IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 28, No. 2, April, 2023, pp. 443-450
Journal homepage: http://iieta.org/journals/isi

Spontaneous Speech and Its Features Are Taken into Account When Creating Recognition Programs

Askhat Yergaliyev[1*], Altynbek Sharipbay[2], Lyailya Baibulekova[3]

[1] Department of Computer Engineering and Information Security, International Information Technology University, Almaty 050040, Republic of Kazakhstan
[2] Department of Artificial Intelligence Technology, L. N. Gumilyov Eurasian National University, Astana 010008, Republic of Kazakhstan
[3] Department of Finance and Accounting, Kenzhegali Sagadiev University of International Business, Almaty 050010, Republic of Kazakhstan

Corresponding Author Email: yergaliyevaskhat@gmail.com

https://doi.org/10.18280/isi.280220

Received: 10 February 2023
Accepted: 23 March 2023

Keywords:
alternative transcription, multilingual corpus, linguistic support of information systems, automatic recognition systems, speech identification, corpus linguistics

ABSTRACT

The relevance of the stated subject of this scientific research is determined by the numerous difficulties of spontaneous speech recognition due to the presence of a complex of unrelated factors, as well as the need to find the best ways to overcome them through recognition programs that are used in the development of a multilingual corpus. This research aims to identify and study the key features of spontaneous speech that should be considered when creating recognition programs. The basis of the methodological approach in this scientific study is a combination of methods combining a systematic analysis of the principles of building a multilingual corpus with an analytical study of the principles of operation of spontaneous speech recognition programs. In the course of this scientific study, results were obtained that indicate a significant increase in the practical efficiency of automatic speech recognition systems when introducing sound signal correction algorithms, which opens up additional opportunities for developing such software. In addition, the results of this scientific study clearly demonstrate the significant impact of the accuracy of the effects of speech recognition programs on the quality of development of multilingual corpora, which include relatively large volumes of texts. The practical significance of the results obtained in the course of this scientific work, as well as the conclusions formulated on their basis, lies in the possibility of their use in the development of a multilingual corpus and spontaneous speech recognition programs for their subsequent use in various information systems, to obtain results related to the need accurate decoding of automatic speech, taking into account all its characteristic features.

1. INTRODUCTION

The problem of this research work is the presence in the spontaneous speech of characteristic features associated with the company in it of emotions of varying severity, all kinds of noise pauses, and other artifacts that complicate the recognition process, require timely identification and must be taken into account when developing modern technologies for recognizing spontaneous speech. Speech, for their subsequent use in creating a multilingual corpus. Currently, the problem has been studied relatively poorly due to the relatively small number of available scientific publications that objectively reveal the stated topic. For this reason, new, modern studies of the features of spontaneous speech are of particular relevance to take them into account when designing speech recognition programs. In a joint scientific study, Leontyeva and Kipyatkova [1] considered the main features of spontaneous speech that are important in developing programs and systems for automatic speech recognition. According to the authors, when developing intuitive speech recognition models, it is advisable to use noise models and non-phonemic elements. Scientists note that in the future, the development of systems for automatic recognition of spontaneous speech, capable of taking into account its specific features, will eliminate some of the restrictions imposed on the dialogue with the user, which will ultimately make human-machine interaction more natural and productive.

This topic is being developed by Beskrovnykh and Markasova [2] in a scientific study of short tongue protrusion in spontaneous speech. The researchers concluded that short bows of the language in automatic speech make it difficult to recognize, being the equivalent of what was not expressed due to several internal and external reasons. The authors note that the ability of short protrusions of the language to connect opposing positions at the level of replicas of different participants, as well as to combine sound speech and silent reaction into a single text, allows us to consider it as a means of micromimic design of vertical connections in discourse.

At the same time, Sidorova [3], in a scientific study of the prospects for building an integrated approach to the study of the lexical characteristics of a text, draws attention to the fact that any text, as a source and method of transmitting information, must be comprehensively studied, both to form a correct assessment of the level of presentation of the material and to ensure high efficiency of its automatic processing and support of search language services. In his other scientific study, devoted to the study of several problematic aspects of the development of linguistic support for information systems based on ontological knowledge models, the author notes that the successful application in practice of technologies for developing the linguistic support of modern information systems makes it possible to create effective speech recognition programs that can take into account all its features and support at the proper level, the chain of creation and practical application of linguistic resources for the analysis of large text fragments in automatic mode. In addition, according to the scientist, such systems should include tools for creating a linguistic ontology, with the help of which direct knowledge holders can set up the content processing of documents – experts and linguists who do not have exceptional programming skills [4, 5].

