Peach signal control software is an important part of our lives

Dilobar O'tkirovna Abdullayeva	Tashkent University of information technologies named after Mukhammad al-Khwarizmi, 108 Amir Temur avenue, Tashkent, 100084, Uzbekistan. Email: fevraldilobar@gmail.com
Miraziz Muxtorovich Ismoilov	Tashkent technical school of information technologies, 7 Mustakillik Avenue, Tashkent, 100000, Uzbekistan. Email: azizismoilov2022@mail.ru
Akmal Tulkin ug'li Djurakulov	Samarkand branch of Tashkent University of information technologies named after Mukhammad al-Khwarizmi, 47A Sh.Mirzo Street, Samarkand, 140100, Uzbekistan. Email: akmaljoraqulov1@gmail.com
Nigora Muzaffar qizi Abdieva	Samarkand State Institute of Foreign Languages, 93 Bustonsaroy street, Samarkand, 140104, Uzbekistan
Durdona Yodgorbekovna Khasanova	Samarkand State Institute of Foreign Languages, 93 Bustonsaroy street, Samarkand, 140104, Uzbekistan.

ABSTRACT

This article introduces voice user interfaces to speech recognition programs, such as voice dialing, for example, "call home", call forwarding, for example, "I want to call", control of domicile devices, keyword search, for example, finding a podcast with specific words, entering simple data, for example, entering a credit card number, preparing structured documents, for example, radiology report, speaker features, speech-to-text processing, for example, word processors or electronic Mail messages and airplanes are usually called direct voice access. Optimal use of high-speed programmable digital signal processors generally demands familiarity with machine architectural features and hence, the production of programs whose structure reflects and exploits those features. In contrast, it is apparent that little effort has been made to develop programming techniques that fully realize the signal processing computational capability of standard minicomputers. In this paper, it is shown that a functional high-level language signal processing program can easily be modified so as to produce a similar program that, when executed, automatically generates another program containing precomputed algorithm sequencing and data access information. The generated program will then utilize central processor arithmetic and logical capability only for data-dependent computation. In this way, instructions normally associated with computation for program sequencing/control or data access are eliminated, and all benefits of increased algorithm complexity for reduction of data-dependent arithmetic computation are in fact realized as decreased program execution time. Examples are given of Fortran programs which generate Fortran FFT subroutines and, for completeness, assembly language realizations of the Pfeifer/Blankinship autocorrelation algorithm. Results demonstrate that using this technique, standard minicomputers may execute digital signal processing algorithms faster than peripheral processors which normally require standard minicomputers as host processors. The power of speech sounds is the amount of energy that travels through a space of a certain width at a given time. This indicator affects the energy and power of speech. The timbre of speech sounds is formed by the main tone. This parameter is different for different people so this parameter is in human speech recognition systems. In most cases, the movement of wheelchairs is controlled by disabled people using a joystick or by an accompanying person. Significantly disabled patients need alternative control methods without using the wheelchair joystick because it is undesirable or impossible for these patients. In this article, we present the implementation of a robotic wheelchair based on a powered wheelchair that is controlled not by the joystick but by the onboard computer that receives and processes data from the extended brain-computer interface (extended BCI). Under this term we understand the robotic complex control system with simultaneous independent alternative control channels. In this robotic wheelchair version the BCI works with voice and gesture control channels.

Keywords:	Signal generators, Digital signal processing, Signal processing algorithms, Microcomputers, Arithmetic, Computer aided instruction, Digital signal processors, Production, Standards development, High level languages, Extended brain-computer interface, robotic wheelchair, control channel, robotics

Introduction

Technical projects on the development of robotic wheelchairs have been carried out since the last century. Modern mobile robotic complexes (MR) which include robotic wheelchairs are complex heterogeneous hardware and software systems and they should provide a certain level of comfort and reliability of control answering the fields of their application. Under the term MR we understand the robotic system having an onboard powerful, versatile, and inexpensive miniature computer with a modern CPU providing the ability to connect the modern peripherals to the system without any restrictions, unlike the microcontroller capabilities. Because of this, it is possible to use the maximum possible set of software, more memory and parallel programming techniques to achieve the real-time mode (RTM).

