

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Speech Research in TUSUR

Roman V. Meshchryakov

*Tomsk State University of Control Systems and
Radioelectronics (TUSUR)
Russia*

1. Introduction

In article results of researches in TUSUR are show. Researches proceed more than 20 years. The basic consumers of production are medical institutions

Researches in the field of speech technologies in TUSUR are spent more than 20 years. Among the basic directions definition of the announcer, speech synthesis on printing text are allocated the adaptive analysis of a speech signal for the decision of problems of the analysis and speech recognition. Important applied value is medical appendices for support of the doctor-logopedist, and also on carrying out of procedures of rehabilitation of vocal and speech functions at the person.

2. Processing of a speech signal

For processing of a speech signal the hierarchical model was used. The information is used from various levels of speech system hierarchy. It allows to use the information presented in various scales for the full-function analysis and synthesis of a speech signal.

Elements at each level are thus allocated, their properties, possible transformations are defined. It is besides defined that elements at the bottom levels of hierarchy of speech system under certain conditions form configurations which are elements of top levels.

Thus conditions of a regularity of the configurations defining calculations are formed. In turn at synthesis of a speech signal elements of top levels will be transformed to configurations of elements of the bottom levels of hierarchy of speech system.

2.1 Speech signal analysis

The acoustic model of speech perception system has been developed for carrying out of researches of a speech signal. Its maintenance is represented by original mathematical model and algorithms of processing. Masking is simultaneous allows to eliminate redundancy in the speech signal, connected with features of delay of reaction sensor's systems after peaks in frequency area. Frequency masking allows to eliminate surplus information on a frequency spectrum which isn't perceived by the person since it disappears behind other frequencies.

In a complex of simultaneous and frequency masking the volume of the processed information on a speech signal at what the information isn't lost is considerably reduced, and its quality only improves, raises it informative.

In fig. 1 results on research of a site of a speech signal are presented. At the left results FFT, on the right – are presented at use of the developed model. It allows to spend segmentation of a speech signal on vocal and non-vocal sequence. They are used for the parametrical description of a speech signal. Besides, on vocal sequence fundamental frequency with split-hair accuracy in comparison with other algorithms is defined.

