06.05.2014	06.06.2014

9. Schedule of work implementation

№	Title	Term of implementation	Mark of instructor
1.	Analysis of existing methods speech recognition in information systems	01.03.2014	
2.	Models and algorithms for speech recognition in lis	01.04.2014	
3.	Software implementation of voice processing queries	01.05.2014	
4.	Security of life activity	01.06.2014	

Graduate ____________________ «____» ____________ 2014

(signature)

Supervisor _____________________ «____» ___________2014

(signature)

TABLE OF CONTENTS

INTRODUCTION

One of the promising ways of organizing human-machine interaction is the transfer of a computer system instructions user in the format of voice commands Voice interface is necessary component when it comes to creating favorable conditions of life for people with disorders of the musculoskeletal system. Such systems eventually enter into everyday life in the process of implementing the concept of so called "smart homes." Furthermore, possible applications and their production in the complexes control actuators. In the development of this research direction is contributed by such scholars as Rabiner, laid the scientific foundations of statistical speech recognition methods, Wilpon, Lee, Higgins, made a significant contribution to the development of methods Speech recognition commands. Vintsyuk, Karpov, Ronzhin involved speech recognition. Analysis of their work revealed that organization for human-computer interaction using voice teams speech recognition system (SRS) must meet the following requirements:

- Ability to work in real time.
- Sufficient recognition quality (at least 95% correct recognizes commands in the absence of the noise component - signal / noise ratio 25dB).
- Expandable vocabulary SRS without reprogramming.

The latter requirement is due to the fact that to improve the reliability speech recognition systems are often created with carefully selected closed vocabulary of commands that includes fine tuning of grammatical construction and selection of specific words in the composition of the teams.

Existing methods to recognize voice commands do not meet all the stated requirements. This fact determines the relevanceresearch in this direction.

Object of study - the voice signal.

Subject of research - models, methods and algorithms of speech recognition systems, human-machine interaction.

The aim - improving efficiency speech recognition in the dynamically extensible vocabulary of commands. Objectives of the study.

1. Analysis of existing models, methods and algorithms for speech recognition in order to identify the extent of their compliance with modern requirements and the selection of prototypes for their own research.

2. Development of models, methods and algorithms for speech recognition, ensuring the achievement of the following indicators to recognize voice commands:- Speed, adequate for use in real-time (two times faster than real-time for the dictionary in 10 teams):

- High quality of recognition (95% correctly recognized voice commands in the absence of the noise component of the signal / noise February 5 dB);

- Ease of modification commands dictionary: ability to add new words and commands without having to reprogram the system.

3. Programme Implementation of the proposed algorithms and spend experimental studies supporting their effectiveness.

Methods of study. We used the methods of probability theory, stochastic processes, mathematical analysis, digital signal processing, Fourier spectral analysis, optimization theory (dynamic programming) and formal language theory.

The practical value of the results. Recognition method in comparison with the approach using a single method of recognition, you can:
- Loosen the dependence of the rate of recognition of the number of words in the dictionary of commands:

- Use the commands that are poorly recognized recognition algorithm keyword.

Application of the algorithm of recognition of keywords using the proposed likelihood function can improve the computational efficiency of recognition through early clipping unpromising options.

Structure and workload. This work consists of an introduction, four chapters, conclusions, applications and bibliography. The total volume of work is __ pages, including figures and tables.

CHAPTER I. ANALYSIS OF EXISTING METHODS SPEECH RECOGNITION IN INFORMATION SYSTEMS

1.1. The overall structure of the recognition system

Recognition problem considered required extensive research in all areas of the design of speech systems, whereby there have been many projects and techniques. However, in the most general case, the structure of voice control system can be represented in accordance with the stages of recognition.

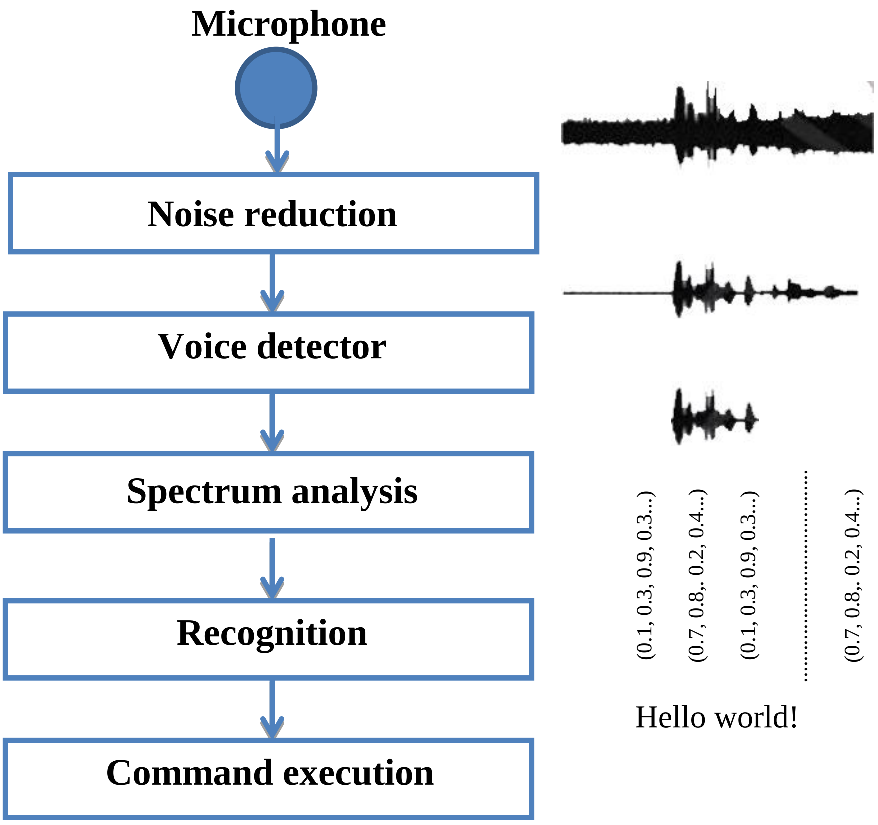


Figure 1.1.1 Structure of a speech recognition system

Let us consider the scheme. Signal is digitized using a microphone. For speech recognition applications is sufficient to digitize a speech signal at 8 kHz. This is due to the limited frequency range of the analog telecommunication lines. However for high quality speech recognition requires approximately 16 kHz frequency 135] (the perceived frequency signal to 8 kHz). Further increase in sampling frequency has no special meaning, as a hook at frequencies above 8 kHz speech signal does not carry useful information, and the system will perceive noise. Bit digital signal quality for recognition must choose at 16 bits per sample.

After digitization in most systems the signal is first fed into the module for noise reduction improves the quality of the signal is due to the removal of noise and distortion introduced by the channel. The detector then selects portions of the voice signal containing speech and these sections are directly entered into the flow measurement signal, which signal performs prefiltering one filter a coefficient (weighting filter) to compensate for physiological characteristics of the vocal tract.

Next is the spectral analysis of the signal which is to extract informative features describing speech signal. To increase the stability of speech recognition systems to the spectral signal conversion can be made adaptive signal filtering.

In step parameterization and selects the transformation parameters obtained in the measurement step. This began to happen association spectral measurements with their first and second derivatives, information about the signal strength, frequency, pitch, etc. One vector. The result is a set of parameters used for recognition.

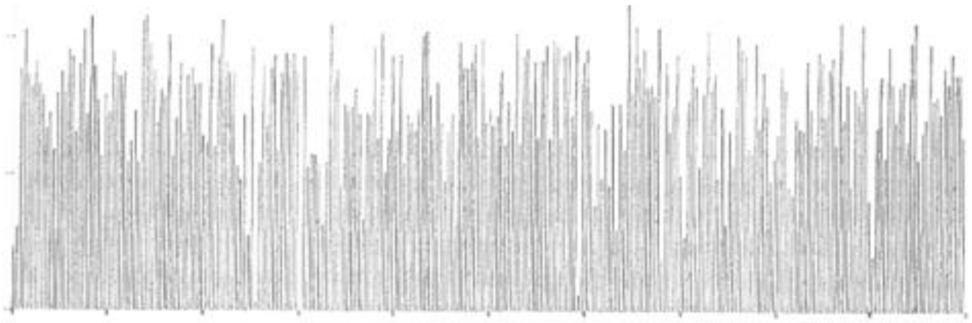
After parameterization of the measurement results of the speech signal produced their initial statistical treatment is to eliminate the correlation of the received parameters and vector quantization. Vector quantization is in spatial association within the group of vectors which are close to each other, and this replacement hruppy its mean value. Thus, reducing the amount of produced information necessary for the recognition subsystems.

At the last stage directly on speech recognition based on data obtained after the statistical processing stage. Today being used in mainly two ways of recognition: speech recognition using HMM , and neural networks. In this paper, as method for research and experimentation method was chosen for HMM. It due to the fact that:

- method provides a powerful mathematical tool for speech recognition;

• There is a deep practical work - developed powerful training and recognition algorithms that provide effective training on large speech databases, as isolated words recognition and continuous speech without adaptation to the speaker.

Sound vibrations formed the entire frequency range, similar to that called noise. Interpretation of the word, taken in the technique differs from generally accepted. A high-pitched whistle (published, for example, the old monitor) can be considered noise in the everyday sense. But this sound has a clearly defined frequency range, and hence it can not be considered as noise in the technical sense of the word.



Picture 1.1.2 Frequency spectrum of the noise

Noise issues moving air - no matter Duno-venie whether this person or the rustle of wind in the microphone. We can say that the gentle sounds of the flute to some extent extracted from the noise produced by air blown man.

Since the noise contains all frequencies, flute can allocate it necessary and enforce them.

If we analyze the discrete values (counts) noise level (and not its frequency spectrum), it turns out random sampling. Good source of noise is high quality random number generator.

In order to obtain a clear spectral characteristics of the sound they need to clean from unnecessary noise.

The input signal is processed by a digital audio filters to get rid of the noise occur when you write the formula.

$$X_i = (X_i - 0{,}9 * X_{i-1})\left[0{,}54 - 0{,}46 * \cos\left((i-6) * \frac{2 * \pi}{180}\right)\right]$$

where Xi - a set of discrete values of the audio signal.

After processing in the signal sought record start and stop, as well as noise is filtered out, the beginning of the fragment will be characterized by a surge signal, if we seek to X0. Respectively, if we look down with Xn, the splash will characterize the end of the fragment. Thus we obtain the start and end of the fragment in an array of discrete signal values. In nonmathematical form, this means that we find the word into the microphone to tell the user that must be averaged with other characteristics of the voice.

In addition to the pitch man feels, and other characteristics of sound - the volume. Physical quantities that most closely matches the volume - this shock pressure (for sounds in the air) and amplitude (digital or electronic submission of sound).

If we talk about the digitized signal, the amplitude - the value you whipping. Analyzing millions of discrete values of the level of the same sound can be said about the peak amplitude, i.e. the absolute value of the maximal makobtained from the discrete values of the sound level. To avoid the distortion caused by signal distortion limit when digital sound recording (this distortion occurs if the value of pi kovoy amplitudy beyond the boundary defined by the format of storage data) optionally sary to pay attention to the value of the peak amplitude. Thus it is necessary to keep the ratio sig-nal/shum the highest attainable level.

The main reason for the different sound volume is time-private pressure exerted by them on the ears. We can say that the pressure waves have different power levels. The wave carrying a large capacity greater force impact on the mechanism of the ears. Electrical signals traveling on wires also transmit power. By wire sound is usually transmitted in the form of alternating voltage and instantaneous power of the sound is proportional to the square of the voltage proportional to determine the full capacity over a period of time, it is necessary to sum all the values of instantaneous power during this period.

The language of mathematics is described by the integral $\int \upsilon_t^2 dt$ where υ_t this voltage at a given time.

Since you are using a sound representation of discrete values, you do not need to take the integral. Simply add up the squares of reference frames Comrade. Mean of the squares of discrete values proportional to the average power.

Since the instantaneous power depends on the square of instantaneous amplitude, it makes sense to similarly have gathered a similar relation between the average amplitude and average power. The way in which this can be done is to determine the average amplitude (rms). Instead, to calculate the average value of the amplitude directly, we first squaring the obtained values, we calculate the mean value of the resulting sets, and then extract it from the root. RMS method applies in the case where it is necessary to calculate the mean for the rapidly changing values. Algebraically, this is expressed as follows schim oorazom: let us N values and x (i) - is the amplitude of the i-th discrete value.

