UDC: 621.391 DOI: 10.18372/2073-4751.82.20366

Lavrynenko O.Yu., Ph.D.,
orcid.org/0000-0002-7738-161X,
e-mail: oleksandrlavrynenko@gmail.com

N-PHONEME MODEL OF SPEECH SIGNAL RECOGNITION BASED ON HIDDEN MARKOV PROCESSES

State University "Kyiv Aviation Institute"

Introduction

In modern telecommunication networks and systems, speech signal processing and recognition systems are occupying an increasingly important niche [1], as the demand for multimedia services is growing steadily, leading to higher requirements for the quality and efficiency of services provided. It is worth noting that the simplest and most natural means of exchanging information and issuing commands for humans is speech. However, automatic speech recognition and understanding is very difficult. The speech signal has undeniable advantages that make it an effective means of transmitting information in a variety of areas. Speech is a fairly fast means of communication, it is easily transmitted through communication channels, while leaving the eyes and hands free to move [2]. However, the implementation of this method of information exchange between humans and machines is currently limited by insufficient study of the process of analyzing and classifying voice commands, which is manifested in the insufficient quality of the existing mathematical models of this information process. Equally important is the problem of developing robotic manipulators [3] that operate in aggressive environments. At the same time, an urgent scientific task is the development of noise-resistant methods for recognizing speech commands [4] for effective robot control, for example, in places dangerous for humans.

Literature review and problem statement

Based on the analyzed literature [5-10], there are currently two main approaches to speech signal recognition: phonemic [5] and formant [6]. In the formant approach, recognition is performed by formant frequencies. Acoustic vibrations of four frequencies, which are formed in the resonant cavities of the speech tract, are actively involved in the formation of speech. Modern speech recognition systems perform spectral analysis [7], which allows to extract the most informative components from the speech signals. These are formant frequencies and noise. In addition to spectral analysis, more advanced methods are used, such as phoneme recognition or wavelet transform [8]. This paper focuses on the consideration and study of the phoneme recognition method. Among the phoneme methods, the most effective and widespread today is the three-phoneme method of speech signal recognition [9]. However, existing methods no longer meet modern requirements for accuracy and efficiency of recognition, their maximum recognition probability is limited to 95% accuracy, which is currently not satisfactory when it comes to the development of authentication and security systems based on speech recognition [10].

Therefore, there is a need to develop new models for recognizing fused speech and voice commands. The analysis of shows that an urgent scientific task is to increase the probability of recognizing fused speech, which can be achieved by improving existing or developing new models and algorithms for recognizing speech signals and commands, which is the focus of this work.

Summary of the main material

This paper presents a four-phoneme model of speech signal recognition from the point of view of the theory of hidden Markov processes. To form a database of standards, it is necessary to identify the structural units of the language of the spoken words - in this case, phonemes. In this paper, we have developed an improved algorithm that searches for interphoneme transitions and a set of segments of predicted phonemes. As a result of the experiments, it turned out that the deformations of the speech rate are clearly nonlinear. For nonlinear matching of the speech signal, gradient descent methods, Markov modeling, and dynamic programming algorithms are used most widely [11].

Thus, a speech signal X_L, which is subsequently subject to recognition, can be described by a sequence of n-measurable elements x_i:

$$X_L = (x_1, x_2, x_3, \ldots, x_i, \ldots x_L),$$

where L is the length of the speech signal.

Phonemes are realizations of the first level of the hierarchy in a speech recognition system. They can be represented as a sequence of elements

$$J_{bm} = (j_{b+1}, j_{b+2}, \ldots, j_m),\ 0 \le m < b \le 1.$$

According to the representation hierarchy system, phonemes are followed by syllables, then words, and then sentences. Images of elements of higher levels of the hierarchy are given by transcriptions in the alphabet of images of the level lower than the element according to the expression

$$k^2 = (k_1^1, k_2^1, k_3^1, \ldots, k^1, \ldots, q(k^2)),$$

where k^2 is the phoneme transcription of the top-level element from the dictionary K^2, $k^2 \in K^2$, $q(k^2)$ is the transcription length.

The number of states determines the complexity of the model, which defines phoneme segments. The most common are models with up to three states.

This paper first proposes to use the theory of hidden Markov processes [13] to describe a model with N-states.