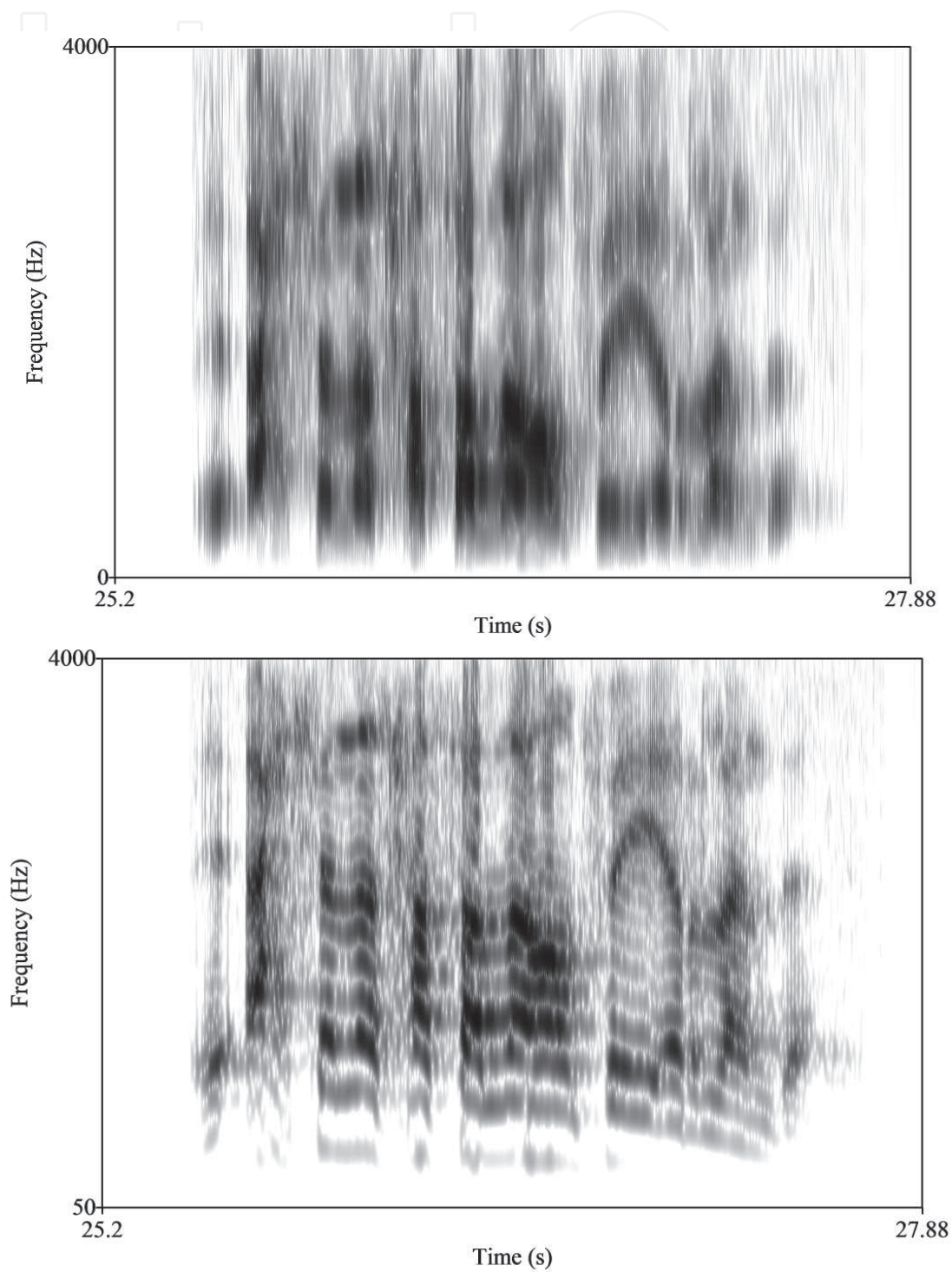


Fig. 1.

2.2 Medical research

Medical researches go in a direction of quality improvement of diagnosis statement and times reduction of patients rehabilitation. The first developed information system shows to the doctor-logopedist of speech signal feature (on a spectrum) and prompts infringements of speech formation bodies of patient.

The second component allows to form with use of biological feedback esophagus voice. To patients cut throat and form artificial vocal folds in a gullet. In fig 2 anatomic features after removal of a throat and to formation of a new vocal path are presented