The team of researchers represented by Prokofyeva et al. [6], in a joint scientific study, considered the problematic aspects of emotion recognition by speech features. The scientists concluded that the development of automated spontaneous speech recognition systems requires the combined efforts of specialists from various fields of knowledge, in particular, psychologists, psycholinguists, and computer technology specialists, to create effective strategies for their subsequent use in multiple areas of everyday life.

This research aims to study spontaneous speech's main features and their consideration when creating recognition programs. This is especially important given the widespread use of speech recognition programs in modern computer technology and the need to develop practical spontaneous speech recognition algorithms for their subsequent use in this kind of software.

2. MATERIALS AND METHODS

The main advantages of the proposed methods compared to other related works are the combination of methods of systematic analysis of the fundamental principles of building a multilingual corpus with an analytical study of the principles of spontaneous speech recognition programs. The application of system analysis of the basic principles for developing a multilingual corpus made it possible to obtain a definition of such and classify multilingual corpora. An acoustic and linguistic analysis were used to identify patterns and features of the speech data. This involved analyzing features such as pitch, duration, and spectral content, as well as syntactic and semantic features of the language. Machine learning and statistical analysis techniques were applied to the speech data to develop and evaluate speech recognition algorithms. In addition, the fundamental principles for the classification of multilingual corpora were presented by their main features. The primary scientific research was preceded by the creation of a theoretical base, which included an analysis of the investigation of authors who studied the problematic aspects of the development of multilingual corpora and spontaneous speech recognition programs for their subsequent use in various technical fields and applied areas of language knowledge.

An analytical study of the principles of operation of spontaneous speech recognition programs made it possible to classify the main features that impede the functioning of automatic recognition programs. At the same time, the use of this method made it possible to determine spontaneous speech as one of the leading indicators of the general culture and intellect of the speaker. At the same time, an experiment was carried out, which consisted in comparing the measurement of the accuracy of the functioning of the spontaneous speech recognition program when the algorithm for correcting the sound stream was introduced into it. The experiment fixed the speech recognition accuracy indicators without using the audio stream correction algorithm and its use. The rational stream of the spontaneous speech consisted of over 28000 speech utterances, which occupied more than 15 hours of the excellent track. The results of the experiment are presented in the corresponding table.

The chosen combination of scientific research methods determined the presence of the following stages of this scientific work. At the first stage of this scientific research, the concept of a multilingual text corpus was defined, and the basic principles for the classification of multilingual corpora were. At the same time, the types of multilingual corpora are defined, the definition of the essence of the concept of spontaneous speech and its characteristic features, which are essential from the point of view of the speech recognition programs' efficiency given.

At the next stage of this scientific work, the factors that impede the work of spontaneous speech recognition programs related to the main features of it were identified. The sequence of development of unique algorithms for transcribing word forms is determined, considering sound reduction and assimilation. A schematic model of such an algorithm is presented, and a model of its functioning is described in detail. In addition, at this stage of the scientific research, a scientific experiment was conducted, the essence of which was to compare the measurement of the accuracy of the spontaneous speech recognition program without using the audio signal correction algorithm and with its use. The issue of the need to implement such an algorithm in automatic speech recognition programs is considered in detail. The results of this experiment are presented in the corresponding table.

At the final stage of this scientific work, an analytical comparison of the results obtained during it with the results and conclusions of other researchers who studied topics related to the declared one was carried out. This made it possible to refine the results obtained and form final decisions based on them, which act as their logical reflection and sum up the entire complex of scientific research on the features of spontaneous speech taken into account when developing programs for its recognition.

3. RESULTS

To date, corpus linguistics is actively developing, which involves the development, formation, and search for opportunities for the practical application of multilingual text corpora. A multilingual text corpus is a carefully sorted and specially processed collection of texts in various languages by established rules. In the future, these texts will be used as a system base in linguistic research [7]. Such studies involve conducting a statistical analysis of the occurring speech turns and set expressions, testing the accepted statistical hypotheses in the field of corpus linguistics, and searching for evidence of the studied linguistic rules. The multilingual corpus is one of the study objects in the corpus linguistics section. The classification of multilingual corpora can be carried out by the basic principles, among which it should be highlighted: tasks that are supposed to be solved by creating a corpus; languages in which the texts are presented; genre focus; dynamism. It is customary to divide multilingual text corpora into two main types: parallel, including many texts in different languages, as well as their translations; pseudo-parallel, including exclusively original texts in many languages. To account for variations in spontaneous speech across different languages and cultures requires a cross-linguistic and cross-cultural approach that involves collecting and analyzing spoken language data from a variety of linguistic and cultural groups. This can help to develop speech recognition models that are more accurate and robust across different languages and cultures.