In Fig. 1 shows an example of a model with one state where $S_b(k^1)$ is the initial state of the model, $S_e(k^1)$ is the final state of the model, $S(k^1)$ is the basic state of the model for the phoneme k^1, $p(j/k^1)$ is the element generated by the transition to one beat.

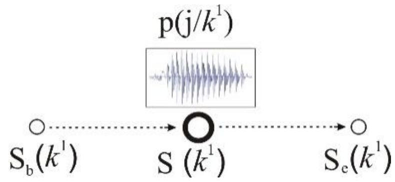

Fig. 1. Model with one state

The probability of the segment $J_{\mu\nu}$ with the independence of observations of the elements j can be calculated using the formula:

$$P(J_{\mu\nu}/k^1) = \begin{cases} \prod_{i=\mu+1}^{\nu} p(j_i/k), & if \quad T_{\min}(k^1) \le \nu - \mu \le T_{\max}(k^1), \\ 0, & if \quad T_{\min}(k^1) \ge \nu - \mu \ge T_{\max}(k^1) \end{cases},$$

where $T_{\min}$ and $T_{\max}$ are the length limits of the generated phoneme segment k^1 [12].

In this model, the state $S_1(k^1)$ has the probability distribution $p_1(j/k^1)$ and the length constraints of subsegments $(T_{\min 1}(k^1), T_{\max 1}(k^1))$, the state $S_2(k^1)$ has the probability distribution $p_2(j/k^1)$ and the length constraints of subsegments $(T_{\min 2}(k^1), T_{\max 2}(k^1))$, and the state $S_b(k^1)$ has the probability distribution $p_3(j/k^1)$ and the length constraints of subsegments $(T_{\min 3}(k^1), T_{\max 3}(k^1))$, $j \in J$. Based on the initial conditions, the following restrictions are imposed on their lengths:

$$T_{\min,2}(k^1) = T_{\min}(k^1),$$

$$T_{\min,1}(k^1) + T_{\min,3}(k^1) = T_{\min}(k^1),$$

$$T_{\max 1}(k^1) + T_{\max 2}(k^1) + T_{\max 3}(k^1) = T_{\max}(k^1).$$

Depending on which trajectory of the model (in the direction of the arrows) $S_1 -> S_2 -> S_3$, $S_1 -> S_2$, $S_2 -> S_3$, $S_1 -> S_3$ we choose, the probability of segment $J_{\mu\nu}$ will be calculated as follows:

$$P(J_{\mu\nu}/k^1)=\left\{\begin{array}{c}\left\{\begin{array}{c}\prod\limits_{i=\mu+1}^{\nu}p_1(j_i/k^1),\quad if\quad T_{\min 1}(k^1)\le\nu-\mu\le T_{\max 1}(k^1),\\ \prod\limits_{i=\mu+1}^{\nu}p_2(j_i/k^1),\quad if\quad T_{\min 2}(k^1)\le\nu-\mu\le T_{\max 2}(k^1),\\ \prod\limits_{i=\mu+1}^{\nu}p_3(j_i/k^1),\quad if\quad T_{\min 3}(k^1)\le\nu-\mu\le T_{\max 3}(k^1),\\ \max\sum\limits_{q\le\nu\le z}\prod\limits_{i=\mu+1}^{\nu}p_1(j_i/k^1)\prod\limits_{i=\mu+1}^{\nu}p_2(j_i/k^1)\prod\limits_{i=\mu+1}^{\nu}p_3(j_i/k^1),\\ if\quad T_{\min 1}(k^1)+T_{\min 2}(k^1)+T_{\min 3}(k^1)\le\nu-\mu\le T_{\max 1}(k^1)+T_{\max 2}(k^1)+T_{\max 3}(k^1)\end{array}\right\}\\ 0,\quad if\quad T_{\min}(k^1)\ge\nu-\mu\ge T_{\max}(k^1)\end{array}\right\}$$

where,

$$q=\max(T_{\min 1}(k^1);\ (\nu-\mu)-T_{\max 3}(k^1))+\mu,$$

$$z=\min(T_{\max 1}(k^1);\ (\nu-\mu)-T_{\min 3}(k^1))+\mu.$$

At present, systems using the three-phoneme model of speech signal recognition [14], which replaced the two-phoneme model, are quite well studied, as the efficiency and accuracy of the two-phoneme model did not satisfy.

