HISTORY AND DEVELOPMENT OF SPEECH SCIENCE AND TECHNOLOGY IN USSR

Boris Lobanov

United Institut of Informatics Problems NAS Belarus
lobanov@newman.bas-net.by

Abstract: In this article, the history and development of speech science and technology in USSR are described. The prehistory of speech research is summarized, and the organization and the main directions of research in the USSR since the 60's are described. The review of the speech research activity in Soviet Belarus is given and the main achievements in the field of speech synthesis are described.

1 Introduction

In 1991, as the world watched in amazement, the Soviet Union disintegrated into fifteen separate countries. On August 19 of 1991 a group of “hard-line” Communists organized a coup d’état. After three days of massive protest, the coup organizers surrendered, realizing that without cooperation with the military, they did not have enough forces to overcome the power of the entire population of the country. These events were anxiously watched by hundreds of participants of the Congress of phonetic sciences (including the author of this article) which was held at this time in Aix-en-Provence, France.

In the Soviet Union, science and technology was an important part of national policy, practice and identity. Moreover, both science and technology were closely associated with the ideology and practical work of the Soviet state. This was one of the main reasons for the lack of links with the world of science. Soviet scientists had very limited access to foreign literature. They had no practice in foreign languages and rarely participated in international conferences. Their scientific works were usually published in Russian, and therefore were almost inaccessible to foreign scholars. However, there were a large number of scientific areas in which Soviet scientists succeeded. This also applies to speech science and technologies. The purpose of this article is to open slightly the "Iron Curtain" over the history and development of speech science and technologies in the USSR.

2 Prehistory *

The world’s earliest attempts to produce synthetic speech were made over two hundred years ago [1]. During the reign of Catherine II St. Petersburg Academy of Sciences announced a competition to create a talking machine. In 1779, Russian Professor Christian Kratzenstein explained physiological differences between five Russian vowels (/a/, /e/, /i/, /o/, and /u/) and made an apparatus to produce them artificially. He constructed acoustic resonators similar to the human vocal tract and activated the resonators by vibrating reeds like in music instruments. The basic structure of resonators is shown in Figure 1. The sound /i/ is produced by blowing into the lower pipe without a reed causing a flute-like sound.

---

*) The material of this section is partly based on the Internet sources [2]

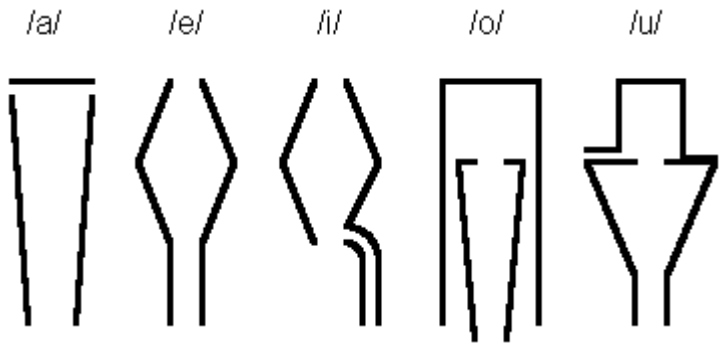
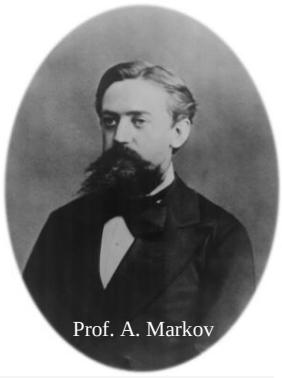


Figure 1 – Kratzenstein's resonators [2].

It is worth to note also distinguished contribution of Russian mathematician Andrey Markov (1856 – 1922). He developed the theory of stochastic processes [2], which is now called a Markov chain. Nowadays a hidden Markov model, that is a Markov chain for which the state is only partially observable, is the most effective instrument for speech recognition and synthesis.


Prof. A. Markov

The next notable attempt to synthesize speech in Russia already belongs to the 30s of XX century, and is associated with the development of the sound film and electronic music. In Moscow the electronic music studio engineer Scriabin Museum EA Sholpo decided that soundtracks could be created artificially. He “painted” sound waves on a specially designed large scale, photographing them frame by frame, and played them through a finished film projector. Although the work was very time-consuming and inefficient, Sholpo voiced a few cartoons in this way (by the device that he had created - variafon).

Murzin who knew of Sholpo, as they say, chose another way. He chose a method of speech synthesis using a Fourier series - which is a sum of elementary spectral components in musical acoustics known as a "pure tone." Bank "pure tones" Murzin designed as a glass disc that is very similar to a modern CD. It was covered with a photographic emulsion, and using a special machine, it was recorded in concentric rings 144 photo-optical sound tracks "pure tone". The process of sound synthesis is shown in Fig. 2.

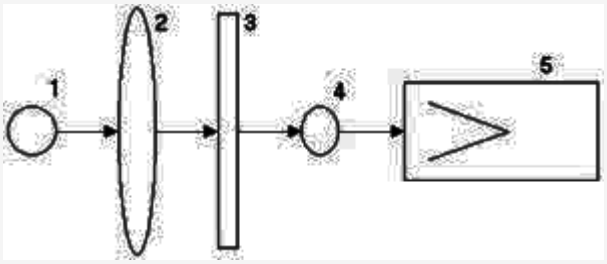


Figure 2 – Functionally-optical circuit synthesizer ANS

The light from the source (1) passes through a rotating disk photo-optical oscillator (2) and intensity-modulated audio tracks. Between the disk and photocell (4) there is a mask (3) with holes for the selection of rays with desired tracks only. The photocell is followed by normal path amplification for projectors (5).

In the early 40's in the Soviet Union the first papers on speech coding and recognition emerged. In 1941-1942 in blockaded Leningrad L. L. Myasnikov completed his doctoral

thesis on work, related to the system of recognition of isolated spoken sounds - all the vowels and some consonants. In 1943, in the journal "Successes of physical sciences" he published an article, which for the first time described a device for Russian speech phonemes recognition. The works of Myasnikov outpaced the first foreign works on speech recognition for 10 years.

Immediately after the end of the Second World War, the major investments in this area of science were made directly by Stalin. In Moscow, the organization "RI-100" (the "sharashka") was established, where a large group of mathematicians, physicists and engineers were solving the problems of coding of the speech signal, the voice identification and recognition of keywords in speech. This period of speech research is well described by Solzhenitsyn's novel "The First Circle". In the late 60's the author of this article had the luck to cooperate with several prototypes of the heroes of this novel.

In the early postwar years, the speech research was conducted mainly in the interests of multiplexing communication. To address these issues at Leningrad Institute of Telecommunications a strong team of scientists led by L.A.Varshavsky and V.S.Fedorovich was created. Either within or in close collaboration with the same team many future leaders of the Soviet speech scientists began their research works. Among them are L.A. Chistovich, L.V. Bondarko, V.I. Galunov, M.F. Derkach et al.

3 The review of the USSR speech research activity from the 1960's*

A sharp increase of interest in the problem of speech recognition and synthesis both in the USSR and abroad occurred in the early 60's, with the development of computer technology. Computer companies realized that in the long run the machine would include a variety of means of computer-human communication in its most convenient form - a speech dialogue. To solve this problem was a very tempting goal, and it seemed not very difficult to achieve.

The illusion of an easy solution was for all those who were not familiar with the problem. After a report on modest results of the studies, one famous Soviet cybernetician who was not engaged in a speech signal field of science, said, "The fact that our speech researchers can not teach a machine to recognize it is a misunderstanding. Mankind has used speech for millions of years, and they cannot understand how it is produced. Need to talk with familiar German Acousticians, they certainly can do it".

Under the influence of the public order and illusion of problems easiness hundreds of research groups around the world began to work with speech signals. In the 1970s, there were more than 150 teams in the USSR. Until the late 60's, when the main results related to the speech of psychoacoustics and phonetics were made, the Soviet scientific school had a leading position in the world.

Many Soviet researchers started acquaintance with the problem on the book by Sapozhkov MA "The speech signal in cybernetics and communication" (1963). Outstanding contribution to the development of speech researches in the USSR was made by the scientists from Pavlov's Institute of Physiology, Academy of Sciences of the USSR (L. Chistovich, V. Kozhevnikov, V. Lublinskaya, et al.). Their collective monograph "The Speech. Perception and production "(1968) was a handbook for anyone who studied physiology and psychoacoustics of speech.

Phonetic and linguistic problems of speech communication were fruitfully studied by scientific groups of Leningrad, Moscow, and Odessa universities (L.Bondarko, L. Verbitskaya, L. Zlatoustova, R. Potapova, T. Brovchenko, E. Nushikyan, etc.).

---

*) The material of this section is partly based on the article [3]

The development of various aspects of speech technology was conducted in many organizations. The groups with the most members in this area were:

- Novosibirsk Institute of Mathematics and Novosibirsk State University (N. Zagoruiko, G. Voloshin, V. Velichko, A. Kel'manov, etc.),
- Kiev Institute of Cybernetics (T. Vintsyuk, E. Ludovik, V. Bogino, etc.),
- Minsk Institute of Technical Cybernetics (B. Lobanov, N. Degtyarev, B. Panchenko, etc.),
- Moscow Computing Center of the USSR Ac. of Sc. (V. Trunin-Donskoy, V.Chuchupal, etc.),
- Moscow Institute for Information Transmission Problems (I. Turbovich, G.Tsemmel, V. Sorokin, A. Knipper, etc.),
- Moscow Institute of Radio Communications (A. Pirogov, G. Slucker, etc.).

Coordination of the work carried out through official channels by the Council of speech recognition and synthesis of the USSR Academy of Sciences Presidium (Academician Yury Zhuravlev as chair). However, the most important role was played by an unofficial body, which was, – the All-Union Seminar on the problem of automatic recognition of auditory images (in Russian abbreviation - ARSO), which was convened annually, and then every two years between 1965 and 1992.

The idea of the ARSO was born during a narrow workshop in 1963 in Novosibirsk, attended by Bondarko, Golubtsov, Zagoruiko, Kozhevnikov and Chistovich, representing the group of mathematicians, engineers, linguists and psychologists. It was decided to organize a school-seminar, where the "school" section included leading specialists in different aspects of the problem who would make summary reports and give educational lectures, while the "Seminar" section implied the participants who would report on results of their latest researches.

Various speech research centers (see fig. 3, 4) took part in the ARSO organization in different years, namely:

ARSO-1 (Novosibirsk, Russia, 1965),
ARSO-2 (Trakaj, Lithuania, 1966),
ARSO-3 (Novosibirsk, Russia, 1967),
ARSO-4 (Kiev - Kanev, Ukraine, 1968),
ARSO-5 (Sukhumi, Georgia, 1969),
ARSO-6 (Tallinn, Estonia, 1971),
ARSO-7 (Alma-Ata, Kazakhstan, 1973),
ARSO-8 (Lvov, Ukraine, 1974),
ARSO-9 (Minsk, Belarus, 1976),
ARSO-10 (Tbilisi, Georgia, 1978),
ARSO-11 (Erevan, Armenia, 1980),
ARSO-12 (Odessa, Ukraine, 1982),
ARSO-13 (Novosibirsk, Russia, 1984),
ARSO-14 (Kaunas, Lithuania, 1986),
ARSO-15 (Tallinn, Estonia, 1989),
ARSO-16 (Suzdal, Russia, 1991),
ARSO-17 (Izhevsk, Russia, 1992).

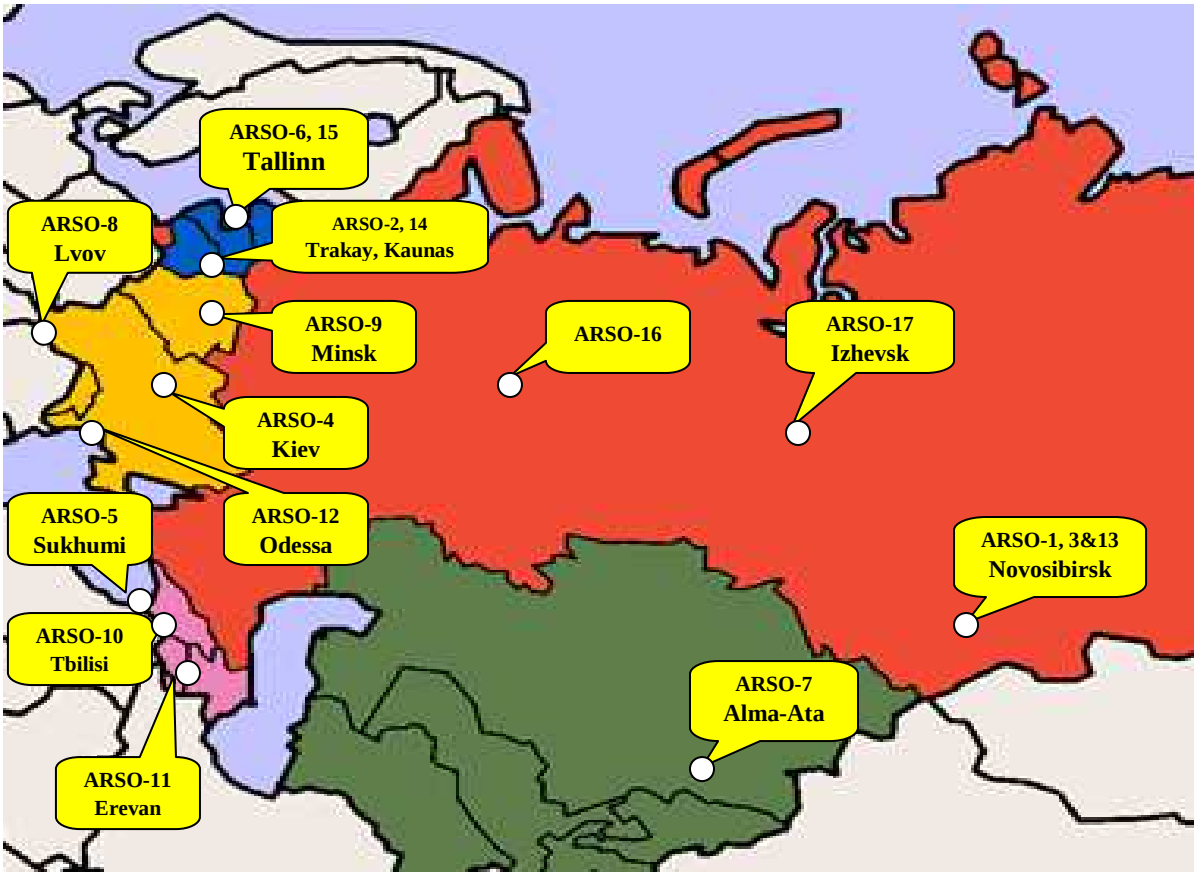


Figure 3 – Map of the ARSO (from 1965 to 1992)

Figure 4 – Members of the ARSO-9 program committee in Minsk
(From left to right: Boris Lobanov, Taras Vintsuk, Valentina Ignatova and Valery Trunin-Donskoy)

Many scholars from different cities participated in the ARSO. For example, on ARSO-9 (Minsk, Belarus, 1976) the total number of registered participants was 239. They represented 73 organizations from 30 cities of the USSR, including:

- Moscow - 74
- Leningrad - 37
- Minsk - 33
- Kiev - 17
- Lvov - 6;

- Penza, Tomsk, Dnepropetrovsk – at 5 persons (15 in total);
- Novosibirsk, Tbilisi, Tallinn, Izhevsk - at 4 person (of 20);
- Kaunas, Odessa, Yerevan, Kuibyshev, Yaroslavl, Poznan (Poland) - at 3 people (18 in total);
- Vilnius, Gorky, Kharkov, Taganrog, Ryazan - at 2 people (10 in total);
- Kazan, Zhitomir, Riga, Frunze, Severomorsk, Aralyk, Warsaw (Poland) - at 1 person (7 in total).

Among the ARSO-9 participants were 15 DrSc (7%) and 75 PhD (32%). 20% of the ARSO-9 present were female participants, and 70% of them had linguistic education.

