The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2023

Volume 23, Pages 220-224

ICRETS 2023: International Conference on Research in Engineering, Technology and Science

The Effectiveness of the Implementation of Speech Command Recognition Algorithms in Embedded Systems

Kamoliddin Shukurov
TUIT named after Mukhammad al-Khwarazmi

Umidjon Khasanov
TUIT named after Mukhammad al-Khwarazmi

Boburkhan Turaev
TUIT named after Mukhammad al-Khwarazmi

A'lokhan Kakhkharov
TUIT named after Mukhammad al-Khwarazmi

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.

Keywords Human-computer interface (HCI), Hardware-software platform, Embedded system

Introduction

Speech command control is widely used in various fields, speech control of devices and equipment, voice assistant in call centers, smart home and Internet of Things (IoT), speech identification, speech control interface for people with disabilities and can be applied other fields (Mazo et al., 1995). 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 (Ngoset et al., 2011; Bahoura & Ezzaidi, 2013; Mosleh et al., 2010; Veitchet al., 2011; Melnikoff, et al., 2001; Vargas et al., 2001; Tamulevicius et al.,2010; Sujuan et al., 2008).

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.

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

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 (Мусаев, 2017; Мусаев, et al., 2019; Алимурадов & Чураков, 2015; Musaev, et al., 2020; Мусаев & Рахимов, 2018).

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 (Musaev et al., 2014; Мусаев & Кардашев, 2014).

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 (Мусаев & Кардашев, 2014).

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.

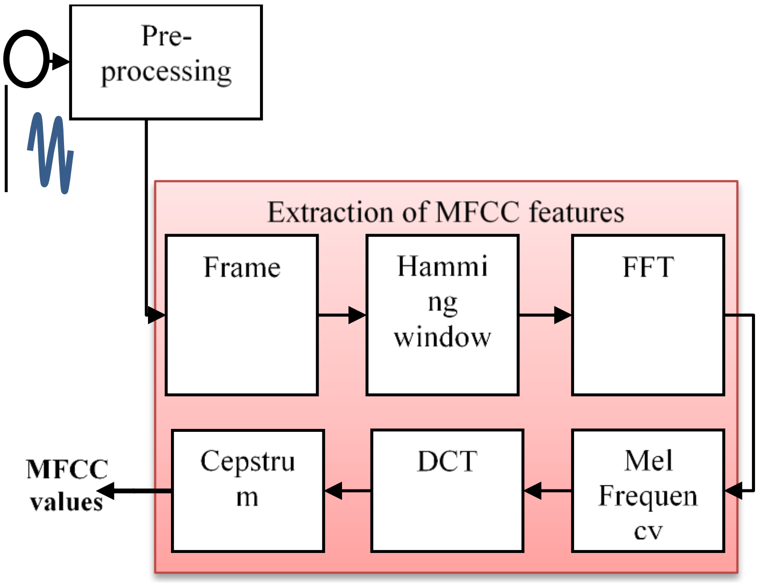

Figure 1. Modern methods of feature extraction of speech signals

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) [20]. 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-meanings.

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.
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.

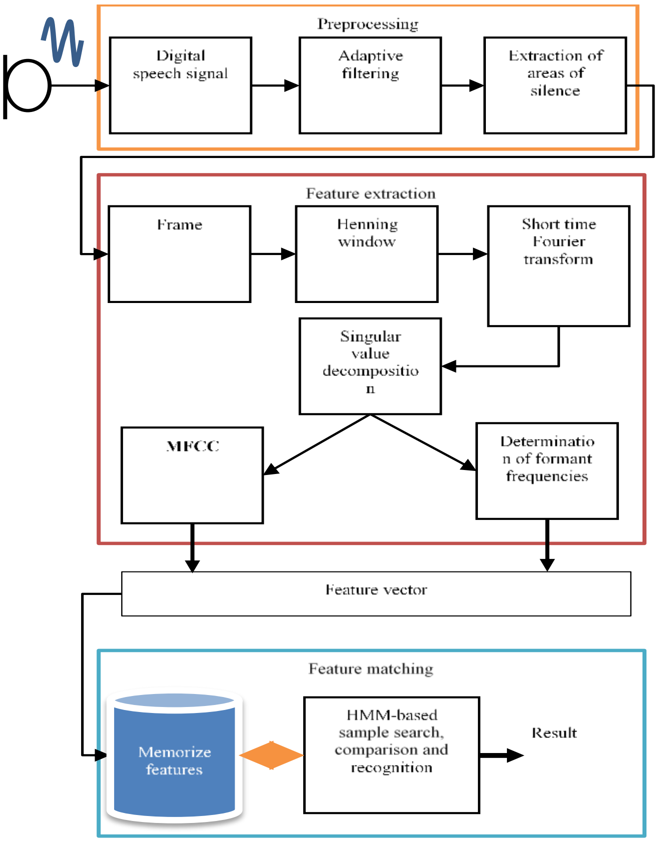
Preprocessing
Digital speech signal
Adaptive filtering
Extraction of areas of silence
Feature extraction
Frame
Henning window
Short time Fourier transform
Singular value decompositio n
MFCC
Determinatio n of formant frequencies
Feature vector
Feature matching
Memorize features
HMM-based sample search, comparison and recognition
Result

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

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.

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).

