

etc.) while the presence of others is mandatory (for example, an external punctuation mark). We will need 5 types of pauses used in speech:

p_s - pause between syllables when pronouncing a word;

p_w - a pause between words when reading a sentence (corresponds to a space between words);

p_i - a pause marking the internal punctuation mark;

p_e - a pause marking the external punctuation mark;

p_p - a pause marking the end of a paragraph.

A conceptual model for synthesizing speech in the text in the form of a flowchart is shown in Figure 1. The synthesizer operates as follows. After entering the next sentence, it is analyzed by the composition of its elements. If the next element is a word, then in block 1 it is divided into syllables indicating the stressed syllable and then it is voiced using the syllable-sound base.

If the next element is a number, then it in block 2 is converted into text and then its sounding occurs through block 1.

If the next element is a symbol, then its scoring takes place in block 3 by extracting the corresponding sound from the "symbol-sound" base.

If the next element is a space, an internal or external punctuation mark, or the end of a paragraph mark, then the corresponding pause is extracted from the corresponding block.

The speech synthesizer, presented in the form of a block diagram, suggests that it is based on the principle of concatenation of voiced syllables.

Since the syllable acts as the main sound unit of speech, for the implementation of the synthesizer it is required to describe the variety of all syllables of the corresponding natural language.

Since each syllable, presented in the form of a chain of letters, needs its sound image, the creation of a "syllable-sound" base is required.

