NOVATEUR PUBLICATIONS
JournalNX- A Multidisciplinary Peer Reviewed Journal
ISSN No: 2581 - 4230
VOLUME 10, ISSUE 5, May - 2024

DIGITAL TECHNOLOGIES AS A FACTOR IN THE DEVELOPMENT OF SPEECH TECHNOLOGIES BASED ON SIGN LANGUAGE

Abduraxmanova Shaxnoza
Tashkent State Pedagogical University named after Nizami,
PhD in Pedagogy, Assistant Professor, Department of Information Technology
E-mail: ashahnoza_78@mail.ru

Abstract:
This article discusses how to improve the efficiency of the use of speech technologies based on digital technologies. This will provide a new level of digital services provided to the population and organizations in the self-service mode, as well as increase the efficiency of the use of speech technologies for people with hearing impairment.

Keywords: speech technologies, digital technologies, speech portals, human-machine systems.

Introduction

According to the World Health Organization, about 5% of people in the world (328 million adults and 32 million children, a total of 360 million people) suffer from hearing loss. Disabling hearing loss is defined as hearing loss in the better ear that is greater than 40 dB in adults and 30 dB in children.
According to statistics, one in three people over 65 years of age suffers from disabling hearing loss. Most of these people live in areas with low social security, and in this age group the highest prevalence is found in South Asia, the Asia-Pacific region and sub-Saharan Africa.
Since "deaf" people hear very little or do not hear at all, they often use sign language to communicate [1].
The World Health Organization estimates that by the middle of the 21st century, about 1 billion people, or one in ten, will have problems associated with hearing loss.
Summarizing and systematizing the state and trends in the development of digital and service systems and taking into account the tasks of overcoming unequal access to information for different segments of the population, which are important in the modern information society, a concept is proposed for attracting modern speech, digital and telecommunication technologies to create new interfaces for accessing data and services, including remote ones. The implementation of this concept will lead to the introduction into practice of speech portals and self-service telephone services based on speaker-independent recognizers and computer telephony tools.
Companies that work with a large number of clients and want to introduce new interface technologies are primarily interested in using computer speech recognition. Among these companies:
- Service Centre;
- telecommunications operators, mobile operators, Internet providers;
- inquiry services of railway stations and airports, services for ordering tickets and ordering a taxi;
- government agencies, e-government services;
- banks, large financial structures, insurance companies;
- various reference and consulting services;

- travel companies;
- large trade enterprises, shops, order delivery services [3].

Here, the economy and the convenience of customers come to the fore everywhere. A separate area of application of speech technologies is the needs of people with disabilities. For such people, speech technologies make it possible to create services that will help them receive information, education, professional knowledge and successfully participate in social and working life.

In America, self-service telephone services have become quite widespread, and this market is developing successfully. For example, according to BCC Research, a company specializing in market research and forecasting the success of new technologies [http://www.bccresearch.com], the speech recognition technology market will grow annually by 12.1% between. The market size in 2022 is estimated at 184.9 billion US dollars. The projected annual increase in sales of software, both basic and applied, is estimated at 8.6%, the same indicator for hardware - at 10.8%.

The most effective way of teaching the correct speech of people with hearing impairments is classes with experienced deaf teachers. Advances in the development of digital technologies have led to the emergence of computer simulators to help people with impaired articulation and hearing. This is facilitated by the presence in such simulators of the visual feedback function, which is especially important for hearing-impaired and deaf people.

The principle of operation of the simulators is the spectral transformation of the input audio signal, the identification of its characteristics, comparison with the standard and the return and presentation of the result. In a number of simulators for the purpose of analyzing the original signal, systems of automatic speech recognition are used. But all these products were stand-alone software that had to be purchased and installed on the user's computer [11].

In Uzbekistan, speech technologies are still not widely used. The following reasons are the most significant to explain this lag:

- the absence on the market until recently of reliable recognizers for the Uzbek and Russian languages, focused on use in public service centers;
- low quality of voice transmission in old generation telephone networks;
- lack of traditions of communication with automated systems among the population;
- distrust of information and service providers in the reliability of the new technology [8].

All these reasons are quite objective. Indeed, the reliability of recognizers has always been a worrying factor when deciding on the implementation of a speech recognition application, since the mathematical models underlying the operation of most modern industrial recognizers are stochastic in nature and fundamentally cannot provide one hundred percent recognition reliability. In addition, the Uzbek and Russian languages are difficult to implement reliable recognizers due to a much larger number of word forms than in many other languages.

The most effective way of teaching the correct speech of people with hearing impairments is classes with experienced deaf teachers. Advances in the development of digital technologies have led to the emergence of computer simulators to help people with impaired articulation and hearing. This is facilitated by the presence in such simulators of the visual feedback function, which is especially important for hearing-impaired and deaf people.

The principle of operation of the simulators is the spectral transformation of the input audio signal, the identification of its characteristics, comparison with the standard and the return and presentation of the result. In a number of simulators for the purpose of analyzing the original signal, systems of automatic speech recognition are used. But all these products were stand-alone software that had to be purchased and installed on the user's computer [10].

Placing such programs on the Internet and providing online access to them would significantly increase the number of people who have the opportunity to independently train the correct pronunciation of sounds, syllables, words and phrases from home computers. For these purposes, work has begun on a sound simulator for the "Uzbekistan sign language (USL)" project for people with hearing impairments.

Methods

Much attention is paid to the development of the interface in modern human-machine systems, since the quality of the system is ultimately evaluated by the quality of the interaction between the person and the system.

A human-machine interface is a set of rules and means for organizing human interaction with an automated system. A unimodal speech interface is understood as an interactive human-machine interaction, where the input of a request and the output of a response are carried out using speech [2]. The use of speech was expected to make human-computer interaction more natural and efficient. Indeed, many applications with a speech interface have recently been introduced, but we have to admit that speech has not yet become a common means for input and output of data or information. It turned out that the speech interface is much more critical and difficult to develop than the graphical interface, which, in turn, only seems simple to implement [6].

At first, it seemed that constructing a simple conversational interaction between a human and a computer was a trivial task. However, it soon became clear that most of the created applications based on computer speech technologies worked unsatisfactorily and caused dissatisfaction and even irritation of customers. It turned out that the creation of effective speech applications requires the developer to have knowledge in various areas, including the properties of recognizers, the nature of speech, the basics of linguistics, and the characteristics of human behavior in different situations, and its inherent limitations in the perception of concepts and speech. In addition, the developer needs the ability to choose the voice that would best represent the company or a particular application to customers, choose the style of communication, find a balance between simplicity and flexibility of the dialogue, find the best ways to detect and correct errors. Therefore, it is believed that the development of a speech interface is a combination of precise knowledge and art [7].

Results and Discussion

The reliability of the recognizer is characterized by the average value of the probability of correct recognition of a simple element of the dialogue p, which we will call the "internal" recognition probability. When clients interact with the system, especially the so-called beginners who do not have experience with such an automaton, a "human factor" appears, which is reflected in the average recognition probability. The average probability of correct recognition of a simple element, which takes into account both the average internal probability of recognition and the influence of the human factor, will be called the "nominal" probability and denoted by p_0.

Identification and correction of errors is carried out by calling the confirmation and re-questioning procedure. Before asking again, the system usually asks the client to speak more clearly, or gives an example of an answer, or offers to choose from a narrower set. It is assumed that this will increase the probability of recognizing the questioned element of the dialogue by reducing the influence of the "human factor" to the value of p_{01} at the first re-question, to the value of p_{02} at the second re-question, etc., in the limit up to the value of the internal recognition probability. Let us take this position as a more general case in the subsequent analysis of the process of asking again in a dialogue of simple and composite elements [5].

Denote by p_{0j} the probability of correctly recognizing an element of the dialogue during the j-th questioning. Then the probability of correct recognition of a dialogue element with n repeated questions is determined by the formula:

$$P_n = p_0 + \sum_{j=1}^{n} p_{0j} \cdot \prod_{k=0}^{j-1} \left(1 - p_{0k}\right), \quad n=1, 2, 3\ldots, \qquad (2)$$

where p_0 is the nominal probability of correct recognition of a dialog element.

Under conditions of homogeneity of the recognition process, when the probability of recognition at each re-question is equal to p_0, the probability of correct recognition of a dialog element is represented by a truncated geometric distribution:

$$P_n = \sum_{j=0}^{n} p_0 \left(1 - p_0\right)^j = p_0 \cdot \frac{1 - \left(1 - p_0\right)^{n+1}}{1 - \left(1 - p_0\right)} = 1 - \left(1 - p_0\right)^{n+1}, \quad n=0,1, 2, 3\ldots. \qquad (3)$$

The value p_n, calculated by formula 3, is the lower estimate of the probability of correct recognition of a dialogue element with n repeated questions. The numerical values of this probability for various values of the nominal probability of element recognition and the number of re-questions are given in Table 1.

Table 1 **Probability of correctly recognizing a simple dialogue element**

p_0	P_1	P_2	P_3
0,6	0,84	0,936	0,974
0,7	0,91	0,973	0,9919
0,8	0,96	0,992	0,9984
0,9	0,99	0,999	0,9999
0,95	0,9975	0,9999	0,9999

It can be seen from the table that for $p_0 > 0.7$, one or two repetitions can provide a probability of recognizing a dialogue element of more than 0.99.

A composite element, according to the definition, contains semantic concepts (components) that are recognized independently. The recognition probability of each component (p_{0l}) is similar to the nominal recognition probability of a simple element. In this case, the nominal recognition probability of the entire composite element p_{0c} is determined by the relation:

$$p_{0c} = \prod_{l=1}^{M} p_{0_l} \quad (4)$$

The probability of correct recognition of a composite element depends on the procedures for identifying and correcting recognition errors implemented in the dialogue process, which can be implemented in three ways:

- confirmation request for the entire constituent element with subsequent adjustment by asking the entire constituent element again;
- confirmation request for the entire component element with subsequent adjustment by asking each of the components of this element;
- request for confirmation for each component with subsequent adjustment by asking each component again.

Conclusion

Measures to improve the quality of speech recognizers improve the satisfaction of users of speech applications, however, the possibility of incorrect recognition by the system of the client's statements remains, and in connection with this, the problem of optimizing scenarios for dialogues with repeated questions and algorithms for organizing procedures for detecting and correcting errors remains. Below, we consider mathematical models for the analysis and optimization of scenarios and algorithms for managing a dialogue with repeated questions and for estimating the main parameters of such dialogues. It is advisable to use these models at the system design stage, when decisions are made regarding the structure of dialogue elements and confirmation procedures. Calculations based on such models make it possible to choose the most appropriate from numerous scenarios and dialogue control algorithms and thereby limit the number of system behavior options that are subject to subsequent labor-intensive testing. [9].

The classification of dialogs is preliminarily made, which is based on the following features: types of dialog elements, location of error detection and correction procedures, method of implementing error detection and correction procedures.

The technique for performing a comparative analysis of variants consists in using the principle of quantification modified in relation to a given task, when the variants are compared on a set of identical typical quantitative values of a number of parameters. For example, a comparative assessment of the probability of successful completion of the dialogue is carried out with the same values of the probabilities of correct recognition of speech blocks and the same allowable number of re-questions for the compared scenarios. The need for such assumptions is dictated by the large dimension of the models and a wide range of parameters. Quantification is a fairly popular method for quantitatively analyzing the quality of interfaces.

Two main quantitative criteria for assessing the quality of speech dialogue human-machine interaction were chosen as characteristics for comparing options: the probability of successful completion of the dialogue and the duration of the dialogue. The search for the optimal scenario is carried out on the basis of minimizing the estimates of the duration of the dialogue with a given probability of its successful completion.

References:

1. Глухота и потеря слуха [Электронный ресурс]. 2021. URL: https://www.who.int/ru/news-room/fact-sheets/detail/deafness-and-hearing-loss (Дата обращения: 13.01.2022).
2. Азимбаев, Д. Ж. Искусственный интеллект и машинное обучение / Д. Ж.Азимбаев, И. А. Куан, И. В. Гулида // Вестник современных исследований. - 2019. - № 1.3 (28). - С. 6-7. – https://elibrary.ru/item.asp?id=36885190
3. Борисова, В. В., & Налбандян, О. Г. (2016). Информационная система «речевые технологии» как эффективный инструмент диагностики и коррекции речевых нарушений детей младшего возраста. Сборник материалов Ежегодной международной научно-практической конференции «Воспитание и обучение детей младшего возраста» (No. 5, pp. 125-127). ООО Мозаика-Синтез.
4. Abduraxmanova Shaxnoza Abduxakimovna, & Saydivosilov Saidiabzal Anvar ugli. (2023). The need to develop the digital technology skills of future computer science teachers in Uzbekistan. World Bulletin of Management and Law, 23, 64-67. Retrieved from https://scholarexpress.net/index.php/wbml/article/view/2874
5. Леонтьев, Н. А., & Ефремов, Д. Д. (2020). ОБЗОР РЕЧЕВЫХ ТЕХНОЛОГИЙ, ПРИМЕНЯЕМЫХ ДЛЯ ЗАЩИТЫ ИНФОРМАЦИИ.Инновационное развитие: потенциал науки и современного образования, 83-84.
6. Молчанова, А. А., & Петрушенко, Р. В. (2016). Речевые технологии-следующий уровень сервиса. Технические науки в России и за рубежом (pp. 6-8).
7. Saidiabzal, S. (2023). Comparative analysis of programming languages used in education. образование наука и инновационные идеи в мире, 22(1), 151-153.
8. Харламов, А. А., & Чучупал, В. Я. (2019). СОВРЕМЕННОЕ СОСТОЯНИЕ И ТЕНДЕНЦИИ РАЗВИТИЯ РЕЧЕВЫХ ТЕХНОЛОГИЙ. Цифровая обработка сигналов и её применение–DSPA-2019 (pp. 19-25).
9. Bakiyeva, Z. R. (2023, May). Theoretical principles of teaching computer animation to students in an electronic learning environment. In Proceedings of International Conference on Modern Science and Scientific Studies (Vol. 2, No. 5, pp. 5-8).
10. Zakirova, F. M., & Abdurakhmanova, S. (2020). The use of multimedia technologies for the development of intellectual skills of students. Theoretical & Applied Science, (9), 24-29.
11. Bakiyeva, Z. (2022). Teaching the steps of creating animation to students in higher education institutions. Академические исследования в современной науке, 1(17), 226-227.