In the formation of these sounds, different speech organs perform their functions based on different formations. That is, which one vibrates or closes as needed. Speech pre-processing subsystem Speech signal pre-processing includes the following steps: - the process of access to the speech signal; - selection of the limit of the speech signal; - digital filtering; - Interrupting the speech signal with overlapping frames; - window signal processing; - spectral transformation; - frequency spectrum normalization. Audio input is done in real-time via WAV files encoded on the sound card or PCM. 8 kHz sampling rate and 16-bit quantization are typical in speech data transmission, storage, and processing systems parameters. Working with files has been introduced to facilitate multiple repetitions of neural network processing, which is especially important for training. Separation of speech signals The following features of a speech signal are used to separate sections containing only speech from the input signal: • short-term energy of the speech signal; • number of intensity zeros (instantaneous frequency); • Distribution density of pause report value. The transient energy and intensity zeros of the audio signal are used to output speech from the input signal at the same time. You can also remove the waste pause using the Gaussian distribution method. Digital filtering Usually different sounds are accompanied by a useful signal. Noise negatively affects the quality of speech recognition systems, so it must be combated. Two types of digital filters are used to reduce the noise level in a small system: • a conductive filter; • Pre-filtering Information about the amplitude of the speech signal and the shape of the envelope is insufficient to separate lexical elements from the speech. Depending on different conditions, the envelope shape of the speech signal can vary widely. To solve the recognition problem, it is necessary to select the basic speech features used in the later stages of the recognition process. Primary features are determined by analyzing the spectral properties of the speech signal. Information about the amplitude of the speech signal and the shape of the envelope is not sufficient to distinguish lexical elements from speech. Depending on different conditions, the envelope shape of the speech signal can vary widely to solve the recognition problem, it is necessary to select the basic speech features used in the later stages of the recognition process. Primary features are determined by analyzing the spectral properties of the speech signal. Spectral analysis of the speech signal In processing systems, the analog speech signal is transmitted to the input of the microphone, from which an electrical signal is obtained. Then the signal is sampled in time and the amplitude is quantified Speech detection This interdisciplinary subfield of Computer science and computational linguistics develops methodologies and technologies that allow recognition and translation of spoken language into text by computers. Also known as automatic speech detection (ASR), computer speech detection, or text speech (STT). It includes knowledge and research in the fields of computer science, linguistics, and computer engineering. Some systems for speech recognition require a "report" (also called a "record"), where A separate speaker reads the text or isolates the dictionary system. The system analyzes a person's specific voice and uses it in precise tuning to recognize that person's speech, resulting in increased accuracy. Systems that do not use training are called "speaker-independent" [1] systems. Systems that use training are called "speaker dependent".

Speech recognition programs include voice user interfaces, such as voice dialing (e.g., "call home"), call forwarding (e.g., "I want to call"), and domain tools. management, keyword search (e.g., finding a podcast where specific words are spoken), entering simple information (e.g., entering a credit card number), preparing structured documents (e.g., radiology report), identifying speaker features, [2] word-to-word processing of speech (e.g., word processors or emails and aircraft is usually referred to as direct voice input. Additional voice recognition [3] or speaker ID [4] refers to the identification of the speaker, not the speaker. or confirmation can be used for cooling.

Theory

The Nontraditional Methods of MR Control

Under the traditional method of MRs control we understand the way to control using the commands passed from the operator to the MR's control system via some interface. Under the non-traditional control methods we understand in this article the following: BCI, voice control, control with gestures.

The BCI is an interface that provides a direct transmission of the information from the brain to the computing device as described by Tromov and Skrugin [5]. Recently different companies has developed the portable BCI such as NeuroSky, MindFlex, Emotiv, as described by Stamps and Hamam [6]. Some of these neurocomputing interfaces allows not only to obtain the EEG (electroencephalogram) data, but also to obtain data about the emotional state of the operator, for example used by Chepin et al. [7].

The voice control is a way of interaction between a man and a computer by the voice. This method of control is based on the processing of audio signals coming from a microphone. Speech recognition system using phonemes and grammar follows.

The gesture recognition system was developed in the "Robotics" laboratory of the NRNU MEPhI. The algorithm determines the fact of the hand getting in the graphics region (specified area) corresponding to a particular gesture. This system is installed into the wheelchair to control its movement by the hand gestures and finger movements of the patient.

Extended BCI

The main scientific and engineering idea of the described project is the development of the control system. The general ideology of the project is based on the concept of "extended BCI" proposed, by which we mean the presence in the control system the following robot-control channels: BCI, voice control, control with gestures.

The term "extended BCI" was introduced in order to emphasize that in addition to the control channel based on the parallel BCI there are other ones, not so common channels. BCI is the main control method, but in practice there are situations when a particular patient more effectively controls the wheelchair using the voice commands and/or gestures. With this approach, it is necessary to solve the problem of choosing the most correct control channel. In addition, the architecture should implement the possibility of taking into account the disease peculiarities of the particular patient and develop a decision-making mechanism based on the analysis of the information from all control channels.

The Task of Decision-Making

The concept of "extended interface" includes several different control channels of the MR by the operator. The general scheme of the decision-making system based on the data from the extended BCI interface is presented in Figure 1. When using more than one data channelto control the MR it is needed to solve the problem of the decision-making. In general, the problem looks as follows: there are several control channels and it is necessary to decide what command should be executed at a given time. The decision-making system should have the following features:

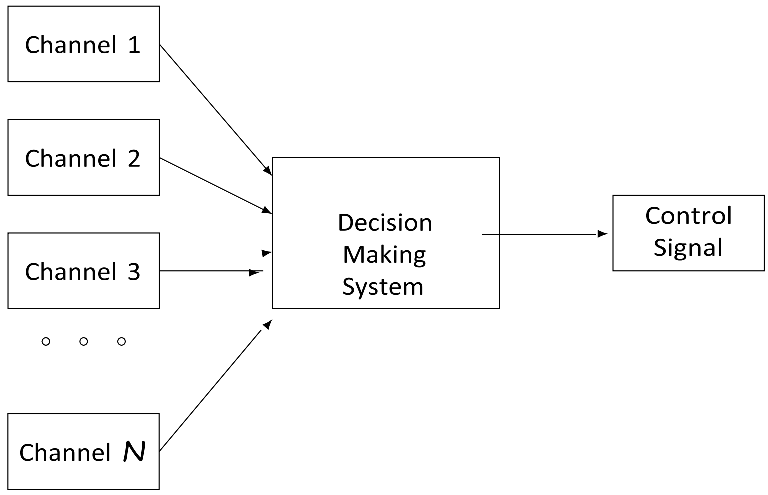

Figure 1: General problem statement

To be deterministic. If we know the specificity of working with the extended BCI-interface system components it is possible to make the deterministic decision-making system taking into account the mentioned features of the components.

The ability to have the varying credibility degree for the control system channels . Since the channels have different degree of credibility, it is logical that the information coming to the decision-making system should have the different values.

Support an asynchronous data input. Despite the fact that the part of the data is received synchronously and during a long period of time, the solutions based on asynchronously incoming information having a greater relative value should be taken timely.

Work with continuous processes. For example, when working with thought-images it is important to record not only the state but also the dynamic characteristics of the process, and the previous state. Thus, the developed method of decision-making should be similar to the automaton with a memory.

The decision-making system prototype satisfying all requirements was implemented using channels accuracy based priority accounting.

Implementation

The current "chair" hardware-software complex consists of: The wheelchair with the electric drive "Titan" LY-103-120, which is designed for the independent movement in the premises and on roads with hard-surface for the disabled people with diseases of the musculoskeletal system and injuries of the lower extremities (Figure 2a). Instead of joystick control the chair has a control unit, which has an interface with the onboard computer port.

The neural interface on the basis of the Emotiv Epoc (Figure 2b) and the software module. The Emotiv Epoc BCI allows one to obtain information not only about the fact that the user thought about the thought-images, but also the quantitative assessment of this fact (power).

b)

a) c)

Figure 2: a) The "chair", b) BCI Epoc Emotiv, c) Intel RealSense

The basic equipment for the operator video interface with the robot for the development of a system of gesture recognition (a stereo camera for fixing the movements and gestures of the operator and the set of individual cameras). The arm tracking is made using an Intel RealSense camera that is shown in Figure 2c.

The basic equipment for the operator audio interface with the chair. The voice recognition takes place with the help of English phonetics of the Sphinx-4 library developed. The dictionary consists of the words matched the available phonetics.

Possibilities

By the early 2010s, speech recognition was clearly different from recognition, also known as voice recognition. Speaker recognition and speaker independence were considered major achievements. Until then, the systems required a "training" period. The 1987 ad for the doll had a caption that read, "Finally, a doll that understands you." - despite being described as "which children can be taught to respond to their own voices." [8].

In 2017, Microsoft researchers reached a historic humane stage in transcribing conversational telephony speech into the widely used Switchboard function. Several in-depth study models have been used to optimize the accuracy of speech detection. Speech recognition error rates were reported to be as low as 4 professional human transcriptions working together on the same indicator, funded by the IBM Watson speech team in the same task. Both acoustic modeling and language modeling important parts of modern statistical-based speech detection algorithms. Hidden Markov models (HMM) are widely used in many systems. Language modeling is also used in many other natural language processing programs Classification of documents or statistical machine translation gives the probability for the vector. Each word, or (for general speech recognition systems), each has a phoneme, a different output distribution; the hidden Markov model for a sequence of words or phonemes is accomplished by combining hidden Markov models trained for individual words and phonemes.

The key elements of the most common HMM-based approach in speech recognition are described above. Modern speech recognition systems use various combinations of a number of standard methods to improve the results of the basic approach described above. Usually, a large vocabulary system is needed for context-dependent phonemes (so different left and right contextual phonemes have different concepts for HMM cases); it uses cepstral normalization to normalize for a different speaker and recording conditions; To normalize the speaker it is possible to use the normalization of the length of the vocal tract for the normalization of the male-female to adjust the overall speaker of the maximum probability linear regression. Features would be called delta and delta-delta coefficients to obtain speech dynamics and can also be used in the heteroscedastic linear discriminant analysis or can skip and use delta-delta coefficients after attachment and based projection, possibly heteroscedastic linear discriminant analysis or conversion of a global semi-linked co-operation (as well as the maximum probability of linear conversion to MLLT). Many systems use discriminatory teaching methods that abandon an absolutely statistical approach to assessing HMM parameters and instead optimize measurements related to certain classifications of learning data. Examples are maximum reciprocal information (MMI), minimum classification error (MCE), and minimum telephone error. Speech decoding term), perhaps to find the best way for the Viterbi algorithm, and here there is a choice between creating a hidden Markov model that dynamically contains acoustic and language model data and combining it pre-statically, or approach). A possible improvement in decoding is the evaluation of these good candidates so that we can choose the best one according to that particular candidate instead of retaining the best candidate and using the higher score function. The candidate package can be stored as a list approach) or as part of models. Re-scoring is usually done by trying to minimize (or approach) it: Instead of accepting the first sentence with maximum probability, we try to obtain a sentence that minimizes the expectation of a given loss function on all possible transcriptions (i.e. we obtain a statement that minimizes the average distance) to other possible judgments drawn with an approximate probability). Loss function

Dynamic change of time is a historically used approach to speech recognition but is now largely due to a more successful HMM-based approach. Dynamic change of time can vary in time or speed An algorithm for measuring the similarity between two sequences. For example, similarities in walking style can be detected even if a person is walking slowly in one video and walking faster in another, or if there is acceleration and deceleration during one observation. DTW applies to video, audio, and graphics - any data that can actually be converted to a line image can be analyzed using DTW.The popular program is automatic speech recognition to overcome different speech speeds. In general, this is a method that allows a computer to find the optimal fit between two sequences (e.g., time series) given with certain constraints. That is, the sequences are “linear” that correspond to each other. This sequence alignment method is often used in the context of latent Markov models in English. When used to predict the probability of a speech feature segment, neural networks allow for natural and effective discriminatory training. However, despite their effectiveness in classifying short-time units such as individual phonemes and individual words, [9] early neural networks have rarely been successful for continuous recognition tasks due to their limited ability to model temporal connections. one is the use of neural networks as pre-processing, modification of properties, or reduction in size ed [step before recognition on the basis of HMM. However, recently LSTM and similar repetitive neural networks and time-delayed neural networks have demonstrated improved performance in this area.Deep neural networks and denoising Autoencoders are also under investigation. A deep neural network (DNN) is an artificial neural network with several latent layers of units between the input and output layers. [10] Similar to shallow neural networks, DNNs can model uncomplicated linear relationships. DNN architectures create composite models, here in addition Ayers allows the compilation of features from lower layers, which provides greater learning ability and thus the ability to model complex patterns of speech data. [11] The success of DNNs in recognizing large lexical speech in 2010 where large output layers of DNN were adopted based on context-dependent HMM cases constructed on the basis of decision trees. See full reviews of this development and modern level in October 2014 in Microsoft Research’s latest Springer book. Automatic speech recognition and the impact of various machine learning paradigms, in particular, see the in-depth study, one of the basic principles of misinformation an in-depth study is the elimination of handcrafted feature engineering and the use of raw properties. This principle was first successfully studied in in-depth autoencoder architecture on “raw” spectrogram or linear filter-bank properties, showing its superiority over Mel-Cepstral properties by removing multiple variable changes from spectrograms. includes z. The real “raw” features of speech, the waveforms, have recently led to very large-scale speech detection results. Since 2014, there has been a great deal of interest in the “end-to-end” ASR. Traditionally on a phonetic basis (i.e., all) HMM-based model) approaches are separate components and pronunciation, acoustic, and language models. End-to-models together explore all the components of speech recognition. This is important because it simplifies the learning process and the distribution process. For example, an n-gram language model is required for all HMM-based systems, and typically an n-gram language model takes up a few gigabytes of memory, making it unwise to place them on mobile devices. Consequently, modern commercial ASR systems are hosted in the cloud by Google and Apple (as of 2017) and require a local network connection, unlike a device. Direct study of all components of speech recognition, including pronunciation, acoustics, and language model possible. This means that there is no need to carry a language model, which is very practical for applications with limited memory during deployment. By the end of 2016, the models in focus had achieved significant success, including surpassing the CTC models (with or without an external language model). Various extensions have been proposed since the original LAS model. Direct extraction of sub-word units proposed by Hidden Sequence Decompositions (LSD) and more natural than Google Brain English characters; Oxford to manage above human level The LAS has been expanded to "Watch, Listen, Participate, and Spell" (WLAS). . After the voice command, the system has a "listening window" during which it can receive speech recognition. can be used to listen to music from a built-in flash drive. Voice recognition skills vary depending on the make and model of the vehicle. Some recent [when?] Car models offer speech recognition in natural language instead of a strict set of commands that allow the driver to use complete sentences and common phrases. With such systems, the user does not need to memorize fixed command words. Speech recognition from the front is when the provider instructs the speech recognition engine that the recognized words are displayed when they are spoken and the dictator is responsible for editing and signing the document. Backward or delayed speech recognition is a digital dictation system in which the provider assigns an "a", a voice speech recognition machine, and the featured project document is edited to the editor along with the original audio file and the report is completed. Delayed speech detection is now widely used in the field. brings great financial benefits to physicians who use them in accordance with their standards. These standards require that large amounts of data be stored by EMR (now more commonly referred to as “.” (Electronic Health Records or EHR). The use of speech recognition is naturally more appropriate when creating a story text as part of a radiological/pathological interpretation, a development note, or summary: ergonomic achievements of speech recognition to enter structured individual data (e.g., numerical values or codes)) register or a controlled dictionary) is relatively minimal for people who can see and can control the keyboard and mouse.

Results and Conclusion

The “chair” project has received quite a wide coverage in Russian and the foreign mass-media: a few reports on some Russian TV-channels, in particular, on the NTV [12], as well as in the press of Spain and Spanish-speaking countries, for example in El Diario de Hoy [13], El Universal [14] and in China (Science and Technology Daily [15]).

References

1. "Speaker Independent Connected Speech Recognition- Fifth Generation Computer Corporation". Fifthgen.com. Archived from the original on November

11, 2013. Retrieved June 15, 2013.

2. P. Nguyen (2010). "Automatic classification of speaker characteristics". International Conference on Communications and Electronics 2010. Pages 147-152. doi: 10.1109 / ICCE.2010.5670700. ISBN 978-1-4244-7055-6. S2CID 13482115.
3. Abdiev, J., & Safarov, O. (2022). BASALT FIBER-BASIC (PRIMARY) CONCEPTS. *Web of Scientist: International Scientific Research Journal*, *3*(4), 212-240.
4. Ortikovich, U. J., Muzaffar o'g'li, A. J., & Rafikovich, M. R. **Fex V**3− x **O4**MAGNETIC AND ELECTRICAL PROPERTIES OF SPINEL.
5. Raufovich, M. R., & Muzaffar o'g'li, A. J. (2022). THE ROLE OF TEACHER AND STUDENT COOPERATION IN THE MODERN EDUCATIONAL PROCESS OF GOOGLE DOCS. *Web of Scientist: International Scientific Research Journal*, *3*(1), 579-589.
6. Saidov, A. S., Saparov, D. V., Usmonov, S. N., Kutlimratov, A., Abdiev, J. M., Kalanov, M., ... & Akhmedov, A. M. (2021). Investigation of the Crystallographic Perfection and Photoluminescence Spectrum of the Epitaxial Films of (Si2) 1-x (GaP) x Solid Solution, Grown on Si and GaP Substrates with the Crystallographic Orientation (111). *Advances in Condensed Matter Physics*, *2021*.
7. LEYDERMAN, A., Saidov, A. S., USMONOV, S., Abdiyev, J. M., & Suyarov, Q. T. (2021, February). INFLUENCE OF WEAK GRADED GAP ON INJECTION DIFFUSION REGIMES OF CURRENT TRANSPORT IN SEMICONDUCTOR PN-STRUCTURES. In *Congress Dates* (p. 40).
8. KUŞÇULU, N. (2021). Congress Dates.
9. Leiderman, A. Y., Saidov, A. S., & Abdiyev, J. M. (2021). INJECTION DIFFUSION PROCESSES IN THE WEAK LINEAR GRADED-BAND SEMICONDUCTOR pn-STRUCTURES. *Euroasian Journal of Semiconductors Science and Engineering*, *3*(1), 3.
10. Leiderman, A. Y., Saidov, A. S., & Abdiyev, J. M. INJECTION DIFFUSION PROCESSES IN THE WEAK LINEAR GRADED-BAND SEMICONDUCTOR pn-STRUCTURES. *Journal of Semiconductors*, *3*(1), 3.
11. Urinov, K. O., Jumanov, K. A., Khidirov, A. M., Urinov, S. K., Abdiyev, J. M., Jumaboyev, T. A., & Eshmirzayev, M. R. (2020, April). Magnetocaloric effect in polycrystalline cobalt. In *Journal of Physics: Conference Series* (Vol. 1515, No. 2, p. 022079). IOP Publishing.
12. Jurabek Abdiev, Omadjon Safarov, & Hasan Julanov. (2022). Study of the properties of polymer composites – reinforcement based on glass and basalt fibers. *Eurasian Scientific Herald*, *7*, 77–88.
13. Nizom Abdurazzokovich Taylanov, Sunatillo Xudoyor o'g'li. Urinov, & Jurabek Muzaffar o'g'li Abdiev. (2022). A Fourth-Order Runge-Kutta Method for Numerical Solution of the Kuramoto-Sivashinsky Equation. *Eurasian Scientific Herald*, *7*, 58–61.
14. "British English definition of voice recognition". Macmillan Publishers Limited. Archived from the original on September 16, 2022. Retrieved 21 February 2012.
15. "Voice recognition, definition of". WebFinance, Inc. Archived from the original on 3 December 2011. Retrieved 21 February 2022.