



ISSN: 2350-0328

**International Journal of Advanced Research in Science,
Engineering and Technology**

Vol. 8, Issue 10 , October 2021

The effectiveness of the implementation of speech command recognition algorithms in embedded systems

Shukurov Kamoliddin

Senior teacher, Department of Artificial Intelligence, Faculty of Computer engineering, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan.

ABSTRACT: Speech is the basis of human-computer interfaces, smart house, IoT and control systems. Implementing a real time voice control system through speech commands recognition in different environments requires simple algorithms and special purpose systems. The article analyzes the existing speech commands recognition algorithms in embedded systems. A simple and efficient algorithm for voice control systems through limited speech commands has been proposed.

KEY WORDS: Speech commands recognition, human-computer interface (HCI), hardware-software platform, embedded system, short time Fourier transform (STFT), Singular value decomposition (SVD), Mel Frequency Cepstral Coefficients (MFCC), Vector quantization (VQ), Dynamic time warping (DTW), Hidden Markov model.

I. INTRODUCTION

Speech command control is widely used in various fields, speech control of devices and equipment, voice assistant in call centres, smart home and Internet of Things (IoT), speech identification, speech control interface for people with disabilities and can be applied other fields [1]. In this case, the processing of speech signals involves complex recognition algorithms.

Implementation of such human-computer interfaces requires special hardware and software platforms. However, with the increasing complexity of speech processing algorithms and the number of speech commands, the computing resource and memory size of the hardware-software platform increase dramatically. In addition, most human-computer interfaces work in real time. When recognizing speech commands in different environments, speech signals are affected by external noise and interference. Real-time mode, in turn, requires speed, compactness, and ease of use [2, 3, 4, 5, 6, 7, 8, 9].

The limitations and requirements for computing resources mentioned above require the implementation of special simplified algorithms and hardware-software solutions for processing speech signals.

II. LITERATURE SURVEY

The simplest view of a point signal for analysis is that of extracted words. Usually, speech words and their dictionaries will be limited. In modern complex speech recognition systems, the recognition of key words is carried out mainly. This is very useful in speech command system control systems.

Speech recognition algorithms and models have been developed for the Romance-Germanic family and for some Asian languages. However, the models and algorithms developed for these languages are not appropriate for Uzbek. In addition, their implementation in hardware and software is still relevant today and requires special approaches [10, 11, 12].

III. METHODOLOGY

A speech signal is a model of a complex dynamic process, the analysis of which relies on several indicators (features) that describe the signal and its fragment. These main features of speech signals are: formant frequency, basic tone frequency, spectral components.

Automatic speech recognition systems are implemented in the following stages: Pre-processing, feature extraction, training, and recognition.

The pre-processing consists of signal reception, conversion from analog to digital, filtering. Feature extraction - removes areas of silence, passes through windows, calculates spectral values and MFCC parameters. Training and Recognition - where each speech word is taught and recognized by machine learning and artificial intelligence algorithms according to the parameters obtained.

Because speech signals are complex signals, there are problems with processing these signals, storing them in memory, and transmitting them through communication channels. To solve these problems, scientists have proposed approaches such as signal compression in the processing of speech signals, extraction features from the signal, working with signal spectral values. All of this is aimed at simplifying the signal processing process [13, 14, 15, 16].

The method of finding cepstral coefficients (MFCC - Mel frequency cepstral coefficients) for extraction certain features in speech signals is widely used, and this method is very common in automatic speech recognition. The extraction of the MFCC characteristic properties is determined by calculating the power spectrum, Mel-Spectrum, and Mel-Cepstral (Figure 1). The main advantage of the algorithm is that it allows to recognize and implement speech with a high degree of accuracy [17].

In the first stage of the algorithm for calculating the MFCC features, the speech signal recorded from the microphone is divided into 25 msec frames. With the exception of the first frame, each frame includes the last 10 ms of the previous frame. This process is done until the end of the signal. Since in most cases the sampling frequency of the speech signal is 16 KHz, the frame length is $N = 256$ and the shift length is $M = 160$. When splitting speech signals into frames, it is recommended that the optimal overlap typically cover 50-75% of the frame length. In the second and third stages, a weight window is applied to reduce distortions on the extraction frames and grind them, followed by a spectral replacement procedure. In practice, the use of Hemming window as a window is common.

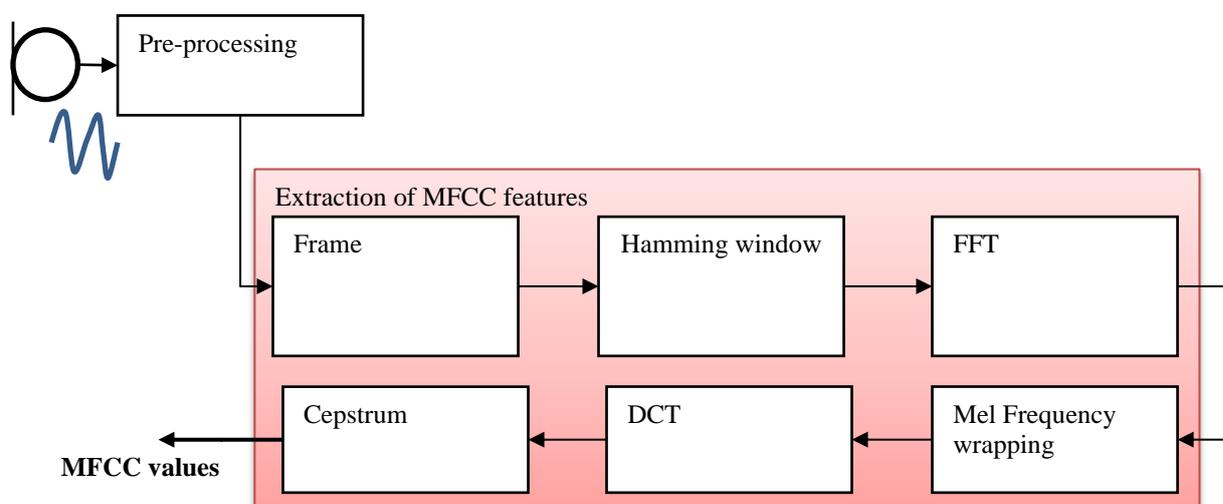


Figure 1. Modern methods of feature extraction of speech signals.

Effective algorithms for compressing audio or low frequency signals are based on discrete orthogonal transforms. One of these transforms is the Singular value decomposition (SVD) [18]. The SVD algorithm for modifying a matrix by singular values is one of the most powerful instrumental tools of linear algebra.

There are many ways to reduce the size of data in machine learning with different efficiencies, including by modifying and displaying one-dimensional space to another dimensional space. These methods include PCA (Principal component analysis), ICA (Independent component analysis), NMF (Non-negative matrix factorization) and K-means.

The operations and algorithms used in all speech recognition require a certain complex number of methods and corresponding processing algorithms. Spectral analysis, wavelet transformation, filtering, scraping, cepstral analysis, etc. are performed on different bases of Fourier including. These algorithms and techniques are more difficult to implement in specialized software or hardware systems. This requires a special approach and the implementation of optimal algorithms.

The sequence of speech command interrupt algorithms, depended or in depended to the speaker in the proposed real-time mode, is shown in Figure 2. Recognition of speech commands by this method is performed in the following algorithmic steps.

1. An analog signal in the form of an incoming speech is converted to a digital signal in the form of a 16 kHz frequency.
2. Speech commands clear from external noise and interference.

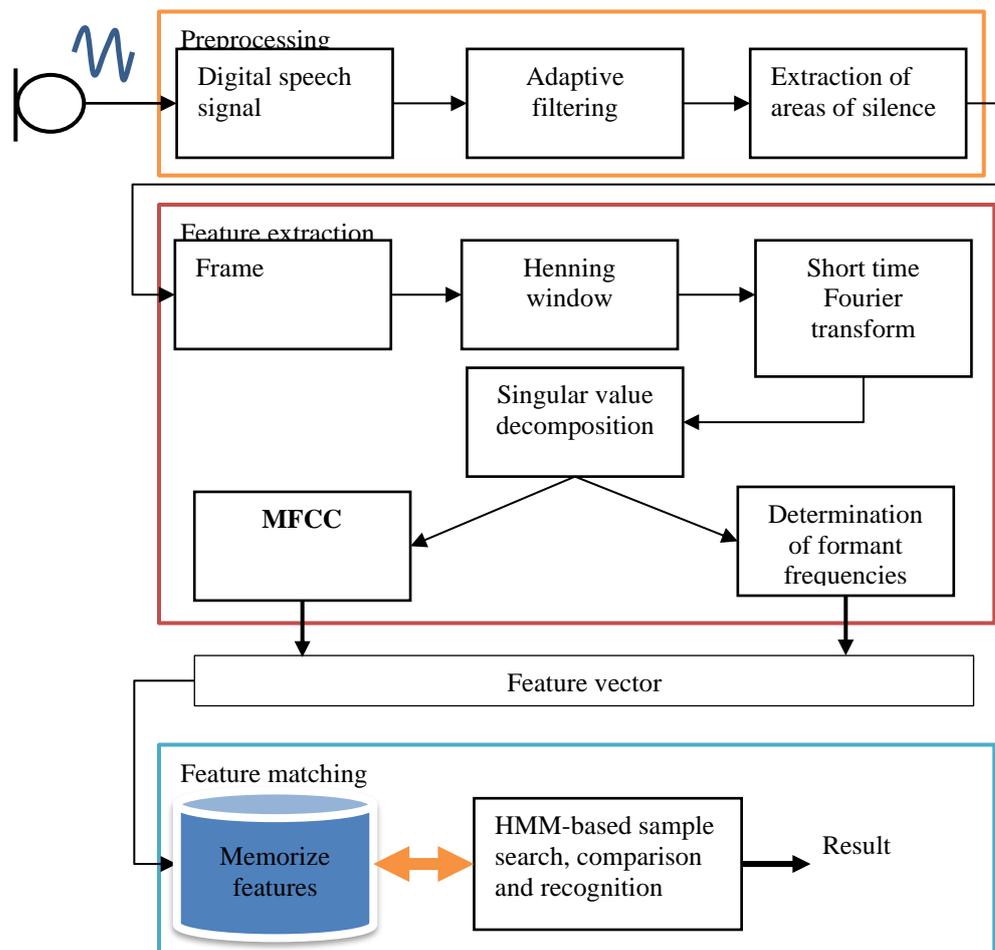


Figure 2. Stages of feature extractions and familiar algorithms from the proposed speech signals.

3. Extraction areas of silence from speech signals.
4. Framing.
5. Henning window.
6. Short time Fourier transform (STFT).
7. Reduce dimensions. The singular value decomposition method was used to reduce the signal size.
8. Determination of formant frequencies.
9. Mel Frequency Cepstral Coefficients (MFCC).
10. Checking and recognizing the conformity of the features.

The process of recognizing speech signals is done through various intellectual processing steps and algorithms. According to the problem statement and speech command recognition requirements, DTW (Dynamic time warping), VQ (Vector quantization), SVM (Support vector machine), HMM (Hidden Markov model), ANN (Artificial neural network) algorithms are common.

The above speech signal processing algorithms can be implemented in the hardware and software part of the computer in different ways using a database of different elements. The elemental database includes a variety of non-programmable and programmable devices.

IV. HARDWARE IMPLEMENTATION

A hardware-software platform is a set of hardware and software designed to perform a specified task. The hardware-software platform is built on the modular principle and consists of a main board block with a special processor, a memory block and I/O block, and external interface blocks (Figure 3).

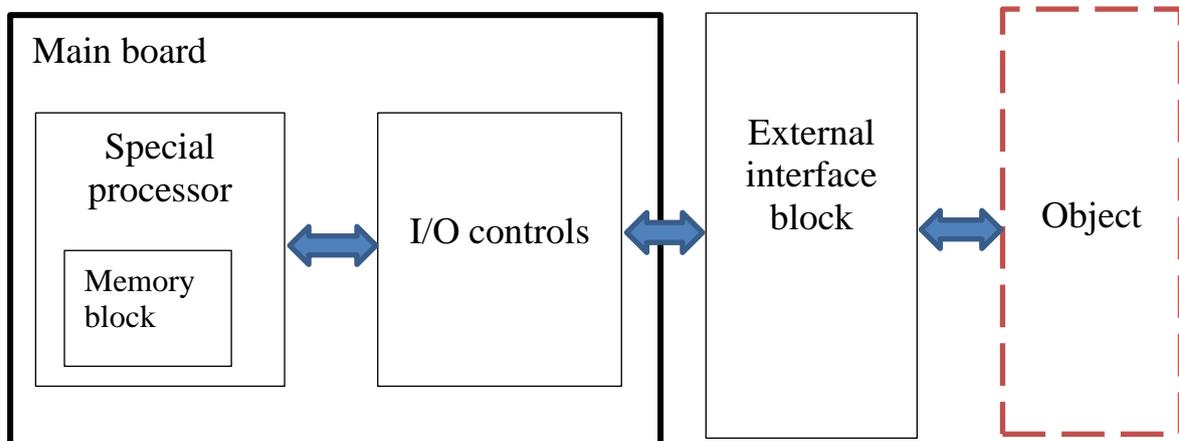


Figure 3. The generalized structure of the hardware-software platform.

The memory block of the hardware-software platform has the ability to be reprogrammed by the user to implement their own algorithm, depending on the type of special processor. An external device or controlled object is connected and addressed to the I/O block via external interfaces. The memory block of the hardware-software platform has the ability to be reprogrammed by the user to implement their own algorithm, depending on the type of special processor. An external device or controlled object is connected and addressed to the I/O block via external interfaces.

This hardware varies depending on the type and architecture of the software processors. The hardware-software platform is called the SoC (system on chip) system, which combines several main components into a single crystal, and in some literatures this system is also called the microcontroller system. The development of technology today has led to the development of systems based on microcontrollers. Today, microcontrollers consist of several processors. In addition, as the efficiency of the microcontroller is increased and closer to the efficiency of the processors, these microcontrollers have the ability to load different installed operating systems. This hardware-software platform can be considered stand-alone or as part of a much larger and more complex system.

Recently, there is a growing interest in the use of user-programmable logic matrix (FPGA - field programmable gate array) and application-specific integrated processor (ASIP) application in real-time, built-in special hardware-software platform systems. For ASIPs, we can cite mobile hardware-software platforms on a single-chip Raspberry Pi, Banana and ARM processor platform. We can see these processors as the optimal solution between features such as size, speed, and flexibility compared to other processors.

It is impossible to build a digital processor that can run any application efficiently, it is physically and logically impossible. The most appropriate and best processor for each task is selected based on the task [19]. Analysis showed that microcontrollers showed very good results during the pre-processing of speech signals. Can easily implement algorithms for digitizing signals, filtering, grinding, extraction silent areas.

At the architecture level of reprogrammable FPGA devices, the implementation of digital signal processing algorithms is envisaged. Therefore, on this hardware-software platform it is possible to easily perform complex operations such as feature extraction, framing, various transforms, compression. However, since there is no analog input, it is definitely used in conjunction with microcontrollers.

The hardware-software platform based on mobile processors has a multi-core processor and the ability to implement built-in real-time operating systems and the ability to perform a complete cycle of processing complex speech signals, given the relative size of memory [20]. The EM3288 module was selected to implement the voice control system through speech. This hardware comes with a 4-core ARM Cortex v7 processor, 2GB of RAM, a graphics processor, and 40 I/O ports. These characteristics allow the processing of speech signals and the implementation of intelligent processing algorithms.

V. EXPERIMENTAL RESULTS

The above-mentioned speech command recognition algorithms for recognizing Uzbek speech commands were implemented in the EM3288 module. These words are: “talaba, malika, samalyot, lola, bola, ona, masala, mamlakat, olti, oltin, lochin, lobar, kitob, maktab, archa, chelak, gul, yulduz, sayyora, viloyat, tegirmon, dala, vagon, bekobod, angren, yangiyo'l, yunusobod, umid”.

Recognition of Uzbek speech commands in the EM3288 module showed the following accuracy (Fig.4).

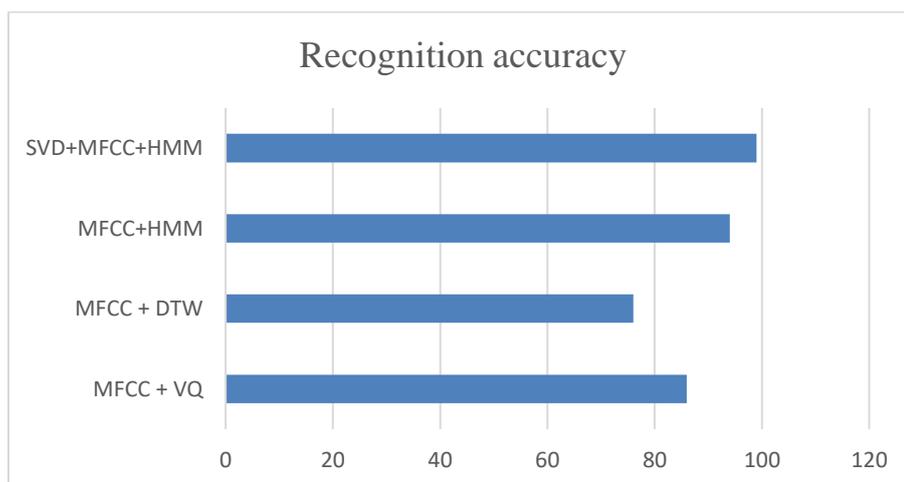


Figure 4. Accuracy of speech command recognition using various algorithms.

From the figure above, it can be seen that the spectral values of the speech signal and the MFCC parameters showed 86%, 76% and 94% accuracy, respectively, when obtained by VQ, DTW and HMM algorithms. The signal parameters obtained by the STFT + SVD algorithms proposed by us showed 98% accuracy when recognized by the HMM algorithm.

**VI. CONCLUSION AND FUTURE WORK**

The built-in EM3288 system comes in handy when implementing speech control systems through speech commands in different environments. Because through this device it is possible to combine algorithms and programs for processing complex speech signals under a single operating system. The HMM model has shown high results in recognizing limited speech commands in embedded systems. Recognition of speech commands in systems installed using the proposed STFT+SVD+HMM algorithm showed 98% accuracy.

REFERENCES

- [1] M. Mazo, F. J. Rodriguez, J. L. Lazaro, J. Urena, J. C. Garcia, E. Santiso, P. Revenga, J. J. Garcia, «Wheelchair for Physically Ultrasonic and Infrared Sensor Control» *Autonomous Robots*, № Vol. 2., pp. pp. 203-224, 1995.
- [2] V. V. Ngoc, J. Whittington, J. Devlin, «Real-time Hardware Feature Extraction with Embedded Signal Enhancement for Automatic Speech Recognition» *Speech Technologies*, Intech, pp. pp. 29-54, 2011.
- [3] M. Bahoura, H. Ezzaidi, «Hardware implementation of MFCC feature extraction for respiratory sounds analysis» M. Bahoura, H. Ezzaidi, *Hardware implementation of MFCC feature ext8th Workshop on Systems, Signal Processing and their Applications*, pp. pp. 226-229, 2013.
- [4] Mosleh M., Setayeshi S., Mehdi Lotfinejad M., Mirshekari A., «FPGA implementation of a linear systolic array for speech recognition based on HMM» *The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, № Vol. 3, pp. pp. 75-78., 2010.
- [5] Veitch R., Aubert L.-M., Woods R., Fischhaber S., «FPGA Implementation of a Pipelined Gaussian Calculation for HMM-Based Large Vocabulary Speech Recognition» *International Journal of Reconfigurable Computing*, 2011.
- [6] Stephen J. Melnikoff, Steven F. Quigley, Martin J. Russell, «Implementing a Hidden Markov Model Speech Recognition System in Programmable Logic» *11-th International Conference on Field Programmable Logic and Applications, Lecture Notes in Computer Scienc*, № Vol. 2147, pp. pp. 81-90., 2001.
- [7] J.J. Rodríguez-Andina, R.D.R. Fagundes, D.B. Junior, «A FPGA-based Viterbi algorithm implementation for speech recognition systems.» *A FPGA-based Viterbi algorAcoustics, Speech, and Signal Processing. Proceedings. (ICASSP '01). 2001 IEEE International Conference*, т. Vol. 2, pp. pp. 1217-1220, 2001.
- [8] G. Tamulevicius, V. Arminas, E. Ivanovas, D. Navakauskas, «Hardware Accelerated FPGA Implementation of Lithuanian Isolated Word Recognition System» *Electronics & Electrical Engineering*, т. 3, № 99, pp. pp. 57-62., 2010.
- [9] Sujuan Ke, Yibin Hou, Zhangqin Huang, Hui Li, «A HMM Speech Recognition System Based on FPGA» *Image and Signal Processing, CISP '08*, т. vol. 5, pp. pp. 305-309., 2008.
- [10] Musaev M. M., «Sovremennyye metody sifrovoy obrabotki rechevyx signalov» *Vestnik TUIT*, т. 2, № 42, 2-13 str., 2017 [In Russian].
- [11] Musaev M., Khujayorov I. and Ochilov M., «The Use of Neural Networks to Improve the Recognition Accuracy of Explosive and Unvoiced Phonemes in Uzbek Language» в *Information Communication Technologies Conference (ICTC)*, Nanjing, China, 2020.
- [12] Musaev M. M., Raximov M. F. «Algoritmy parallelnoy obrabotki rechevyx signalov» *Vestnik TUIT*, т. 2, № 46, str. 2-13, 2018 [In Russian].
- [13] M.M. Musaev, U.A. Berdanov, K.E. Shukurov, «Hardware and software solution signal compression algorithms based on the Chebyshev polynomial» *International Journal of Information and Electronics Engineering*, т. Vol. 4, № No. 5, pp. 380-383, 2014.
- [14] Musaev M.M., Kardashev M.S., «Spektralny analiz signalov na mnogoyadernyx protsessorax» *Jurnal «Tsifrovaya obrabotka signalov»*, № 3, str. 82-86., 2014 [In Russian].
- [15] Musaev M.M., *Uskorennyy spektralny analiz signalov na mnogoyadernyx protsessorax*, Saarbrucken: Palmarium Academic Publishing, 2014, str. 98 [In Russian].
- [16] Shukurov K.E., Abdurazzoqov F.B., «Signallarga spectral ishlov berish algoritmlari tahlili» v Shukurov K.E., Abdurazzoqov F.B. *Signallarga spectral Respublika ilmiy-texnik anjumanining ma'ruzalar to'plami, 1-qism.*, Nukus, 2015 [In Uzbek].
- [17] Mermelstein, S. B. Davis, «Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences» *IEEE Trans. on Acoustics, Speech and Signal Processing*, № vol. 28, pp. pp. 357-366, 1980.
- [18] Theodoridis Sergios, *Machine Learning: A Bayesian and Optimization Perspective*, 2nd edition, Academic Press, 2020, p. 1160.
- [19] Shukurov K.E., Zoxirov Q.R., «Obzor sovremennix apparatnix obespecheniy obrabotki signalov» v Fan, ta'lim va ishlab chiqarish integratsiyasida axborot-kommunikatsiya texnologiyalarini qo'llashning hozirgi zamon masalalari. Nukus, 2019 [In Russian].
- [20] Shukurov K.E., «Raspberry pi qurilmasida o'zbek tili nutq buyruqlarini tanib olish tizimini amalga oshirish» "TATU XABARLARI" Toshkent axborot texnologiyalari universitetining ilmiy-texnika va axborot-tahliliy jurnali, т. № 2, № 54, pp. 45-61, 2020 [In Uzbek].



ISSN: 2350-0328

**International Journal of Advanced Research in Science,
Engineering and Technology**

Vol. 8, Issue 10 , October 2021

AUTHOR'S BIOGRAHY



Shukurov Kamoliddin Elbobo o'g'li is from Uzbekistan and was born on 1 April, 1984. He obtained bachelor of electronics and automation from the Tashkent State Technical University (TSTU), Uzbekistan, in 2004, master in computer science and engineering from Lipetsk State Technical University (LSTU), Russia, in 2008. Author of many scientific papers. Currently, he is a researcher and Senior-teacher at the department Artificial intelligence, Tashkent University of Information Technologies (TUIT), Uzbekistan since 2013. His current interest is in the field parallel information processing, computer architecture and real time systems. He is currently researching digital processing of speech and digital image processing