But over time, the three-phoneme model also ceased to meet the requirements for accuracy and efficiency of speech signal recognition. The paper investigates a four-phoneme model of speech signal recognition.

Fig. 2 shows a four-phoneme model of speech signal recognition.

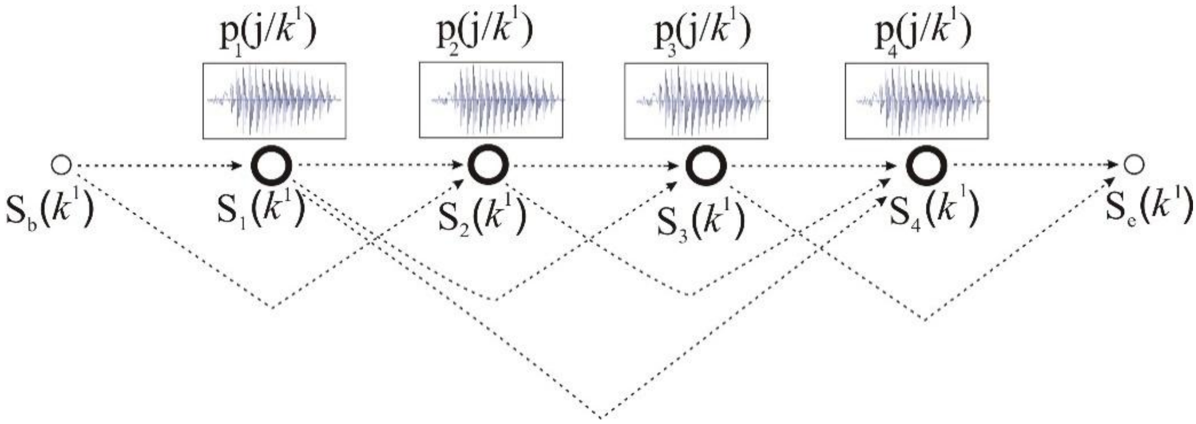

Fig. 2. Model with four states

As in the models with fewer states, each state has a probability distribution: $S_1(k^1)$, $S_2(k^1)$, $S_3(k^1)$, $S_4(k^1)$ and sub-segment length constraints: $(T_{\min 1}(k^1),T_{\max 1}(k^1))$, $(T_{\min 2}(k^1),T_{\max 2}(k^1))$, $(T_{\min 3}(k^1),T_{\max 3}(k^1))$, $(T_{\min 4}(k^1),T_{\max 4}(k^1))$ respectively. Based on the initial conditions $(T_{\min}(k^1),T_{\max}(k^1)$, restrictions are imposed on their lengths:

$$T_{\max 1}(k^1)+T_{\max 2}(k^1)+T_{\max 3}(k^1)+$$
$$+T_{\max 4}(k^1)=T_{\max}(k^1).$$

To theoretically calculate the effectiveness of the developed model, it is necessary to calculate the probability of a segment appearing. Depending on the path through the model (by arrows) $s_1->s_2->s_3->s_4$, $s_1->s_2->s_3$, $s_1->s_2$, $s_2->s_3->s_4$, $s_3->s_4$, $s_1->s_3->s_4$, $s_1->s_3$, $s_1->s_4$, $s_2->s_3$, $s_3->s_4$, $s_1->s_2->s_4$, the probability of segment $J_{\mu\nu}$ will be calculated in accordance with one of the following principles. The probability of the $J_{\mu\nu}$ segment is calculated as follows:

$$P(J_{\mu\nu}/k^1)=\begin{cases}\prod_{i=\mu+1}^{\nu} p_1(j_i/k^1), \ \ if \ \ T_{\min 1}(k^1)\le\nu-\mu\le T_{\max 1}(k^1),\\ \prod_{i=\mu+1}^{\nu} p_2(j_i/k^1), \ \ if \ \ T_{\min 2}(k^1)\le\nu-\mu\le T_{\max 2}(k^1),\\ \prod_{i=\mu+1}^{\nu} p_3(j_i/k^1), \ \ if \ \ T_{\min 3}(k^1)\le\nu-\mu\le T_{\max 3}(k^1),\\ \prod_{i=\mu+1}^{\nu} p_4(j_i/k^1), \ \ if \ \ T_{\min 4}(k^1)\le\nu-\mu\le T_{\max 4}(k^1),\\ \max \sum_{q\le\nu\le z}\prod_{i=\mu+1}^{\nu} p_1(j_i/k^1)\prod_{i=\mu+1}^{\nu} p_2(j_i/k^1)\prod_{i=\mu+1}^{\nu} p_3(j_i/k^1)\times\\ \times\prod_{i=\mu+1}^{\nu} p_4(j_i/k^1), \ \ if \ \ T_{\min 1}(k^1)+T_{\min 2}(k^1)+T_{\min 3}(k^1)+\\ +T_{\min 4}(k^1)\le\nu-\mu\le T_{\max 1}(k^1)+T_{\max 2}(k^1)+T_{\max 3}(k^1)+T_{\max 4}(k^1)\\ \max \sum_{q\le\nu\le z}\prod_{i=\mu+1}^{\nu} p_1(j_i/k^1)\prod_{i=\mu+1}^{\nu} p_2(j_i/k^1)\prod_{i=\mu+1}^{\nu} p_3(j_i/k^1), \ \ if\\ T_{\min 1}(k^1)+T_{\min 2}(k^1)+T_{\min 3}(k^1)\le\nu-\mu\le T_{\max 1}(k^1)+\\ +T_{\max 2}(k^1)+T_{\max 3}(k^1)\\ \ldots\\ \max \sum_{q\le\nu\le z}\prod_{i=\mu+1}^{\nu} p2(j_i/k^1)\prod_{i=\mu+1}^{\nu} p_4(j_i/k^1)\prod_{i=\mu+1}^{\nu} \ , \ \ if\\ T_{\min 2}(k^1)+T_{\min 4}(k^1)\le\nu-\mu\le T_{\max 2}(k^1)+T_{\max 4}(k^1)\\ \max \sum_{q\le\nu\le z}\prod_{i=\mu+1}^{\nu} p3(j_i/k^1)\prod_{i=\mu+1}^{\nu} p_4(j_i/k^1)\prod_{i=\mu+1}^{\nu} \ , \ \ if\\ T_{\min 3}(k^1)+T_{\min 4}(k^1)\le\nu-\mu\le T_{\max 3}(k^1)+T_{\max 4}(k^1)\end{cases}$$

Based on the studied four-phoneme model of speech signal recognition, a recognition algorithm is proposed, which includes the following steps:

The algorithm has the following steps:

1. Preparation of a dictionary;
2. selecting a set of phonemes and creating a transcription of words from the dictionary;
3. creating a database index from four phonemes to transcriptions;
4. training an acoustic model based on the accumulated speech signals.

Now let's describe the stages of recognition:

1. apply a phonetic transcriber to the input signal to obtain a phoneme sequence;
2. divide the sequence of phonemes into groups of four phonemes with a shift of one phoneme;
3. create queries to the four-phoneme database;
4. obtain transcription lists by querying the database index;
5. organize the transcriptions by their rank;
6. select the first N transcriptions with the highest ranks as a sub-dictionary for recognition;
7. recognize the input speech signal in a limited sub-vocabulary.

The recognition process is shown in Fig. 3.