Constructing a multilingual corpus involves collecting a large amount of text or speech data in multiple languages and then organizing and annotating the data for use in natural language processing applications such as machine translation, speech recognition, and sentiment analysis. The general steps for constructing a multilingual corpus are: define the scope and goals of the corpus; collect the data; preprocess the data; annotate the data; align the data; evaluate the corpus; publish the corpus. Overall, constructing a multilingual corpus requires careful planning, data collection, and annotation, as well as rigorous evaluation and publication of the corpus to ensure its usefulness for natural language processing applications.

Spontaneous speech as an object of study in linguistics is considered in the context of the prospects for its subsequent processing in special recognition programs. Spontaneous, extemporaneous speech is delivered under conditions of constantly changing external circumstances and communicative conditions. Due to the above cases, automatic speech should be attributed to one of the critical indicators of the language competence of the person pronouncing it, as well as a manifestation of his general cultural level. Creating a program for automatic recognition of spontaneous speech to create a multilingual corpus involves using approaches that go beyond the trivial adaptation of known models of foreign languages. Automated text processing, both in any new, previously unknown language and in a familiar language, is associated with numerous difficulties that are caused by the characteristics of a particular language, as well as the characteristics of spontaneous speech in this language. There are several unrelated factors that can make spontaneous speech recognition difficult, including:

1. Spontaneous speech is often accompanied by environmental noise, such as traffic, machinery, or other people talking. This can interfere with speech recognition algorithms and make it harder to accurately transcribe speech.
2. People from different regions or with different cultural backgrounds often have unique accents and dialects, which can be challenging for speech recognition software to understand. These variations in pronunciation and intonation can make it difficult to accurately transcribe spoken language.
3. People may speak at different rates, use fillers such as "um" or "ah," or pause frequently during speech. These variations in speaking style can make it challenging for speech recognition algorithms to accurately transcribe speech.
4. Disfluencies such as stuttering, repetitions, or false starts can make it difficult for speech recognition software to accurately transcribe speech. These interruptions can cause the algorithm to misinterpret what is being said.
5. Speakers may have different voices, pitches, and speaking styles, which can make it difficult for speech recognition software to accurately transcribe speech. The algorithm must be trained to recognize and differentiate between different speakers.
6. Spontaneous speech may contain informal language, slang, and colloquialisms that may not be recognized by speech recognition algorithms. Additionally, grammar and syntax variations can further challenge recognition accuracy.

All of the above factors significantly complicate the work of the spontaneous speech recognition program and reduce its overall performance [1, 8]. This predetermines the need to develop unique algorithms for transcribing word forms contained in spontaneous speech for the purpose of its subsequent synthesis, as well as the assimilation of pronounced sounds. Recognition programs use several techniques to overcome difficulties with spontaneous speech, including:

1. Speech recognition programs use noise reduction algorithms to filter out background noise and isolate the speaker's voice.
2. Acoustic modeling techniques use statistical models to capture variations in pronunciation, speaking style, and dialects. This allows the algorithm to recognize speech patterns more accurately.
3. Language modeling techniques use statistical models to analyze the structure and context of spoken language. This allows the algorithm to recognize informal language, slang, and colloquialisms more accurately.
4. Speaker adaptation techniques allow speech recognition programs to adapt to a specific speaker's voice and speaking style. This can improve recognition accuracy and reduce errors.
5. Disfluency handling techniques enable speech recognition programs to recognize and handle interruptions, repetitions, and other disfluencies that are common in spontaneous speech.
6. Vocabulary and grammar adaptation techniques allow speech recognition programs to adapt to the specific vocabulary and grammar used by a particular speaker or in a particular context. Overall, recognition programs use a combination of these techniques to improve recognition accuracy and overcome the difficulties associated with spontaneous speech.

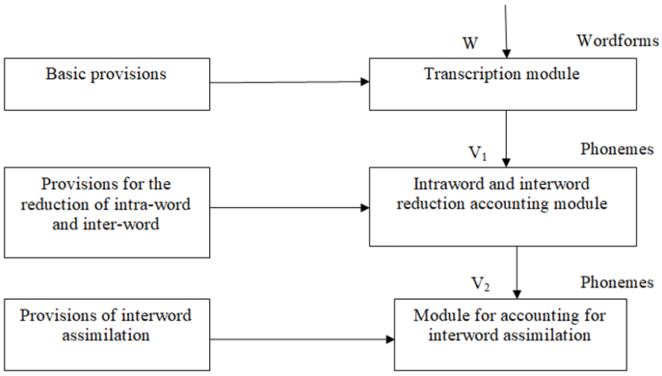
W
Wordforms
Basic provisions
Transcription module
V1
Phonemes
Provisions for the reduction of intra-word and inter-word
Intraword and interword reduction accounting module
V2
Phonemes
Provisions of interword assimilation
Module for accounting for interword assimilation