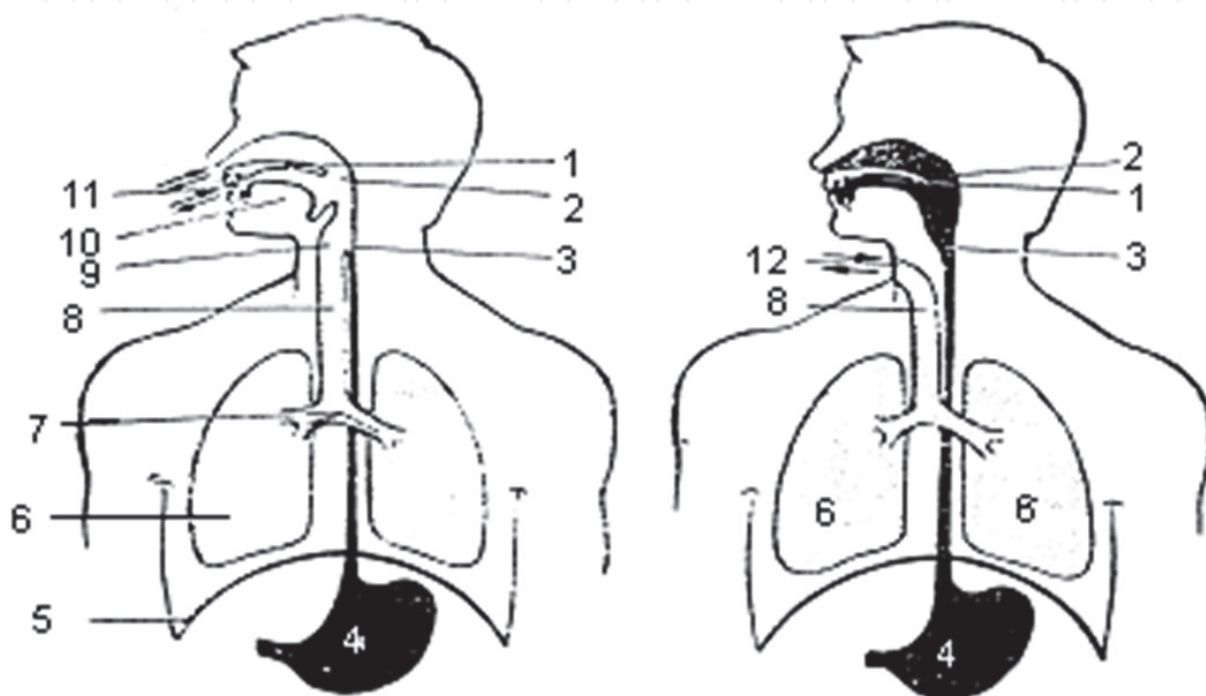


Fig. 2.

In fig. 3 basic elements of man-machine system are presented. In it the logopedist forms the task for the patient. The patient spends trainings on formation or restoration of vocal function about use of a hardware-software complex. The hardware includes a microphone with which help the data and the monitor on which process of interaction with the patient and its results is displayed are entered. Program and algorithmic maintenance forms contains mathematical model and functions of processing of a speech signal, interaction with the patient, and as history of speech rehabilitation.

Using biocontrol at first it is formed phonation, then the patient is trained to operate vocal folds (stability of fundamental frequency, duration phonation, etc.).

In figure 4 the photo from scientific research institute of oncology of Russian Academy of Medical Science on which base medical researchers are conducted is presented. Active participation is accepted by professor Balatskaja L.N., professor Chojnzonov L.T., logopedist Krasavina E.A. In TUSUR Bondarenko V. P, Kornilov A.J., Kotsubinsky V.P. were engaged and engaged in Konev A.A., Kostjuchenko E.J., Kvasov A.N. in the given problem

In figure 5 results of change of frequency of the basic tone at stages of vocal rehabilitation are presented (a) - Fundamental frequency, b) - Deviation fundamental frequency, c) - Sound duration [a]).

