

**R.K. Potapova**

**THE PRIORITY TRENDS IN DYNAMICS  
OF MODERN APPLIED SPEECHOLOGY**

*Moscow State Linguistic University  
Russia, 119837, Moscow, Ostozhenka, 38  
Tel.: (095)201-56-97; Fax: (095)246-28-07  
E-mail: [mglu@online.ru](mailto:mglu@online.ru)*

The aim of the paper was to study several main trends in dynamic of evolution of modern speechology on the basis of the last scientific events on the domain of speechology and new technologies. The main focus is some topics which include conceptual models for natural spoken language, integrating of diverse knowledge in the speech recognition and understanding process, speech and multi-lingual/multi-modal systems and so on.

In the last 10 years, we have encountered an impressive growth of the entire field with expanded interdisciplinary activities and a dominance of speech technology which has penetrated phonetics and linguistic departments and congresses [1].

Science is like a biological system, and the growth of science shows similar traits. In nature, the species develop generation after generation towards a more perfect form. The same is true of the development of speech communication systems. We are eager to build into them human functions of speech and hearing and understanding. However, in many respects the progress has not been as fast as we have hoped. Although speech technology is striving ahead with an increasing effort of people and funding, and with hard work and not without partial success, this is still along a path wrecked with technical failures and frustrations. We seldom reach the level of performance we are aiming at, and everything takes longer time to complete than anticipated [2].

Last year some important international events on the domain of speechology gave our phonetics community the opportunity to come together: the XIV<sup>th</sup> International Congress of the Phonetic Sciences (ICPhS, August, San Francisco), the European Conference on Speech Communication and Technology (Eurospeech'99, September, Budapest), the Conference of International Association on Forensic Phonetics (IAFP'99, July, York), the International Workshop «Speech and Computer» (SPECOM'99, October, Moscow). These conventions have been interpreted as impressive and encouraging symbols of the power and dynamics of our discipline [3].

The XIV<sup>th</sup> International Congress of the Phonetic Sciences covered numerous phonetic areas, including sociophonetics, speech technology, clinical phonetics, forensic phonetics, language acquisition, language teaching, sound change, prosody, intonation, speech production research and much more! [4].

The volume «A Guide to the History of the Phonetic Sciences in the United States» was published and covering the following areas:

(a) - thematic historical essays on: acoustic phonetics;

American dialects; American dictionaries and on spelling reform; clinical phonetics; field phonetics; speech perception research; physiological phonetics; speech technology; and forensic phonetics.

(b) - The history of phonetic research at eleven laboratories: Bell Telephone Labs; University of California at Berkeley; Univ. of Florida; Haskins Labs; Univ. of Iowa; Univ. of Michigan; MIT; Ohio State Univ.; the Speech Communication Research Laboratory at Santa Barbara, etc.

The European Conference on Speech Communication and Technology -Eurospeech'99 was the sixth in a biennial series launched in 1989 and sponsored by ESCA, the European Speech Communication Association. After Paris, Genoa, Berlin, Madrid, and Rhodes, the conference was hosted by the Technical University of Budapest [5].

The plenary paper by F. Jelinek (USA) presented a material entitled "Putting language into language modeling" which dealt with the structured language model (SLM), a new approach of language modeling in large-vocabulary speech recognition systems.

Although the majority of the papers were presented in the classical domains of speech recognition and speech synthesis, there were sessions about topics such as articulatory measurements

and modeling, first and second language learning, speech disorders and speech for the disabled, audio-visual speech, speech corpora, speech and noise, speech and the Internet, topic detection and tracking, and education in speech communication.- Five sessions dealt with various aspects of prosody in speech synthesis, dialogue, and speech recognition, A special session on education, the "Education Arena", was organized like an exhibition with presentations of educational speech processing software covering all aspects of speech communication, such as basic speech signal analysis, simulation of articulatory movements, or developing small applications in speech recognition.

The scope of the conference was not at all limited to speech technology or applied research. A large number of papers dealt with basic research in speech communication and phonetics. Two of the five plenary papers were presented by phoneticians: "The controversial connection between speech production and perception: Theories versus facts"; "How speech works: Questions and preliminary answers". The main topic of this talk was how to cope with the massive variability in the speech signal, and the author pleaded for measures that would try to make sense of this variability instead of trying to make it disappear.

Speech recognition however was the primary topic, with the largest number of presented papers at the conference. And so it was quite natural that the last plenary paper "Perspectives of speech technologies research dealt with this topic. Many improvements have been reported in acoustic modeling, language modeling, and search. These findings may be small but they add up. The probabilistic framework has been successful and is being used in other fields. Acoustic modeling is still a bottleneck and better features are needed. At the first Eurospeech 10 years ago, a speech recognition system could deal with a 1000-word vocabulary and read speech; today's systems approach 64000 words or more with a natural-language grammar and spontaneous speech [5].

