

ALGORITHM AND SOFTWARE FOR AUTOMATIC RECOGNITION OF UZBEK SPEECH

S.Beknazarova, N.Xalikova, I.Ismonov

(DSc of "Audiovisual technology" department, student of Tashkent University
of Information Technology named after Muhammad Al-Khwarezmi)

E-mail: missnancy5505@gmail.com

ABSTRACT

The article describes an implementing a real time speaker identification system by voice for embedded and general purpose computers. A review and analysis of existing speaker identification algorithms are made. The speaker's input speech is recorded in the system, go through the preprocessing stage, extract features and voice parameters for further identification. To recognize the speaker by voice parameters, the Vector quantization and Hidden Markov model algorithms are used.

Keywords: *speaker identification, pre-processing, filtering, feature extraction, recognition, Vector quantization, Hidden Markov model.*

АННОТАЦИЯ

В статье описывается реализация системы идентификации говорящего по голосу в реальном времени для встроенных компьютеров и компьютеров общего назначения. Сделан обзор и анализ существующих алгоритмов идентификации говорящего. Входная речь говорящего записывается в систему, проходит этап предварительной обработки, извлекает характеристики и параметры голоса для дальнейшей идентификации. Для распознавания говорящего по параметрам голоса используются алгоритмы векторного квантования и скрытой марковской модели.

Ключевые слова: *идентификация говорящего, предварительная обработка, фильтрация, выделение признаков, распознавание, векторное квантование, скрытая марковская модель.*

INTRODUCTION

Analysis of speech signal processing areas today focuses on speech command recognition, synthesis of hardware control commands, transmission of speech signals via IP telephony channels and protection of transmitted speech, concise speech recording and speech biometrics, criminology requires new research into the problems of analysis and synthesis of speech signals in the field of human identification (identification). In solving the problem of spectral transformations, it is difficult to implement without new approaches and approaches to filtering in different

environments, parameterization of signals, algorithms for feature extraction and methods of recognition [1]. Biometric technology is based on measuring the unique features of a person. Biometric systems are used where identification of person is required. One of the most common biometric characteristics of a human being is his or her voice. Identifying a person by voice or speech involves a set of technical, algorithmic, and mathematical methods that involve complex steps from sound recording to preprocessing of speech data [2,3].

RELATED WORKS

Identification of the speaker by voice has been studied by scientists from different countries for many years and is still ongoing. Because, depending on the application, new approaches are required for speaker identification. Inverted MFCC methods with a recognition accuracy of 98%, the author of [4] works to identify the speaker based on VQ and Kekre's Median Codebook Generation Algorithm with an accuracy of 84%. The author of [5] have design of speaker verification using dynamic time warping (DTW) on graphical programming for authentication process with 84% accuracy. In [6] Speaker Verification using Convolutional Neural Networks (CNN) are implemented. The authors of [6] are implemented real time speaker identification on FPGA, but here the emphasis is mainly on the preprocessing stage of signal processing, but the recognition accuracy is not given. For the implementation of the speaker identification system in biometric systems, control of technical units and voice identification in smart homes. The above algorithms are not sufficient. Because in real conditions our voice is always accompanied by different noises and the implementation of a complex recognition and identification algorithm in Embedded devices requires new approaches and optimal algorithms.

MAIN PART

Voice or speech identification is a separate scientific discipline that is part of the processing of speech signals [7]. Human speech identification prohibits access to various information resources and physical objects [8], the management of voice services in mobile communication in systems based on telecommunication channels, protection against fraud through the introduction of voice identification, can be widely used in investigative processes, in the protection of verbal information and in the identification of offenders through digital speech in digital criminology. In addition, a person has the opportunity to learn about his age, gender and ethnicity or dialect, emotional state through speech [9]. Especially nowadays, there is great interest in the application and development of voice identification and speech