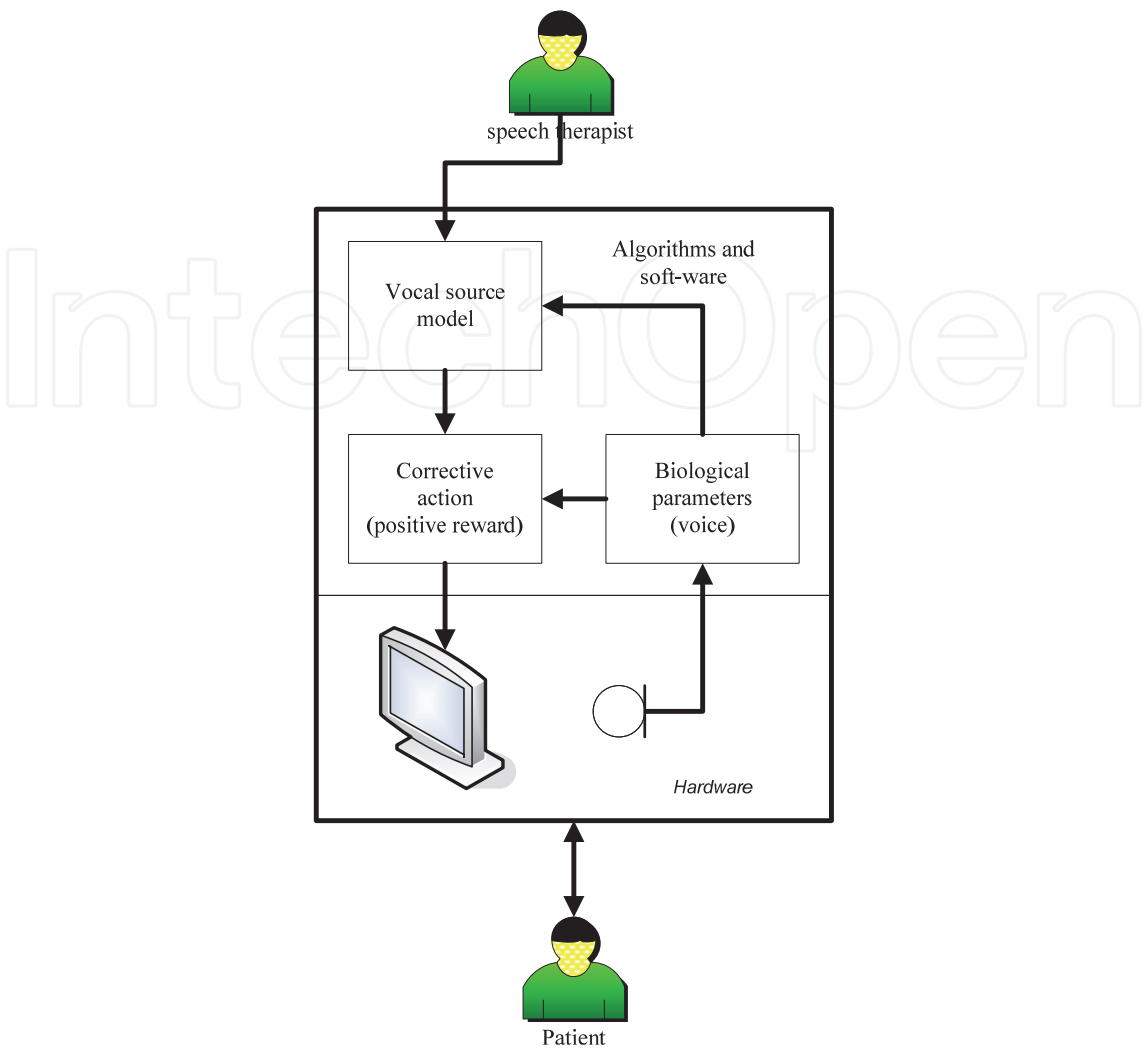


Fig. 3.



Fig. 4.