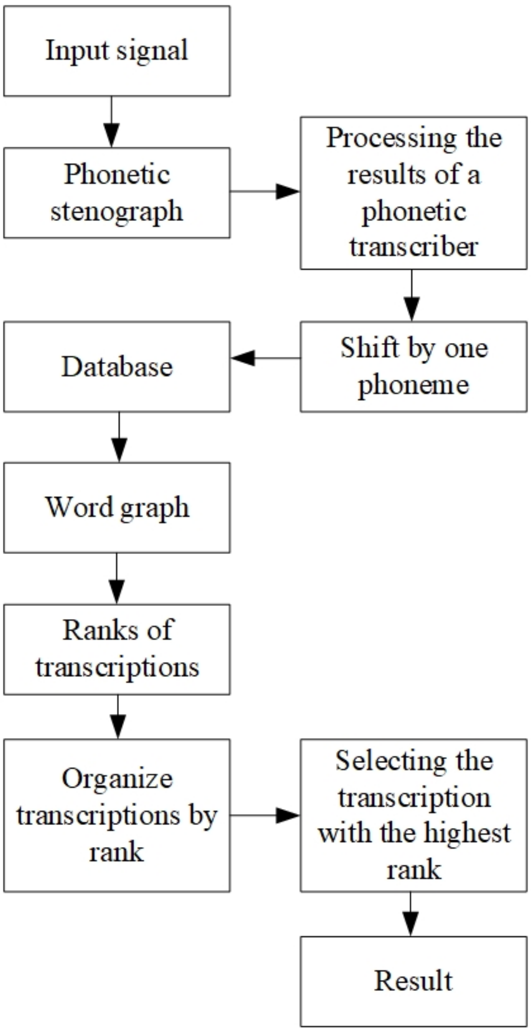

Fig. 3. The process of speech signal recognition

If there is a coincidence of transcriptions of the same word generated by different phoneme quadruplets, then the ranks of these transcriptions are increased by one. For each moment of time, you can count the number of transcriptions of the words that are included in this time period.

To reduce the complexity of the word graph, the algorithm uses the N constraint on the number of words at any given time.

Since the word graph is formed from left to right, it can be generated in real time, with a delay equal to the maximum transcription length.

Results of the research

In this section, we consider the effectiveness of speech signal recognition using the four-phoneme method, where we set the parameters of speech signal recognition efficiency and investigate the recognition probabilities with the introduction of distortion factors into the speech signal, namely, amplitude and phase noise. First, a sample of voice commands was created for their further recognition.

A Mackie Carbon microphone was used for recording. The microphone is connected to the sound card of a Manli C-Media 8738 PCI-E (5.1) computer (M-CMI8738-PCI-E). The study of voice command recognition speed was conducted for vocabularies of different sizes (10, 20, 30, 50, 100, 125, 140 commands). The maximum size of the dictionary was chosen based on the considerations that 140 commands is a sufficient number to control the robotic system.

The results of the voice command studies, namely the dependence of recognition time on the size of the dictionary, are shown in Fig. 4.

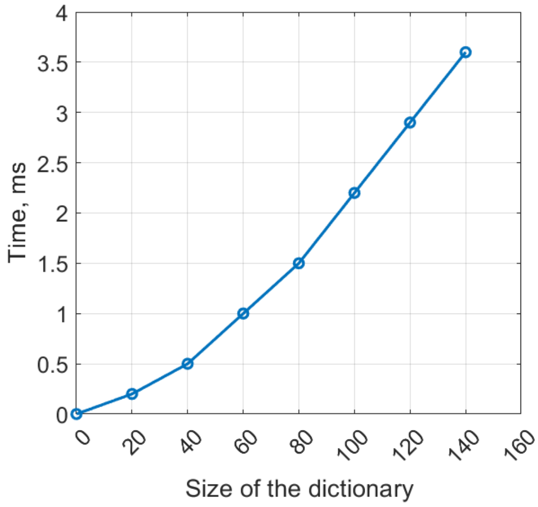

Fig. 4. Dependence of command recognition time on the size of the dictionary

The results allow us to conclude that the time required to recognize voice commands depends on the size of the dictionary, and the larger the dictionary size, the longer it takes to recognize.

To investigate the probability of recognizing fused speech using the four-phoneme method, a sample of 100 samples was recorded.

Based on the obtained results of fused speech recognition, Fig. 5 shows the dependence of the recognition probability on the size of the dictionary and Fig. 6 shows the dependence of the recognition time on the size of the dictionary.

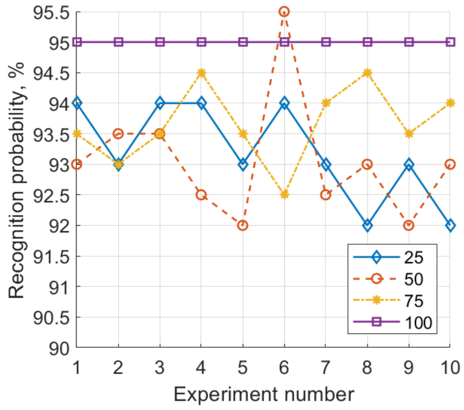

Fig. 5. Dependence of recognition probability on the size of the dictionary in the recognition of fused speech

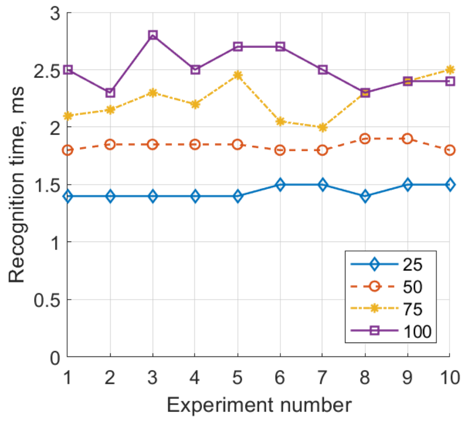

Fig. 6. Dependence of the time of fused speech recognition on the size of the dictionary

The graphs above show that as the vocabulary increases, the recognition probability and recognition time increase.

We studied the recognition of fused speech under conditions of amplitude signal distortion. The results are shown in Fig. 7. and Fig. 8.

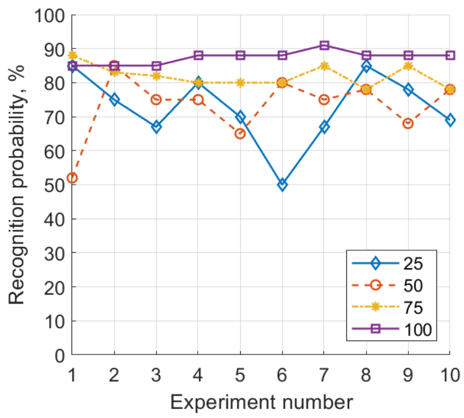

Fig. 7. Dependence of the probability of recognizing a fused speech on the size of the dictionary with an amplitude-distorted signal

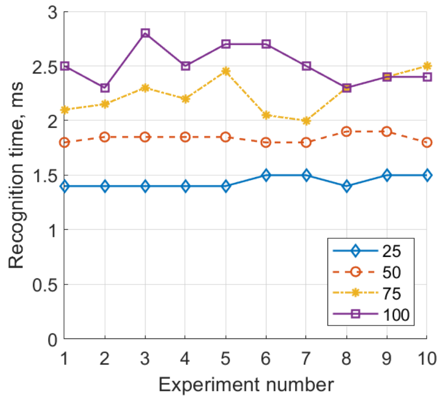

Fig. 8. Dependence of the fused speech recognition time on the size of the dictionary with an amplitude-distorted signal

The graphs above show that as the vocabulary increases with an amplitude-distorted signal, the recognition probability decreases compared to recognition without distortion, and the recognition time increases.

We have studied the recognition of fused speech under conditions of phase signal distortion, and the results are shown in Fig. 9. and Fig. 10.

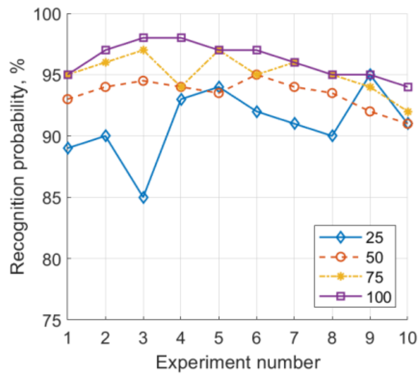

Fig. 9. Dependence of the probability of recognizing a fused speech on the size of a dictionary with a phase-distorted signal

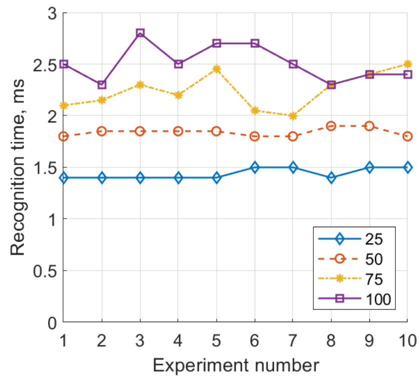

Fig. 10. Dependence of the fused speech recognition time on the size with a phase-distorted signal

The graphs above show that under the conditions of a phase-distorted signal, with an increase in the vocabulary, the recognition probability decreases compared to recognition without distortion, but this probability is higher than recognition with amplitude noise, and the recognition time increases. The studies of the recognition probability for voice commands and fused speech using the four-phoneme method showed that it is possible to achieve a recognition probability of 98%. Based on the results, it can be concluded that the recognition probability improves with the increase in the size of the dictionary, but at the same time, the recognition time increases.

Conclusions

This paper solves the urgent scientific problem of increasing the probability of recognizing commands and fused speech in radio engineering devices and telecommunications under the influence of distorting factors by developing new recognition models.

The article reviews methods and ways of representing speech signals, considers models for recognizing speech signals with a different number of states, and investigates the problems that arise when recognizing speech signals using existing methods. It is established that in order to improve the efficiency of recognition in radio engineering devices and telecommunications, it is necessary to develop new or improve existing models of phoneme recognition of speech signals.

The paper proposes to use hidden Markov processes to conduct a probabilistic description of the one-, three-, and four-phoneme model of speech signal recognition, which makes it possible to theoretically estimate the probability of recognition using each of the models. On the basis of a comparative analysis, the four-phoneme model of speech signal recognition was investigated, which, by improving the three-phoneme model by adding one more state to the model, allows, unlike other models of speech signal recognition, to increase the probability of their recognition.

The probability of recognizing speech signals and commands using the four-phoneme method is established, and it is shown that its application in practice with the help of the developed software allows to achieve a probability of 98%. The influence of amplitude and phase distortion of the speech signal on the recognition probability was studied, which showed that the recognition probability decreases when amplitude interference (recognition probability is 81.7%) and phase interference (recognition probability is 92.3%) are introduced into the speech signal. A comparative analysis of the four- and three-phoneme models is carried out, which shows that the recognition probability error of the four-phoneme model is 40% less than that of the three-phoneme model.

References

1. Shahamiri S. R. Speech Vision: An End-to-End Deep Learning-Based Dysarthric Automatic Speech Recognition System. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2021. Vol. 29. P. 852–861. DOI: 10.1109/TNSRE.2021.3076778.

2. Shi Y. et al. Keyword Guided Target Speech Recognition. *IEEE Signal Processing Letters*. 2024. Vol. 31. P. 1945–1949. DOI: 10.1109/LSP.2024.3432324.

3. Yu Z., Wang H., Ren J. RealPRNet: A Real-Time Phoneme-Recognized Network for "Believable" Speech Animation. *IEEE Internet of Things Journal*. 2022. Vol. 9(7). P. 5357–5367. DOI: 10.1109/JIOT.2021.3110468.

4. Zhu D. et al. TWLip: Exploring Through-Wall Word-Level Lip Reading Based on Coherent SISO Radar. *IEEE Internet of Things Journal*. 2024. Vol. 11(19). P. 32310–32323. DOI: 10.1109/JIOT.2024.3427329.

5. Hsiao C.-H. et al. A Text-Dependent End-to-End Speech Sound Disorder Detection and Diagnosis in Mandarin-Speaking Children. *IEEE Transactions on Instrumentation and Measurement*. 2024. Vol. 73. P. 1–11. DOI: 10.1109/TIM.2024.3438853.

6. Kurtoğlu E. et al. ASL Trigger Recognition in Mixed Activity/Signing Sequences for RF Sensor-Based User Interfaces. *IEEE Transactions on Human-Machine Systems.* 2022. Vol. 52(4).

P. 699–712. DOI: 10.1109/THMS.2021.3131675.

7. Lee S. et al. IR-UWB Radar-Based Contactless Silent Speech Recognition of Vowels, Consonants, Words, and Phrases. *IEEE Access*. 2023. Vol. 11. P. 144844–144859. DOI: 10.1109/ACCESS.2023.3344177.

8. Xue J. et al. Auffusion: Leveraging the Power of Diffusion and Large Language Models for Text-to-Audio Generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing.* 2024. Vol. 32. P. 4700–4712. DOI: 10.1109/TASLP.2024.3485485.

9. O'Shaughnessy D. Review of Methods for Automatic Speaker Verification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2024. Vol. 32. P. 1776–1789. DOI: 10.1109/TASLP.2023.3346293.

10. Rademacher P., Wagner K. Efficient Bayesian Sequential Classification Under the Markov Assumption for Various Loss Functions. *IEEE Signal Processing Letters.* 2020. Vol. 27. P. 401–405. DOI: 10.1109/LSP.2020.2973854.

11. Marie B. et al. Phase-Sensitive Optical Time-Domain Reflectometry-Based Audio Excitation Signal Demodulation and Reproduction. *IEEE Transactions on Instrumentation and Measurement*. 2025. Vol. 74. P. 1–14. DOI: 10.1109/TIM.2025.3529561.