The International Association for Forensic Phonetics (IAFP) in 1999 held its annual conference in York, UK [6]. The first conference on Forensic Applications of Phonetics was held at this same venue in 1989 and IAFP was officially founded some two years later after the third annual conference in 1991.

The International Association of Forensic Linguistics delivered a paper presenting the latest advances in computer programmes for the detection of plagiarism and disputed authorship. This focused upon an approach developed at the University of Birmingham involving automated comparative analysis of lexico-syntax across documents. The programme was seen to have applications in both the academic context (identification of plagiarized essays by university students) and the forensic field (e.g., determination of common authorship of threatening letters).

Subsequent contributions were organized around three main themes: (i) *The Evaluation of Evidence by Lay-Witnesses (non-phoneticians)*; (ii) *Methodological Issues in Forensic Speaker Identification*; (iii) *Technological Aids for Speech Processing and Interpretation* [6].

**Evaluation of Lay-Witness Evidence:** research in the field of the comparative accuracy of lay-listeners and automated expert systems to determine whether tape recorded samples of speech emanated from a common source; the results of experiments in which was attempted to establish whether there were any correlations between the accuracy of earwitness identifications of speakers and various aspects of the witness's musical abilities; experiments which attempted to assess the degree to which priming could assist in improving subjects' abilities to interpret what was said in recordings masked by noise.

**Methodological Issues in Forensic Speaker Identification:** the nature of the forensic speaker identification task and the ways in which the strategies and questioning techniques adopted by cross-examining barristers could pose difficulties for experts in giving a full and proper representation of their findings and conclusions; epistemological and methodological issues in the use of Decision Theory in moving from findings in the form of acoustic correlations to conclusions about common - or different - speaker identity; a range of auditory phonetic and acoustic methods ( $F_0$  estimation; formant averaging; LTAS) had been deployed in giving rise to their eventual conclusion; voice disguise is a persistent and difficult problem for forensic speaker identification experts. The patterns of disguise have to date been only poorly documented; different channel characteristics; also present difficulties in the forensic comparison task; the effects of three transmission media (orthophonic, telephonic and GSM - the euro-standard grande system mobile); expert evidence of speaker identification on samples

in Albanian (Macedonia); fundamental frequency as a function from a range of factors (background noise, alcohol consumption, transmission channel, and so on); experimental data on the effects of physical exertion (and its simulation) on fundamental frequency; the effects of compression and signal coding of speech, the ways in which these processes may affect individual voice features were described; the application of an auditory-phonetic approach with auditory analysis of consonant and vowel pronunciations.

**Technological Aids: Speech Processing and Interpretation:** database of information relevant to forensic phonetics and incorporated into an electronic (computer-based) encyclopaedia for use by Russian experts in the field, the theoretical background to this work and a practical demonstration were given [7]; an introduction on the basic principles and procedures of enhancing the intelligibility of noisy and difficult forensic recordings via the sound filtering technology produced by the Digital Audio Corporation; sound processing equipment.

The International Workshop for SPEECH AND COMPUTER (SPECOM'99) was held in Moscow and the aim of SPECOM'99 was to discuss the most important topics of man-computer interaction by voice and more perspective applied areas of speech dialogue [8].

TOPICS to be covered included:

- ⇒ Models for spoken language semantic interpretation.
- ⇒ Integration of diverse knowledge in the speech understanding process.
- ⇒ Speech-to-speech translation.
- ⇒ Telephone natural language response generation.
- ⇒ Speech dialogue models creation.
- ⇒ Multi-lingual and multi-modal systems.
- ⇒ Speech recognition for dialogue systems.
- ⇒ Conceptual approaches to dialogue educational systems; speech units.
- ⇒ Applied systems and new technologies; transmission methods.
- ⇒ Speech science - speech technologies today.
- ⇒ Text -to- speech synthesis.
- ⇒ Speech signal databases.
- ⇒ Phonetic aspects of dialogue with computers.
- ⇒ Systems of identification and verification of speaker.
- ⇒ Perspectives of evolution of speech technologies.

This year 82 people from many countries including Germany, France, UK, Greece, The Netherlands, Japan, Korea, Slovakia, Romania, Ukraine, Mexico and Russia participated in the workshop.