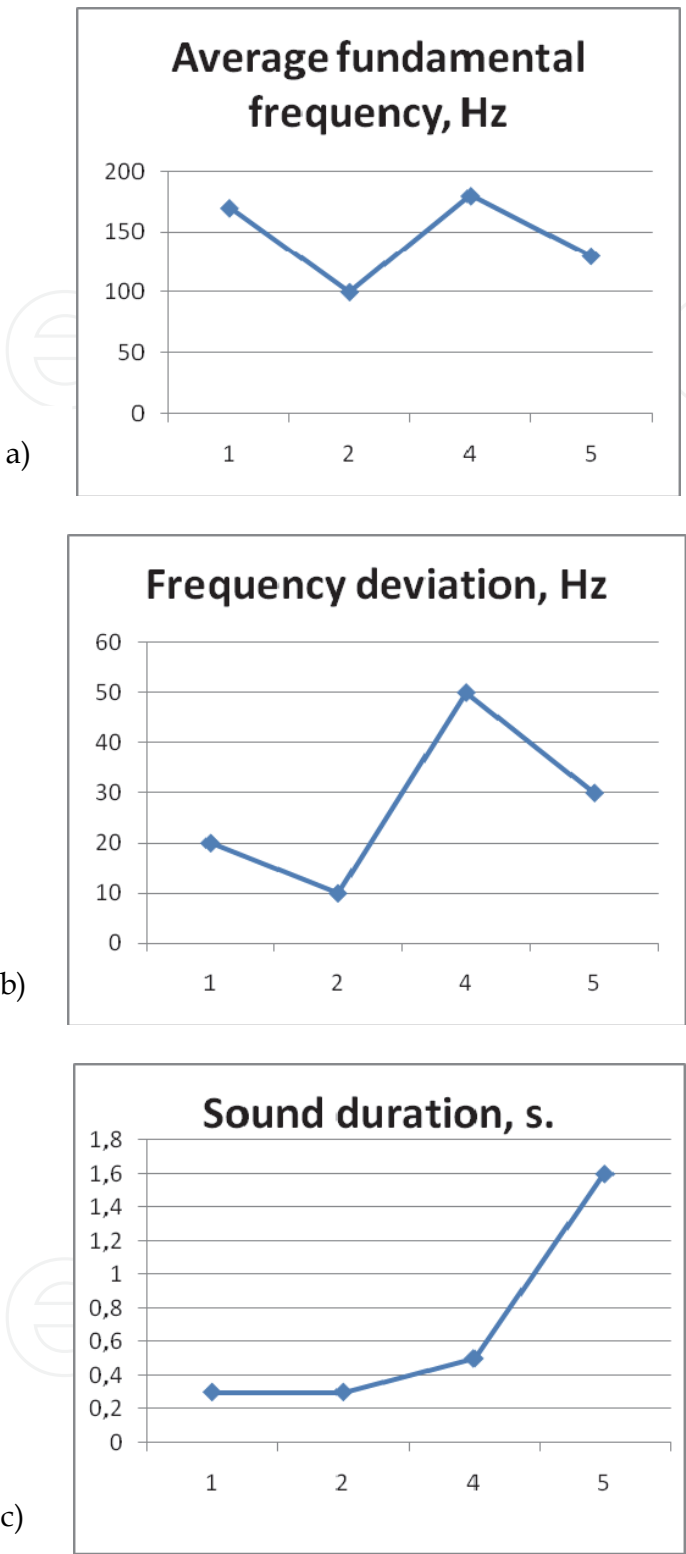


Fig. 5.

It is obvious that the base investigated characteristic is fundamental frequency. Measurements at various stages logorestore experts were spent:

1. prior to the beginning of complex rehabilitation;
2. in the beginning of complex rehabilitation;

3. after the termination of vocal rehabilitation;
4. in 6 months after vocal rehabilitation;
5. in 1 year after vocal rehabilitation.

On a figure 5a) are resulted for patient *M* change of fundamental frequency on stages, on a figure of 2 deviations of fundamental frequency from average value and on a figure 3 duration of a said sound [a]. In the resulted case the most widespread kinds of changes of investigated characteristics are reflected. The patient learns to say more long phrases and in wide range to fundamental frequency control. It is obvious that use of new methods of the speech signal analysis is expedient for using at tracing of positive dynamics of vocal rehabilitation. At various stages logorestore therapies at the person are formed the new mechanism of speech formation. However, as already it has been noted above, at throat removal vocal folds speech formation a path of the person leave. At vocal rehabilitation in the first physiological narrowing of a gullet the pseudo-vocal crack which in the subsequent to be analogue of vocal folds of a throat is formed. One of the most important characteristics which are subject to research is the fundamental frequency defining frequency of fluctuations of vocal folds and defining voice quality.

Efficiency of vocal function restoration has made 92,6 % in terms from 8 days till 22 days. The estimation of application efficiency of the developed technique is proved by objective methods of functional researches: acoustic, electro-miography and studying of life quality.

Acoustic parameters esophagus voices remain in a database after each vocal training. The detailed analysis of spectral components of a formed voice is carried out and dynamics of the basic indicators is traced. Duration phonation, depending on quantity of trainings has on the average increased with 80 ms to 850 ms. Dynamics of fundamental frequency by results of trainings has made from 40 Hz to 120 Hz.

2.3 Speech synthesis

For synthesis of a speech signal the system model speech formation the person is used. Modeling of formation of the message by the person and control of speech formation bodies is spent. Processes of a breath, an exhalation, phonation, fluctuations of vocal folds, changes of articulation bodies are modeled.

As base the sound system of Russian language is accepted and process of transformation of the information from pragmatically level on level of a physical signal is modeled. Each of the models used at speech synthesis imposes the restrictions. It is caused by that from the entrance text elements of various levels are allocated. Further in system there is a generation of elements of new type and their configuration. For intonation realization it is necessary to generate predicted change of a prosody, and then to operate process of its updating.

All external and internal data used for formation of speech synthesis is reflected in a Fig. 6. In blocks of the left part all data on which basis configurations are under construction are allocated and signs are defined. In an average part there is an information on a stage of transformation and result of its performance. Blocks of the right part are rules of language and enter directly into blocks of the left and central part in the form of tables, rules, conditions of a regularity, algorithms of processing. The given scheme is used as methodological bases of systems of synthesis of speech.

As the given model describes predicted parameters of model of generation of a speech signal, in the resulted scheme there are no some feedback. They are added at a stage of direct management by lungs and speech formation a path in the course of generation of a speech signal.