Decibels can only be used to compare two signals. However, measurement of sound in decibels was so convenient ethyl that use some sound ka honors reference standard. This standard is very close to the quietest sound that can only hear people. The loudest sound that is able to hear people louder standard approximately 120 dB (a million million times louder than standard) - its volume is almost the next working volume of a jet engine. Adapted to human hearing the perception of sounds in a wide range of volume.

Decibel scale is also used to measure the loss of sound. If two different sounds with the same energy Propus-tit through some ronnuyu electric circuit or algorithm of digital audio processing, the output of one sound may be 6 dB weaker than the other.

Decibel scale is used to measure the level of noise or distortion that have been added (unintentionally) to any signal.

There are several reasons why using measurements in decibels, it is possible to approximate well the way a person feels volume. First, the sense of hearing a person very close to the logarithm: the perceived difference between the two sounds and the volume depends on the relationship, rather than the difference of the power of each of the th sounds. Although it is not quite correct, it would be nice to be considered as the minimum decibel perceived change in volume.

Another aspect for which measurements in decibels give an accurate picture of human perturbations oschu - is that the perceived loudness is highly dependent on the relative tive power. In particular, the known acoustic illusion called masking. If the sound is formed by two independent components, and one of these components is much louder than the other, then a quieter part will often inaudible. In fact, the human ear "tuned" to the level of a loud sound and a quiet sound is heard much quieter than it actually is. This is especially true in situations where these sounds are very close pitch.

Masking effect - an important tool in modern audio technology. Identifying and selectively discarding the faint sounds that will mask Roval louder, you can simplify the whole sound and ensure that it will be easier to handle. Good understanding of the masking effect will reveal the most audible components of complex sound: it requires to understand what sounds the most large amplitudes are not necessarily heard us better than anyone else.

There are several factors that influence our perception gromkos ty. First, the volume is partially dependent on the pitch. Human ear is more sensitive to a certain average frequency range. Its sensitivity decreases progressively to lower or higher tones. As a result, if you take the average pitch of the sound and the sound of high-pitched, which will Odie tical power, the louder the sound will seem mean-tone.

In addition, the complex sounds a person hears sounds worse of simple tones. In particular, it is very difficult to hear high-frequency noise. Digital conversion method, called erosion (dithering) allows you to convert certain types of errors in less perceptible high frequency noise.

1.2. Methods of the spectral representation of the speech signal

In step signal spectral analysis carried selection of informative signs that describe the speech signal, which is then directly used in the recognition process.

The quality of this method depends strongly on the quality of the whole system: the more informative signs are, the higher the quality of recognition and stability of the system to signal distortion and noise.

There are six major classes of spectral analysis algorithms, currently used in speech recognition systems. Method for estimating signal using filter bank was historically the first method of measurement and parameterization of the speech signal. Linear prediction methods were introduced in the 1970s and dominated until the beginning of the 1980s Currently, widespread methods Fourier transform, linear prediction and cepstral transformation. Cepstral transformation is the de facto standard in the recognition systems, so this method was chosen for prototyping and conduct numerical experiments.

Consider these methods, starting from a set of digital filters and methods of allocating blocks the signal for further conversion of these units by appropriate means.

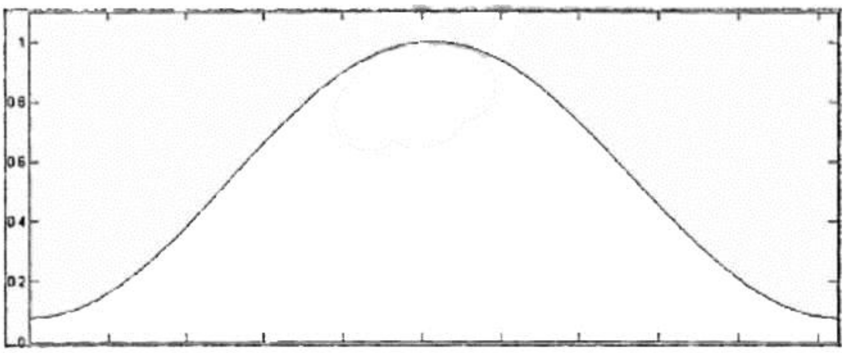
The speech signal is continuously changing due to the work of human vocal tract and thus is non-stationary. However, at time intervals up to 100 ms it moleno considered stationary. It is said that the speech signal is quasi-stationary. In other words, at small time intervals, it can be reliably considered stationary, so the parameters describing the speech signal (the Fourier spectrum, cepstral coefficients, the coefficients of the linear transform), wavelet transform coefficients except, not computed continuously, and this short period of time, called a window. In this case, since all the transformations are calculated on a time range from minus infinity to plus.

Thus, since all the transformations calculated for the range of from minus infinity to plus infinity, in order to overlay the windows do not distort the signal characteristics, the signal is multiplied by the so-called windowing or weighting function which decreases rapidly at the boundaries of the interval in which conversion is performed.

Windows theory was at one time a very active area of research in the field of digital signal processing. There are many types of windows, including rectangular, Hamming, Hanning, Blackman, Kaiser and Bartlett. Currently speech recognition is used primarily Hamming window.

$$w(n) = 0.54 - 0.46\cos(2\pi n\text{-}1) \qquad (1.2.1)$$

for $0 < n < N$ and $w(\text{n}) = 0$ in all other cases; N - the duration of the window in samples.



Picture 1.2.1 The function Hamming

Window need for weighting samples toward the center thereof. This characteristic, together with overlapping assay described datee, performs an important function for smooth changing parameter estimates. It is very important that the width of the main part of the window in the frequency response was minimal or that the process of applying the window could have a decreasing effect on the spectral analysis of the segment.

Parameters are calculated frame by frame. Frame size T_f enforces as the time (in seconds) during which the parameter set is correct. Frame period determines the time between successive calculations of parameters. In practical systems, the frame duration is selected in a range og 10 ms to 20 ms. Values in this range are the choice of optimal solutions between the frequency spectrum changes and the complexity of the system. Corresponding frame length depends entirely on the speed of pronunciation (velocities of the changes in the vocal tract).

Besides, the important interval at which the parameters are calculated

- The number of samples used to calculate N, known as the window size. The window size is T_w, usually measured in units of time (seconds). Average size of the window controls, or smoothed, used in calculating the sum of the parameters.

The duration of the frame and the window size is usually associated in pairs: a window size of 30 ms is used in conjunction with a frame length of 20 ms, while the window size of 20 ms is used in the frame length of 10 ms. Generally, as the short duration of the frame is used to capture the fast dynamics of the spectrum, the window size should be as small parts of the spectrum that were not overly smoothed.

This type of analysis is often called the overlapping analysis, because for each new frame is changed only a fraction of the signal. Amount of overlap to some extent controls the speed parameter changes from frame to frame. Percentage of overlap can be calculated as:

$$O = \frac{T_w - T_f}{T_w} \times 100\%, \qquad (1.2.2)$$

where T_w - window size in seconds, T_f - frame length. If $T_w < T_f$, the percentage overlap zero.

The combination of frame 20 ms and 30 ms window size corresponds to 33% overlap. Some systems use a 66% overlap. One reason for such a large amount of overlap is to reduce the amount of noise in the measurements made by such artifacts as the placement of windows and non-stationary noise channel. On the other hand, excessively smoothed estimates may conceal the true signal variations.

Digital filter bank - one of the basic concepts in speech processing. Filter bank can be regarded as a simplified model of the initial stages of the human auditory system.

We can set the frequency distribution of acoustic / person on a scale of perceived frequency follows.

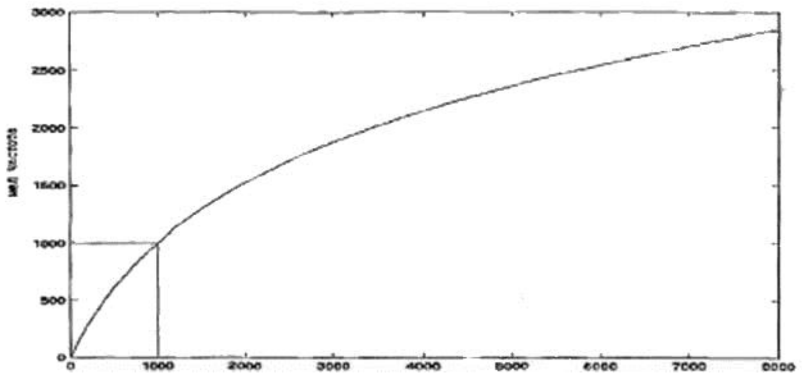
$$\text{Bank} = 13arctg(\frac{0.76f}{1000}) + 3.5arctg(\frac{f}{7500})^2 \qquad (1.2.3)$$

Units such a scale called the critical frequency bands or Watcom-scale.

Increasingly popular approach to this type of distribution in speech recognition is known as Mel-scale:

$$Mel = 2595\lg(1+\frac{f}{700}) \qquad (1.2.4)$$

Mel-scale attempts to display the perceived tone frequency or height on a linear scale. On the frequency range 0 .. 1000 Hz range is linear, after 1000 Hz – logarithmic (picture 1.2.2).



Picture 1.2.2 Mel-scale.

Expression for the bandwidth of each filter is as follows:

$$bw = 25 + 75[1+1/4\ (f/\ 1000)^2]^{0.69} \qquad (1.2.5)$$

This transformation can be used to calculate the band on bookcase perceived frequency filter at a given frequency to Bark-or-Me1 shkacham. Vagk-scale, and

Mcl-scale can be regarded as a transformation into a linear frequency scale, sensuously perceived scale.

Thus is a set of linear phase FIR filters linearly arranged along a Bark-or Mel-scale. The bandwidth of the filters is selected in accordance with the formula (1.2.5) corresponding to the central frequency of the day calculated from the formula (1.2.3) or (1.2.4).

Short-term Fourier transform. Many speech recognition systems are used as parameters of the speech signal, its spectrum, calculated after the application of a Hamming window. The spectrum is obtained by performing a discrete Fourier transform on the signal:

$$y_n = \sum_{k=0}^{N-1} x_k e^{-i\frac{2\pi}{N}kn} \qquad (1.2.6)$$

Where x_k samples of the speech signal, converted Hamming window, N - number of samples in the interval window to - sample number signal, n - number of frequencies in the discrete spectrum. For speech recognition systems are only important amplitude of the complex spectrum obtained by the formula (1.2.6).

The spectrum obtained after the Fourier transformation can be interpreted as the output of the Fourier filter bank.

For calculating the discrete Fourier transform is applied fast Fourier transform algorithm proposed by Cooley and Tukey.

In all the methods of the most important means of continuous analysis of stationary signals is the Fourier transform of continuous-time (CTFT). This approach is a classic, well-proven, so was chosen as a method of parameterization for research. Additionally, currently gaining popularity alternative method based on the continuous wavelet transform.

Spectral Signal Conversion

Since any sound is decomposed into sine waves, we can construct the frequency spectrum of the sound. The frequency spectrum of the sound wave is a graph of amplitude versus frequency.

Frequency - the number of complete cycles are stacked in one second; it is associated with a time period necessary for one cycle. The vertical scale represents amplitude which corresponds to the reference voltage, a current or air pressure.

Since any sound is decomposed into sine waves, we can construct the frequency spectrum of the sound. The frequency spectrum of the sound wave is a graph of amplitude versus frequency.