recognition methods in the areas of voice control and speech recognition in smart home, smart and safe city systems [10]. It should also be noted that the use of a combination of different technologies and methods is required to ensure the safest and most accurate identification and authentication, especially when performing important work. In other words, in these cases, it is advisable to use biometric methods of speech identification with special input / output devices with fast memory and high-performance microprocessors. Implementing a voice or speech identification system in a real-world setting (other than a recording studio or laboratory) can present serious challenges and barriers: - in such an identification, various changes and noises occur in the input signal due to the peculiarities of the equipment and devices for recording, processing and storing information. - external acoustic noises inevitably affect the speech signal. herefore, it is difficult to demonstrate a sufficiently high efficiency in the processing and analysis of speech data with external noise or in real conditions. High performance speech identification systems in the laboratory can show much lower reliability in real conditions. Identifying a person through speech or voice requires sophisticated hardware and software solutions that incorporate a complex set of technical, algorithmic, and mathematical methods. Identifying a speaker through speech is divided into 2 parts: - the method of identification, which depends on the text spoken by the speaker; - the method of identification, which is not depend on the text spoken by the speaker.

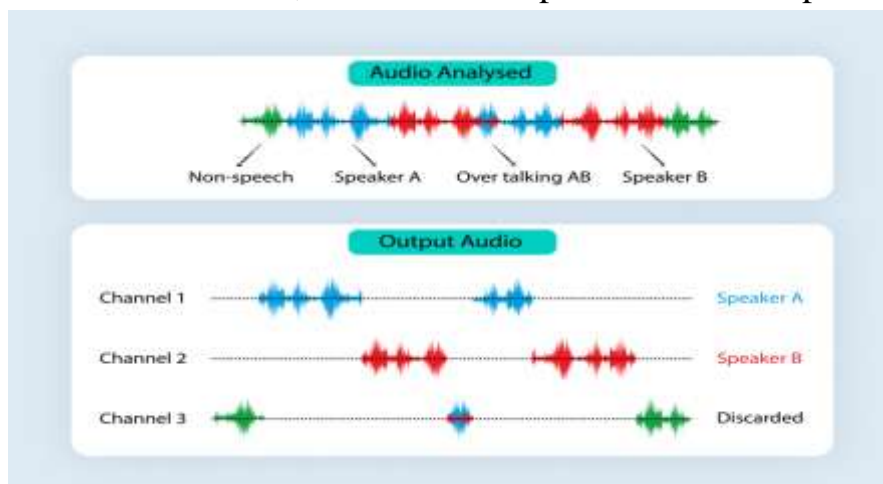


Fig. 1. Methods of speaker identification by voice. In the non-text-based method of identifying a speaker through speech, the speaker's speech is converted to a digital signal and the formant frequencies are determined from that signal after preprocessing. As mentioned above, the human voice is a biometric property, and each person's voice has its own formant frequency. In this paper, we will look at text depended identification. This is because most voice control and certain security

systems use a specific command word or a specific keyword to identify a person through speech. The following figure illustrates the steps in implementing the structure of a speech identification system.

The obtained features are compared with the previous features of the speakers stored in the memory of computer(or embedded system) using intelligent algorithms VQ or HMM. 5. If the features of the speaker's voice match with the features stored in the memory, the result is that the speaker has been identified. Digitization of the input signal from the microphone. The signal read from the microphone is converted from analog to digital according to Kafelnikov's formula. There were no additional requirements for the microphone in this system, and the existing microphone of the personal computer was used. Input speech is defined as a sampling frequency of 16000Hz. Pre-processing stage. At the preprocessing stage, adaptive filtering is initially performed [11]. Speech sounds are a low frequency signal. Typically, human speech sounds range from 80-180 Hz in men to 165-280 Hz in women. The frequency of common speech signals is in the range of 80 to 3000 Hz. In addition, the addition of a number of external interferences to normal speech sounds can lead to poor signal quality. The frequency of speech signals also varies depending on the person's condition and movement. Filtering variable speech signals is a very complex process, and filtering using simple digital filters is inefficient. Filtering speech signals using adaptive filters serves to improve signal quality. In implementing this system, the VQ algorithm was used to identify the speaker and determine the compatibility of the feature. The VQ algorithm is relatively simple compared to other classification algorithms and is easy to identify in a speaker and implement in speech-dependent speech control hardware and software systems. In addition, the features of this algorithm have the ability to compress and save computer memory.

CONCLUSION

Biometric identification is evolving in several ways. Methods of identification of speakers through biometric parameters, such as hand writing or by voice, becomes a popular. Hence, the identification of a speaker by voice or speech is widely used in smart home, safe city, smart car and remote voice control systems. These systems require the development of complex mathematical hardware and software solutions. Intellectual algorithms with varying degrees of complexity are used to identify the speaker by speech. The results showed that the identification of the speaker can be easily performed in hardware and software by taking the speech signal MFCC information features and classifying it using machine learning algorithms. Moreover, the accuracy level of this algorithm is not less than the accuracy level of other

complex algorithms. For implementing a real-time speaker identification system on a hardware software platform (embedded system), with a less computing resources is more suitable VQ algorithm. If the task of identifying the speaker requires a relatively easy algorithm.

REFERENCES

[1] Shukurov K.E. Raspberry pi qurilmasida o'zbek tili nutq buyruqlarini tanib olish tizimini amalga oshirish.// TATU xabarlarlari 2(54)/2020. 45-61 b.

[2] Sahoo, J. K. Deepak R. "Speaker recognition using support vector machines." International Journal of Electrical, Electronics and Data Communication, ISSN: 2320-2084 Volume-2, Issue-2, Feb.-2014.

[3] Singh, S. and E. Rajan. "Vector Quantization Approach for Speaker Recognition using MFCC and Inverted MFCC." International Journal of Computer Applications 17 (2011): 1-7.

[4] H. B. Kekre, V. A. Bharadi, A. R. Sawant, O. Kadam, P. Lanke and R. Lodhiya. "Speaker recognition using Vector Quantization by MFCC and KMCG clustering algorithm," 2012 International Conference on Communication, Information & Computing Technology (ICCICT), 2012, pp. 1-5, doi: 10.1109/ICCICT.2012.6398146.

[5] Barlian H., Dahniyal S. "Design of Speaker Verification using Dynamic Time Warping (DTW) on Graphical Programming for Authentication Process." JITECS Volume 2, Number 1, 2017, pp 11-18 [8] Hossein, S. "Speaker Verification using Convolutional Neural Networks." <https://doi.org/arXiv:1803.05427v2>

[6] S.Gourav, S.Goutam, "Real Time Implementation of Speaker Identification System with Frame Picking Algorithm." Procedia 17 Bulletin of TUIT: Management and Communication Technologies Shukurov K. E. 2021 4(46) Computer Science Volume 2, 2010, Pages 173-180. doi:10.1016/j.procs.2010.11.022

[7] Ramos-Lara, R., López-García, M., Cantó-Navarro, E. et al. Real-Time Speaker Verification System Implemented on Reconfigurable Hardware. J Sign Process Syst 71, 89– 103 (2013). <https://doi.org/10.1007>

[8] Flanagan, Dzh.L. Analiz, sintez i vospriyatiye rechi / Dzh.L. Flanagan; per. s angl. A.A. Pirogova. – M. : Svyaz', 1968. – 396 s.

[9] Mariethoz, J. "Speaker Verification Based on User-Customized Password." / J.Mariethoz, B. Herve, M.F. BenZeghiba // IDIAP Research Report 01-13. – Martigny, 2001. – 22 p.

[10] Pellandini, F. "GSM Speech Coding And Speaker Recognition" / F. Pellandini, M. Ansorge, A. Dufaux [at al.] // International Conference on Acoustics,

Speech, and Signal Processing (ICASSP): Book of abstracts. – Istanbul, 2000. – vol. 2. – pp.1085–1088.

[11] Amrouche, A. “Effect of GSM speech coding on the performance of Speaker Recognition System.” / A. Amrouche, A. Krobba, M. Debyeche // 10th International Conference on Information Sciences Signal Processing and their Applications (ISSPA): Book of abstracts. – Kuala Lumpur, 2010. – pp. 137–140.