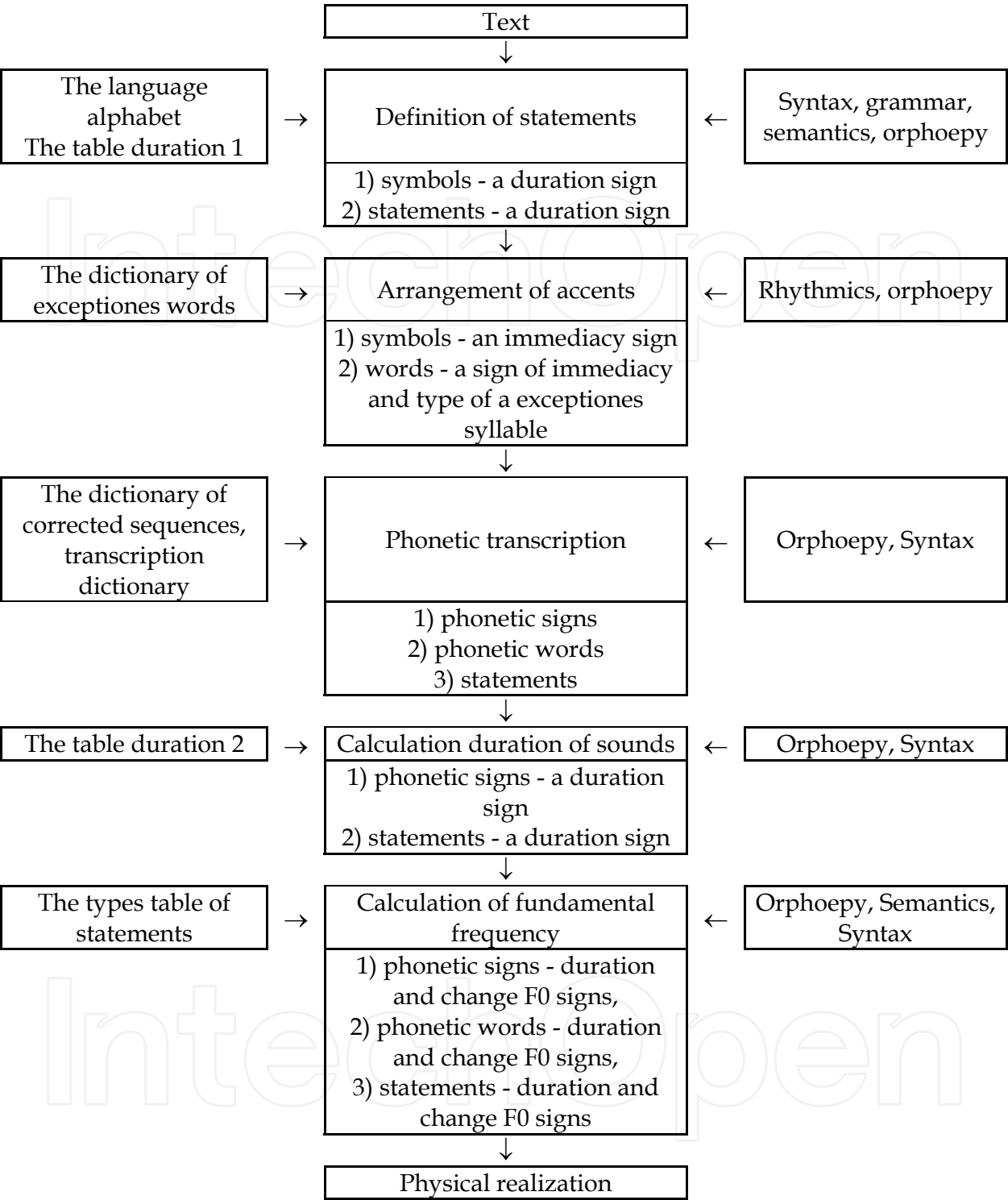


Fig. 6.

On the basis of the scheme resulted in figure 6 and the received practical results it is possible to make recommendations about the organization of knowledge bases for language systems:

1. Any speech system should be preliminary considered for revealing of the most informative blocks influencing result. By results of the analysis the criteria differentiating the information on significance values should be created. In particular, it is necessary to use criterion of efficiency of the transferred information.