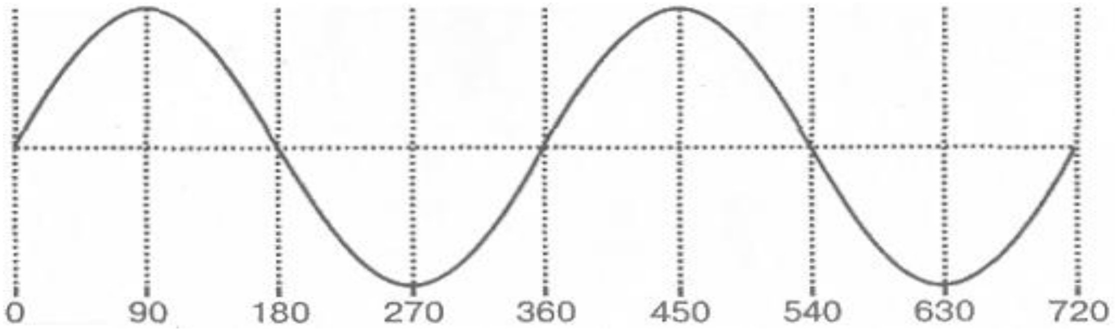
Frequency - the number of complete cycles are stacked in one second; it is associated with periodMatematicheski sinusoid described functions sin () or cos (). Simple function sin (t) has an amplitude equal to unity, the period is equal to 2 seconds and the corresponding frequency equal to 1/2 second cycles. You can convert this record into a more useful form: Asin (2 π) which corresponds to a sine wave with an amplitude A and frequency f.

Here we assume that t is the time (in seconds-dah), f - frequency values tion. When working with digital signal as t is more convenient to use the number of reference. In this case, record Asin (2 π) is a sine wave with amplitude A and frequency π, where S - the sampling frequency. Further, we will work at any given time with groups of N samples and we will be interested in a certain frequency, so I use the records like sin (2 π / N) and cos (2 π / N), which represent the wave of unit amplitude and frequency equal π / N.

The amplitude and frequency do not give the full picture. Delay time can cause displacement waves relative to each other, as shown in Figure 4.3. While these shifts are measured as time delays, it is more convenient to represent them as a fractional part of the period, called phase.

nd time required for one cycle. The vertical scale represents amplitude which corresponds to the reference voltage, a current or air pressure.

Since closely related sinusoids with circles, the phase measured in degrees. One complete cycle - is 360 °. Her timing, the horizontal axis is marked in degrees phase. As 360 ° rotation returns to its original position, so the change in phase by 360 ° leaves the signal unchanged.



Picture 1.2.3. Sine, a marked phase in degrees

Phase changes often occur because of time delays. For example, each cycle of 1000 Hz signal is a 1/1000 second. If you hold the signal to 1/2000 seconds (half), you get a 180-degree shift in phase. Note that this effect is based on the relationship between the frequency and the time delay. If the 250 Hz signal is delayed by the same 1/2000 seconds, that is implemented 45-degree phase shift.

If you add together the two sine waves of the same frequency, we get a new sine wave of the same frequency. This is true even if the two original signals have different amplitudes and phases. For example, Asin (2 ft) and Bcos (2 ft) - two sinusoids with different amplitudes and phases, but I c at a frequency equal.

To measure the amplitude of a frequency necessary to multiply the existing signal at the same frequency sinusoid and add these samples.

If these calculations are repeated for different values off, it is possible to measure the amplitude of all frequencies in the signal. For any integer A_f less easily defined AF value representing the amplitude of the corresponding frequency as a proportion of the total signal. These values can be calculated using the same formula:

$$A_f = \sum_{t=0}^{N-1} s_t * \cos(2\pi t f / n)$$

If we know the values of Af we can recover samples. To restore the signal must combine all zancheniya for different frequencies.

To maintain accurate inverse Fourier transform, in addition to the amplitude and frequency is necessary to measure the phase of each frequency.

This requires complex numbers. You can change the toboggan previously described method of calculation so that it will give a two-dimensional result. Simple komi1 complex number - is a two-dimensional value, but however it is and simultaneously the amplitude and phase.

1.3 Conclusions to Chapter I

In this chapter, the analysis of methods of modeling and automatic speech recognition in the context of the task of developing the voice control. This is primarily due to the lack of a mathematical model of the semantics of the speech signal. Research of modern methods of construction of speech recognition systems has allowed to identify the main (components modules) speech recognition systems, as well as to conclude that the recognition of continuous speech most successfully solved using probabilistic based on hidden Markov models.

CHAPTER II. MODELS AND ALGORITHMS FOR SPEECH RECOGNITION IN LIS

2.1. Speech recognition algorithms

Both acoustic modeling and language modeling are important parts of modern statistically-based speech recognition algorithms. Hidden Markov models (HMMs) are widely used in many systems. Language modeling is also used in many other natural language processing applications such as document classification or statistical machine translation.

Dynamic time warping (DTW)-based speech recognition

Dynamic time warping is an approach that was historically used for speech recognition but has now largely been displaced by the more successful HMM-based approach.

Dynamic time warping is an algorithm for measuring similarity between two sequences that may vary in time or speed. For instance, similarities in walking patterns would be detected, even if in one video the person was walking slowly and if in another he or she were walking more quickly, or even if there were accelerations and decelerations during the course of one observation. DTW has been applied to video, audio, and graphics – indeed, any data that can be turned into a linear representation can be analyzed with DTW.

A well-known application has been automatic speech recognition, to cope with different speaking speeds. In general, it is a method that allows a computer to find an optimal match between two given sequences (e.g., time series) with certain restrictions. That is, the sequences are "warped" non-linearly to match each other. This sequence alignment method is often used in the context of hidden Markov models.

Neural networks

Neural networks emerged as an attractive acoustic modeling approach in ASR in the late 1980s. Since then, neural networks have been used in many aspects of speech recognition such as phoneme classification, isolated word recognition, and speaker adaptation.

In contrast to HMMs, neural networks make no assumptions about feature statistical properties and have several qualities making them attractive recognition models for speech recognition. When used to estimate the probabilities of a speech feature segment, neural networks allow discriminative training in a natural and efficient manner. Few assumptions on the statistics of input features are made with neural networks. However, in spite of their effectiveness in classifying short-time units such as individual phones and isolated words, neural networks are rarely successful for continuous recognition tasks, largely because of their lack of ability to model temporal dependencies. Thus, one alternative approach is to use neural networks as a pre-processing e.g. feature transformation, dimensionality reduction, for the HMM based recognition.

2.2. Speech recognition based on HMM

Hidden Markov models

Modern general-purpose speech recognition systems are based on Hidden Markov Models. These are statistical models that output a sequence of symbols or quantities. HMMs are used in speech recognition because a speech signal can be viewed as a piecewise stationary signal or a short-time stationary signal. In a short time-scale (e.g., 10 milliseconds), speech can be approximated as a stationary process. Speech can be thought of as a Markov model for many stochastic purposes.

Another reason why HMMs are popular is because they can be trained automatically and are simple and computationally feasible to use. In speech recognition, the hidden Markov model would output a sequence of *n*-dimensional real-valued vectors (with *n* being a small integer, such as 10), outputting one of these every 10 milliseconds. The vectors would consist of cepstralcoefficients, which are obtained by taking a Fourier transform of a short time window of speech and decorrelating the spectrum using a cosine transform, then taking the first (most significant) coefficients. The hidden Markov model will tend to have in each state a statistical distribution that is a mixture of diagonal covariance Gaussians, which will give a likelihood for each observed vector. Each word, or (for more general speech recognition systems), each phoneme, will have a different output distribution; a hidden Markov model for a sequence of words or phonemes is made by concatenating the individual trained hidden Markov models for the separate words and phonemes.

Described above are the core elements of the most common, HMM-based approach to speech recognition. Modern speech recognition systems use various combinations of a number of standard techniques in order to improve results over the basic approach described above. A typical large-vocabulary system would need context dependency for the phonemes (so phonemes with different left and right context have different realizations as HMM states); it would use cepstral normalization to normalize for different speaker and recording conditions; for further speaker normalization it might use vocal tract length normalization (VTLN) for male-female normalization and maximum likelihood linear regression (MLLR) for more general speaker adaptation. The features would have so-called delta and delta-delta coefficients to capture speech dynamics and in addition might use heteroscedastic linear discriminant analysis (HLDA); or might skip the delta and delta-delta coefficients and use splicing and an LDA-based projection followed perhaps by heteroscedastic linear discriminant analysis or a global semi-tied covariance transform (also known as maximum likelihood linear transform, or

MLLT). Many systems use so-called discriminative training techniques that dispense with a purely statistical approach to HMM parameter estimation and instead optimize some classification-related measure of the training data. Examples are maximum mutual information (MMI), minimum classification error (MCE) and minimum phone error (MPE).

Decoding of the speech (the term for what happens when the system is presented with a new utterance and must compute the most likely source sentence) would probably use the Viterbi algorithm to find the best path, and here there is a choice between dynamically creating a combination hidden Markov model, which includes both the acoustic and language model information, and combining it statically beforehand (the finite state transducer, or FST, approach).

A possible improvement to decoding is to keep a set of good candidates instead of just keeping the best candidate, and to use a better scoring function (rescoring) to rate these good candidates so that we may pick the best one according to this refined score. The set of candidates can be kept either as a list (the N-best list approach) or as a subset of the models (alattice). Rescoring is usually done by trying to minimize the Bayes risk (or an approximation thereof): Instead of taking the source sentence with maximal probability, we try to take the sentence that minimizes the expectancy of a given loss function with regards to all possible transcriptions (i.e., we take the sentence that minimizes the average distance to other possible sentences weighted by their estimated probability). The loss function is usually the Levenshtein distance, though it can be different distances for specific tasks; the set of possible transcriptions is, of course, pruned to maintain tractability. Efficient algorithms have been devised to rescore lattices represented as weighted finite state transducers with edit distancesrepresented themselves as a finite state transducer verifying certain assumptions.

For simplicity, consider the example of the Markov model for the sound. This model consists of a sequence of states indicated that involve probabilistic instant! " transitions depicted with arrows and probability. Only transitions to the next state and looping. At any time, the model performs vsroyatiospgy transition from one state to another or in the same state, thus there is an acoustic radiation output vector with the output probability distribution corresponding to this state. These probabilities are called emission probabilities. Then some statement describes the sequence of acoustic parameter vectors Y = {y_1, y_2, …, y_n), you can simulate a sequence of discrete stationary states Q = {q1, q2, .., qK}, K < N. with instantaneous transitions between these states and the sequence emitted at the same acoustic vectors Y = {y_1, y_2, …, y_n)

Thus, the hidden Markov model consists of a Markov chain with a finite number of states and transition matrix (transitive) that determine the probability of length of stay in a given state. Markov chain models the temporal changes of the speech signal, as well as a finite set of emission veroyatnosteyb which allow to model the spectral variations in the signal. This approach defines two concurrent stochastic processes, one of which is the basic and unobservable - a sequence of HMM-co states, and we can only judge him by another random process by a sequence of observations.

To use the HMM system is necessary to make some simplifying, but very important assumptions about the speech signal:

- successive observations are statistically

and hence the probability of a sequence of observations sst simply the product of the probabilities of individual observations;

- Although it is a non-stationary process, it is modeled by a sequence of vectors of observations, which are piecewise stationary process;

- proper Markov assumption, ie the assumption that the probability of remaining in a state at time / is dependent only on the state in which the process was at time.

Now consider a simple recognition system. Ideal to have a HMM for each of the possible expressions. However, it is obvious that this is done only for very limited tasks, such as recognition of isolated words in a small vocabulary. Therefore a smaller speech units, which from a linguistic point of view correspond to the phonemes of the language.

Thus, learning is to select the model parameters according to some optimality criterion. Unfortunately, there is no known analytical expression for these parameters. Furthermore, in practice, having a sequence of observations as training data, it is impossible to specify the optimum method of estimating parameters. However, using the iterative procedure, such as the Baum-Welch algorithm or. equivalently, the EM method (mathematical expectation-modification), or gradient methods, which can choose the model parameters so as to maximize the probability of locally.

Should markings. These algorithms belong to the class learning algorithms "unsupervised" because they produce an unobservable parameter estimation of the probability distribution, without requiring pre-marking. At the recognition stage unknown utterance X is necessary to find the most suitable model M.

The method of finding the best model based on dynamic programming algorithm called Vitsrbi.

Education and recognition due to the choice of some optimality criterion. There are several such criteria. They all have a physical meaning and are used in practice. Selected optimality criterion (eg, maximum likelihood or maximum a posteriori probability) affects the model parameters such as the amount of training data and the requirements for computational resources, the accuracy of recognition, the ability to collate data from the training sample. One of the best criteria can be considered as a Bayesian classifier based on a posteriori probability (or classifier to the maximum a posteriori probability, the MAP estimator) that the sequence of acoustic vectors X was generated by Mi model with multiple parameters. Using Bayes' rule can be written in the form of expression.

Acoustic modeling task is to estimate the probability densities are generally independent of other models. Since the probability of M due to it only depends on the parameters and the model. dropping as in the expression can be rewritten as where the set of parameters associated with the model Mi. Thus, both the training and recognition requires estimates of the probability that a sequence is called a global likelihood parameter vectors X at a given

Voice control methods based on HMM

Task in voice recognition is the selection and flow of sound (both voice and not) a predetermined set of voice commands. An example of this command is the phrase "robot gripper to open." The system should not respond to other portions of the speech signal, including those which contain parts of the predefined commands.

Thus, in terms of the CFG and in the context of the problem of voice control audio signal can be represented as follows: (using extended Backus - Naur Form):

Plot audio signal noise = {| silence | extraneous speech | command}.

where the team - the model used voice commands, which can be represented in the following form:

Because the portion of the audio signal may include not only commands for recognition, but also noise and extraneous speech direction voice recognition is considered one of the most complex of Speech recognition directions. Ego due to the fact that in the process recognition not only need to choose the most appropriate phrase According to the dictionary, but also to give an accurate estimate of the likelihood recognition, namely, whether the expression is the recognized fact that Pronunciation, or not. Second chaeg is most difficult since

accurately assess the likelihood impede the following factors:

• various individual characteristics of people: specificity pronunciation, accent, accents, hezitatsin;

• spontaneous speech, which is different type of utterance (type utterance - a way to implement speech phoneme model word). It is generally known that the spontaneous speech inherent usually use of incomplete type casting. If we compare

spontaneous and prepared speech, the prime listening shows that the number of sites in a partial type casting spontaneous speech is greatly increased.

Thick line indicates the most likely sequence of HMM input to the algorithm will yield a list of probabilities of generating each HMM model for each of the discrete-time state. Since keyword can begin and end anywhere signal, this method iterates through all possible pairs of start and end of keyword occurrences and using the Viterbi algorithm calculates the most likely path (P) for the keyword and this segment as if the keyword present there in.

Recognition algorithms for keyword, word recognition seems built into a foreign language. On this basis, the model methods aggregates treated this foreign speech by explicitly modeling an alien speech using secondary models. For this recognition dictionary added "generalized" word. The role of these words is that any signal segment unfamiliar words or nonverbal acoustic event was detected by the system as a chain of one or generic words. For each generic words will be created and trained acoustic model on the body danpyh tagged with the relevant segments of the signal. Besides vocabulary of the system can be expanded to model some common nonverbal speech events (cough, filled pauses).

At the output of the decoder outputs a string consisting of vocabulary words (keywords) and generalized words. Generalized words then discarded, and the rest of the chain recognition result.

When using this method raises the question of the optimal choice of the alphabet generalized words. Explained by the fact that the space of acoustic events modeled alternative models are very large and complex, so the training target and alternative models plays an important role in the effectiveness of the method. As a result, the preparation of models of filler becomes non-trivial process, aimed at a specific set of commands.

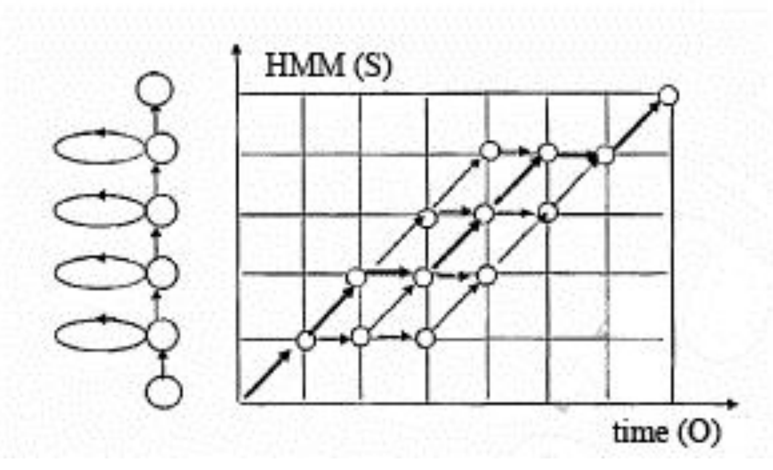
As a tool to solve this problem using algorithms recognition keyword (keyword spotting) and measures the likelihood based on continuous speech recognition methods discussed. Among search algorithms can be divided into two teams of different approaches:

1. Entering search keywords using the values of the local similarity measures (eg, likelihood estimates). The most common method is the sliding window (sliding window).

2. Based on complete alien speech modeling method models filler (filler models)

Sliding window method

Sliding window method is one of the most common methods for solving the problem of determining the occurrence of a word based on the use of units of local similarity measures. The essence of the sliding window method [16. 71] is to determine the keyword occurrences using an algorithm Vitsrbi (Vitcrbi). which is widely used for continuous speech recognition (CSR). This algorithm solves the following problem: given a vector of observations (O) is required to determine the most appropriate sequence of HMM and the transitions between their states for this observation vector (Figure). We will call such a sequence putem1. So on ris.risunok shows all possible ways for the sector signal and the specific sequence HMM; thickened line indicates the most likely path.



Picture 2.2.1 Example of the Viterbis algorithm!

Thick line indicates the most likely sequence of HMM input to the algorithm will yield a list of probabilities of generating each HMM model for each of the discrete-time state. Since keyword can begin and end anywhere signal, this method iterates through all possible pairs of start and end of keyword occurrences and using the Viterbi algorithm calculates the most likely path (P) for the keyword and this segment as if the keyword present in it:

For each found probable path, keywords used, the likelihood function w (confidence measurement or Confidence), based on triggered if the path value, calculated in accordance with the applicable method of assessing the way more than a predetermined value.

Model method placeholders

Recognition algorithms for keyword, word recognition seems built into a foreign language. On this basis, the model methods aggregates [72. 73] treated this foreign speech by explicitly modeling an alien speech using secondary models. For this recognition dictionary added "generalized" word. The role of these words is that any signal segment unfamiliar words or nonverbal acoustic event was detected by the system as a chain of one or generic words. For each generic words will be created and trained acoustic model on the body danpyh tagged with the relevant segments of the signal. Besides vocabulary of the system can be expanded to model some common nonverbal speech events (cough, filled pauses).

At the output of the decoder outputs a string consisting of vocabulary words (keywords) and generalized words. Generalized words then discarded, and the rest of the chain schigaegsya recognition result.

When using this method raises the question of the optimal choice of the alphabet generalized words. Oto explained by the fact that the space of acoustic events modeled alternative models are very large and complex, so the training target and alternative models plays an important role in the effectiveness of the method. As a result, the preparation of models of filler becomes non-trivial process, aimed at a specific set of commands.

Analysis of the methods considered

Worth avenge that both methods in principle equivalent, as cases, the solution adopted is based on the value of the likelihood of the observed data. In the first case, the value used for the likelihood of finding similarity measures and the decision on the basis of their values, in the second case likelihood value is used implicitly - in the unlikely word Recognition will be replaced by another, more believable.

In order to compare different detection algorithms keyword formulate a list of the main characteristics and performance of this class of algorithms: dnktoronezavisimost, speed, quality recognition, the ability to modify the dictionary commands work in noisy conditions.

The main drawback of this method is that the sliding, it enumerates all possible occurrences of the keyword and is applied to each of the possible commands from the vocabulary of commands, which creates a large computational complexity. Additionally, another significant drawback is the poor quality of recognition of some of the commands that caused by the following reasons:

• Parts contain complex commands for phoneme recognition language (for example, [with] well recognized, while the are not succumb to recognize and therefore the value of the likelihood function can vary significantly);

• there are defects in some models of phonemes obtained by virtue of imbalance speech databases (RDB), which produced training, or due to improper training process.

Neelie second limitation can be eliminated by proper selection keyword and quality SRMs then not change the computational complexity succeed. Thus, the method can only be used in voice control with a small vocabulary of commands that do not require the mode real-time or in systems which have significant computing resources (supercomputers, etc.).

But, despite the fact that the method is the aggregate models preferable in terms of quality and speed of recognition, a method of sliding the window is a very important feature - the method does not require additional training models of filler (non-trivial process aimed at specific instruction set) that lets you easily change the dictionary of recognizable commands, allowing the system to adapt to changing requirements or be used for the solutions by a new range of problems.

Thus, from the standpoint of ease of use, the first method (sliding window) looks more attractive because it requires no additional training models of filler, which is non-trivial process. In addition, the method allows to easily change the dictionary of recognizable commands, which is very important, allowing the system to adapt to changing requirements or be used to solve a new range of problems. The disadvantage is high computational complexity and low quality of recognition of certain commands.

In contrast, the same method of sliding window method aggregates models shows good speed and quality by optimizing method set under a given set of voice commands: specially selected generic words.

Criterion for the possibility of modifying the dictionary commands is important at this time, which makes it more relevant research and development of new occurrences of the search algorithm based Keyword use values of the local similarity measures.

Proceeding from the above, the main objective of this work is to improve the efficiency and quality of speech recognition in the RAF with a dynamically extensible vocabulary of commands.

In accordance with this purpose were as follows:

Modelling, methods, and speech recognition algorithms that achieve the following indicators to recognize voice commands:

• speed, adequate for use in real-time (two times faster than real-time for the dictionary in 10 teams);

• high quality of recognition (95% correctly recognized voice commands in the absence of the noise component - S / N ratio 25dB);

• easy modification dictionary commands: the ability to add new words and commands without pereprodammirovaniya system.

2.3. Conclusions to Chapter II

In this chapter, the analysis of methods of modeling and automatic speech recognition in the context of the task of developing the voice control. Demonstrated that the development of speech recognition system is a complex time-consuming task. This is primarily due to the lack of a mathematical model of the semantics of the speech signal. Research of modern methods of construction of speech recognition systems has allowed to identify the main components (modules) speech recognition systems, as well as to conclude that the recognition of continuous speech most successfully solved by using a probabilistic approach based on hidden Markov models.

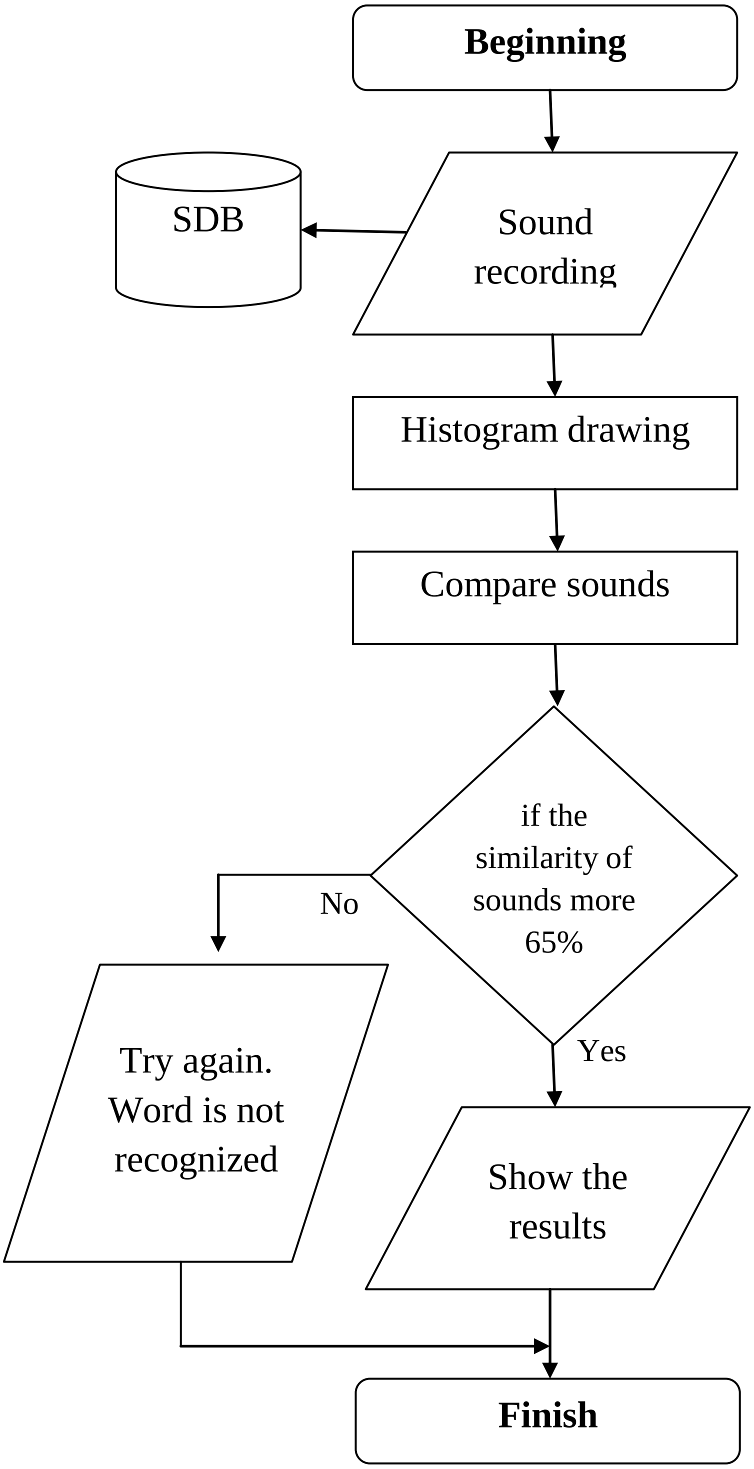
In addition, research has allowed to distinguish two main classes of methods to search for and recognition of speech commands:

1. Entering search keywords using the values of the local similarity measures (eg, likelihood estimates). Most extended method is the sliding window

2. Based on complete alien speech modeling method models aggregates.

CHAPTER III. SOFTWARE IMPLEMENTATION OF VOICE PROCESSING QUERIES

3.1. Algorithm of software



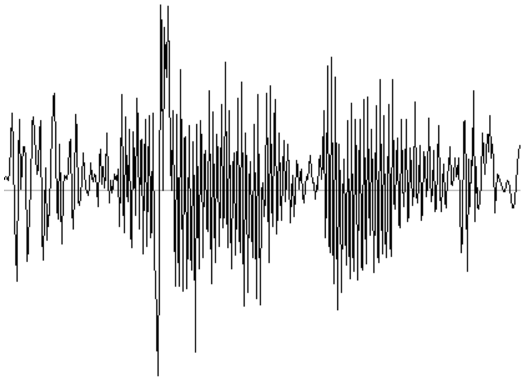
In previous algorithm given speech recognition. Primarily recorded sound and connecting with databases.

Sound recorded with microphone. Microphone used during conversion usually makes no desirable impurity in a signal, such network noise (sound with a frequency of 50 Hz electric wires), loses some of the bass and treble and nonlinear claim tion. AC - converter also makes its own distortions due to nonlinear linear transfer function and DC offset fluctuations. The weakening of the bass and treble often causes problems with pro sequential algorithms parametric spectral analysis.

The main objective of the digitization process is to obtain data of the speech signal with a high signal / noise ratio. At the present time tele tional systems provide this ratio of about 30 dB for proposals under voice recognition, which is more than enough to get such high-performance applications. Changes in converters, and background noise channels, however, cause some problems.

Then draw a histogram of the recorded sound. In this part of the use HMM.

For example recorded voice Alisher Navoiy



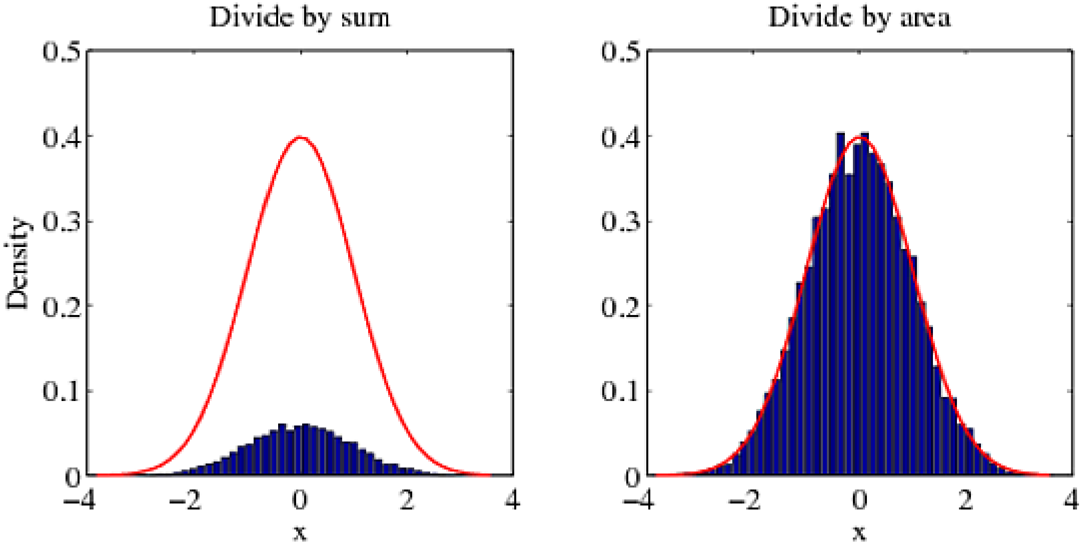
Picture 3.1.1 Histogram of the recorded voice

Process of creating histogram from a normal distribution

```
[f,x]=hist(randn(10000,1),50g=1/sqrt(2*pi)*exp(-0.5*x.^2);
METHOD 1: DIVIDE BY SUM
figure(1)
bar(x,f/sum(f));hold on
```

```
plot(x,g,'r');hold off
METHOD 2: DIVIDE BY AREA
figure(2)
bar(x,f/trapz(x,f));hold on
plot(x,g,'r');hold off
```

Another Method (more straight forward than Method 2) to normalize the histogram is divide by "sum(f*dx)" which expresses the integral of the prob density function.



```
METHOD 3: DIVIDE BY AREA USING sum()
figure(3)
dx = diff(x(1:2))
bar(x,f/sum(f*dx));hold on
plot(x,g,'r');hold off
```

Next step compares sounds with using php code. In this steps used Hamming algorithm.

```
<?php
   // ini_set('memory_limit', '256M');
   class imagediff
```

```
{
    private $image1;
    private $image2;
    function  __construct($img1, $img2)
    {
        $this->image1['path'] = realpath($img1);
        $this->image2['path'] = realpath($img2);
        if($this->image1['path'] === false || $this->image2['path'] === false)
        {
            throw new Exception('Image "'.htmlspecialchars( $this->image1 ? $img2 : $img1 ).'" not found!');
        }
        else
        {
            $this->image1['type'] = $this->imagetyte($this->image1['path']);
            $this->image2['type'] = $this->imagetyte($this->image2['path']);
        }
    }
    private function imagetyte($imgname)
    {
        $file_info = pathinfo($imgname);
        if(!empty ($file_info['extension']))
        {
            $filetype = strtolower($file_info['extension']);
            $filetype = $filetype == 'jpg' ? 'jpeg' : $filetype;
            $func = 'imagecreatefrom' . $filetype;
            if(function_exists($func))
            {
                return $filetype;
            }
            else
            {
                throw new Exception('File type "'.htmlspecialchars( $filetype ).'" not supported!');
            }
        }
        else
        {
            throw new Exception('File type not supported!');
        }
    }
    private function imagehex($image)
    {
        $size = getimagesize($image['path']);
        $func = 'imagecreatefrom'.$image['type'];
        $imageres = $func($image['path']);
        $zone = imagecreate(20, 20);
        imagecopyresized($zone, $imageres, 0, 0, 0, 0, 20, 20, $size[0], $size[1]);
        $colormap = array();
        $average = 0;
        $result = array();
        for($x=0; $x<20; $x++)
        {
            for($y=0; $y<20; $y++)
            {
                $color = imagecolorat($zone, $x, $y);
                $color = imagecolorsforindex($zone, $color);
```

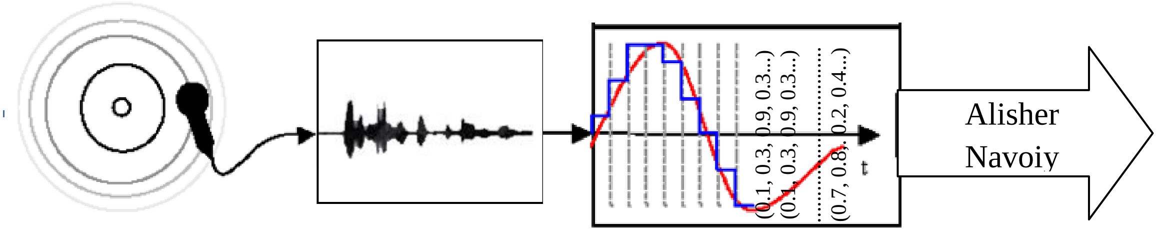
```
                    $colormap[$x][$y]= 0.212671 * $color['red'] + 0.715160 * $color['green'] + 0.072169 * $color['blue'];
                    $average += $colormap[$x][$y];
                }
            }
            $average /= 400;
            for($x=0; $x<20; $x++)
            {
                for($y=0; $y<20; $y++)
                {
                    $result[]=($x<10?$x:chr($x+97)) . ($y<10?$y:chr($y+97)) . round(2*$colormap[$x][$y]/$average);
                }
            }
            return $result;
        }
        public function diff()
        {
            $hex1 = $this->imagehex($this->image1) ;
            $hex2 = $this->imagehex($this->image2);
            $result=(count($hex1) + count($hex2)) - count(array_diff($hex2,$hex1))-400;
        return $result / ( ( count($hex1) + count($hex2) ) / 2 );
        }
    }
?>
```

In the last steps we can see the results. In this module, we used this function which finds histograms with basis that percentage of similarity more than others

```
        $dir = opendir('baza');
        $count = 0;
        while($file = readdir($dir)){
    if($file == '.' || $file == '..' || is_dir('baza' . $file)){
        continue;
    }
    $count++;
}

        for($i=0; $i<=$count-1; $i++){
        $baza = glob("baza/*.png");
    $diff = new imagediff('simple.png', $baza[$i]);
    //print ($diff->diff() * 100 ).'%';
        $prot[$i] = $diff->diff() * 100;
    //echo $prot[$i].'<br>';
        }
        $max = $prot[0];
        for($i=0; $i<=$count-1; $i++){
        if($prot[$i]>$max){
        $max = $prot[$i];
        $recog = basename($baza[$i], ".png");
        $recogimg = $baza[$i];
        }
        }
?>
```

Structure of software

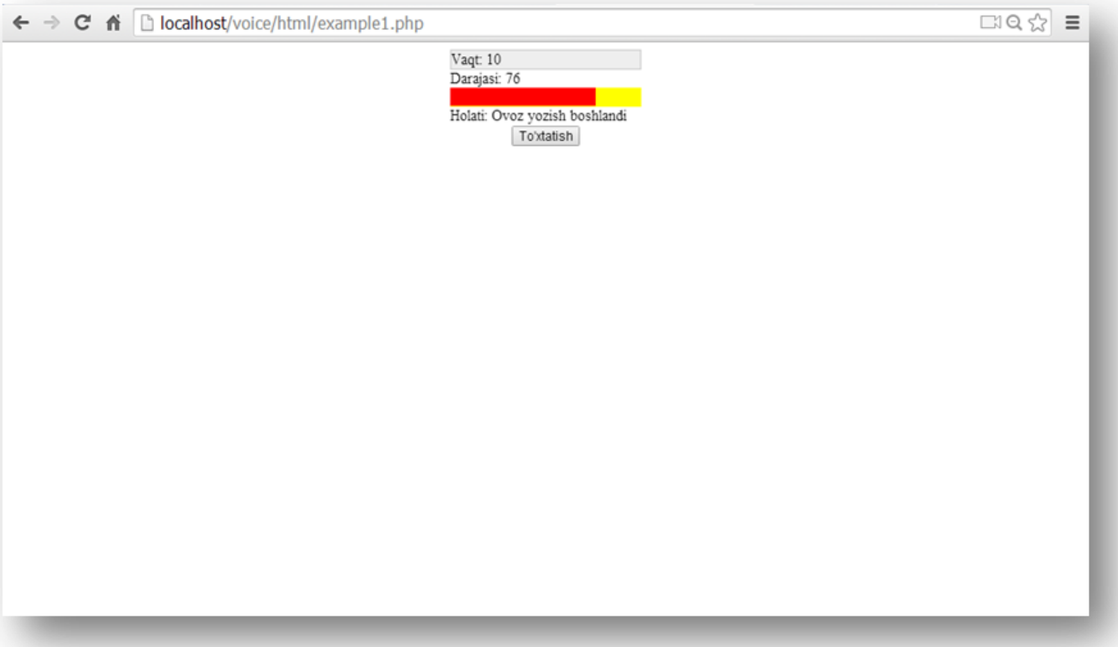


Picture 3.1.1 Structure of software

The software consist of 3 modules:

1. Sound recording module
2. Histogram drawing
3. Compare sounds

Sound recording module



Vaqt: 10
Darajasi: 76
Holati: Ovoz yozish boshlandi

Picture 3.1.2 Sound recording module

This module is written by JavaScript and flash programming. For recording sounds we use the function “start”:

```
//function call to start a recording
```

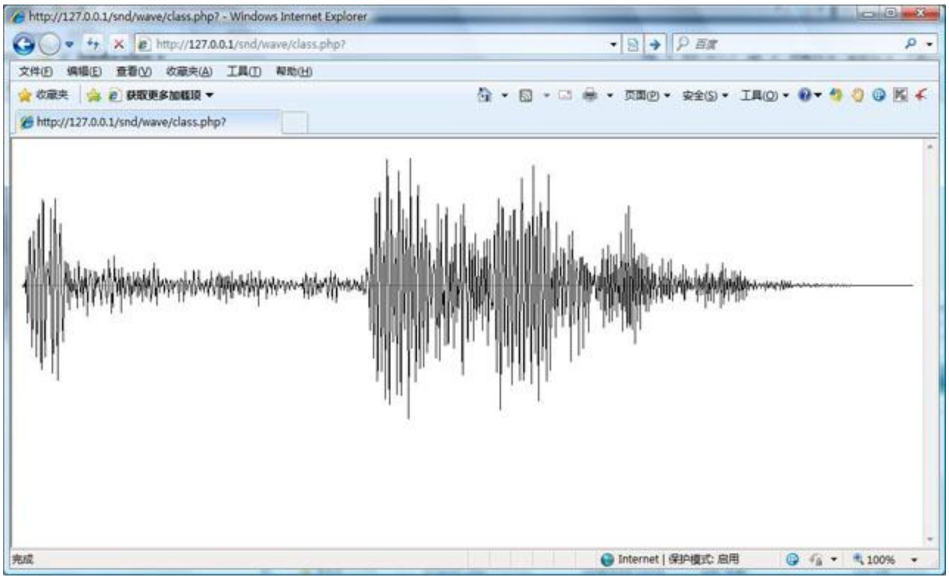
```
    $.jRecorder.record = function(max_time){
                                                    //change z-index to make it top
    $(  '#' + jRecorderSettings['recorderlayout_id'] ).css('z-index', 1000);
    getFlashMovie(jRecorderSettings['recorder_name']).jStartRecording(max_time);
   }
```

For stopping and send data used function “stop” and “send”:

```
//function call to stop recording
    $.jRecorder.stop = function(){
        getFlashMovie(jRecorderSettings['recorder_name']).jStopRecording();
   getFlashMovie(jRecorderSettings['recorder_name']).jSendFileToServer();
        }

    //function call to send wav data to server url from the init configuration
        $.jRecorder.sendData = function(){
    getFlashMovie(jRecorderSettings['recorder_name']).jSendFileToServer();
        }
```

Histogram drawing module


http://127.0.0.1/snd/wave/class.php? - Windows Internet Explorer
http://127.0.0.1/snd/wave/class.php?
文件(F) 编辑(E) 查看(V) 收藏夹(A) 工具(T) 帮助(H)
收藏夹
获取更多加载项
页面(P) 安全(S) 工具(O)
http://127.0.0.1/snd/wave/class.php?
完成
Internet | 保护模式: 启用
100%

Picture 3.1.3 Histogram drawing module

In this module drawing histograms from recorded sounds.

I wrote this module is written by PHP and it uses a class which named example_draw.

Here given the code of the class which creates for drawing histograms:

```
class wave {
    var $fp, $filesize;
    var $data, $blocktotal, $blockfmt, $blocksize;
    function __construct($file) {
        if(!$this->fp = @fopen($file, 'rb')) {
            return false;
        }
        $this->filesize = filesize($file);
    }
    function wavechunk() {
        rewind($this->fp);
        $riff_fmt = 'a4ID/VSize/a4Type';
        $riff_cnk = @unpack($riff_fmt, fread($this->fp, 12));
        if($riff_cnk['ID'] != 'RIFF' || $riff_cnk['Type'] != 'WAVE') {
            return -1;
        }
        $format_header_fmt = 'a4ID/VSize';
        $format_header_cnk = @unpack($format_header_fmt, fread($this->fp, 8));
        if($format_header_cnk['ID'] != 'fmt ' || !in_array($format_header_cnk['Size'], array(16, 18))) {
            return -2;
        }
```

```
			$format_fmt=		'vFormatTag/vChannels/VSamplesPerSec/
VAvgBytesPerSec/vBlockAlign/vBitsPerSample'.($format_header_cnk['Size'] ==
18 ? '/vExtra' : '');
			$format_cnk = @unpack($format_fmt, fread($this->fp,
$format_header_cnk['Size']));

			if($format_cnk['FormatTag'] != 1) {
				return -3;
			}
			if(!in_array($format_cnk['Channels'], array(1, 2))) {
				return -4;
			}

			$fact_fmt = 'a4ID/VSize/Vdata';
			$fact_cnk = @unpack($fact_fmt, fread($this->fp, 12));
	}
```

Comparing sounds by module

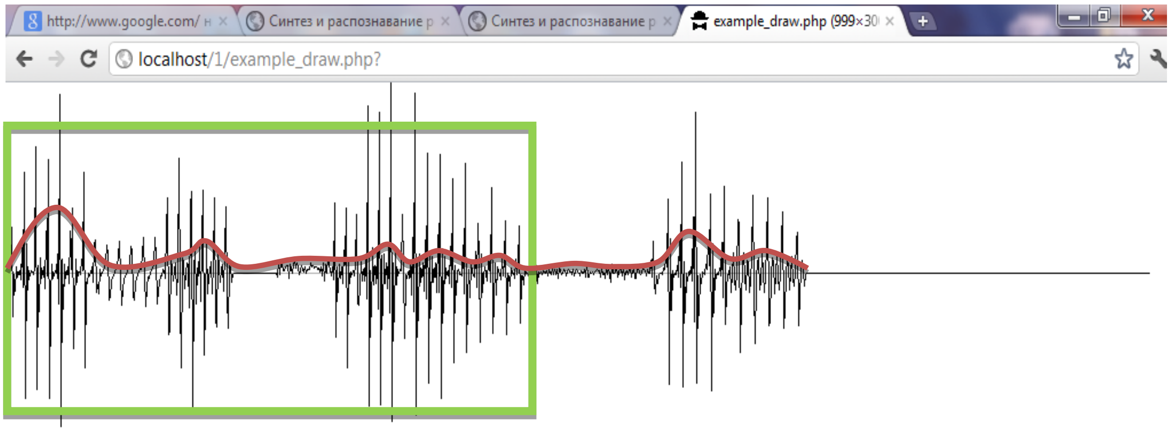
Comparing sounds by module we can see the results. In this module we use this function which finds histograms with basis that percentage similarity more than others.

```
        $count = 0;
        while($file = readdir($dir)){
     if($file == '.' || $file == '..' || is_dir('baza' . $file)){
        continue;
     }
     $count++;
```

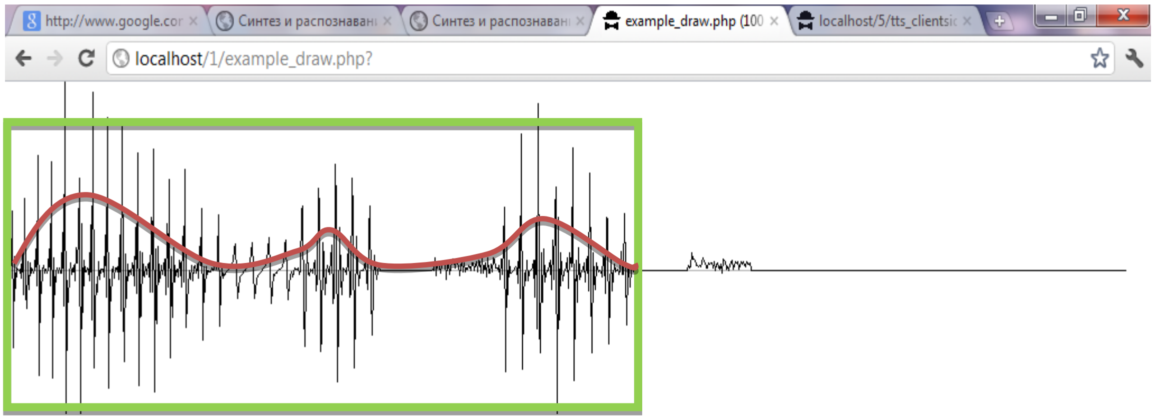
```
}

        for($i=0; $i<=$count-1; $i++){
        $baza = glob("baza/*.png");
    $diff = new imagediff('simple.png', $baza[$i]);
    print ($diff->diff() * 100 ).'%';
        $prot[$i] = $diff->diff() * 100;
    //echo $prot[$i].'<br>';
        }
        $max = $prot[0];
        for($i=0; $i<=$count-1; $i++){
        if($prot[$i]>$max){
        $max = $prot[$i];
        $recog = basename($baza[$i], ".png");
        $recogimg = $baza[$i];
        }
        }
?>
```

In this modules compared histograms which recorded and used this function which finding histograms with basis that percentage similarity more than others.



Picture 3.1.4 Histogram of the word education



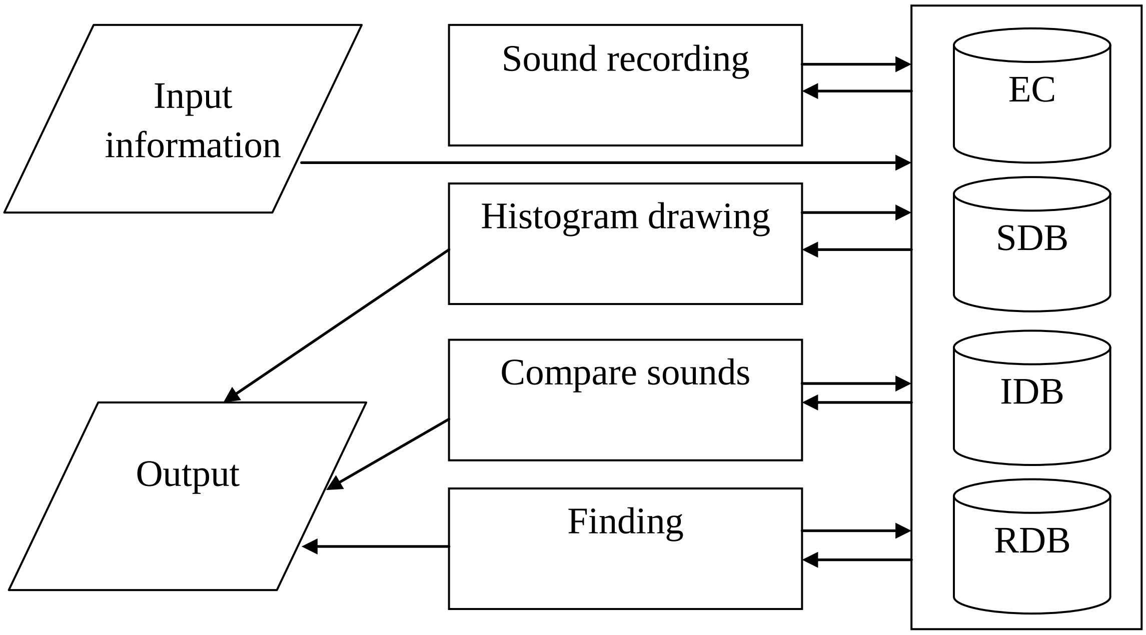
Picture 3.1.5 Histogram of the word educate

There is compered two word histograms. Similarity of histograms equal to 88%.

3.2. Functional structure of the software

The functional structure –means something which reflects relation between functional elements of the process or activity. Elements mean, A functional structure is a structure that consists of activities such as coordination, supervision and task allocation. The functional structure determines how the function performs or operates. In a functional organizational structure the organization is grouped based on functional areas, such as IT, finance, and marketing. Some argue that functional departmentalization allows for greater operational efficiencies in that employees with shared skills and knowledge are grouped together by functions performed. Functional organization is a common type of Functional structure in which the organization is grouped based on specialization by function. Functions can be structured in various ways, and the structure of an Function can determine the modes in which it operates and performs. The functional structure is one structure with associated advantages and disadvantages.

In functional management, the organization is grouped by areas of specialty within different functional areas (such as IT, finance, operations, and marketing), which some refer to as ‘silos,' referring to an image of these areas as vertical and disconnected . Correspondingly, the company's top management team typically consists of several functional heads such as the chief financial officer and the chief operating officer, and communication generally occurs within each functional department, transmitted cross-department through the department heads.



Recorded sounds in the function input information and it will write this sound to SDB (Sound Data Bases)

Function Sound recordings recorded voice commands and send to function histogram drawing.

Sound histograms drawed in the functions histogram drawing and send results to data bases.

In the function compare sounds compared two sounds histograms and results send to RDB (results data bases) and Outputs.

With function Findings found similarity of sounds more 65% and send results to Output

3.3. Conclusions to Chapter III

While writing this chapter, we can get this results:

1. Created algorithm of voice recognition in LIS
2. Created 3 modules and software
3. Created functional structure of the software

CHAPTER IV. SAFETY OF VITAL ACTIVITY

4.1. Ergonomics

Ergonomics (from Greek ergon - work and nomos - law) - a scientific discipline that studies human beings in its activities related to the use of machines. The goal of ergonomics - the optimization of working conditions in the "man-machine" (SChM). Ergonomics pertains to technology and human conditions of its operation. Ergonomic engineering is the most general indicator of the properties and other indicators of technology.

Ergonomics - the science of how people with their different physical characteristics and ways of functioning interact with the operating environment (equipment and machines they use). The goal of ergonomics is to provide comfort, efficiency and safety in the use of computers at the stage of development of keyboards, computer boards, furniture and others working to eliminate physical discomfort and health problems in the workplace. Due to the fact that more and more people are spending a lot of time in front of computer screens, scientists from many fields including anatomy, psychology and the environment, are involved in the study right from the point of view of ergonomics, work environment.

Most furniture manufacturers do not take into account the individual characteristics of the human body in the design of computer workstations. Construction ergonomically equipped places may require additional expenses. If your budget allows, buy ergonomically designed furniture such as chairs, shelves and tables that can be tailored to your individual physical data.

The objectives of ergonomics as an applied discipline are:

- Design the system "man-machine" that is, the distribution of functions between man and machine;
- design of the workspace so that the physical environment consistent with the characteristics of a person;

- designing environment in accordance with the requirements of the operator;
- Design work situations (working hours, rest breaks, etc.).

Psychology, as it follows from the above, is almost an integral part of ergonomics, the crucial problem of the organization of the system "man-machine" by:

- the distribution of functions between man and machine;
- analysis of the functions performed by a person in the "man-machine";
- Information system design, selection of a sensitive channel;
- design of controls;
- the design of workplaces;
- providing facilities maintenance vehicles;
- recruitment and training.

Accounting, ergonomic requirements must be implemented at all stages of the project, and includes:

- Development professiogram defining the goals and objectives of employment, its physiological characteristics, demands on man and technology.
- Analysis and specification of purpose, principles of operation and design technology, its characteristics for the purposes of employment.
- The distribution of functions between man and technology on the basis of quality assessment tasks man and machine and the overall efficiency of the system.
- Establishment of a sequence of operations performed by a man and a determination of the amount and form of presentation of information.
- Orientation assessment, time and accuracy requirements for human activities.

Based on the cited papers is determined by: the composition of experts, their function and organization of work, the composition of displays, controls, jobs and controls; arrangement of displays and controls, posting jobs in production areas.

Relationship with the environment and the working environment.

Workplace - this is the area in which work activity is performed performer or group of performers. Jobs can be individual or collective, universal, specialized and specific.

General requirements to be met in the design of workplaces, the following:

- Adequate working space for human rights;
- optimal working posture;
- adequate physical, visual and auditory communication between man and machine;
- Optimal placement of the workplace in the room;
- acceptable level of the factors of production conditions, the optimal placement of information and motor field;
- a means of protection against occupational hazards.

The construction zone should provide easy access and optimal motor field of the workplace and the optimal coverage of the information field workplace. Viewing angle to the horizontal should be 30-40 °.

Selecting the operating position must take into account the efforts expended by a person, the magnitude of movements, the need for movement, the tempo of operations. Selection of working postures should take into account human physiology, and workplace settings determined by the choice of body position at work (sitting, standing, alternately).

Jobs for work "sitting" are organized under light to moderate work, and with a heavy - working posture - "standing."

The design of the equipment and the work area must be such as to permit the regulation of individual elements to ensure optimum working position.

Equipment design should ensure that it complies anthropometric and biomechanical characteristics of a person on the basis of consideration of the dynamics change the size of heat when it is moved, the range of motion in the joints.

To take into account in the design of equipment anthropometric data should:

- determine which people, which is intended for equipment;
- select a group of anthropometric characteristics;
- establish the percentage of employees who must meet the equipment;
- determine the boundaries of the interval size (forces) to be implemented in hardware.

In the design used anthropometric dimensions of the body, and take into account differences in body size between men and women, national, age, professional. To determine the boundaries of intervals, which take into account the percentage of the population uses a system pertseteley. Construction equipment must allow for at least 90% of consumers.

For operation in the "sitting" used various working seat. Distinguish workers seat. Long and short-term use. General requirements for the seats durable following: sitting posture should provide that reduces the statistical work of muscles, to create the conditions for the possibility of changing the working posture, no hindrances in the systems of the body, to ensure the free movement relative to the work surface, have adjustable parameters, have a semi-soft upholstery. For short-term use is recommended hard chairs and stools of various types.

With the increasing mechanization and automation of production processes are particularly important means of displaying information about the object. The widespread use of received information model, that is organized according to specific rules about the state of the control object. For information models must meet the following requirements:

- The content of the information model should be adequate to display the object management;
- Information Model should provide an optimal balance of information;
- shape and composition information model must be consistent with the work process and the possibilities of man to receive the information.

Practice allows you to map out the sequence of the development of the information model: the definition of the tasks of the system, the sequence of their solutions and sources of information, an inventory control objects and their attributes, the distribution of objects in order of importance, the distribution of functions between automatic and man, the choice of a coding system objects and the drawing up of the overall composition of the model; definition of executive actions of man.

In the process of designing the information model defined location of the media in the workplace, selected label dimensions and layout. Display means are placed in the field of view with the optimum angles and observation areas. Dimensions are determined by observation of signs with the maximum accuracy and speed of perception and brightness character, the values of contrast, the use of color. Considered optimum brightness values, which provide maximum contrast sensitivity. Its value will be greater, the smaller the size of the object of discrimination. The optimum region the contrast value is equal to 60-90%.

In my eyes there is a certain inertia, which requires taking into account the exposure time and the visual signal timing for a sense of separateness signals one after the other. In most cases, the exposure time signal must be at least 50 ms. Each species has its own area of indicators used: Backlit indicators are used to display high-quality information that requires immediate attention of the operator, dial indicators are used to read measured values, integral indicators for combining information about several parameters.

The structure and dynamics of the managed object is usually represented by a chip. In some cases, a display for displaying information and the perception of its team of operators.

In the design of the workplace should be considered rules of economics movement: the work of the two arms of the movement must be simultaneous and symmetrical, the movement should be smooth and rounded, rhythmic and familiar to the employee. Equipment design should take into account the rules on speed and accuracy of movement of workers. For example, the most rapid movement to itself, in the horizontal speed of the hands more than the vertical, the accuracy of motion in the sitting position rather than standing up etc. The controls used in the workplace, must comply with the general requirements ergonogetiki: the direction of motion controls should correspond to the movement of the associated indicator, matching the location of the control sequence of the operator's ease of use, the creation of government agencies in the mechanical resistance, etc. Besides, for each type of pressure corresponds to a different area of use and the special requirements for size, shape, force, etc.

On the workstation operator-communicator (the operator in the control room) are generally used:

- Display means for individual use (imaging units, signaling devices, etc.);
- Controls and input (remote display, keyboard control, individual controls, and so on);
- Device information and communication (modem, telegraph and telephones)
- Documenting and storage device information (printing apparatus, recording, etc.);
- Auxiliary equipment (office equipment, storage for media, local lighting devices).

At the workstation must be provided information and structural compatibility of the technical means of anthropometric and psycho-physiological characteristics of the person.

If your work area should be taken into account not only the factors that reflect the experience, level of training, individual and personal property of the operators, signalers, but the factors that characterize the compliance forms, methods of presentation and data entry capabilities psychophysiological person.

When optimizing the procedures of interaction operators, signalers with the technical means in terms of automation ergonomic factors are the principal causing probability characteristics and hard work. These factors are sensitive to variations in the properties of the individual personality of the operator.

Workshop furniture should be comfortable to perform the planned work operations. The construction work of furniture: table, chair is essential for a healthy environment and a high-efficiency work. Workshop furniture is constructed taking into account the anthropometric data of human, technical, aesthetic, and economic factors.

Included working furniture has important industrial design chair as it determines the position worker and therefore energy consumption and the degree of fatigue. Operating the seat should have the required dimensions corresponding anthropometric rights and be mobile. The most comfortable seats and chairs with adjustable back tilt and height of the seat. By varying the height of the seat from the floor and back angle, you can find a position that most closely matches the labor process and the individual characteristics of the employee.

As a rule, all the surfaces of written and desktops should be at elbow height at working man's position. When you select a table height should be considered a man sitting during work or cost.

Inconvenient table height reduces efficiency and causes rapid fatigue. The lack of sufficient space for knees and feet causes constant irritation of the employee. Minimum working height of the table should be at least 725 mm. Practice shows that the average height for the working height of the desktop received 800 mm. For another employee growth can change the height of a chair or working status of its steps so that the distance from the object to the processing of eye working height was equal to about 450 mm.

Accommodation facilities and the operator's seat in the work area to provide convenient access to the major functional units and units of equipment for technical diagnostics, maintenance inspection and repair, and the ability to quickly take up and leave the work area, with the exception of accidental actuation of the controls and data entry; convenient posture and pose recreation. In addition, the layout must meet the requirements of integrity, and compactness of the technical and aesthetic expression of the working posture.

The display should be placed on a desk or table so that the viewing distance to the screen does not exceed 700 mm (optimal length 450 - 500 mm). Display adjustment should be located so that the angle between the center line of the screen and the horizontal view was 200. The horizontal viewing angle of the screen should not exceed 600. Remote display must be placed on a table or stand, so that the keypad height relative to the floor was 650 - 720 mm. When placing the console on a standard table height of 750 mm is necessary to use a chair with adjustable seat height (450 - 380 mm) and a footrest.

Document (blank) for operator input of data is recommended to have a distance 450 - 500 mm from the eyes of the operator, especially the left, the angle between the display screen and the document in the horizontal plane 40 should be 30 °. Angle of the keyboard should be equal to 15 °.

Screen display and keyboard instruments display panel should be arranged to drop the brightness of surfaces, independent of their location relative to the light source does not exceed 1:10 (recommended value

1: 3). At nominal values of the brightness of the screen image 50 - 100 cd/m2 luminance of the document should be 300 - 500 lux.

The workplace should be equipped in such a way that the movement of the worker would be the most rational, less tedious.

Device documentation and other infrequently used technical tools is recommended to have the right of an operator in the zone of maximum reach and means of communication to the left to release the right hand to take notes.

4.2. Psychophysiological load per person.

In the section of psychophysiological stress the most important is stress and fatigue.

Under stress is understood the emotional state that arises in response to all sorts of extreme exposure.

When stress ordinary emotions are replaced by anxiety, causing disturbances in physiological and psychological terms. This concept was introduced by Hans Selye to refer to non-specific response of the body to any adverse effects. His research showed that the various adverse factors - fatigue, fear, hurt, cold, pain, humiliation, and more in the body cause the same kind of comprehensive response regardless of what kind of stimulus acts on it at the moment. Moreover, these stimuli need not exist in reality. A man reacts not only to the actual danger and the threat or reminder of her.

Human behavior in situations of stress is different from the affective behavior. Under stress a person can usually control their emotions, to analyze the situation, make appropriate decisions.

Currently, depending on the stress factor identify different types of stress, including the pronounced physiological and psychological. Psychological stress, in turn, can be divided into information and emotional. If a person is unable to cope with the problem, do not have time to make the right decisions at the required rate with a high degree of responsibility, ie, when there is information overload may develop informational stress. Emotional stress arises in situations of danger, resentment, etc. Hans Selye identified in the development of stress three phases. The first stage - the alarm reaction - the mobilization phase defenses, which increases the stability with respect to a particular traumatic stress. In this case, there is a redistribution of body reserves: our primary objective is due to minor problems. The second step - the stabilization of parameters derived from the balance in the first phase, fixed at a new level. Externally, the behavior is not very different from the norm, as if everything is adjusted, but internally is overrun adaptive reserves. If the stressful situation persists, there comes the third stage - exhaustion, which can lead to a significant deterioration of health, various diseases, and in some cases death.

Stages of development of the state of stress in humans:

- build-up of tension;
- proper stress;
- Reduction of internal tension.

In its first phase duration is strictly individual. Some people "plant" for 2-3 minutes, and another increase in stress can take place over several days or even weeks. But in any case, the state and behavior of the person who is in stress, change pas' opposite sign. "

So, quiet reserved person becomes fussy and irritable, he may even become aggressive and violent. And the person in real life lively and agile, it becomes dark and taciturn.

In the first stage of stress weakens a person self-control: it gradually loses the ability to knowingly and intelligently regulate their own behavior.

The second stage of the stress state is manifested in the fact that man is a loss of effective self-conscious (full or partial). "The Wave" destructive stress damaging to the human psyche. He can not remember what he said and did, or be aware of their actions, rather vague and incomplete. Many then noted that under stress they have done that in a tranquil setting would not have done. Usually all later regret it very much.

Also, like the first, the second phase in duration strictly individual - from several minutes or hours - several days or weeks. Having exhausted its energy resources (achieving higher voltage observed when a person feels the devastation, fatigue and

Stress conditions significantly affect the activities of man. People with different features of the nervous system to react differently to the same psychological burden. In some people there is increased activity, mobilization of forces, improve business performance. On the other hand, the stress can cause disruption of the sharp reduction of its effectiveness, and total inhibition of inactivity.

Human behavior in a stressful situation depends on many factors, but primarily on the psychological preparation of a person, which includes the ability to quickly assess the situation, the instantaneous orientation skills in unexpected circumstances, a strong-willed discipline and determination, experience, behavior in similar situations.

Methods of dealing with stress

Stress - the feeling that one experiences when she believes that it can not effectively cope with the situation.

If the situation is causing stress depends on us, a more rational to focus on how to change it. If the situation is not up to us to accept and change your perception, your attitude to this situation.

In most situations, the stress goes through several stages.

1. Phase anxiety. This mobilization of energy resources of the body. Moderate stress useful in this step, it leads to higher efficiency.

2. Phase resistance. This is a balanced spending reserves. Outwardly, everything looks normal, people effectively solves the problems faced by them, but if this step takes too long and is not accompanied by relaxation, then, the body works hard.

3. Phase depletion (distress). Man feels weakness and fatigue, reduced performance, dramatically increases the risk of disease. Short time this can still fight at will, but then the only way to restore power - it's a solid rest.

One of the most common causes of stress - the contradiction between reality and perceptions of man.

Stress response is equally easy to run as real events, and existing only in our imagination. In psychology this is called "the law of the emotional reality of the imagination." As psychologists have calculated, about 70% of our experiences come about events that do not exist in reality, but only in the imagination.

By the development of stress can lead not only negative but also positive life events. When something changes dramatically for the better, the body also reacts to this stress.

Usually, the fatigue is understood the reduction in the workability caused by previous work, which has a temporary character. If it occurs during mental activity, talk about mental fatigue. State of fatigue is manifested in changes physiological processes, reducing productivity and techno-economic indicators, change in mental status.

Psychologists say that the development of fatigue, the person has a special psyche, which is called the fatigue - a subjective reflection arising ing processes in the body, leading to fatigue. It appears long before the loss of productivity lies in the fact that there is a special experience painful stress and uncertainty. Manage feels that he could not continue to work properly. Thus there is a disorder of attention - in the development of fatigue, people are easily distracted, becomes sluggish, inactive, or, on the contrary, it appears chaotic mobility insta bility. There are disturbances in the sensory area - for fatigue changes work receptor, for example, there is a visual fatigue - decreased ability to process information coming through the visual analyzer, with the durationtion manual work is reduced tactile and kinaesthetic sensitivity. Lead to abnormalities in the motor area: a slowing of movements, movements appear haste, rhythm disorder, weakening the accuracy and co-ordination of movements, de-automatization of movements. There are defects in memory and thinking, weakened the will, determination, endurance, self-control. With strong fatigue, somnolence.

Intensity of change depends on the depth of fatigue. For example, significant changes in mental status almost there, and with fatigue all these changes is extremely pronounced.

Due to changes in the mental state of a number of physiologists has isolated lyat 3 stages of fatigue. Stage 1: When her with the feeling of fatigue significantly, labor productivity is not reduced. Stage 2 - characterized by a significant reduction

tion of labor and severe mental changes. The third stage, which some scholars regard as acute fatigue, accompanied by the expression experience fatigue.

Utation can be physical (musclenym) or neuropsychiatric (central). Both forms of fatigue combined with hard work, and they can not be strictly separated from one another. Heavy physical work leads primarily to muscle fatigue, and enhanced mental functions or monotonous work is tiringtion of central origin. It should be a clear distinction between exhaustion and fatigue, caused need for sleep.

In addition, determine the primary Utition, which is developing quite rapidly at the beginning of the work shift and is a recognized com insufficient consolidation of skills, it can be overcome in the process, resulting in an "second wind" - a significant increase workable STI. Secondary, slowly progressive fatigue actually tiringtion, which occurs after about 2.5-3 hours from the beginning of the work shift, and to remove it needs rest.

Fatigue or chronic fatigue - another type of fatigue. It is due to the lack of proper rest between each working day, is regarded as a pathological condition. Manifests the general decline in productivity, increased incidence, the slowdown in the cultural and technical level and skills of running, decreased creativity and mental capacity, changes in the cardiovascular system.

According to K.K Platonov are four degrees of fatigue restarting, lung, and severe, each of which requires appropriate methods of struggle. So, to relieve fatigue suf beginning precisely regulate the regime of work and rest. Mild fatigue optionally sary to wait for release and use it effectively. In marked overworked SRI urgent needs rest, better organized. In severe pereutomtion to treatment.

Conclusions to Chapter   IV

In this Chapter learn eronomics and Psychophysiological load per person. I understood importance of this.

CONCLUSION

1. Analysis of methods of modeling and automatic speech recognition in the context of the task of developing the voice control;
2. Research of modern methods of construction of speech recognition systems ;
3. Entering search keywords using the values of the local similarity measures (eg, likelihood estimates). Most extended method is the sliding window;
4. Based on complete alien speech modeling method models aggregates;
5. Created algorithm of voice recognition in LIS;
6. Created 3 modules and software;
7. Created functional structure of the software.

REFERENCES

1. *Bah! L.R. and Jelinek F.* Decoding for channels with insertions, deletions, and substitutions with applications to speech recognition // IEF.E Trans. Informat. Thcoiy. 1975. Vol. IT-21, pp. 404-411.
2. *Baker J.K.* The DRAGON system - An overview // lEEETrans. on Acoust. Speech Signal Process. 1975. Vol. ASSP-23. No. 1. pp. 24-29
3. *Baum L.F., Petrie T.* Statistical inference for probabilistic functions of finite state Markov chains *U* Ann. Math. Stat. 1966. Vol.37. pp. 1554-1563.
4. *Baum L.E., Egon J.A.* An inequality with applications to statistical estimation for probabilistic functions of a Markov proccss and to a model for ecology // Bull. Ainer. Meteorol. Soc. 1967. Vol. 73. pp. 360-363.
5. *Baum L.E., Petrie T.. Soules G.,* and Weiss N. A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains // Ann. Math. Stat. 1970. Vo! 4!. No. I. pp. 164-171.
6. *Jelinek F.* A fast sequential decoding algorithm using a stack // IBM J. Res. Develop., 1969. Vol. 13. pp 675-685.
7. *Jelinek F., Bald L.R., and Mercer R.L.* Design of a linguistic statistical decoder for the recognition of continuous speech /IEEE Trans. Informat. Theory. 1975. Vol. IT-21, pp. 250-256.
8. *Levinson S.E., Rabiner L.R., and Sondhi MM.* An introduction to the application of the theory- of probabilistic function of a Markov process to automatic speech recognition // Bell Syst. Tech. Journal, Apr. 1983. Vol. 62, no.4, pp. 1035-1074.
9. *Bourlard H.. Morgan N.* Hybrid connectionist models for continuous speech recognition // In: C.H. I ce, F.K. Soong, K.K. Paliwal (Eds), Automatic Speech and Speaker Recognition: Advanced Topics, The Kluwcr International Scries in Engineering and Computer Scicncc, Kluwcr Academic Publishers, Boston, USA 1996.

10. *Bourlard H., Morgan N.* Conncctionist Speech Recognition. A Hybrid Approach // The Kluwer International Series in Engineering and Computer Science, Vol. 247, Kluwer Academic Publishers, Boston, 1994.

11. *Higgins A.* Keyword recognition using template concatenation. Acoustics. Speech, and Signal Processing, IEEE International Conference on ICASSP, 1985.

12. *Rumelhart D. Ii., Hinton G.* A'.. *Williams R. J.* Learning internal representations by error propagation // In: Rumelhart, D. E., G. E. Hinton, (cds), Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol.l Foundations., chapter 8. Bradford Books/MIT Press. Cambridge, MA. 1986 ISBN 0- 262-18120-7.

13. *Watbel A., Hanazawa T.* Phoneme Recognition Using Time-Delay Neural Networks // IEEE Transaction on Acoustic Spccch Signal Processing Vol. 37, 1989, pp. 328-339.

14. *Almeida L.H.* A Learning Rule for Asynchronous Perceptrons with Feedback in a Combinatorial Environment // In: Г' International Conference on Neural Networks.

15. *Hu-Hua Liu.* Environmental Adaptation for Robust Speech Recognition. The Ph. D. diesis. Carnegie Mellon University. USA. 1994.

16. *Richard C. Rose, Douglas B. Paid.* A hidden markov model based keyword recognition system IEEE. ICASSP 90, vol. l.pp. 129-132, Apr. 1990.

17. *Goodwin MM.* Adaptive Signal Models: Theory. Algorithms, and Audio Applications. The Ph. D. thesis. University of California. USA. 1997.

18. *Morcnn P.* Speech Recognition in Noisy Environments. The Ph. D. thesis. Carnegie Mellon University. USA. 1996.

19. *Churbanov A,* Wintcrs-Hilt S. Implementing EM and Viterbi algorithms for Hidden Markov Model in linear memory. The Research Institute for Children, 2008.

20. *Stew Young.* The application of hidden Markov models in speech recognition. Foundations and Trends in Signal Processing archive Volume 1 , Issue 3 (January 2008). Pages: 195-304.
21. *./. A. Bilmes,* “Graphical models and automatic speech recognition” in Mathematical Foundations of Speech and Language: Processing Institute of Mathematical Analysis Volumes in Mathematics Series. Springer-Verlag, 2003.
22. *S. S. Chen and R. Gopinath,* "Gaussianization," in NIPS 2000, Denver, CO, 2000.
23. 103 *S. S. Chen and R. A. Gopinath,* "Model sclcction in acoustic modelling." in Proceedings of Eurospeech, pp. 1087-1090, Rhodes. Greece, 1997.
24. *L. Deng. A. Acero. M. P/umpe, andX. D. Huang,* "Large-vocabulary' speech recognition under adverse acoustic environments." in Proceedings of ICSLP, pp. 806- 809, Beijing, China, 2000.
25. www.google.com
26. www.wikipedia.org
27. www.github.com
28. www.code.google.com