12. Lehmann F., Pieczynski W. Suboptimal Kalman Filtering in Triplet Markov Models Using Model Order Reduction. *IEEE Signal Processing Letters.* 2020. Vol. 27. P. 1100–1104. DOI: 10.1109/LSP.2020.3002420.

13. Liu Q. et al. Modular End-to-End Automatic Speech Recognition Framework for Acoustic-to-Word Model. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2020. Vol. 28. P. 2174–2183. DOI: 10.1109/TASLP.2020.3009477.

14. Truong L. V. Replica Analysis of the Linear Model With Markov or Hidden Markov Signal Priors. *IEEE Transactions on Information Theory*. 2023. Vol. 69(12). P. 7953–7975. DOI: 10.1109/TIT.2023.3299490.

Lavrynenko O.Yu.

N-PHONEME MODEL OF SPEECH SIGNAL RECOGNITION BASED ON HIDDEN MARKOV PROCESSES

This paper solves the urgent scientific problem of increasing the probability of recognizing commands and fused speech in radio engineering devices and telecommunications under the influence of distorting factors by developing new recognition models. It is proposed to use hidden Markov processes to conduct a probabilistic description of the one-, three-, and four-phoneme model of speech signal recognition, which makes it possible to theoretically estimate the probability of recognition using each of the models. On the basis of a comparative analysis, the four-phoneme model of speech signal recognition was investigated, which, by improving the three-phoneme model by adding one more state to the model, allows, unlike other models of speech signal recognition, to increase the probability of their recognition. The probability of recognizing speech signals and commands using the four-phoneme method is established, and it is shown that its application in practice with the help of the developed software allows to achieve a probability of 98%. The influence of amplitude and phase distortion of the speech signal on the recognition probability was studied, which showed that the recognition probability decreases when amplitude noise (recognition probability is 81.7%) and phase noise (recognition probability is 92.3%) are introduced into the speech signal. A comparative analysis of the four- and three-phoneme models is carried out, which shows that the recognition probability error of the four-phoneme model is 40% less than that of the three-phoneme model.

Keywords: *speech signals, phoneme speech recognition, hidden Markov processes, influence of interference, speech recognition probability.*

Лавриненко О.Ю.

N-ФОНЕМНА МОДЕЛЬ РОЗПІЗНАВАННЯ МОВНИХ СИГНАЛІВ НА ОСНОВІ ПРИХОВАНИХ МАРКІВСЬКИХ ПРОЦЕСІВ

В даній роботі вирішено актуальну наукову задачу підвищення ймовірності розпізнавання команд та злитної мови в радіотехнічних пристроях та засобах телекомунікацій в умовах дії спотворюючих факторів шляхом розробки нових моделей розпізнавання. Запропоновано з допомогою використання прихованих Марківських процесів проводити ймовірнісний опис одно-, трьох- та чотирьохфонемної моделі розпізнавання мовних сигналів, що дає можливість теоретично оцінити ймовірність розпізнавання з використанням кожної з моделей. На основі порівняльного аналізу було досліджено чотирьохфонемну модель розпізнавання мовних сигналів, яка за рахунок вдосконалення трьохфонемної, шляхом додавання ще одного стану до моделі, дозволяє, на відміну від інших моделей розпізнавання мовних сигналів, підвищити ймовірність їх розпізнавання. Встановлена ймовірність розпізнавання мовних сигналів і команд з використанням чотирьохфонемного методу, і показано, що його застосування на практиці за допомогою розробленого програмного забезпечення дозволяє досягнути ймовірності на рівні 98%. Проведено дослідження впливу на ймовірність розпізнавання амплітудного та фазового спотворення мовного сигналу, які показали, що ймовірність розпізнавання зменшується при внесені амплітудної завади (ймовірність розпізнавання становить 81,7%) та фазової завади (ймовірність розпізнавання 92,3%) у мовний сигнал. Проведено порівняльний аналіз чотирьох- та трьохфонемних моделей, в результаті чого показано, що помилка ймовірності розпізнавання чотирьохфонемної моделі на 40% менша ніж у трьохфонемної.

Ключові слова*: мовні сигнали, фонемне розпізнавання мови, приховані Марківські процеси, вплив завад, ймовірність розпізнавання мови.*