Mathematical Models of the Modern Educational Space: Virtual Communication, Processing of Natural Language Information, Normalization of Speech Signal*

Ravshanbek Zulunov[1,†], Hryhorii Hnatiienko[2,†], Vladyslavi Hnatiienko[2,†] and Larysa Myrutenko[2,*,†]

[1] *Tampere University of Applied Sciences, 4 Kalevantie, 33100 Tampere, Finland*
[2] *Taras Shevchenko National University of Kyiv, 64/13 Volodymyrska str., 01601 Kyiv, Ukraine*

Abstract

This paper reviews and improves the tools for researching the modern educational space. Particular attention is paid to the means of communication, which largely shape the educational space. In particular, the methodology of applying natural language information processing tools in the study of educational space is considered. Today, speech recognition is often used in video surveillance and access control systems, as well as in various mobile and cloud platforms. A speech recognition system is a technology that can convert human speech into text. It can work autonomously, or it can learn the pronunciation of a particular user. Voice recognition is a part of speech recognition technology. Voice identification is used in biometric verification to restrict access to personal files.

Keywords

virtual communication, text data, application methodology, natural language information, speech, speech signal, discrete signal, signal normalization

1. Introduction

Text data is an attribute of our civilization: we see it when we read books, newspapers, and other printed materials, search for information on the Internet, use Facebook and Twitter, communicate with each other on various forums, and so on. The amount of this data is growing exponentially. About 80% of text data is unstructured text. These are Wikipedia articles, web pages, blogs, emails, social media posts, e-books, etc. It is impossible to read and process all of this textual data, and to extract the most useful information from it, it needs to be structured, organized, systematized, etc. Thus, there is a need for tools that help people process unstructured texts more efficiently [1, 2]. Therefore, the involvement of computers in solving such tasks is quite natural [3].

In addition, the article deals with speech, speech detection, and speech signal normalization. Speech recognition is evolving nowadays. Today, speech recognition is often used in video surveillance and access control systems, as well as in various mobile and cloud platforms. A speech recognition system is a technology that can convert human speech into text [4, 5]. It can work autonomously, or it can learn the pronunciation features of a particular user. Voice recognition is a part of speech recognition technology [6, 7]. Voice identification is used in biometric verification to restrict access to personal files. The system memorizes a person's voice and distinguishes it from other voices.

CPITS 2025: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, February 28, 2025, Kyiv, Ukraine
* Corresponding author.
† These authors contributed equally.
zulunovrm@gmail.com (R. Zulunov); g.gna5@ukr.net (H. Hnatiienko); vladgnat1483@gmail.com (V. Hnatiienko); myrutenko.lara@gmail.com (L. Myrutenko)
0000-0002-2132-0834 (R. Zulunov); 0000-0002-0465-5018 (H. Hnatiienko); 0009-0000-2678-5158 (V. Hnatiienko); 0000-0003-1686-261X (L. Myrutenko)

2. The current state of research on the problem

Speech is a historically formed form of communication between people through language structures created based on certain rules. If air is used as a conductive medium for the transmission of information (communication), then speech is obtained—a sound vibration characterized by frequency and amplitude. Speech is an information carrier signal used by a person to transmit messages. By its physical nature, it is an acoustic signal that changes continuously over time. To emphasize the essence of this signal and distinguish it from other types of signals, speech is called a speech signal in the technical literature. In addition, the terms “speech”, “speech signal” and “spoken speech” are used interchangeably, except when it is necessary to emphasize the meaning of a separate term [8].

Speech, as a mode of communication rooted in history, facilitates interaction among individuals through language structures shaped by specific rules. When air serves as the conduit for information transmission, speech emerges as a manifestation—a dynamic sound vibration defined by its frequency and amplitude. This auditory phenomenon, inherently tied to the physical realm, operates as an information carrier signal, enabling individuals to convey messages effectively.

The very essence of speech lies in its acoustic nature– a continuous modulation of sound over time. This continuous evolution of sound defines the dynamic quality inherent in speech, making it a nuanced and adaptive means of expression. In the realm of technical literature, the term "speech signal" is employed to accentuate the unique characteristics of this form of communication. This categorization helps to distinguish speech signals from other types of signals that may exist in various communication modalities [9].

Within the technical discourse, the interchangeable use of terms such as “speech,” “speech signal,” and “spoken speech” prevails, demonstrating the fluidity in referencing this complex form of communication. However, a nuanced approach is adopted when precision is paramount, warranting the emphasis of a specific term to convey a distinct facet of the communication process.

In delving into the intricacies of speech, it is crucial to recognize its dual identity as both a historical construct and a contemporary tool for conveying information. The structured rules that underpin language systems have evolved, shaping speech into a multifaceted vehicle for human expression. By employing air as the medium for communication, speech transcends its historical roots, embracing technological and scientific dimensions that characterize it as a signal—imbued with meaning and purpose in the intricate tapestry of human interaction.

Speech recognition technology, or Speech-to-Text (voice-to-text), appeared at the end of the last century, but programs learned to efficiently convert human speech into text only in the 2000s, as IT technologies and machine learning developed. Today, speech recognition systems are widely used in everyday life and business, because it significantly saves resources.