Figure 1. Algorithm for transcribing word forms, taking into account the provisions of sound reduction and assimilation
Source: Created by the authors based on Ref. [1]

Transcription algorithms for word forms contained in spontaneous speech are complex software tools designed to distinguish between the correct sounding of word forms and then determine their semantic meaning in the given context [9, 10]. Such algorithms include a given set of modules, the tasks of which are to accept a particular set of word forms, with their subsequent transformation into separate phonemes, and the implementation of the reduction principle. The end result, which is pursued at all stages of the process of functioning of a given algorithm, is to obtain a set of transcriptions that correctly describe the semantic essence of the spontaneous speech stream processed using the recognition program. Figure 1 shows a schematic representation of the word form transcription algorithm, considering sound reduction and assimilation provisions.

The word forms w entering the transcription module are converted into phonemes V1 and V2 by applying the introductory phonetic data processing provisions. Afterward, phonemes suitable for reduction are selected in the intraword and interword reduction modules. Then the selected phonemes are processed, selecting all possible combinations of their forms, which are automatically generated. After that, all generated combinations of phoneme forms are automatically processed separately. The result of the described process is obtaining a particular set of typical transcriptions (V1, V2, ...) of the original word forms (W1, W2, ...) when creating which all the features and options for reducing automatically generated phonemes were taken into account [1]. Spontaneous speech can rightly be classified as a unique linguistic phenomenon for which certain deviations from the norm are typical [11]. Among the characteristic features of spontaneous speech, which can be attributed to the category of deviations from the norm, expressed in a calm, smooth flow of speech flow, one should includeи:

– independent corrections by the speaker of the sounds, words, and phrases he utters;
– false starts, impulsive, thoughtless beginning of a sentence or a single phrase;
– failures in the pronunciation of some grammatical forms, as well as the established order of words and phrases;
– frequent, spontaneous breaking off of words and phrases without their expressed logical conclusion;
– constant reservations and problems with the pronunciation of individual sounds;
– speech repetitions;
– stammering without bringing the expressed thought or consideration to its logical conclusion.

When developing spontaneous speech recognition programs, one should consider the need to use elements that can detect deviations from the normal flow of smooth speech, as well as external interference and extraneous noise that interfere with the high-quality implementation of the automatic recognition [12-14]. To implement this task, it is advisable to introduce unique sound correction algorithms into the program that can provide a qualitative account of the characteristic features of the flow of spontaneous speech and refine the data at the program’s output. Correction of the sound stream of spontaneous speech through the use of this algorithm provides recognition of almost any fragment of spontaneous speech, achieves higher sound quality, and, consequently, improves the overall quality of spontaneous speech recognition [15]. In order to obtain experimental confirmation of the effectiveness of the audio stream correction algorithm, a test was conducted using the accumulated language base, which includes over 28,000 speech fragments (approximately 15 hours of the audio track). The results of the experiment are presented in Table 1.

Table 1. Comparison of the change in the accuracy of the spontaneous speech recognition program

Recognition objects	Number of words	Number of records	Recognition accuracy without using an algorithm, %	Recognition accuracy using the algorithm, %	Decrease in error, %
Titles of works of art	156	3244	96.22	99.12	55.14
Topics of scientific works	122	2756	98.17	98.87	28.33
Names of geographical routes	89	8678	97.07	98.30	19.46
Description of geographical objects	144	7542	95.87	98.18	65.43
Name of city toponyms	78	6532	96.55	91.18	21.44

As can be seen from the data presented in Table 1, when using each object of recognition, there is a decrease in the error when introducing the specified audio signal correction algorithm by an average of 19-65%. At the same time, the accuracy of recognition of the sound signal of spontaneous speech increases when using the algorithm for correcting the sound signal. The experiment results indicate the effectiveness of the practical implementation of algorithms of this type in speech recognition programs, which necessitates further research into the prospects for their development. The development of modern spontaneous speech recognition programs is associated with the need to consider its characteristic features, which are of fundamental importance in ensuring the final product's high quality. In this case, absolutely all features of the sound signal of spontaneous speech received at the input of the transcription module, as well as the nature of the transformation of word forms included in this sound signal into phonemes, should be taken into account [16]. With any changes like the incoming sound signal, the operator of the speech recognition program must make all the necessary adjustments to the program, taking into account the features of changes of this kind and the specified parameters for the clarity of recognition at the output. The creation of such a software system for the recognition of spontaneous speech implies the need to test this software, taking into account the above features and the ability of the developed software to account for and recognize them effectively. Figure 2 shows a diagram of a test software model for evaluating the effectiveness of the designed spontaneous speech recognition software.