After disintegration of the USSR the unique ARSO (ARSO-17, Izhevsk, Russia, 1992) was held with only a few dozen participants. For comparison, during the heyday of ARSO-13 in 1984, in Novosibirsk about 800 participants gathered, which was comparable in size to such an international conference as the 1st Eurospeech 1989 - Paris, France. At this ARSO 291 reports at 14 sections were submitted, namely:

1. Phonetics and prosody - 42
2. Speech production and perception - 21
3. Models of speech dialogue systems - 27
4. Models of speech signals preprocessing - 7
5. Vocoders - 15
6. Phonemic features extraction- 32
7. Segmentation of speech - 16
8. Speech Synthesis - 25
9. Speaker identification and verification - 12
10. Recognition of speech commands - 28
11. Continuous speech and keywords recognition – 21
12. Hardware for speech signals analysis – 8
13. Systems for the speech study - 16
14. Application of speech systems - 21

This section list shows that in spite of the minimal contacts with Western scientists because of the "Iron Curtain", Soviet scientists developed a very wide range of speech communications problems, and they obtained many original scientific results. For example, G. Slutsker [4] and T. Vintsuk [5] in 1968 developed independently the DTW-algorithm of speech recognition, which later became widely used in the world. Based on research results obtained in the 80's industrial samples of speech recognition systems, TTS-synthesizers and voice verification systems were created.

It was a romantic period of search for new scientific results in the study of speech. The ARSO was the central event in the speech community, a regular showcase of the latest achievements. In contrast to large conferences with diverse subjects, which every time gathered new people, the ARSO maintained high stability of key players. "ARSO-members" knew each other well, and it excluded the possibility of false advertising of questionable results. Both developers and customers of speech systems were able to make objective judgments about the team's work level and abilities. After discussions, ARSO-members could adjust the direction of their researches.

Essentially, the APCO played an important role of the informal but effective All-Union Center for the coordination of scientific works in the field of speech researches and technologies.

4 The review of the speech research activity in Soviet Belarus

4.1 The initial stage of speech researches

The first group of speech researchers in Belarus was organized by the author of this article in 1965 in Minsk Radioengineering Institute. At that time it included also N. Degtyarev , B. Panchenko, M. Fateev, etc., which for a long time worked, and some of them still work in this direction.

The group's early studies were connected with working out of general principles of a speech signals analysis and allocation of informative features that would perform a continuous speech signal as a sequence of phonetic segments. The results of these studies were summarized in the author's PhD-dissertation, "Some questions of the analysis of speech signals", defended in 1968 at Moscow Institute of Radio. The most important results of this work were later published in prestigious international journals [6, 7].

These researches became the basement for the development of a rather simple device for speech commands recognition « SESAME – 2*» was developed for the first time in the USSR. In 1968 it was awarded by a silver medal of the All-Union Exhibition in Moscow.

The device consisted of two parts: the analyzer of speech signal features, such as "voice", "noisy", "vowel", etc., and the decision-making unit on relative positioning and a number of features in the speech command. Sufficiently high reliability of recognition of 20 words (including the names of numbers) was achieved, regardless a speaker's voice , volume and tempo of pronunciation.

During the same period, some specialized devices were developed for experimentally-phonetic researches of speech: the Spectrograph and Intonograph. Thanks to them during the next years, numerous researches were conducted in phonetic laboratories of the Institute of Linguistics of the National Academy of Sciences of Belarus and Minsk Institute of Foreign Languages.

Studies of dynamic spectra of speech gave rise to the development of nonlinear methods for comparing spoken words with their patterns (DTW-method) [5, 6]. In 1969 further development of the DTW method for a very important practical case when borders of recognized words are unknown was proposed [8]. That is to solve the problem of detection and recognition of sound combinations in a continuous speech signal.

The beginning of 1970s is marked by the start of work on the creation of speech synthesizers for the Russian language. An important role in the development of Russian speech synthesis was played of the author's visiting in 1970 the laboratory of Professor Lorenz (Edinburgh University), where one of the first models of formant synthesis of speech signals was developed. It was the formant synthesizer that allowed to obtain by the author high-quality synthesized samples of Russian speech for the first time in history.

---

*) The previous model – «SESAME - 1" – was designed for the purpose of opening the lab door with a voice.