This is a complex multi-stage algorithm, so we will try to describe the general principle of operation. If you tell voice search “Taras Shevchenko”, the phone will not hear the name of the famous writer, but a sound signal without clear boundaries. Based on this continuous signal, the system reconstructs the phrase reproduced by a person as follows:

- First, the device records a voice request, and the neural network analyzes the speech stream. A sound wave is divided into fragments—phonemes.
- The neural network then accesses its templates and matches the phonemes to a letter, syllable, or word. Next, an order is formed from words known to the program, and it inserts unknown words according to the context. The result of combining information from these two stages is the translation of speech into text.

At the dawn of development, the Speech-to-Text process consisted of an elementary acoustic model—human speech was compared with patterns. However, the number of dictionaries in the system was not enough for accurate recognition; the program often made mistakes.

Thanks to the learning ability of neural networks, the quality of speech recognition has increased significantly. The algorithm knows the typical sequence of words in live speech and can perceive the structure of the language—this is how the language model works. Each new processed voice information affects the quality of processing of the next one, reducing the number of errors.

Speech recognition technology allows us to search for the necessary information and create a route using the navigator. Here are a few other areas where using Speech-to-Text has made life easier:

- Telephony. The technology saves not only the caller's time but also the company's resources. Using voice dialing and a robot, customers can order goods, answer surveys, and receive advice without the participation of managers.
- Household appliances and personal computers. Today you can control various devices with your voice: switches, lighting systems, and gadgets. You can train your computer to recognize your voice (with Windows and Mac systems).

Speech recognition allows you to automate many business processes, from sales and customer service control to protection from fraudsters.

Using this technology, analytics of telephone conversations with customers has become easier and cheaper: the system automatically records calls and collects data to increase conversion. For example, the MANGO OFFICE speech analytics system helps you find out which competitors your customers most often compare your product with. You create tags for competitor mentions, analyze conversation reports, and understand how to improve your marketing strategy. You can also analyze the work of employees—mark stop words, and monitor compliance with sales scripts. If you need to transcribe speech from a video, you can download an audio file from it and upload it to a speech analytics service. Speech on video must be clear, so use a microphone when speaking on video.

Another area where speech analytics helps business development is interactive voice systems (IVR). It is an indispensable tool in call center management. Speech-to-Text recognizes the client's speech, and the voice robot automatically selects the necessary information to answer or transfers the call to an operator. The technology reduces the number of abandoned calls, as many people are late or unable to press buttons in the voice menu. Service control services do not need to conduct additional surveys: this can be done automatically, and then analyze the reports. Bank security teams use speech analytics to protect customers' data [10, 11].

3. Mathematical model of virtual communication and some aspects of leadership psychology in virtual space

The famous philosopher Johan Huizinga [12, 13] convincingly proved that play, to a greater extent than labor, was a formative element in human culture, that the most fundamental human activity takes place in the field of play, and that all human culture was a form of play. The role of play has become especially characteristic in the computer age, and for modern civilization, the Internet has become a space that generates new forms of human interaction, new principles for designing their interaction, problems of "virtual" freedom, and many other problems. Various forms of human activity are carried out on the Internet: communicative, cognitive, commercial, and gaming. The world in which we live is a fragment of the real world, which is perceived by humans through the senses and can be described by the following model:

$$F^1 = F^1\left(f_i^1, g_i^1, i \in I^1\right), \tag{1}$$

where F^1 is the function of perception by the senses, f_i^1 are the human senses, g_i^1 are the thresholds of sensory sensitivity that act as filters that allow us to perceive only information that is essential to the situation, I^1 is a set of indices of the senses: vision, hearing, smell, taste, touch, balance, and

the sense of body position in space. In virtual reality, the sensory thresholds of the senses change significantly: some senses become more acute and hypertrophied, and the thresholds of others become zero, meaning that these senses are not used in virtuality.

Virtual reality, created by the Internet environment, is a space for human interaction. Electronic means of communication have become an extension of the human nervous system and contribute to the globalization of society. Virtual reality is characterized by the following features: it imposes a new rhythm of life, contributes to changing people's perception of the world, shifts and even destroys the points of beginning and end of events, violates the principle of irreversibility as a fundamental property of our real space-time, gives the right to make mistakes in the artificial world, promotes anonymity and pseudonymity of communications in the virtual world, allows staging human personality, etc. Virtual communication has some features that distinguish it from real communication. Communication is a prerequisite for managerial actions—managers spend about 80% of their working time on communication. Virtual communication can be described in terms of the following functional dependency:

$$F^2 = F^2\left(f_i^2, i \in I^2\right), \tag{2}$$

where f_1^2 is anonymity, which is an effective way of managing the impression of oneself, contributes to psychological emancipation, non-normativity, and unrestrictedness by the norms of social roles; f_2^2 is in conditions of limited sensory capabilities in virtual reality, the resonance of communication becomes primarily the root cause of emotional intimacy; f_3^2 are the possibility of expressing staged feelings to the interlocutor rather than real ones; f_4^2 are limited possibilities of expressing feelings in the form of emoticons and textual interpretations; f_5^2 is the voluntary nature of virtual contacts, the possibility of their interruption at any time; f_6^2 are the destruction of the stable self-identification and individuality of the interlocutor.

One of the features of the 21st century is the massive creation of virtual organizations. The activity of a virtual organization can be represented as the following function

$$F^3 = F^3\left(f_i^3, i \in I^3\right), \tag{3}$$

where f_1^3 are the mission and vision of the organization; f_2^3 is organizational culture; f_3^3 is organizational structure; f_4^3 is the management structure of the organization; f_5^3 are information flows.

The features of a virtual organization are: functioning in the information space, voluntary participation, independence of the organization's members, free configuration of relationships between members, common values of members in the information environment, limited interaction, territorial distribution of members, etc. [14]. In his theory of organization, Weber pointed out the need for a clear formal fixation of organizational rules and norms [15].

A game is a type of activity characterized by the interaction of players whose actions are limited by rules and aimed at achieving a goal. The game has a strictly regulated hierarchy of players; it is not declarative and indicative, but rigidly implemented and realized. At the same time, the hierarchy in the game does not correlate with the social hierarchy. The impression of a player is formed under the influence of the following factors: the player's self-identification, desired and undesired image, role restrictions, and cultural and ethical values.

To adequately describe the game, we should introduce the concept of a virtual charismatic leader. Leadership in this case has two aspects: virtual power and virtual charisma. It is known that power is one of the main aspects of an organization's existence and provides the leader with 2/3 of the influence necessary for leadership. Knowledge and experience in a real organization, according to expert estimates, affect the result [16]. The activities of a leader in a virtual organization can be

represented in the form of two blocks of key competencies—three managerial and five leadership competencies:

$$F^4 = F^4\left(f_i^4, i \in I^4\right), \tag{4}$$

where f_1^4 are planning (goals, objectives, actions, resources, etc.), f_2^4 are managing subordinates, creating a control system, f_3^4 is exercising control (monitoring activities, identifying problems, and eliminating them), f_4^4 are forming organizational strategy, goals, and organizational culture, f_5^4 are forming communications, coordinating coalitions, managing relationships, f_6^4 are motivation and encouragement, f_7^4 are forming and maintaining values, f_8^4 are training and development.

The formalization of the leader's virtual behavior in the game can be described by introducing a formula:

$$F^5 = F^5\left(f_i^5, i \in I^5\right), \tag{5}$$

where f_1^5 are the natural properties of a person according to formula (1), f_2^5 are behavior as a function of type (2) of the natural properties of a person as a result of socialization, f_3^5 are the virtual environment according to formula (3) that has developed around the game, f_4^5 are the real external environment around the virtually charismatic leader as an individual, f_5^5 are the material and virtual resources available to players, f_6^5 are the leadership qualities described by formula (4), f_7^5 are psychological characteristics, and f_8^5 are the rules of the game.

Humans operate in natural, social, cultural, and other environments. With the development of civilization, human presence in the information environment is increasing. Today, cyberspace has become an integral part of the noosphere. The structure of cyberspace is determined by the processes of creating, storing, transmitting, distributing, processing, consuming, and perceiving information, procedures for interacting with social institutions, norms of cyberspace ethics, etc. Mathematical modeling and decision theory methods can be successfully used to study virtual space [17].

4. Psychological aspects of decision-making in virtual reality

The transition from agrarian to industrial and then to post-industrial society was accompanied by an increase in entropy, a loss of structure in society, and a tendency toward systematic fluidity and disordered interrelationships. Entropy is a measure of unstructuredness, unpredictability, and uncertainty of the values that describe certain objects. In terms of degrees of freedom entropy can be defined as a measure of the connectedness of an object's degrees of freedom [18]. The more degrees of freedom, the more unpredictable an object is, and the higher its entropy. The concept of entropy is associated with disorder, equilibrium, homogeneity, equality, freedom, stability, and ignorance, while the concept of negentropy is associated with order, disequilibrium, heterogeneity, inequality, constraint, instability, and knowledge. Increasing connectivity between individuals in a social system leads to a decrease in total entropy, and the greater the connectivity, the lower the total entropy—it decreases by the amount of predictability that comes from connectivity [18]. Thus, the way to reduce the entropy of a system is to increase the interconnectedness of the elements and the coherence of their actions. Entropy is minimal in a rigidly hierarchical, predictable, well-defined system. When people have some kind of relationship, or ties—playful, business, family, official, friendship, etc.—it becomes much easier to predict the behavior of such a social group. Often, such ties lead to the coherence of actions, which translates into overall systemic behavior.

Game structures in social systems are low-entropy, ordered social integrities. The game space has its unconditional order. A positive feature of the game is that it creates order, and organizes

actions. In an imperfect world, in real life, it creates a temporary, limited perfection [19]. The game gives rise to a special form of human connection with the world in general and the social world in particular. The intensive use of games in various forms by modern people can be considered a subconscious response to the increasing entropy of post-industrial society.

In recent decades, computer games have become extremely popular. In this case, a game is a type of activity characterized by the interaction of players whose actions are limited by rules and aimed at achieving a goal. A computer game is a way of transforming a human personality, self-identification, and creative individuality. Players begin to subconsciously perceive the computer as an extension of their personality in space. The main psychological features of a computer game can be identified [18, 20], which influence decision-making in the game environment. The features of this phenomenon and their detailed description are given in Table 1.