The interaction of the software package with the operator is provided by the capabilities of the graphical user interface, which includes the means of interaction with the main module. The module for interaction with speech recognition systems is also associated with the main module. Its primary function is to ensure the transfer of recorded spontaneous speech directly to the system and to receive spontaneous speech recognition data. In addition, the module for interaction with speech recognition systems performs the functions of obtaining and saving information about the speed of the spontaneous speech recognition process and saving all the received statistical data. Sound recording and playback modules directly interact with the software audio system. Their essential functions are the recording of recognizable spontaneous speech, as well as its subsequent playback directly in the software system. The main module is the central part of the test model for evaluating the effectiveness of spontaneous speech recognition programs. Its primary function is to ensure smooth interaction between all nodes of the software complex and keep it in working condition at all stages of the above technological operations [17].

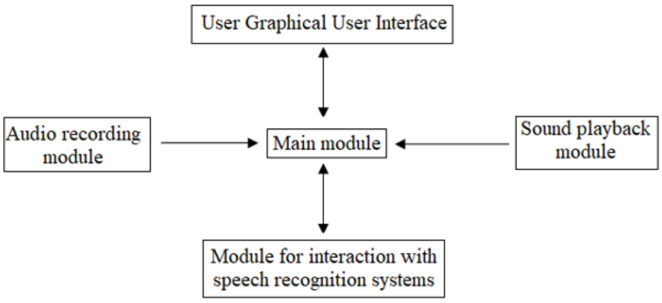
User Graphical User Interface
Audio recording module
Main module
Sound playback module
Module for interaction with speech recognition systems

Figure 2. Model for evaluating the effectiveness of spontaneous speech recognition software
Source: Created by the authors based on Ref. [17]

Testing of spontaneous speech recognition programs should determine the recognition accuracy of both individual phrases, taking into account the sequential conversion of word forms into phonemes entering the transcription module and the entire text as a whole, taking into account the actual capabilities of the software interface itself. In this case, one should also consider the probability of errors when sending commands from the head module to applications and the possibility of inconsistent operation of individual parts of the test program as a whole. For a qualitative solution to such tasks, the operator must constantly monitor the progress of the operations performed and the state of the entire software system as a whole and its individual nodes in particular [18]. The main difficulty in using automated spontaneous speech recognition systems lies in the fact that such speech is characterized by the complete absence of a pre-prepared form. Therefore, it also does not contain any verbal messages or direct participation of persons directly conducting the dialogue. At the same time, such parameters as the pace of the speech flow, the manner in which the speaker pronounces individual phrases and phrases, as well as excessive emotionality against the background of an unpredictable large volume of phrases used, vary unpredictably [19, 20]. All this significantly complicates the development of spontaneous speech recognition programs and predetermines the need to take into account all these features both directly in the software development process and at all stages of its subsequent operation in order to obtain optimal results, expressed in the high final quality of spontaneous speech recognition by modern software, methods.

Improved spontaneous speech recognition technology has a wide range of practical applications across various industries. Spoken language recognition can be used to automate customer service interactions, such as call centers or virtual assistants, which can reduce wait times and improve customer satisfaction. Speech recognition technology can assist healthcare providers with dictation and medical transcription, which can improve the accuracy and efficiency of medical documentation and patient care. Speech recognition technology can be used to create automated language assessment tools that evaluate students' language proficiency, helping educators provide more personalized language learning experiences. Lawyers and legal professionals can use speech recognition technology to transcribe and document legal proceedings, such as depositions and court hearings, which can increase efficiency and accuracy. Police officers and first responders can use speech recognition technology to transcribe interviews and record incident reports, improving the accuracy and timeliness of incident documentation. Spoken language recognition can be used to control Internet of Things devices, such as smart speakers, thermostats, and lighting systems, making it easier for users to interact with their devices hands-free.

However, there are several potential limitations and challenges associated with implementing improved spontaneous speech recognition technology in real-world settings, including:

1. Collecting and processing large amounts of speech data for speech recognition applications raises concerns about data privacy and security, especially when dealing with sensitive personal or medical information.
2. While improved speech recognition technology can better handle variations in pronunciation and speaking styles, it may still struggle to recognize accents or dialects that are outside of the trained models.
3. Speech recognition technology may struggle to accurately transcribe spoken language that contains slang, colloquialisms, or domain-specific terminology.
4. Spontaneous speech recognition can be hindered by environmental noise, such as traffic or other people talking, which can interfere with speech recognition algorithms and reduce accuracy.
5. Implementing improved speech recognition technology can be costly, especially for small businesses or organizations with limited budgets.
6. Introducing new speech recognition technology can require changes to existing systems and processes, which can be disruptive and time-consuming.
7. Some users may be uncomfortable with using speech recognition technology or may prefer other methods of communication.