The first, though not a fully advanced model of the Russian text-to-speech synthesizer, "PhonemePhone-1" (see Figure 5) started speaking in the early 70's. Its successful creation is primarily associated with the development of new methods of hardware implementation of formant synthesis of speech signals. As the result of experimental studies a complete set of phonemes' formant "portraits" was created, which allowed for the first time to create a system of Russian speech synthesis for any text. Later on, an improved version of the synthesizer - "PhonemePhone-2" was constructed. It contained more detailed rules of "phoneme - allophone" transformation.

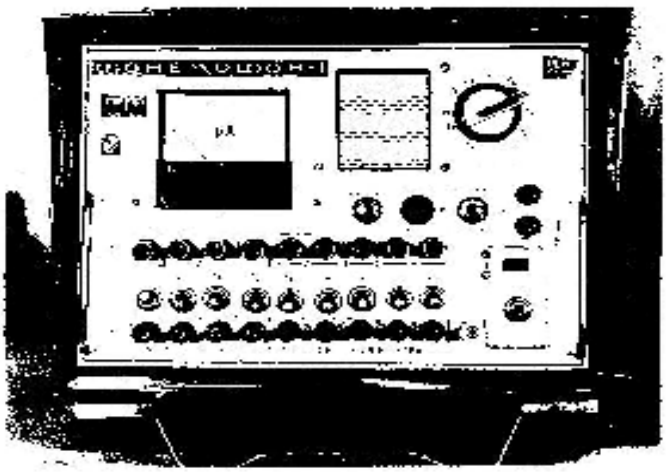


Figure 5 – Synthesizer "PhonemePhone-1"

4.2 Speech Communication Laboratory (1975 - 1987)

In December 1974 at the Minsk Branch of the Institute of Communications (MONIIS) the Speech Communication Laboratory was established, which played a crucial role in the guidance for developing applications for speech technology in Belarus. At the APCO-9, held in 1976 in Minsk, the first prototype of an automated telephone information service with synthesized voice responses was demonstrated . In the late 70's this service was used for a long time in Minsk as a system of automatic telephone call-around of debtors in long. By the mid 80's, this system was implemented in many cities of the USSR - from Brest in Belarus to Kamchatka and Sakhalin in Russia.

Figure 6 – Synthesizer "PhonemePhone-3" in Geneva

However, successful implementation of speech synthesizers in real conditions demanded still much work towards further improvements of the technology itself and the quality of the synthesized speech. The main disadvantage of the first synthesis model" PhonemePhone - 1, 2" was a poor quality of the synthesized speech, due to the usage of the most simplified models of sounds interaction within speech production (effects of coarticulation and reduction), as well as unsophisticated speech intonation models. In the next version - "PhonemePhone - 3" the additional units of articulation and speech intonation modeling were introduced, which significantly increased the quality of the synthesized speech.

In 1979, the "PhonemePhone-3" was exhibited at the World's Fair "Telecom-79" in Geneva (see Figure 6).

Arthur C. Clarke, the famous science fiction writer , after visiting the USSR pavilion and listening to the speech synthesizer, wrote in a book review, "*You*

anticipated my fantasies from the "Space Odyssey - 2001" movie. The Swiss newspaper “Observer” published an article with the following title: "Now Russians study foreign languages with a computer that speaks".

The next landmark in the history of industrial hardware was the creation of a new digital model of a speech synthesizer and recognizer. It became the basement for the speech synthesizer "PhonemePhone - 4" and the voice terminal - "Mars" (Fig. 7), which were constructed for the first time in the USSR in 1983 It became possible due to the efforts of the Design Department, led by V. Afonasev and laboratory staff members N. Degtyarev and V. Shaternik. In the voice terminal "Mars" the functions of speech recognition and synthesis were first integrated. The algorithm of speech recognition was based on the decision-making DTW method with a set of formant parameters of the speech signal. Designs of the "Mars" were performed on the microprocessor. Original technical solutions used in the creation of systems "PhonemePhone" and "Mars" were protected by copyright USSR certificates for inventions.