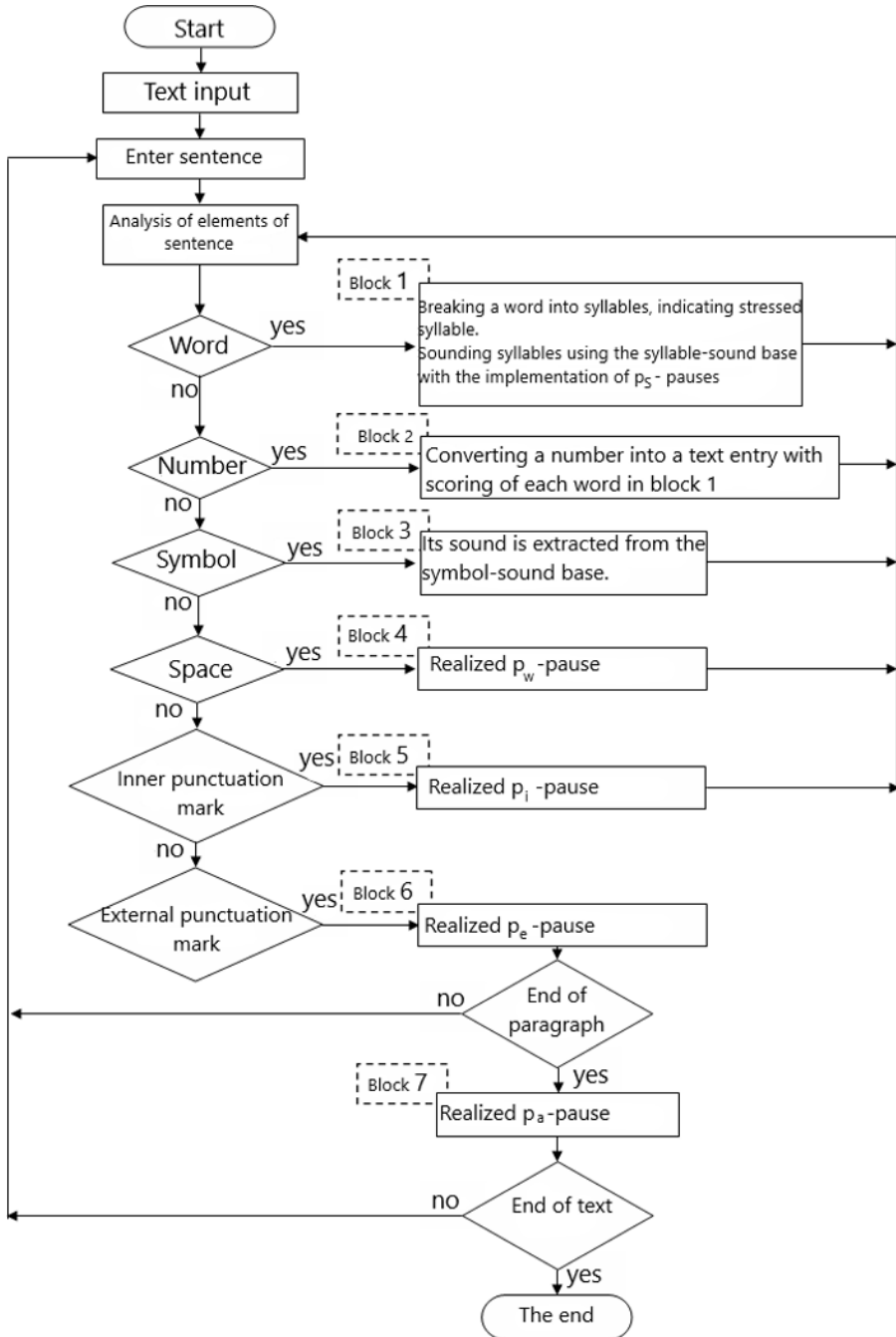


Fig. 1. Schematic diagram of the synthesis of speech in the text.

Since the synthesizer provides for the sounding of numbers and symbols, the corresponding algorithms and programs have been developed for the implementation of the synthesizer to transform the number into text and create a symbol-sound base.

And finally, you need to adjust the duration of the pauses p_s , p_w , p_i , p_e and p_a in such a way as to obtain, as far as possible, a natural and legible synthetic speech

3. Based on detailed elaborations of the conceptual scheme, the Tajik Text-to-Speech synthesizer was created, that is, a set of programs for synthesizing Tajik speech in the text. The synthesizer parameters were tuned by computational experiments. Satisfactory values of pause durations were established:

- for paragraph boundaries $p_a = 900$ ms,
- for offer boundaries $p_e = 600$ ms,
- for commas inside sentences $p_i = 400$ ms,
- for inter-word and inter-syllable pauses, accordingly, $p_w = 200$ ms и $p_s = 20$ ms.

To evaluate the synthesizer's performance, experiments were organized to voice a variety of textual information (fragments from novels, novels, scientific articles, textbooks, newspapers, magazines, Internet sites). The assessment of the completeness of the many syllables used to form synthetic speech was associated with the percentage of spoken words in relation to the total number of words within the selected text fragments. The results of the experiment showed quite satisfactory quality of the Tajik Text-to-Speech software package for scoring the Tajik text. The block diagram of the software package is presented in Figure 2.

In the first block, the "User Interface" consists of two components - "Text Entry" and "Speech", which have one-way communication, that is, the user has the opportunity to enter text information and as a result receive a speech version of the input text. To get the results, block 1 is connected with block 2 in two directions - to provide information for linguistic analysis and to obtain the results of scoring. Block 1 also interacts with block 3 directly to use the necessary data about the system settings (male or female voice selection, volume and speed of scoring).

The second block "Analytical subsystem" consists of two parts - "Linguistic analysis" and "Sound module". The first of them consists of the submodules "Text Validation", "Text Encoding" and "Separating Words into Syllables". "Text Validation" is used to validate input information, which includes text elements such as words, integers, characters, and punctuation marks. This submodule checks text elements, converts integers and characters into a test case, and then passes them for encoding. The coding process implements the submodule of the same name, which converts each word of the input text into an ordered set of zeros and ones, i.e. all words are represented by their syllabic structures. The encoded text is transmitted to the subdivision "Separation of words into syllables." Syllable words are linguistically analyzed and transmitted to the Sound Module. In this module, the formation of sound information occurs using the base "syllable-sound" of the information subsystem, stressed syllables, inter-syllable and inter-word pauses, as well as pauses marking such punctuation marks as a comma and period. The scoring module is the final stage of the analytical subsystem, and the audio version of the text information is sent to the user interface.

The third block, "Information Subsystem," contains databases called "System Settings" and "Syllable-Sound Base". The first of them is used to store temporary system setup data, the second "syllable-sound" base - to store statistical data on sound files of 3259 Tajik syllables. To work with this database, a module is used to provide access, check and select the necessary data.

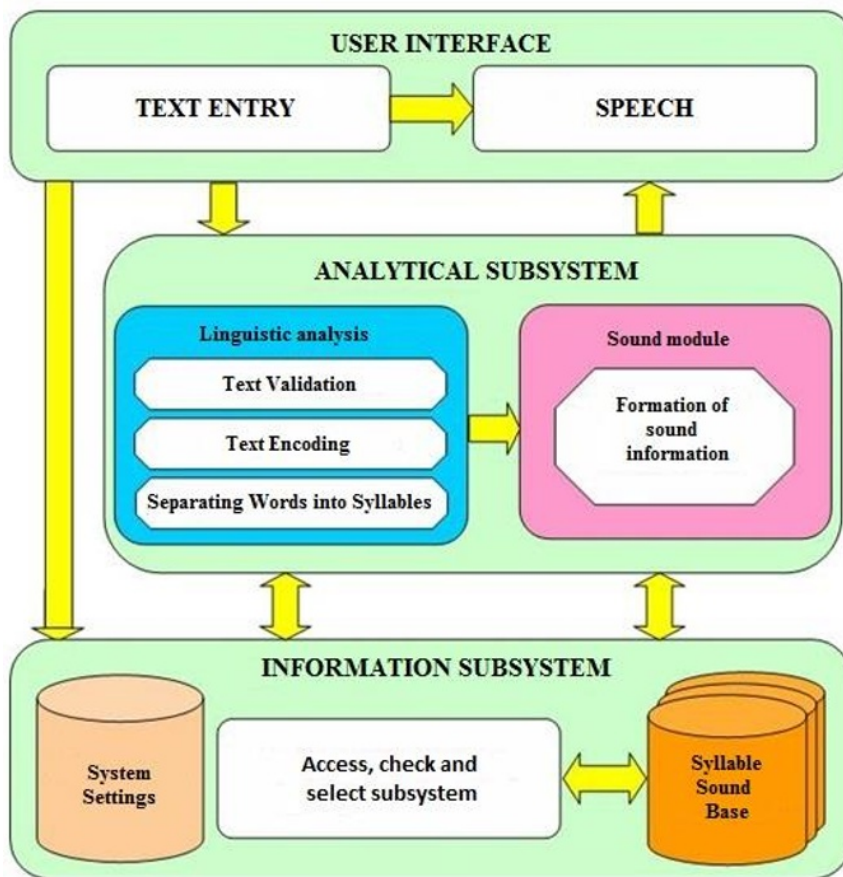


Fig. 2. Block diagram of Tajik Text-to-Speech.

6 Conclusion

Thus, the software package for computer dubbing of the Tajik text Tajik Text-to-Speech [24] and the announcer of the Tajik text Tajik Text-Narrator [25], although they do not completely solve the problem of synthesizing Tajik speech, are still the first software product, satisfactorily performing computer scoring of Tajik texts. At this level of development, the complex can now be used by people with impaired vision. The experiments were carried out at scientific seminars of the Khujand Polytechnic Institute of the Tajik Technical University named after Academician M.S. Osimi. Its participants, at their discretion, entered Tajik texts into the computer and then evaluated the naturalness and intelligibility of the sound of synthetic speech. The general opinion of the seminar - a computer synthesizer, built on the principle of concatenation of 3259 Tajik syllables, quite successfully performs the function of scoring Tajik texts. The synthesizer implements such elements of prosodic synthesis as the arrangement of stresses, taking into account the intonation pause between paragraphs, after the decimal point inside sentences and the point at the end of the sentence. Computational experiments have established the prospect of further development of the Tajik Text-to-Speech and Tajik Text-Narrator software systems into a Tajik speech synthesizer with Russian language.

References

1. J. Beskow, *Talking Heads - Communication, Articulation and animation*, Proceedings of Fonetik-96: pp. 53-56 (1996)
2. M. Cohen, D. Massaro, *Modelling Coarticulation in Synthetic Visual Speech*, Proceedings of Computer Animation 93, Suisse (1993)
3. B. Kröger, *Minimal Rules for Articulatory Speech Synthesis*, Proceedings of EUSIPCO92 (1), pp. 331-334 (1992)
4. J.L. Flanagan, *Speech Analysis, Synthesis, and Perception*, Springer-Verlag, Berlin-Heidelberg-New York (1972)
5. J.L. Flanagan, K. Ishizaka, and K.L. Shipley, *Synthesis of speech from a dynamic model of the vocal cords and vocal tract*, The Bell System Technical Journal, 54(3), pp. 485-506 (1975)
6. D.H. Klatt, *Synthesis by rule of segmental durations in English sentences*, In B.E.F. Lindblom and S. Ohman (Eds.), *Frontiers of Speech Communication Research*, pp. 287-299, Academic (1979)
7. D. Klatt, *Software for a Cascade/Parallel Formant Synthesizer*, Journal of the Acoustical Society of America, JASA, Vol. 67, pp. 971-995 (1980)
8. D.H. Klatt, *The Klattalk text-to-speech conversion system*, IEEE ICASSP-82, pp. 1589-1592 (1982)
9. D. Klatt, *Review of Text-to-Speech Conversion for English*, Journal of the Acoustical Society of America, JASA, **Vol. 82 (3)**, pp. 737-793 (1987)
10. D. Klatt, L. Klatt, *Analysis, Synthesis, and Perception of Voice Quality Variations Among Female and Male Listeners*, Journal of the Acoustical Society of America, JASA, **Vol. 87 (2)**, pp. 820-857 (1990)
11. C.M. Harris, *A study of the building blocks in speech*, Journal of the Acoustical Society of America, 25(5), pp. 962-969 (1953)
12. R.K. Potapova, *The main modern methods of analysis and speech synthesis*, Moscow (1971)
13. R.K. Potapova, *Syllabic phonetics of Germanic languages*, Moscow (1986)
14. Yu.I. Zhuravlev, *Selected scientific papers*, Master, Moscow (1998)
15. V.Ya. Chuchupal, K.A. Makovkin, A.V. Chichagov, *To the question of the optimal choice of the alphabet of the models of sounds of Russian speech for speech recognition*, Artificial Intelligence, **Vol. 4**, No. 1, Kiev, pp. 575-579 (2002)
16. V.N. Sorokin, I. S. Makarov, *The inverse problem for a voice source*, Information Processes, **Vol. 6**, No. 4, pp. 375-395 (2006)
17. V.M. Velichko, N.G. Zagoruyko, A.V. Kelmanov, S.A. Khamidullin et al, *Development of algorithmic, technical, and software recognition tools for isolated speech commands in a limited frequency range*, Report of Novosibirsk State University on research on the subject "M-91-86", state number, 01870014595, Novosibirsk, pp. 89 (1987)
18. L.V. Zlatoustova, S.V. Kodzasov, O.F. Krivnova, I.G. Frolova, *Algorithms for converting Russian spelling texts into phonetic recording*, Moscow, Moscow State University (1970)
19. A.A. Kharlamov, Yu.N. Zhigulevtsev, *Microprocessor-based tools for building embedded speech applications "Artificial Intelligence"*, No. 4 (2006)

20. B.M. Lobanov, *Analysis and synthesis of speech. Collection of scientific works*, Academy of Sciences of the BSSR Institute of Technical Cybernetics. Scientific Ed. B. M. Lobanov, Minsk, p. 86 (1991)
21. B.M. Lobanov, E.B. Karnevskaia, T.V. Levkovskaya, Text-to-speech synthesizer as a computer tool for “cloning” a personal voice, Tr. International Conference Dialogue-2001, Moscow, pp. 265-272 (2001)
22. B. M. Lobanov, L. I. Tsirulnik, D. V. Zhadinets, O. G. Sizonov, Algorithms for synthesizing prosodic characteristics of speech by text in the Multifon system, Joint Institute for Informatics Problems of the National Academy of Sciences of Belarus, Minsk (2007)
23. Z.D. Usmanov, H.A. Khudoyberdiev, *The experience of computer synthesis of Tajik speech in the text (scientific monograph)*, Irfon, Dushanbe, p.145 (2010)
24. Z.D. Usmanov, Kh.A. Khudoyberdiev, *Computer dubbing of Tajik text Tajik Text-to-Speech*, A patent (intellectual product) is registered by the National Patent Information Center of Ministry of Economic Development and Trade of the Republic of Tajikistan, 041TJ from 09/04/2007 (2007)
25. Z.D. Usmanov, H.A. Khudoyberdiev, A.A. Khudoyberdiev, *Announcer of the Tajik text Tajik Text-Narrator*, A patent (intellectual product) is registered by the National Patent Information Center of Ministry of Economic Development and Trade of the Republic of Tajikistan, No. 4201800386 from 07/10/2018 (2018)