4. DISCUSSION

In a joint scientific work, Vashkis and Yevseyenko [21] considered the main aspects of the linguistic support of automated systems. The researchers concluded that there is a steady development of linguistic support as an integral part of modern information systems. According to scientists, it is essential to realize that the effective functioning of information and communication systems can be achieved only in close interaction with the full range of components, the task of which is to ensure the reliable and high-quality operation of the automated control systems used. The researchers' conclusions are confirmed by the results of this scientific work in the context of assessing the role of linguistic support and the work of modern information systems. At the same time, the conclusion regarding ensuring the effectiveness of their functioning seems controversial due to the existing differences in approaches to ensuring the quality work of separately communicative and information systems.

For its part, the raised topic is developed by Granel [22] in the scientific study of several problematic aspects of multilingual information flow management. According to the scientist, multilingual information is in great demand today in today's globalized economy. This is facilitated by global processes in industry and in the world's leading markets, as well as active cooperation between the world's leading powers. The author also notes that multilingual information management builds on previous empirical research, as information and technology are used in the translation community as information assistants between different languages and cultures to increase their productivity and competitiveness in today's market. The conclusion of the researcher regarding the role of the community of translators in providing information assistance seems to be disputable since, in this case, the main emphasis should be placed on the quality of the translation of the text.

At the same time, Schultz and Kirchhoff [23] studied the issues of multilingual speech processing. They concluded that a qualitative result in this process could be achieved only if several factors were taken into account competently. Researchers include the correct choice of voice input algorithms, the ability to exchange system components and information in different languages, and so on. According to the researchers, the declared topic has many unexplored aspects, such as speech synthesis and the problems of its recognition, speech transformation through the use of the capabilities of modern dialogue systems, and automatic identification of languages. The researchers' conclusions complement and expand the results of this scientific work, emphasizing the need for further scientific research in the field of studying the problems of developing spontaneous speech recognition programs and their practical use in developing multilingual text corpora.

In a joint scientific work, a team of research scientists represented by Li et al. [24] considered several problematic aspects of building systems for the automatic recognition of spontaneous speech. A group of authors concluded that the currently used methods of automatic recognition of spontaneous speech often could not cope with external noise, interference, and other artifacts that complicate the course of the recognition process. According to the authors, this fact necessitates the search and implementation of more effective spontaneous speech recognition systems and the improvement of existing ones in order to achieve qualitatively better results in this process. The authors' conclusion repeats and develops the conclusions of the researchers discussed above while additionally pointing to the lack of knowledge of the subject under consideration in general and its aspects in particular, which does not fundamentally contradict the results obtained in this research work.

Ainsworth [25], in his scientific study of modern mechanisms for spontaneous speech recognition, notes that when considering this problem, one should pay attention to related topics, such as the study of auditory systems, speech production options, auditory psychophysics, as well as recognition of vowels and consonants and various features of speech distortions. According to the scientist, only the implementation of an integrated approach to the study of the features of the functioning of speech recognition mechanisms, taking into account the above factors, will make it possible to obtain a result expressed in determining clear prospects for creating high-quality spontaneous speech recognition programs. The conclusions of the scientist do not conflict with the results of this scientific study, at the same time pointing to additional areas of research in the area under consideration, which should be studied in subsequent scientific works to expand and deepen the accumulated methodological base.

For their part, a group of authors was represented by Waibel and Lee [26], who studied the problems of reading in spontaneous speech recognition in joint scientific work, notes that in the past few decades, spontaneous speech recognition has proven its practical effectiveness as an advanced technology that provides high-quality information processing in modern information systems. Scientists emphasize that the current scientific approaches to studying spontaneous speech recognition technologies converge in need to develop multilingual text corpora for their subsequent practical use in complex systems involving the processing of large amounts of data. The researchers' conclusions generally coincide with the results obtained in this scientific work, supplementing them in the context of assessing the need to develop and implement multilingual text corpora in modern information systems.

Dey [27, 28], in a scientific works devoted to the study of the features of intelligent processing of a speech signal, draws attention to the fact that the use of everyday speech analytics capabilities in several systems to perform specific, coordinated actions allows the use of programs for recognizing spontaneous speech in many practice areas. According to the author, sharing data analytics allows you to create networks for collaboration between several participants and organize influential video conferences in a wide variety of areas using the described applications. The conclusions of the scientist highlight the prospects for the practical application of speech recognition programs in modern technological systems, which, in the context of the results of this scientific work, is acceptable only if the rules for operating these systems are observed, otherwise the expected effect will not be achieved.

At the same time, a team of researchers represented by Zhang et al. [29] in a joint scientific study of deep, multimodal affective features for the recognition of spontaneous speech emotions points to the fact that excessive emotions significantly complicate the functioning of spontaneous speech recognition programs. According to a group of researchers, such a feature of spontaneous speech as excessive emotions requires a detailed study with the involvement of specialists in the field of psychology and communicative communication in order to create universal mechanisms for interpreting emotional expressions and their subsequent implementation in the practice of using spontaneous speech recognition programs to obtain more accurate results, process. The conclusions of a group of scientists fundamentally coincide with the results obtained in this scientific work, repeating them in terms of assessing the negative role of emotions in the work of programs for the automatic recognition of spontaneous speech.

Another group of research scientists represented [30], investigated the issues of distinguishing between the acoustic characteristics of spontaneous speech as well as speech read, in addition, during the study, an assessment was made their impact on the performance of speech recognition in general. The scientists concluded that determining the frequency of vowels and consonants, which form the basis of spontaneous speech, is the most critical aspect in terms of correctly interpreting the emotional component of recognizable speech. This will improve the quality of spontaneous speech recognition using available software tools. The researchers note that creating computer algorithms for recognizing spontaneous speech to determine particular vowels and consonants opens up additional opportunities for developing multilingual corpora. The researchers' conclusions fundamentally coincide with the results obtained in this scientific work. Thus, the discussion of the results obtained in this scientific study shows the presence of a relatively poor scientific base in the field of research on the problems of creating spontaneous speech recognition programs and also confirms the high accuracy of the results of this scientific study, since the conclusions and results of other scientific studies are mainly correlated with them.

5. CONCLUSIONS

Spontaneous speech features often make recognizing it challenging with the help of programs specially designed for this, which necessitates such a problem. These include transcription algorithms for developing unique algorithms and software tools that can effectively cope with the word forms contained in spontaneous speech and algorithms for correcting the sound flow of spontaneous speech in general. The use of word form transcription algorithms makes it possible to obtain at the output a particular set of typical transcriptions of word forms coming at the input to the program. In this case, the task is to clarify the semantic essence of words and phrases pronounced in different parts of spontaneous speech and to obtain a clear semantic picture of the final meanings. At the same time, the practical application of algorithms for correcting the sound stream of spontaneous speech can improve the overall quality of the signal by eliminating noise and other external factors that interfere with the typical perception of the speech stream.

They are conducting experimental studies that clearly demonstrated the effectiveness of the practical implementation of algorithms for correcting the sound stream in spontaneous speech recognition programs. A comparison of indicators of changes in the accuracy of spontaneous speech recognition programs clearly demonstrated the facts of increasing recognition accuracy using the specified algorithm. This is of crucial importance from the point of view of the prospects for the development and subsequent implementation of algorithms of this kind, as they have proven their high efficiency. In order to qualitatively take into account the features of spontaneous speech in the development of programs for its recognition, it is necessary to ensure adequate testing of this software through special test programs. This allows you to identify software deficiencies and eliminate them at the design and development stages. Achieving high-quality recognition of spontaneous speech by software methods is possible, provided that all the features of spontaneous speech itself are taken into account, as well as with direct control by the operator of compliance with all established conditions necessary for the smooth functioning of program nodes at all stages of the process.

REFERENCES

[1] Leontyeva, A.B., Kipyatkova, I.S. (2008). Accounting for the features of spontaneous speech when creating automatic recognition systems. News of Higher Educational Institutions, 11(51): 51-56.

[2] Beskrovnykh, V.I., Markasova, Y.V. (2019). Momentary tongue protrusion in spontaneous speech. Steps, 1(5): 54-69.

[3] Sidorova, Y.A. (2019). An integrated approach to the study of the lexical characteristics of the text. Vestnik SSUTI, 3: 80-88.

[4] Sidorova, Y.A. (2013). Development of linguistic support for information systems based on ontological knowledge models. Georesource Engineering, 5(322): 143-147.

[5] Aizstrauta, D., Ginters, E. (2015). Integrated acceptance and sustainability assessment model transformations into executable system dynamics model. Procedia Computer Science, 77: 92-97. https://doi.org/10.1016/j.procs.2015.12.364

[6] Prokofyeva, L.P., Plastun, I.L., Filippova, N.V., Matveyeva, L.Y., Plastun, N.S. (2021). Emotion recognition by speech signal characteristics (linguistic, clinical, informational aspects). Siberian Philological Journal, 2: 325-336.

[7] Hamed, I., Denisov, P., Li, C.Y., Elmahdy, M., Abdennadher, S., Vu, N.T. (2021). Investigations on speech recognition systems for low-resource dialectal Arabic-English code-switching speech. Computer Speech & Language, 72: 101278. https://doi.org/10.48550/arXiv.2108.12881

[8] Aviv, I., Gafni, R., Sherman, S., Aviv, B., Sterkin, A., Bega, E. (2023). Infrastructure from code: The next generation of cloud lifecycle automation. IEEE Software, 40(1): 42-49. https://doi.org/10.1109/MS.2022.3209958

[9] Tettegah, S.Y., Gartmeier, M. (2015). Emotions, technology, design, and learning. Academic Press, London.

[10] Ginters, E., Aizstrauta, D. (2018). Technologies sustainability modeling. Advances in Intelligent Systems and Computing, 746: 659-668. https://doi.org/10.1007/978-3-319-77712-2_61

[11] Shivakumar, P.G., Narayanan, S. (2022). End-to-end neural systems for automatic children speech recognition: An empirical study. Computer Speech & Language, 72: 101289. https://doi.org/10.1016/j.csl.2021.101289

[12] Biswas, A., Yilmaz, E., van der Westhuizen, E., de We, F., Nteisler, T. (2022). Code-switched automatic speech recognition in five South African languages. Computer Speech & Language, 71: 101262. https://doi.org/10.1016/j.csl.2021.101262

[13] Bondarenko, I.N., Gorbenko, E.A. (2018). Forming the powerful microwave pulses using resonator storage. Telecommunications and Radio Engineering (English translation of Elektrosvyaz and Radiotekhnika), 77(15): 1311-1319. https://doi.org/10.1615/telecomradeng.v77.i15.20

[14] Mussina, A., Ceccarelli, M., Balbayev, G. (2018). Neurorobotic investigation into the control of artificial eye movements. Mechanisms and Machine Science, 57: 211-221. https://doi.org/10.1007/978-3-319-79111-1_21

[15] Gusev, M.N. (2014). Methods and models of Russian speech recognition in information systems. Saint Petersburg State University of Telecommunications named after M.A. Bonch-Bruevich, Saint Petersburg.

[16] Tanaka, T., Masumura, R., Oba, T. (2021). Neural

candidate-aware language models for speech recognition. Computer Speech & Language, 66: 101157. https://doi.org/10.1016/j.csl.2020.101157

[17] Alekseyev, I.V., Mitrokhin, M.A., Kolchugina, Y.A. (2018). Software tool for evaluating the effectiveness of speech recognition technologies. News of Higher Educational Institutions, 5: 5-12.

[18] Yang, J., Chen, Z., Qiu, G., Li, X., Li, C., Yang, K., Chen, Z., Gao, L., Lu, S. (2022). Exploring the relationship between children's facial emotion processing characteristics and speech communication ability using deep learning on eye tracking and speech performance measures. Computer Speech & Language, 76: 101389. https://doi.org/10.1016/j.csl.2022.101389

[19] Reddy, D.R. (2005). Speech recognition, 8. Academic Press, London.

[20] Kopnova, O., Shaporeva, A., Iklassova, K., Kushumbayev, A., Tadzhigitov, A., Aitymova, A. (2022). Building an information analysis system within a corporate information system for combining and structuring organization data (on the example of a university). Eastern-European Journal of Enterprise Technologies, 6(2-120): 20-29. https://doi.org/10.15587/1729-4061.2022.267893

[21] Vashkis, I.I., Yevseyenko, I.N. (2020). Linguistic support of automated systems. I-Methods, 1(12): 1-7.

[22] Granel, X. (2014). Multilingual information management. Chandos Publishing, Oxford.

[23] Schultz, T., Kirchhoff, K. (2006). Multilingual speech processing. Academic Press, London.

[24] Li, J., Deng, L., Häb-Umbach, R., Gong, Y. (2015). Robust automatic speech recognition. Academic Press, London.

[25] Ainsworth, W.A. (2019). Mechanisms of speech recognition, 9. Pergamon, Oxford.

[26] Waibel, A., Lee, K.F. (2020). Readings in speech recognition, 8. Morgan Kaufmann, Burlington.

[27] Dey, N. (2019). Intelligent speech signal processing. Academic Press, London.

[28] Dey, N. (2021). Applied speech processing. Academic Press, London.

[29] Zhang, S., Tan, X., Chuang, Y., Zhao, X. (2021). Learning deep multimodal affective features for spontaneous speech emotion recognition. Speech Communication, 127: 73-81. https://doi.org/10.1016/j.specom.2020.12.009

[30] Nakamura, M., Iwato, K., Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. Computer Speech & Language, 22(2): 171-184. https://doi.org/10.1016/j.csl.2007.07.003