Figure 7. – Voice terminal - "Mars"

The beginning of 1984 is notable for the final formulation, theoretical and experimental development of a common lingvo-acoustic approach to the problem of text-to-speech synthesis, its implementation in the form of technical systems and practical application in automated models of management and communications. The results of these studies were summarized in the author's doctoral dissertation, "A method of automatic text-to-speech synthesis for Russian ", defended in 1984 at the Institute of Electronics and Computer Science, Academy of Sciences of the Latvian SSR. Later, the results were adapted for speech synthesis systems for other European languages. In particular, by 1987 thanks to the collaboration with E. Karnevskaya, a professor of Minsk Institute of Foreign Languages, the English version of the synthesizer [9] had been developed and then demonstrated at the World Congress of Phonetic Sciences where it was highly appreciated by the English-speaking experts. Here is a facsimile review about this demonstration by Professor Gunnar Fant, one of the most prominent researchers in the world of speech (see Fig. 8).

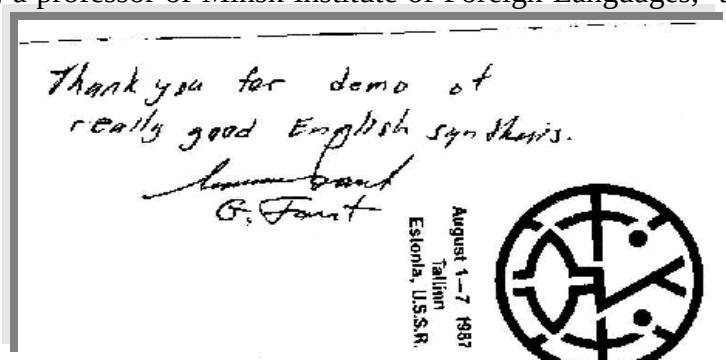


Figure 8 - Facsimile review of G. Fant

4.3 Speech Recognition and Synthesis Laboratory (1988 – 1991)

In 1988, the Laboratory MONIIS was reorganized into the Speech Recognition and Synthesis Laboratory at the Institute of Technical Cybernetics, Academy of Sciences of Belarus (now - the United Institute of Informatics Problems), led by the author of this article who was invited by the Directorate of the Institute for this position. As many still remember well, the end of the 1980s was marked by the appearance of the first personal computers, so naturally the laboratory paid attention to some questions related to equipping computers with voice systems as new information input and output. The formant method, which had a key role in the

development of text-to-speech synthesis systems for a long time, was unsuitable for this purpose because it demanded a huge amount of calculations in real time, which was unrealizable for the PC at that time.

In the late 1980s, a new microwave (MW) method for the synthesis of speech signals was suggested [10]. Instead of formants calculating this method implied using a pre-prepared set of microwaves of a natural speech signal. The set consisted of microwave signal segments of the length equal to the period. Their total number, which was required for generating any sound of speech, reached a few hundred. The MW-method was implemented in the synthesizer of Alexander Ivanov "PhonemePhone-5" in the form of a specialized software, oriented towards working with an internal sound card or with an autonomous device that was connected to the RS-232 (see Fig. 9). Surprising to many, its compactness (only 64K bytes) allowed equipping the first IBM PC-XT and even domestic PC ES1840 with speech synthesis.

The speech synthesizer was asked-for in many practical applications, but especially it was proposed [9] was, and even still is, needed for blind computer users. At the beginning of the 1990s, over a hundred of sets of specialized hardware and software products for the blind have been created and distributed by George Losik in Russia, Ukraine and Belarus. Its quite legible sound can be still heard on the Internet or when purchased on the market CD ROM “Talking mouse”. Later on, versions for the Czech and Polish on the basis of the MW-method were designed (after about 3 months of stay in the country on the invitation of customers), as well as an autonomous single-board module of speech synthesis, the Ukrainian-language version of which for a long time was used by Kiev Metro lines.