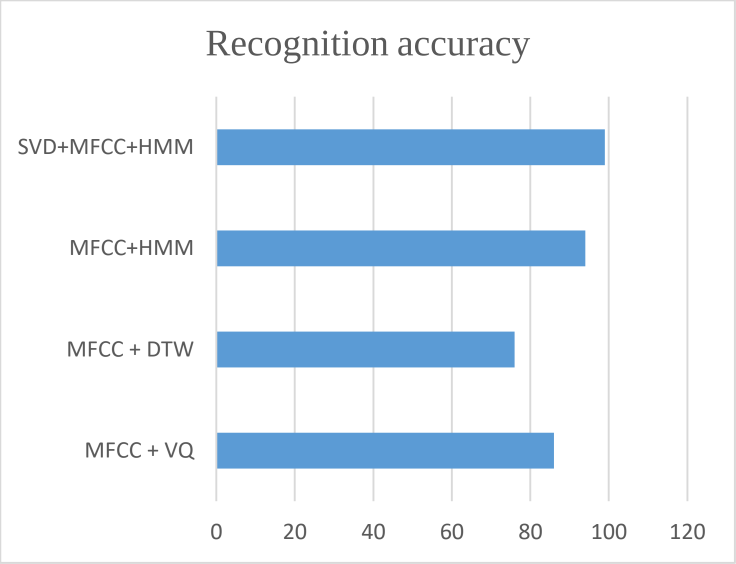

Figure 4. Accuracy of speech command recognition using various algorithms.

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.

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

Алимурадов А.К. & Чураков П.П. (2015). Обзор и классификация методов обработки речевых сигналов в системах распознавания речи. *Измерение.Мониторинг.Управление.Контроль*, *2,12* 27-35. [In Russian].

Bahoura M. & Ezzaidi H. (2013). Hardware implementation of MFCC feature extraction for respiratory sounds analysis. *8th Workshop on Systems, Signal Processing and their Applications (WoSSPA),* 226-229.

Mazo M., Rodriguez F.J., Lazaro J.L., Urena J., Garcia J.C., Santiso E., Revenga P. & Garcia J.J. (1995).Wheelchair for physically ultrasonic and infrared sensor control. *Autonomous Robots,2,* 203-204.

Melnikoff S.J., Quigley S.F. & Russell M.J. (2001). 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*, *2147*, 81-90.

Mosleh M., Setayeshi S., Mehdi Lotfinejad M. & Mirshekari A. (2010). FPGA implementation of a linear systolic array for speech recognition based on HMM. *The 2nd International Conference on Computer and Automation Engineering (ICCAE), 3,* 75-78.

Musaev M.M., Berdanov U.A. & Shukurov K.E. (2014). Hardware and software solution signal compression algorithms based on the Chebyshev polynomial. *International Journal of Information and Electronics Engineering*, *5*, 380-383.

Musaev M., Khujayorov I. & Ochilov M. (2020). The use of neural networks to improve the recognition accuracy of explosive and unvoiced phonemes in Uzbek language. *Information Communication Technologies Conference (ICTC)*, 231-234.

Мусаев М.М. & Кардашев М.С. (2014). Спектральный анализ сигналов на многоядерных процессорах. *Цифровая обработка сигналов,3,* 82-86. [In Russian].

Мусаев М. М. (2017). Современные методы цифровой обработки речевых сигналов. *Вестник ТУИТ*, *42*, 2-13. [In Russian].

Мусаев М. М. & Рахимов М. Ф. (2018). Алгоритмы параллельной обработки речевых сигналов. *Вестник ТУИТ*, 46, 2-13. [In Russian].

Мусаев М.М., Хужаяров И.Ш. & Очилов М.М. (2019). Машинали ўқитиш алгоритмлари асосида ўзбек тили фонемаларини таниб олиш. *Информатика ва энергетика муаммолари*. [In Uzbek].

Ngos V.V., Whittington J. & Devlin J. (2011). Real-time hardware feature extraction with embedded signal enhancement for automatic speech recognition. *Speech Technologies*, *Intech*, 29-54.

Sujuan K., Yibin H., Zhangqin H. & Hui L. (2008). A HMM speech recognition system based on FPGA. *International Congress on Image and Signal Processing (CISP 2008)*, 305-309.

Tamulevicius G., Arminas V., Ivanovas E. & Navakauskas D. (2010). Hardware accelerated FPGA implementation of Lithuanian isolated word recognition system. *Electronics & Electrical Engineering*, *99*, 57-62.

Vargas F.L., Fagundes R.D.R., & Junior D.B. (2001). A FPGA-based Viterbi algorithm implementation for speech recognition systems. *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221),2,* 1217-1220.

Veitch R., Aubert L.M., Woods R. & Fischaber S. (2011). FPGA Implementation of a pipelined Gaussian calculation for HMM-based large vocabulary speech recognition. *International Journal of Reconfigurable Computing*, 1-10.

Author Information

Kamoliddin Shukurov
TUIT named after Mukhammad al-Khwarazmi,
Artificial intelligence department, Uzbekistan
Contact e-mail: *keshukurov@gmail.com*

Umidjon Khasanov
TUIT named after Mukhammad al-Khwarazmi,
Artificial intelligence department, Uzbekistan

Boburkhan Turaev
TUIT named after Mukhammad al-Khwarazmi,
Artificial intelligence department, Uzbekistan

A’lokhan Kakhkharov
TUIT named after Mukhammad al-Khwarazmi,
Artificial intelligence department, Uzbekistan

To cite this article:

Shukurov, K., Khasanov, U., Turaev, B., & Kakhkharov, A. (2023). The effectiveness of the implementation of speech command recognition algorithms in embedded systems. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 23,* 220-224.