The authors of survey lectures represented their works. They were as follows:

Speech technologies and speech science (*V. Galunov, V. Taubkin*); Spoken Dialogue Systems for Call Centre Automation (*George Kokkinakis*); Statistical Pattern Recognition Techniques for Multimodal Human Computer Interaction and Multimedia Information Processing (*G. Rigoll, S. Müller*); The beginning of machine translation in the USSR (retrospective review) (*Yu.N. Marchuk*); The interactive expert system for forensic speaker identification used in Russia (*N.F. Popov, A.N. Linkov (1), An.V. Fesenko, N.B. Kurachenkova, N.V. Baicharov (2), I.P. Karlin, I.N. Timofeev (3), R.K. Potapova (4)*); The Knowledge Based Speech – Input expert System for Russian (*R.K. Potapova*); Realization of some reserves of language and extralinguistic knowledge for the speech dialogue systems improvement (*Yu. Kosarev, J. Savage*).

A lot of interesting problems were represented and discussed on the domain of interaction between speech and computer: e.g., Stochastic Conceptual Model for Spoken Language; Combining Statistics with Semantic Networks in a Real-time Dialogue System; Spoken Dialogue Systems: Toward a Complexity Measure for System's Questions; Acceleration of Lexical Access Caused by Word Familiarity and by Accent Nucleus Perception; Speech Based, Multimodal Human computer Interaction; Improvement of Multi-Band Speech Recognition; Integrating Speech Recognition and Natural Language LTAG Techniques with Weighted Synchronized Automata; An Adaptive Pruning Threshold Algorithm for Efficient Speech Recognition; Development and Evaluation of Digital Map

Retrieval System Using speech Recognition Interface; Acoustic Speech Recognition by Two and three Layered Neural Networks with Competition and Cooperation; Towards a Better Collaboration Between a n-class and a n-gram Language Model; Phoneme-by-Phoneme Recognition for Continuous Speech; Hypertextual Multimedia Database of Russian Phonetics for distance Education; Acoustic Evidence of Tension in Voiceless German Plosives; Usage of the Acoustic-Phonetic Classifier for Optimization of Codecs with a Linear Prediction; Recurrent multiband pitch estimation; A speech interface system for information retrieval tasks on the WWW; Automatic formant tracking by an active contour technique; Speech quality measurements; Pattern Response Generation as a Part of the Model of the Dialogue with the Telephone Query System; Testing of «Text-to-Speech» System; The Application of Text-to-Speech System to Singing voice Synthesis; Main Principles and Overall Structure of TTS system; Prosody Generation and Segment Concatenation in the Time Domain; Recording Concatenative Units for Speech Synthesis using a Reference Pitch Prompt; Design and Implementation of a Russian Telephone Speech Database; Fast and Robust Segment Searching Algorithms; Sound Pressure Distributions and Forensic Phonetics Experts Learning; Speaker Identification Using Discriminative Feature Selection etc.

At present we encounter a polarisation of research efforts [1]. The knowledge based approach aiming at describing the speech code relating abstract linguistic units to speech wave characteristics has been found too difficult to handle, not only for those with a more narrow computer technology background but also for more phonetically oriented people. Our knowledge base is large but fragmental lacking complete generative and quantitative descriptions of any language. Statistical tools, such as neural network and HMM processing, have dominated [1].

In order to reach our ultimate aims in speech recognition as well as in synthesis, we need to penetrate more deeply the domain of the speech code. We take pride in collecting great data bases but up till now the analysis has been rather meagre. My recommendation for progress is to put more emphasis on descriptive speech wave analysis combined with articulatory modeling and resynthesis which will promote the learning process. For this purpose we need improved physiological mapping of the speech organs as well as a closer insight in articulatory movements and in the aerodynamics and acoustics of the production process, an appropriate representation of voice source dynamics and prosody.

## REFERENCES

1. Fant G. Half a century with speech science. TMH-QPSR, <sup>1</sup> 3, 1998, pp. 1-3.
2. Fant G. The role of speech research in the advance of speech technology. STL-QPSR, <sup>1</sup> 4, 1990, pp. 1-6.
3. Koster J.-P. From the President. The Phonetician, <sup>1</sup> 80, II, 1999, p.3.
4. Ohala J. and Bronstein A. XIVth ICPhS in San Francisco. The Phonetician, <sup>1</sup> 80, II, 1999, pp. 18-19.
5. Hess W. Eurospeech in Budapest. The Phonetician, <sup>1</sup> 80, II, 1999, pp. 20-22.
6. French P. Annual Meeting of the IAFP in York. The Phonetician, <sup>1</sup>80, II, 1999, pp.22-25.
7. Potapova R.K., Potapov V.V. Database of forensic phonetics knowledges (as applied to electronic encyclopaedia for Russian experts). Proc. of the Conf. of JAFP, York, 1999. – pp. 6-7.
8. Potapova R.K. SPECOM'99 in Moscow. The Phonetician, <sup>1</sup> 80, II, 1999, p.20.
9. Potapova R.K. Some aspects of forensic phonetics experts learning (on the basis of Russian). Proc. of SPECOM'99, Moscow, 1999. – pp. 208-211.