2. All information necessary for transformation, it is expedient to divide into base data (tables and dictionaries) and rules according to which transformation is made.
3. For the information which can't be formalized it is presented in the form of rules dictionaries are used. The size of dictionaries is limited to demanded quality and admissible quantity of objects. In the given model 3 dictionaries and 5 sets of the rules considering both language, and parameters of the announcer are used.
4. For the information unequivocally defined by object it is expedient to use tables, in conformity with which on demanded object its sign is defined. For example, the table duration.
5. The general parameters of a speech signal should be based on physiological parameters speech formation systems of the person. It allows to adjust system of synthesis of speech on the announcer and to receive natural speech.
6. Introduction of criteria of efficiency is necessary both at the various intermediate stages, and on resultant. Their task can be as in obvious, and implicitly in the form of restrictions. Thus, we receive a control system of speech signal reception.

3. Conclusion

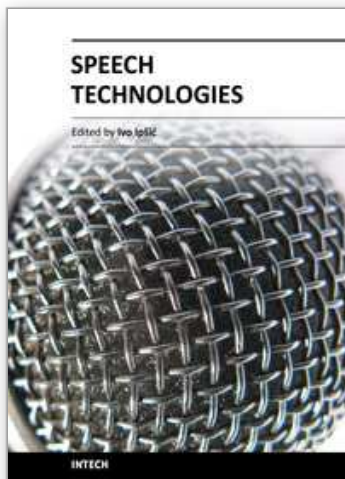
Now there is a development in the field of realization of effective algorithms. Researches on cases of various languages are conducted.

4. Acknowledgment

The collective long time was headed by professor Vladimir Bondarenko. Two students: Evgenie Kostjuchenko and Evgenie Shchipunov have received grants ASA (2008, 2010).

5. References

- Meshcheryakov, Roman & Bondarenko Vladimir* (2008) Dialogue as a basis for construction of speech systems. *Cybernetics and Systems Analysis*, Vol.44, No.2 (March 2008), pp. 175-184, ISSN: 1060-0396
- Meshcheryakov, Roman, Ronzhin Andrey, Karpov Alexey, Zelezny Milos, Hoffmann Ruediger* Development of Multimodal Applications for Disabled People. Proc. of the Eighth All-Ukrainian International Conference, Kyiv, Ukraine, August 28–31, 2006, pp. 163-166.
- Bondarenko V.P., Meshcheryakov R.V., Konev A.A.* Biologic feedback formation by vocal rehabilitation Proceedings of the International Workshop SPEECH and COMPUTER (SPECOM'2006) St.Peterburg, Russia 25-29 June 2006 pp 251-257
- Meshcheryakov R.V., Bondarenko V.P., Kotsubinsky V.P.* Peculiarities of vocal sounds generational speech synthesis by rules Proceedings of the Ninth International Conference "SPEECH and COMPUTER" (SPECOM'2004) Saint-Petersburg, Russia 20-22 September 2004, 726p., pp 575-577
- Taylor Paul*, Text-to-Speech Synthesis: Cambridge University Press, 2009. – 597p.



Speech Technologies

Edited by Prof. Ivo Ipsic

ISBN 978-953-307-996-7

Hard cover, 432 pages

Publisher InTech

Published online 23, June, 2011

Published in print edition June, 2011

This book addresses different aspects of the research field and a wide range of topics in speech signal processing, speech recognition and language processing. The chapters are divided in three different sections: Speech Signal Modeling, Speech Recognition and Applications. The chapters in the first section cover some essential topics in speech signal processing used for building speech recognition as well as for speech synthesis systems: speech feature enhancement, speech feature vector dimensionality reduction, segmentation of speech frames into phonetic segments. The chapters of the second part cover speech recognition methods and techniques used to read speech from various speech databases and broadcast news recognition for English and non-English languages. The third section of the book presents various speech technology applications used for body conducted speech recognition, hearing impairment, multimodal interfaces and facial expression recognition.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Roman V. Meshchryakov (2011). Speech Research in TUSUR, Speech Technologies, Prof. Ivo Ipsic (Ed.), ISBN: 978-953-307-996-7, InTech, Available from: <http://www.intechopen.com/books/speech-technologies/speech-research-in-tusur>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen