Technical science and innovation

Volume 2022 | Issue 1 Article 8

4-15-2022

PREPROCESSING OF SPEECH SIGNALS FOR THE SYSTEM OF RECOGNITION AND SPECTRAL ANALYSIS OF SPEECH

Sherzod Nematov
Tashkent State Technical University, 100095, st.University 2, Tashkent, Uzbekistan, shnematov@hotmail.com

Y Kamolova
Tashkent State Technical University, 100095, st.University 2, Tashkent, Uzbekistan

Recommended Citation

Nematov, Sherzod and Kamolova, Y (2022) "PREPROCESSING OF SPEECH SIGNALS FOR THE SYSTEM OF RECOGNITION AND SPECTRAL ANALYSIS OF SPEECH," *Technical science and innovation*: Vol. 2022: Iss. 1, Article 8.
DOI: https://doi.org/10.51346/tstu-01.22.1-77-0163
Available at: https://btstu.researchcommons.org/journal/vol2022/iss1/8

PREPROCESSING OF SPEECH SIGNALS FOR THE SYSTEM OF RECOGNITION AND SPECTRAL ANALYSIS OF SPEECH

Sh.K. Nematov, Y.M. Kamolova
Tashkent State Technical University
University St,2 100095, Tashkent city, Republic of Uzbekistan

Abstract: *The effectiveness of solving applied problems in the field of speech technologies is determined by the completeness of the use of phonetic information obtained in the study of the properties of natural speech. The representation of a speech signal in digital form opens up wide possibilities for its analysis and processing. Having a digital representation of a speech signal, we can think about metrics, that is, the parameters of this signal, with the help of which the program can recognize sounds, words and sentences with approximately the same result that a healthy hearing aid and a healthy human brain give.*

Keywords: *speech signals, fragments of speech signals, time segmentation, spectral analysis, Fourier transform, Hamming window, PCM-encoded.*

INTRODUCTION. Creating a natural means of human communication with a computer is currently the most important task of modern science, with speech input in the most convenient way for the user. Speech recognition is the task of classifying the patterns of acoustic characteristics of speech signals. Subsystem for the preliminary processing of speech signals.

The preprocessing of the speech signal includes the following stages: speech signal input process; extraction of the speech signal boundary; digital filtering; slicing the speech signal by overlapping frames; signal processing in the window; spectrum conversion; frequency spectrum normalization. Let's consider the stages in detail.

Speech signal input process. Sound is input in real-time via sound card or PCM-encoded WAV files. Sampling rate 8 kHz and 16-bit quantization are typical parameters in speech transmission, storage and processing systems. File handling has been envisaged to facilitate multiple repetitions of neural network processing, which is particularly important in training.

Speech signal boundary extraction

The following characteristics of the speech signal are used to extract sections of the input signal containing only speech:

The short-term energy of the speech signal;

Number of zeros of intensity (instantaneous frequency);

Pause report value distribution density.

The short-term energy of the speech signal and the number of zeros in intensity are used simultaneously to separate speech from the input signal. It is also possible to remove the pause from the output signal using a method based on the normal (Gaussian) distribution [1].

Digital filtering*.* Together with a useful signal a variety of noises usually get in. Noise has a negative impact on the performance of speech recognition systems, so it must be dealt with. Two types of digital filtering are used to reduce noise in the subsystem: band-pass filter; a passband filter; a pre-filter. A passband filter can be thought of as a combination of a low-pass and an upper-pass filter. Such a filter filters all frequencies below the so-called lower pass frequency, as well as above the upper pass frequency.

Pre-filtering is introduced to reduce the influence of local distortion on the characteristic features that will later be used for recognition. For spectral equalization of the speech signal it must be passed through a weighted low pass filter. Slicing the speech signal into overlapping frames

To obtain feature vectors of equal length, the speech signal has to be sliced into equal parts and then transformed within each frame. Overlapping is used to prevent loss of signal information at the edge. The smaller the overlap, the lower the resulting dimensionality of the feature vector specific to the area in question. The overlap is sometimes omitted due to the saving of computational resources, as it slows down the data processing speed significantly. Typically segment lengths should be chosen to coincide with a time interval of 20-30ms. Signal processing in the window

In-window processing is used to reduce the boundary effects resulting from segmentation. It is common practice to multiply the signal by a window function to suppress unwanted boundary effects. The Hamming window is used as the function.

Spectral transformation. Information about the amplitude and shape of the envelope of a speech signal is not enough to extract lexical elements from speech. Depending on different circumstances the envelope shape of the speech signal can vary within wide limits. To solve the recognition problem, it is necessary to identify the primary features of speech, which will be used in later stages of the recognition process.

The primary features are extracted through the analysis of the spectral characteristics of the speech signal. The Fast Fourier Transform (FFT) is used to obtain the frequency spectrum of the speech signal. The FFT is presented to obtain the amplitude spectrum and signal phase information (in real and imaginary coefficients). The signal phase information is discarded and the amplitude spectra are calculated. The logarithm of this value is more often used [2].

$$\mathrm{NS} = \frac{\mathrm{N}}{2} \qquad 1$$

Where -amplitude spectrum of the i-th frequency,

-real coefficient,

- is imaginary coefficient,

N - FFT size,

- size of informative part of spectrum.

MATERIAL AND METHODS. Since audio data does not contain imaginary part, by FFT property the result is symmetrical, i.e., the size of the informative part of the spectrum NS is N/2.

Normalizing the frequency spectrum

All the calculations in neural networks are performed over floating point numbers. So values of parameters of objects classified by neural networks are limited by the range [0.0, 1.0]. To perform neural network spectrum processing, the resulting spectrum is normalized to 1.0. To do this, each component of the vector is divided by its maximum component.

Spectral analysis of a speech signal

In a processing system the analogue speech signal is fed to a microphone whose output is an electrical signal. The signal is then sampled in time and quantized in amplitude. During the quantization process, distortions (quantization errors) occur, which essentially means loss of information.

Initial information is represented as a function of amplitude versus time (usually .wav files). The resulting digital data sequence is further processed to determine the frequency range and other characteristics of the signal from which it can be reproduced. Since the signal is usually noisy, the simplest way to remove noise is to zero out those signal values that are less than some threshold value [3].

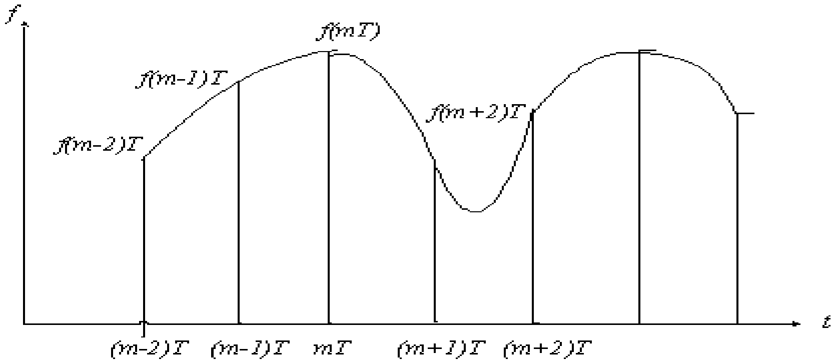

Fig. 1. Signal quantization

The temporal representation of a signal, i.e. the change in the signal as a function of time, allows the amplitude, energy, power and duration to be determined. Signal models as a function of time are used to analysis the shape of signals. Complex signals can be represented as a system of basic functions

$$f(t) = \sum_{k} c_k \ \varphi_k \ (t) \qquad 2$$

where is the interval of signal's existence $t \in [t_1, t_2]$ $[t_1, t_2]$

Given a chosen set of basic functions, the signal f(t) is completely defined by a set of dimensionless coefficients. Such sets of numbers are called discrete signal spectra. The basis function, where used in the Fourier transform.

Apart from the temporal characteristics of the signal, its frequency properties are also important. Frequency representations of a function in the form of a spectrum are used to study them. The spectral representation of a signal is its decomposition into a finite or infinite sum of harmonic signals. Knowledge of the frequency properties of the signal can solve the problem of signal identification (determining its most informative parameters), filtering (identification of a useful signal against the background of interference), the choice of sampling rate of a continuous signal, as this parameter is decisive for the processing equipment.

A set of sinusoidal components of a complex sound, given by means of amplitudes and frequencies of these components represent an acoustic spectrum. For spectral analysis of the signal, the discrete Fourier transform (DFT) and the fast Fourier transform (FFT), which is an accelerated FFT procedure, are used [4].

Consider a finite series of discrete signals f(mT) at m = 0,1, 2..., M-1.

The function F(K) defined by formula (2) is called the discrete Fourier transform for f(mT).

$$F(K) = \sum_{m=0}^{M-1} f(mT) \left(e^{-\frac{2\pi}{M}}\right)^{Km} \qquad 3$$

where, К = 0,1, 2..., M-1, a - is a complex function with an imaginary unit.

$$e^{-2\pi/M}$$

If an FFT is found, it is possible to reconstruct the original signal (inverse Fourier transform) from the discrete values of the signal.

According to Kotelnikov's theorem, an arbitrary signal whose spectrum contains no frequencies above Fv Hz can be fully recovered if the counting values of that signal, taken at equal time intervals of 1/(2-Fv) s, are known.

The inverse Fourier transform is defined by

$$f(mT) = \frac{1}{M}\sum_{K=0}^{M-1} F(K)\left(e^{\frac{2\pi}{M}}\right)^{Km} \tag{4}$$

where m = 0,1, 2..., M-1.

A real speech signal has a finite duration and when represented in the frequency domain its spectrum is unlimited. Therefore, the signal is segmented into portions of the order of 10 ms, where it is considered stationary.

One of the options for speech signal preprocessing is shown in Fig. 2.

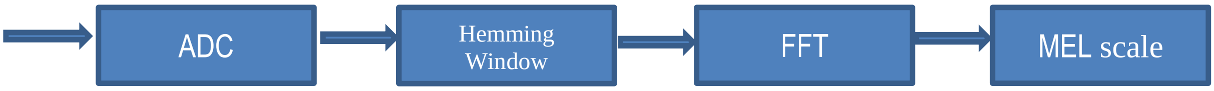

Fig. 2. Speech signal preprocessing

Signal weighting by Hamming window weighting function (fig. 3) reduces spectral distortions of the signal due to boundary conditions. The application of the time window is appropriate for intervals exceeding 15 ms or including several periods of the fundamental tone.

The value of the weighting function is given by the formula:

$$W_n = \{0.54 - 0.46 * cos\left(\frac{2\pi n}{N-1}\right), 0 < n < N \tag{5}$$

Different parts of the spectrum are not equally informative: the low-frequency region contains more information than the high-frequency region. Therefore, the high-frequency part of the spectrum is compressed in the frequency space. The most common method due to its simplicity is logarithmic compression, or Mel-compression

$$m=1125*log(0.0016f+1) \tag{6}$$

where f is frequency in the spectrum, Hz; m is frequency in a new compressed frequency space [5]. Samples of speech signal segments are presented in Fig. 3.

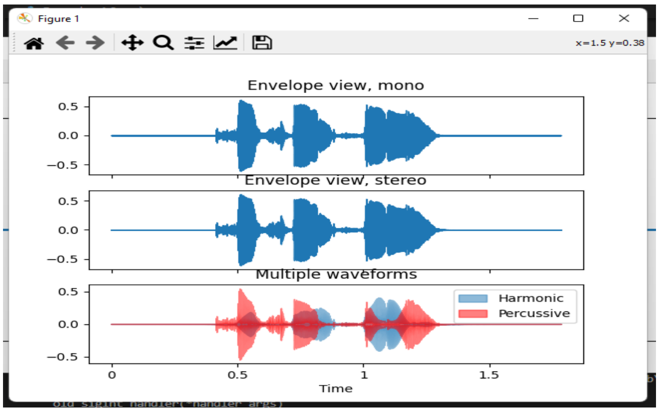

Fig. 3. Speech signal segments: a) segment extracted using Hamming window b) vowel segment

Fig. 4. shows the result of frequency analysis of a 16-bit speech signal with a sampling frequency of 11025 Hz, performed in the Analyze - Frequency Analysis window of the Cool Edit audio editor. This airborne spectrum is generated by the vocal cords and oral sound source through selective resonance that occurs during the transmission of sound along the speech pathway [9].

The speech tract is made up of the larynx, oral cavity, tongue, nasal cavity, etc. The editor allows you to record and play files in different audio formats, edit, convert and mix audio files, generate noise and different tones, perform frequency analysis, etc.

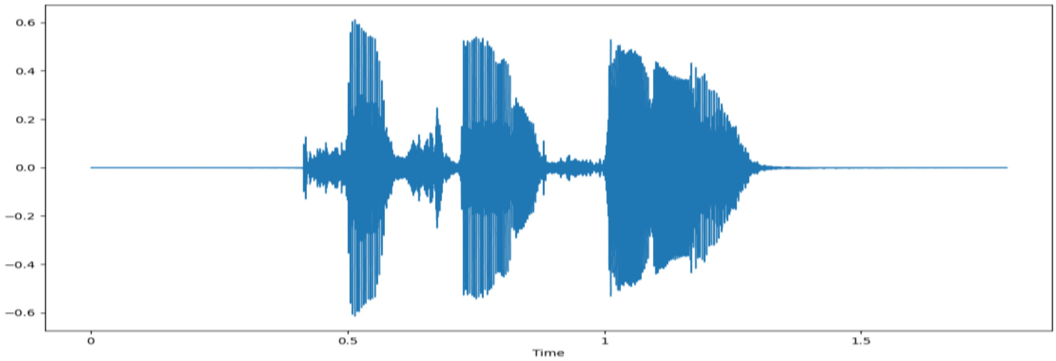

Fig. 4. Spectrum analysis window: Fsh - noise frequency, Fo - base tone frequency, 2F0-5F0 - overtones, 3F0, F1-F3 - formant frequencies.

The speech signal has a number of features that need to be taken into account:

- the properties of the signal are not constant over the section of word length chosen for analysis, it is a non-stationary random process,

- complexity of the signal form (speech is more like noise than a regular signal).

RESULTS AND DISCUSSION. To overcome these difficulties, as stated above, the discrete random process of the digitized speech signal is considered stationary at an interval of the order of 10 ms, since the parameters of the voice path do not change significantly at this interval. This is an experimentally justified time interval [6].

The main task of signal processing is to calculate from the input signal a set of parameters (signs) that contain information about the signal that is used for synthesis and recognition.

Typically, the following signal parameters are defined:

frequency of the main tone to form the trajectory of the main tone period;

short-term energy to synthesize the short-term energy trajectory;

linear prediction coefficients (LPC) to build a transfer function trajectory of the speech path;

formant frequencies to reproduce the trajectory of formant frequencies.

Formants are the maximums of the energy distribution of the audio signal in the coordinates amplitude, frequency, and time. In order to achieve good signal quality, it is sufficient to set the parameters of a few higher formants of the fundamental tone. When high quality is to be achieved, some of the above parameters or their combinations are used. The problem of separating speech from noise is quite difficult because the energy of the speech signal is almost equal to the noise energy when pronouncing some consonants ("f", "p", "t" etc.).

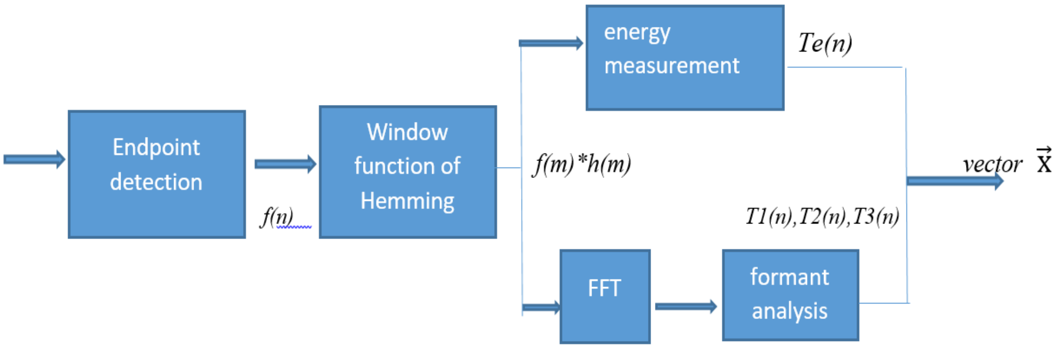

Fig. 5. Signal processing unit

One of the phrase extraction algorithms (proposed by L. Rabiner) is based on the measurement of two simple characteristics - energy and number of zero crossings. The average energy is calculated using a 10 ms window (approximately 110 counts) in which the squares of the counts are summed.

Samples of signals and their spectrograms are given in fig. 6.

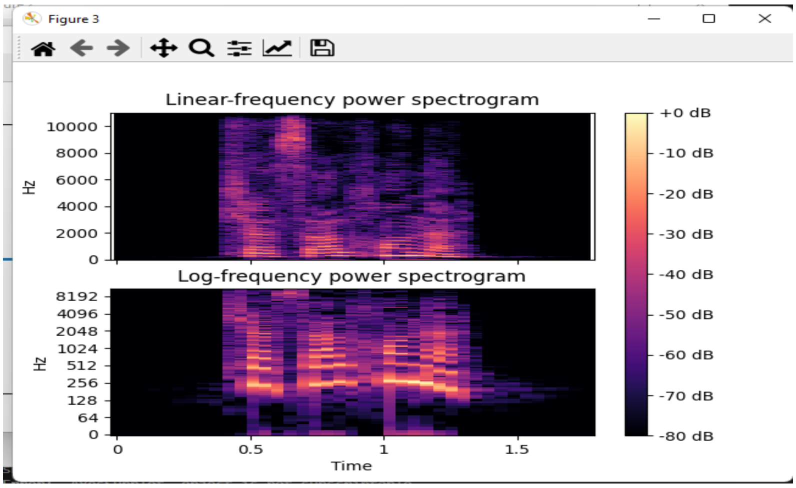

Fig. 6. Signals and their spectrograms.

The fundamental frequency, energy and duration provide the prosodic characteristics of speech [9].

The visualization of the signal parameters in coordinates amplitude, frequency, time is shown in figure 7.

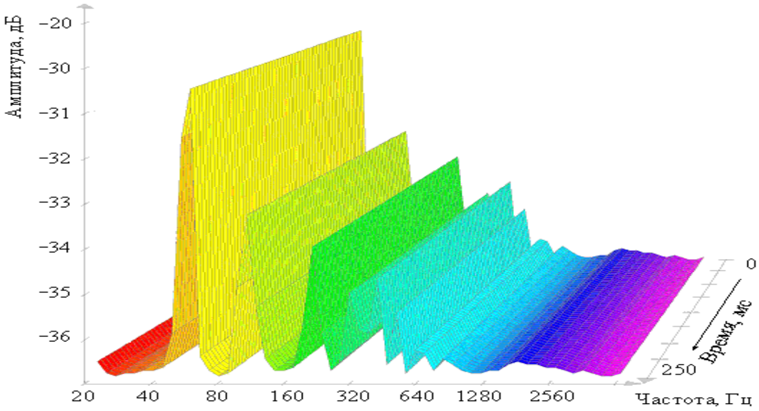

Fig. 7. Signal in the frequency-amplitude-time coordinate.

The fundamental frequency is one of the most important features of a speech signal. There are different ways to estimate it, spectrum analysis can be used in particular. If FFT is found, we can reconstruct the original signal (inverse Fourier transform) using discrete signal values. The structure of the basic tone frequency calculation system is shown in Fig. 8.

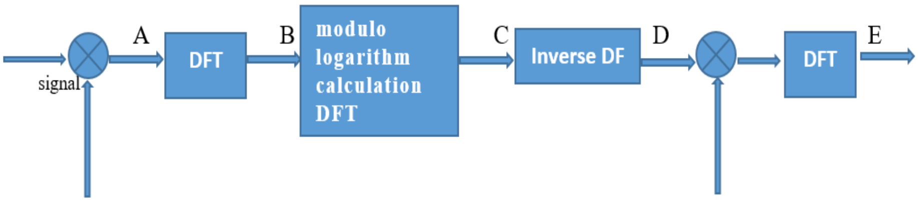

Fig. 8. Principal Tone Frequency Calculation

Since the inverse DFT is linear, the signal at point D (called the signal keystroke at point A) is equal to the sum of the excitation function and the impulse response of the vocal tract. It can be shown that the keystroke at D allows separating the effects of the excitation and the vocal tract characteristics. Indeed, the excitation signal can be viewed as a quasi-periodic pulse sequence with a near linear Fourier transform, with spectral lines corresponding to harmonics of the fundamental frequency. Calculating the logarithm of the modulus does not change the linear nature of the excitation function spectrum [7].

The inverse DFT gives a new quasi-periodic sequence of pulses with pulse intervals equal to the period of the fundamental frequency. Thus, the excitation signal cepstral must consist of pulses located near n = 0, T, 2T,..., where T is the period of the fundamental tone. The impulse response of the vocal tract is typically a non-zero sequence over an interval of 20-30 ms. After computation of the modulus logarithm and inverse DFT we obtain a sequence of a small number of non-zero samples, which is usually less than the number of samples at the fundamental tone period [8].

The result of calculating the cepstral of a vocalized signal is shown in Fig. 9.

Speech frequency phoneme Fourier

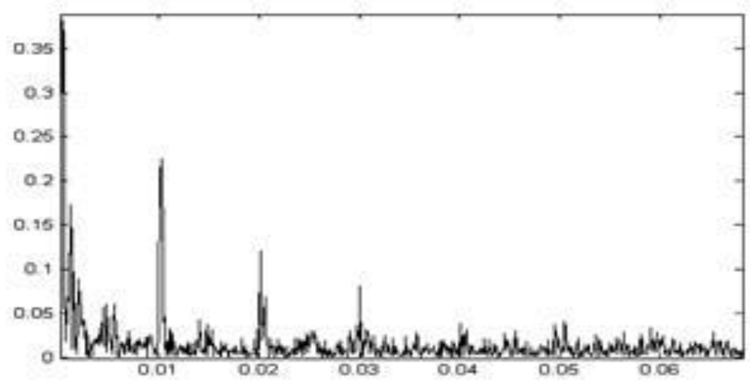

Fig. 9. The keystone of a vocalized signal.

CONCLUSION. Studies have shown that for a vocalized segment of speech, a peak appears in the cestrum, corresponding to the pitch period. For an unvoiced segment, such peaks do not appear in the cestrum. This property of the cestrum can be used to classify sounds as voiced, unvoiced, and to calculate the pitch period of voiced speech.

The cestrum obtained by the method described above is examined in order to find a peak in the range of possible pitch values (4-40 ms).

If the maximum cestrum does not exceed the threshold, then the segment is classified as unvoiced. If the peak in the cestrum exceeds the set threshold, then the segment is classified as voiced, and the peak coordinate gives an estimate of the pitch period, and the pitch frequency is calculated accordingly. Thus, it is possible to construct an efficient algorithm for extracting the fundamental frequency.

References:

1. E. Eificher., B. Jervis. Digital Signal Processing. “*A practical approach”,* **2004**. 992.
2. R. Lawrence., B.H. Juang. Fundamental of Speech Recognition. *“Prentice Hall”,* **1993**.
3. S. Manolakis., G. Dimitris. Applied digital signal processing: theory and practice. “*MIT Press, Cambridge”*, **2012**, ISBN 978-0-521-11002-0 (Hardback)
4. A.V. Sergiyenko. *“Digital signal processing”,* **2002**. 608.

5. N.V. Le., J.P. Panchenko. Pre-processing of speech signals for speech recognition system. “*Young scientist*”, **2011**. 74.
6. I.V. Bocharov. Recognition of speech signals based on the spectral estimation method [Electronic resource]. “Researched in Russia”, **2003**. 1537.
7. I.V. Bocharov., D. Yu. Akatiev “*Access mode: http: // journal.ape.relarn.ru / articles”,* **2003**. 130.
8. A. Dorokhin., D.G. Starushko., E.E. Fedorov., V.Y. Shelepov. Segmentation of the speech signal*. “Artificial intelligence”*, **2000**. 450.
9. N.A. Krasheninnikova. The main factors that interfere with the recognition of speech commands. *“Siberian scientific bulletin”*, **2011**. 188.
10. Sh.K. Nematov., Y.M. Kamolova. Recognition of speech signals based on the method of spectral analysis. “*Journal technical sciences and innovation”,* **2021**. 212.
11. G.S. Khaydarova., Y.M. Kamolova., U.A. Obidova., G.B. Yuldasheva. Rehabilitation of hearing impaired children with gaming programs after cochlear implantation. “*Znanstvena misel journal. Slovenia”*, **2019**. 39.

UDC 681.5

CONTROL, MODELING AND CONTROL OF THE COMBUSTION PROCESS IN GAS COMBUSTION FURNACES

N.R. Yusupbekov, S.M. Gulyamov, A.T. Rajabov
Tashkent State Technical University
University St.,2 100095, Tashkent city, Republic of Uzbekistan

Abstract: *The current state of the theory and scientific - theoretical foundations of automatic control, modeling and control of technological processes of fuel combustion in combustion furnaces are analyzed and he trends for their further improvement and development are indicated. The analysis of methods of diagnostics, control and automatic control of the process of combustion of gaseous fuel in gas-burning furnaces and installations has been carried out, and research objectives have been formulated; Methods for monitoring the ionization characteristics of the primary measuring transducer in the form of ionization transition currents in the “electrode-torch-burner” and “electrode-torch-electrode” circuits are analyzed by using high-temperature measuring electrodes immersed in various zones of the gas-air torch. The relationship between the parameters of the combustion zone of the furnace and the productivity of a two-wire gas burner with the combustion coefficient is revealed. It has been established that the measurement of low-frequency electrical signals of the torch is accompanied by losses, leading to a decrease in the useful signal. It is shown that one of the promising ways to modernize safety automation systems for the fuel combustion process is the development and implementation of monitoring and control devices in the fuel combustion mode, which provides for obtaining the specified characteristics of the combustion process and combustion products in the working volume of the studied gas combustion plant.*

Keywords: *gas combustion plants and furnaces, control and regulation of the combustion process, natural gas, two-wire turbulent gas burner.*