Table 1
The main features of a computer game and their characteristics

Number in order	A feature of a computer game	Comments
1	The game violates the basic property of time—its irreversibility	The source of the game's appeal lies in the change in a person's relationship with time. In virtual reality, there is no such fundamental property of our world as irreversibility. A person can do any action and has the opportunity to go back a few steps back in the game and prevent their own mistake
2	Players can quickly create their new virtual heroic image	At the same time, the player gets used to a different, more dynamic pace of life
3	Virtual reality gives the right to make mistakes in the computer world	Under the terms of the game, you can repeat the attempt to achieve the goal in a short time, while in real life you can wait for years for the opportunity to repeat the attempt
4	The game has simple, tangible, artificial, but understandable rules	The game is much simpler than life. The player is attracted to strict regulation, while in society there is always uncertainty. The friend-or-foe distinction in a game is much more transparent, clear, and less blurred than in the real world. The meaning of the game is always simpler than the meaning of life
5	The game space has communication features	The game space is characterized by a certain linguistic design—slang, jargon, the use of specific terms, a special style of text that eliminates the uninformed
6	Anonymity of a computer game participant	This leads to the fact that all hidden psychological complexes of the player are revealed, which strengthens the power of the irrational over the psyche
7	Anonymity of communication in the game	This enriches the possibilities of self-presentation, allowing the player to create an impression of himself in the virtual world of his choice and be whoever he wants to be

All of the above factors affect decision-making in virtual reality: time reversal is broken, the pace of events accelerates, the cost of a decision-making error is reduced, and all actions and reactions to them are regulated. At the same time, the decision-maker is anonymous and heroized at will, and the virtual scope of his or her activities is orders of magnitude larger than that available to him or her in the real world.

Computer games often serve as a psychological relief, a kind of psychological training. A person is attracted to the game by the desire to relieve irritation, aggression, and the possibility of transferring tension to a new object. In general, computer games are a way of social experience that is important for personal development. Today, many users of modern computer technologies believe that existence is possible only online. If a person is not known in cyberspace, it is as if he or she does not exist in the real world at all.

Play, to a greater extent than labor, was the basis for the formation of human culture [19], the fundamental human activity takes place in the field of play. All human culture has been a form of play. At the same time, in the virtual world, people are mostly interested in themselves: at all times, the most important task is to find themselves and their like-minded people, as well as to study the history of the world [20]. In earlier times, this information was obtained through fairy tales, legends, and memories, later—through fiction, and in the modern world—through computer games. Today, the game, like fiction before, in the period of human development was a guide—a means of determining the direction, a special prism through which the world looks different, and a criterion for the compliance of this world with ideals.

The special normative order, internal ethics, and morality of a computer game are focused on maintaining and strengthening the internal game order [18]. This order is intended to set the system of internal relations and regulate the organization of internal relations to at least prevent the growth of entropy, not to allow entropy indicators to go beyond the permissible limits. The latter means that the internal regulatory system should, on the one hand, provide a certain freedom of adaptive action, and at the same time, limit freedom to ensure internal coherence, orderliness, and integrity. At the same time, such low-entropy orderliness should be supported by the inner world of a person, his or her feeling of being needed, expedient, and demanded by the structure, and this feeling dominates the desire for freedom and equality.

Thus, the total appeal of the modern person to a computer game is a response to the increase in entropy characteristic of modern civilization. Human nature is based on the desire for certainty, and people feel confident when they have some control over their environment. The game is a self-organizing counterpoint to the complex world. A game is a way to obtain order, stability, and certainty, at least in the virtual world, for a while.

5. Methodology for the application of natural language information processing in the study of educational space

5.1. SLAM algorithms

The ability of computers to perform useful tasks related to human language, to perform high-quality text or speech processing, to help in communication between people who speak different languages, and, in general, the ability to communicate between people and machines—all these problems are tried to be solved by Natural Language Processing (NLP) [21–26]. This is a general area of computer science, artificial intelligence, and computational linguistics, the main problem field of which is to ensure interaction between computers and human (natural) languages. That is, it is the processing of language, words, and speech by a computer: how to program computers to process and analyze large amounts of data in a natural language.

A “natural language” refers to a language used for everyday communication between people: English, Ukrainian, Italian, etc. Unlike artificial languages, such as programming languages and mathematical expressions, natural languages live, change, develop, and are passed down from generation to generation title.

5.2. Areas of NLP research

Today, the main areas of application are as follows:

- Information Retrieval
- Information Extraction
- Machine Translation
- Question-Answering Systems
- Dialogue systems
- Speech Recognition
- Natural Language Generation
- Sentiment Analysis.

There are low-level and high-level NLP subtasks. High-level subtasks are built based on low-level ones.

5.3. Low-level NLP subtasks

Among the low-level subtasks, the following main subtasks are usually distinguished:

- Sentence boundary detection or Sentence boundary disambiguation, and abbreviations complicate this task.
- Tokenization is the detection of individual tokens (words, punctuation marks) in a sentence: Lexical analyzer, lexer, or Tokenizer—a program or part of a program that performs tokenization.
- Part-of-speech tagging is the automatic assignment of parts of speech or other forms to elements in a text.
- Lemmatization is a method of morphological analysis that reduces a word form to its original dictionary form (lemma): as a result of lemmatization, inflectional endings are dropped from the word form, and the basic or dictionary form of the word is returned.
- Stemming is the process of reducing a word to its base by dropping auxiliary parts, such as an ending or suffix.
- Shallow parsing or chunking—grouping a sequence of words into a phrase.