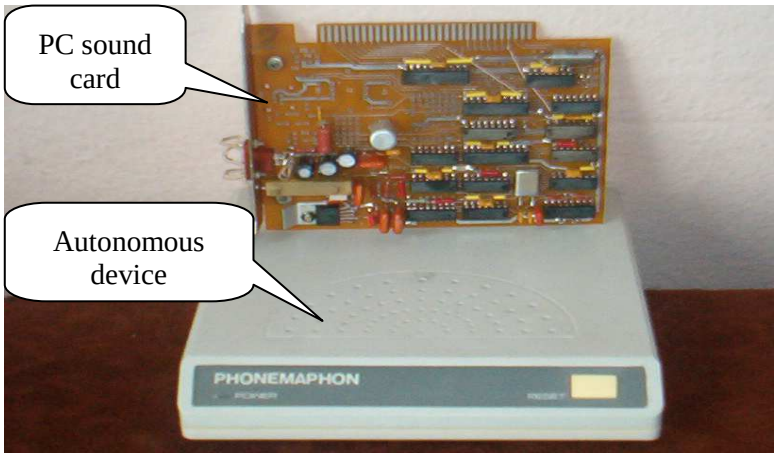


Figure 9 – Microwave speech synthesizer

5 Conclusion

As shown above, in the 60s - 80s in the USSR formed large groups of successful researchers. Largely their scientific achievements were world-class in those years. After the disappearance of the USSR and Republics’ withdrawal from it, the scientific community was disintegrated. In addition, the process of liberalization pursued by the Russian and other CIS countries, virtually instantaneously rose "iron curtain", through which many of the leading scientists and specialists left with their experience in leading research centers: the U.S., Belgium, Israel, Australia, etc.

Many of the above-mentioned schools of the USSR did not recovered from the brain drain of leading professionals. The situation was also worsened by all the difficulties of economic liberalization of the early 1990's. But despite all this, some of them managed to survive that period thanks to market mechanisms which were running at that time. It can be explained by the fact that in the 1990s it became possible to establish and develop well-functioning private companies, including those in the field of high technology. A brilliant example of such a successful and big company is the Speech Technology Center (STC) in St. Petersburg (see Fig. 9).

Figure 9 – Employees of the CSTs on vacation (http://speechpro.com/company)

In present, a range of research institutes and universities in St. Petersburg, Moscow, Kiev and Minsk continue to work successfully in the field of speech technologies, although the number of their workers is not so quantitatively significant.

Acknowledgements

This work was partially enabled by a grant of the "Kulturstiftung des Freistaates Sachsen". My sincere thanks are to the Kulturstiftung as well as to Dr. Klaus Mauersberger and Kirsten Vincenz from the Custody of the TU Dresden who initiated my research stay in Dresden.

And last not least, I should like to thank Prof. Ruediger Hoffmann for his attention to this work and for many useful discussions.

Literatur

1. Schroeder M. A Brief History of Synthetic Speech. Speech Communication, – 1993, vol. 13, pp. 231-237.
2. Markov A. Distribution of the law of large numbers to values that depend on each other. // Proceedings of the Physical and Mathematical Society of Kazan University. - Series 2. – 1906, Vol. 15 - pp. 135-156. (in Russian)
3. Zagoruiko N. ARSO and speech technologies // Proceedings of the IM SB RAS, "Computer systems", – 1995, Novosibirsk, pp. 3-31. (in Russian)
4. Slucker G. Non-Linear Method of Speech Signals Analysis // Proceedings of the Moscow Research Institute of Radio, N2, Moscow, 1968 – pp. 36-48. (in Russian)
5. Vitsuk T. Spoken words recognition by Dinamic Programming Methods // Cibernetics, N1, Kiev. 1968 – pp. 81-88. (in Russian)
6. Lobanov B. More About Speech Signal and the Main Principles of its Analysis // IEEE Transactions on Audio and Electroacoustics.- 1970, N 3. – pp. 316-318
7. Lobanov B. Classification of Russian Vowels Spoken by Different Speakers // The Journal of the Acoustical Society of America.- 1971, N 2 (2). – pp. 521 – 524
8. Lobanov B., Slucker G., Tizik A. Recognition of Sound Combinations in the Current Speech Signal // Proceedings of the Moscow Research Institute of Radio, N4, Moscow, 1969.- pp. 67 - 75 (in Russian)
9. Lobanov B. The Phonemophon Text-to-Speech System // Proceedings of the XI-th International Congress of Phonetic Sciences, Tallin, 1987.- pp. 100 - 104
10. Lobanov B. Karnevskaya E. MW - Speech Synthesis from Text // Proceedings of the XII International Congress of Phonetic Sciences.- Aix-en-Provense, Franse 1991.- pp. 387 - 391