5.4. High-level subtasks

Today, scientists most often refer to the following areas of applied research as high-level tasks:

- Spelling/grammatical error identification.
- Named-entity recognition.
- Word sense disambiguation is an unsolved problem of natural language processing, which consists of the task of choosing the meaning (or sense) of a multivalent word or phrase depending on the context in which it is found.
- Relationship extraction between named objects. Given a piece of text, you need to determine the relationships between named objects.

5.5. Areas of research on educational space

The main components of the methodology for applying natural language information processing in educational space research can be developed and implemented in several ways. These components of the methodology and their detailed characteristics are summarized in Table 2. The style of texts within tables should be normal.

Table 2
The main components of the methodology for applying natural language information processing in educational space research

Component of the methodology	Detailed description of the methodology components
Component 1	Monitoring of existing measures of similarity between texts and development of new measures of similarity if necessary
Component 2	Generating text annotations using 7 approaches and determining the similarity measures between the generated annotations to identify the tools that best generate text annotations in the selected field of knowledge
Component 3	Generation of abstracts of theses defended in Ukraine since 1991
Component 4	Identification of research areas in the field of education based on the analysis of abstracts of dissertations
Component 5	Clustering of research areas in the educational space based on automatic detection of these areas by abstracts of dissertations
Component 6	Generating annotations of dissertations that are in the public domain
Component 7	Determination of the similarity measures of dissertation annotations made by the dissertator and automatically generated by different approaches
Component 8	Construction of membership functions for annotations created by the dissertation based on calculated similarity measures with automatically generated annotations
Component 9	Automation of research on the level of internationality of educational and scientific events
Component 10	Classification of publications submitted to the scientific event according to the declared tracks (sections) of the educational and scientific event
Component 11	Automated determination of the level of novelty of scientific results and the quality of educational scientific activity of scientists by the formula for determining the best teacher
Component 12	Dynamic automatic supplementation of the scientific performance of teachers of higher education institutions
Component 13	Automatic determination of the number of self-citations of higher education teachers in scientific texts that are in the public domain
Component 14	Building graphs related to mutual citations of different authors in scientific papers

Component 15	Investigation of cycles in references graphs in cases of indirect (cyclic) references
Component 16	Study of the level of cooperation of teachers of higher education institutions, i.e. the ratio of the total number of scientific papers written in co-authorship to the total number of co-authors and the number of published scientific papers

5.6. Prospects for further research

Modern research is increasingly focusing on:

- Unsupervised and semi-supervised learning algorithms (unsupervised and partially supervised learning algorithms). Such algorithms are capable of learning from data that has not been manually annotated or using a combination of annotated and unannotated data.
- Deep learning techniques are being developed that give good results in language modeling and parsing.
- Creating an intelligent system for analyzing the tone or sentimental analysis of Ukrainian texts.
- A spam filtering system for Ukrainian-language emails.
- A system of grammatical and morphological analysis of Ukrainian texts.
- A system of rhyming words to create poems or songs in Ukrainian.
- A system for determining verse size (rhythm) in Ukrainian poems using neural networks.
- A system for automatically generating questions to Ukrainian-language texts.
- An intelligent system for Ukrainian language stemming.
- Building a lemmatizer for the Ukrainian language.
- A tokenizer for the Ukrainian language, taking into account the ambiguity of punctuation.
- Assessment of psychological qualities of a person based on texts based on the big five.
- Analysis of the tone of Ukrainian-language texts based on the big five.

6. Speech signal and its normalization

6.1. Methods

Most signals (including speech) are analog, so they are converted to discrete signals by analog-to-digital transformation (ADT) for processing in digital computers. Using this procedure, a set of $[n]$ samples obtained at Δ instantaneous values of a continuous signal devoid of physical nature is obtained, and their maximum and minimum values are determined by the ADT bit depth. For example, if the ADT bit depth is 2 bytes, then it $\left[2^{16-1}, 2^{16-1}-1\right]$corresponds to the range of all values in the samples determines the storage. The sampling frequency is the inverse of the sampling phase Δt. According to Kotelnikov's theorem, only such an analog signal can be losslessly recovered from a discrete signal, the high frequency of whose spectrum is equal to half of the sampling rate [26]:

$$f_s > 2 \cdot f_n$$

Digital signal processing tools are used to describe and transform discrete signals. The most important CIA procedure is the Discrete Fourier Transform (DFT):

$$S[m] = \sum_{i=1}^{N} s[n] \cdot e^{\frac{j 2\pi n m}{N}}, m = 1, \ldots, N$$

where N is the number of N-DFT constructed samples; j is an imaginary unit.

6.2. Results and Discussions

DFT allows go to the frequency, i.e. divide it into a set of harmonics and find the amplitude (energy)) of a harmonic as a function of its frequency. Fig. 1 shows a portion of the speech signal with the vowel "a" in the time domain. At the same time, to abstract from the ADT bit depth, the samples of the digitized signal are usually described in relative values [27]: or in fractions of the maximum value (this is for 2 bytes or in decibels). The first method of presentation is used in this work. n=1024 reference fields were allocated to find the DFT; the result is shown in Fig. 2. In this case, the frequency of the harmonic is horizontally, vertically—$|S[\ ]|$, which is the amplitude of the harmonic.

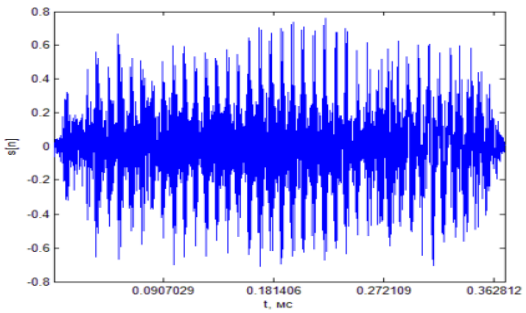

Figure 1: Section of speech with vowel sound "A"

Speech is a non-stationary signal, that is, its characteristics change over time. These changes can be visualized by plotting DFT modules for successive parts (frames) of the speech signal. Fig. 3 shows the waveform of the word "Forward". The resulting image is called a spectrogram (Fig. 4). Figs. 2, 3, and 4 show that frequencies up to 8 kHz consume the most energy [28]. Therefore, a typical choice of sampling rate when digitizing a speech signal is 16 kHz.

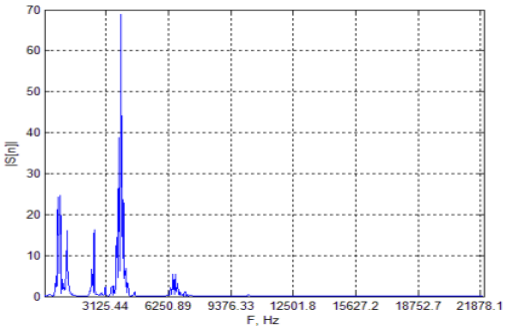

Figure 2: DFT for a portion of the speech signal with the vowel "A"

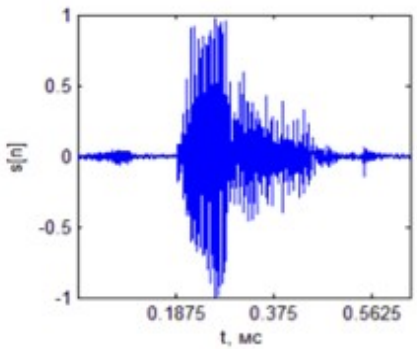

Figure 3: The waveform of the word "Forward"

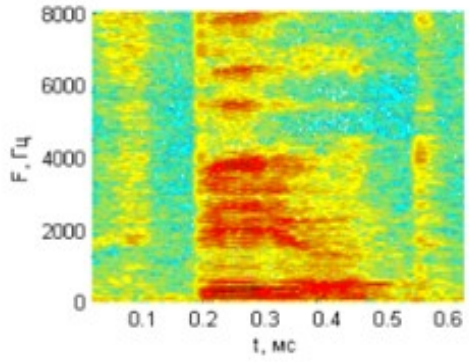

Figure 4: Spectrogram of the word "Forward" signal

Just as the words in written speech are formed from a limited set of characters—the alphabet of the language, the spoken speech also includes a limited set of sound "letters" in all their variability. The minimal semantically distinct unit of speech is the phoneme. The Ukrainian language has 38 phonemes, of which there are 6 vowels and 32 consonants. Unfortunately, there is no further uniform classification of phonemes, so Fig. 4 shows one of the combined options, which includes intersecting classes, for example, voiced (voiced) and fricative, deaf (voiceless) and explosive, etc. [29].

When recording a speech, some factors affect the amplitude of the audio signal: the speaker's voice pitch, his distance from the microphone, etc. These factors lead to a large variability in the pitch of the speech signal. This phenomenon is especially noticeable when using heterogeneous recording equipment. The amplitude normalization procedure is used to eliminate volume dispersion. With this technique, the signal amplitude is within the limits [Δ/2, −Δ/2] (Figs. 5 and 6). Sampling of the normalized signal [29] is carried out according to the following formula:

$$S[n]=\frac{\Delta}{max\ |S[m]|}\times s[n],$$

this earth a Δ is abscissa axis relatively symmetrical and has been normalization zone width (for example, in Figs. 5 and 6: Δ=1).

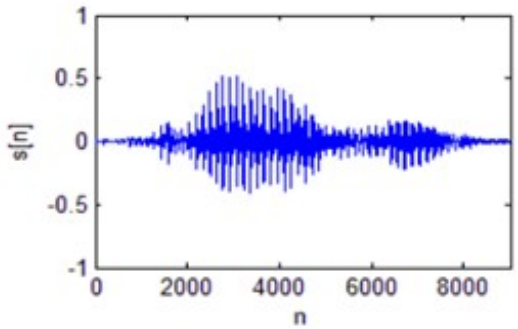

Figure 5: Example 1 of normalization after digitizing a speech signal [30]

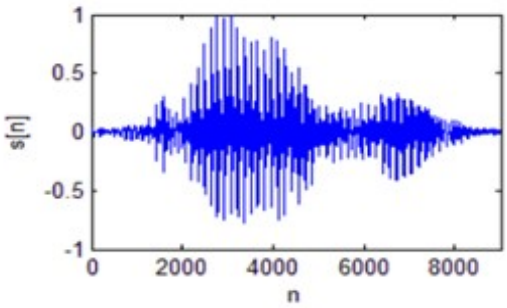

Figure 6: Example 2 of normalization after digitizing a speech signal [30]

Voice height variability evaluation for one or one how many the words pronunciation of doing Q examples seeing get out in length N of samples q is an example for average sound height value $M\ q$ and Q examples for average MQ the find:

$$M[q]=\sum_{n=1}^{N}|S[n]|, q=1,\ldots,Q;$$

$$M_Q=\frac{1}{Q}\sum_{q=1}^{Q}M(q).$$

From this after, each one of the example sound height average from value relative we calculate the deviation:

$$D(q)=\left|1-\frac{M(q)}{M_q}\right|.$$

Graphical illustrations for the initial values are shown in Figs. 7 and 9. The normalized signals are shown in Figs. 8 and. 10.

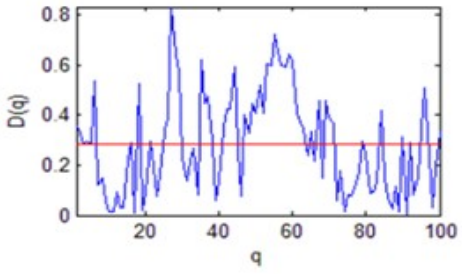

Figure 7: Initial example 1 energy

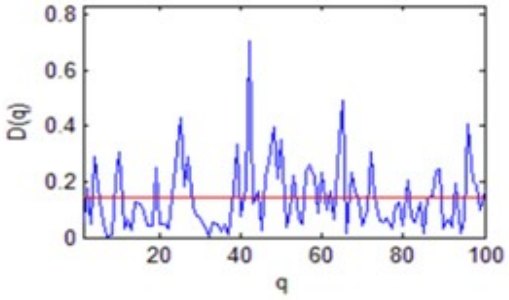

Figure 8: Normalized for the average energy from the exclusion of values to example 1

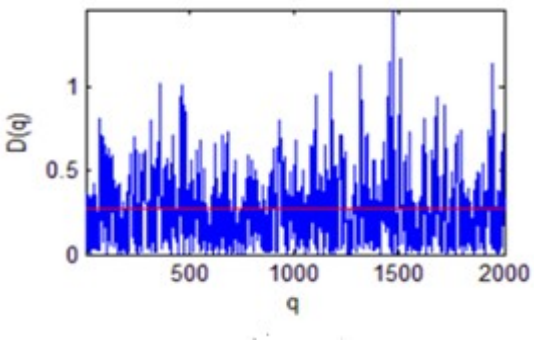

Figure 9: Initial example 2

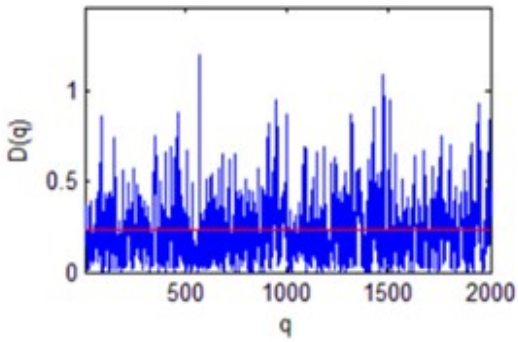

Figure 10: Normalized for the average energy from the exclusion of values to example 2

Above from formulas apparently as a result of the sample absolute value also, in the example this samples the number too effect does, therefore for collection of h size variability evaluation needed. Length approx one different has been one different to the class about examples and in general of the basis sound in height changes in the Figs. 7 and 8. For $Q = D\ (q)$ is suitable respectively from normalization before and then say "three." Note 100 made example shown. Voice height original examples 28.5% for and normalized ones 14.3 % for organize did in Figs. 9–10. $Q = 2000$ different pronunciations of examples speech base for $D\ (q)$ is displayed. Voice height original base 25.8% and normalized for 23.11 % organize did [31, 32].

Conclusions

From research apparently, as from normalization use each always different words for sound in height changes reduces. These are the results the only one is available in the equipment one different in the circumstances collected speech base for taken, therefore for normalization, in general, is the insignificant role played. However, with every difference in the circumstances speech signal acceptance when done, is real of the system in operation normalization necessary.

Voice-to-text technology simplifies everyday tasks and helps advance many professional fields. In business, Speech-to-Text is used to effectively interact with clients and quickly process large amounts of data. Analytics and voice robots reduce costs, increase the average bill, and study the real needs of customers. Speech analytics automates call control and saves time. You increase sales conversion, improve the quality of service, and receive feedback from the market in understandable language.

Many challenges still exist, but significant progress has been made in the field of natural language processing in recent years. Today, the maturity of natural language processing is encouraging more and more companies to use natural language processing in their products or their internal organization.

Declaration on Generative AI

While preparing this work, the authors used the AI programs Grammarly Pro to correct text grammar and Strike Plagiarism to search for possible plagiarism. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

[1] A. S. Pillai, R. Tedesco, Introduction to machine learning, deep learning, and natural language processing, 1st Edition, CRC Press, 2023.

[2] N. Tmienova, B. Sus, System of Intellectual Ukrainian language processing, in: Selected Papers of the 18th International Scientific and Practical Conference "Information Technologies and Security" (ITS 2019), vol. 2577, 2019, 199–209.

[3] S. Bird, E. Klein, E. Loper, Natural language processing with Python, Published by O'Reilly Media, 2009.

[4] O. Ilarionov, et al., Intelligent module for recognizing emotions by voice, Adv. Inf. Technol. 1 (2021) 46–52. doi:10.17721/AIT.2021.1.06

[5] B. W. Schuller, Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends, Commun. ACM 61(5) (2018) 90–99. doi:10.1145/3129340

[6] X. Huahu, G. Jue, Y. Jian, Application of speech emotion recognition in intelligent household robot, in: International Conference on Artificial Intelligence and Computational Intelligence, vol. 1, 2010, 537–541.

[7] K. Sailunaz, et al., Emotion detection from text and speech: a survey, Soc. Netw. Anal. Min. 8(1) (2018) 1–26.

[8] H. A. Bourlard, N. Morgan, Connectionist speech recognition: A hybrid approach, Kluwer Academic Publishers, Norwell, MA, USA, 1993.

[9] G. Cheng, et al., An exploration of dropout with Lstms, in: Interspeech, 2017, 1586–1590. doi:10.21437/Interspeech.2017-129

[10] T. Babenko, H. Hnatiienko, V. Vialkova, Modeling of the integrated quality assessment system of the information security management system, in: 7th International Conference "Information Technology and Interactions", 2020, 75–84.

[11] H. Hnatiienko, et al., Application of cluster analysis for condition assessment of Banks in Ukraine, in: 8th International Scientific Conference "Information Technology and Implementation", vol. 3179, 2022, 112–121.

[12] J. Huizinga, Homo Ludens. Experience in defining the game element of culture, 1994.

[13] M. Petrushkevych, Carnival features of communication in new media: Challenges of mass culture, sociocultural challenges of modernity: The need for theoretical understanding, Ostroh, 2022, 103–138.

[14] D. Palko, et al., Cyber security risk modeling in distributed information systems. Appl. Sci. 13 (2023) 2393. doi:10.3390/app13042393

[15] S. P. Robbins, T. A. Judge, Organizational behavior, 18th Edn. New York, NY: Pearson, 2019.

[16] H. Hnatiienko, et al., Application of expert decision-making technologies for fair evaluation in testing problems, in: 20th International Scientific and Practical Conference "Information Technologies and Security" (ITS 2020), vol. 2859, 2021, 46–60.

[17] H. Hnatiienko, et al., Methods of identifying the correlation of Ukrainian scientific paradigms based on the study of defended dissertations, in: 10th International Scientific Conference "Information Technology and Implementation" (IT&I 2023), vol. 3646, 2023, 64–75.

[18] M. Kushnir, Economy and society, Max Weber, trans. from German, Vsevit, 2013.

[19] L. D. Bevzenko, Social self-organization. Synergetic paradigm: possibilities of social interpretations, 2002.

[20] L. Bevzenko, Agents of social change in a crisis society: Options for problematization and outlines of the conceptual framework of the study, Sociol. Theor. Meth. Mark. 4 (2020) 111–132.
[21] O. Romanovskyi, et al., Accuracy improvement of spoken language identification system for close-related languages, Advances in Computer Science for Engineering and Education VII, vol. 242 (2025) 35–52. doi:10.1007/978-3-031-84228-3_4
[22] I. Iosifov, et al., Transferability Evaluation of speech emotion recognition between different languages, Advances in Computer Science for Engineering and Education 134 (2022) 413–426. doi:10.1007/978-3-031-04812-8_35
[23] I. Iosifov, O. Iosifova, V. Sokolov, Sentence segmentation from unformatted text using language modeling and sequence labeling approaches, in: IEEE 7th International Scientific and Practical Conference Problems of Infocommunications. Science and Technology (2020) 335–337. doi:10.1109/PICST51311.2020.9468084
[24] O. Iosifova, et al., Analysis of automatic speech recognition methods, in: Cybersecurity Providing in Information and Telecommunication Systems, vol. 2923 (2021) 252–257.
[25] O. Romanovskyi, et al., Automated pipeline for training dataset creation from unlabeled audios for automatic speech recognition, Advances in Computer Science for Engineering and Education IV, vol. 83 (2021) 25–36. doi:10.1007/978-3-030-80472-5_3
[26] M. Gales, S. Young, The application of hidden Markov models in speech recognition, foundations and trends in signal processing, 1(3) (2008) 195–304. doi:10.1561/2000000004
[27] A. F. Voloshin, G. N. Gnatienko, E. V. Drobot, A method of indirect determination of intervals of weight coefficients of parameters for metricized relations between objects, J. Autom. Inf. Sci. 35(1–4) (2003). doi:10.1615/JAutomatInfScien.v35.i3.30
[28] H. Hnatiienko, et al., Method for determining the level of criticality elements when ensuring the functional stability of the system based on role analysis of elements, in: Cybersecurity Providing in Information and Telecommunication Systems, vol. 3654, 2024, 301–311.
[29] R. Zulunov, et al., Detecting mobile objects with AI using edge detection and background subtraction techniques, in: E3S Web of Conferences, vol. 508, 2024, 03004.
[30] R. Zulunov, et al., Building and predicting a neural network in PYTHON, in: E3S Web of Conferences, vol. 508, 2024, 04005.
[31] V. V. Byts, R. M. Zulunov. Specification of matrix algebra problems by reduction, J. Math. Sci. 71 (1994) 2719–2726.
[32] U. Akhundjanov, et al., Handwritten signature preprocessing for off-line recognition systems, in: E3S Web of conferences, vol. 587, 2024, 03019.