FOREWORD

Corneliu BURILEANU

Faculty of Electronics, Telecommunications and Information Technology, University "Politehnica" of Bucharest

Two years ago, at the 2005 edition of our Conference, we pointed out some main directions in the domains of spoken language technology and human-computer dialogue: development in speech recognition software, text-to-speech synthesis, voice access interfaces, multimodality in human/computer dialogue, wearable computers, portable messaging and information appliances, spontaneous speech recognition, multi-channel signal processing, speech technology in biometric security.

Speaking about past events, "IEEE Computer" Journal is offering us a very interesting memento under the generic title "32 & 16 Years Ago".

For example, in volume 38 (no. 8) issued in August 2005, we find the following "news": "*August 1973: **Voice Synthesis**. A totally new concept has been developed by Master Specialties Company in Costa Mesa, California. This company has achieved a technique for digitizing and storing whole-words in metal oxide semiconductor read-only memories. Until this technique was developed, it was totally impractical to consider storing whole-words in such a device because of the amount of storage for each word and the cost of storage space. A typical word requires about 40,000 bits of information to directly convert it to a digital signal. By analyzing completely plotted audio wave forms, engineers at Master Specialties Company have been able to develop a proprietary method for converting the analog audio signal of a word into a digital signal, requiring as little as 8,000 bits of memory storage, while still remaining all of the natural voice qualities for reproduction. This makes individual word storage practical and opens up a whole new field of applications for electronic voice response*."

Then, on October 2005 (vol. 38, no. 10), about the events 16 years ago: "*October 1989: **Voice Recognition**. Just about everyone talks to their computer. I've been muttering to computers since the first time I sat down in front of a TTY. Before that, I muttered to card punches. Lately, however, it seems my PC is beginning to understand me. The ability to enter commands by voice has tremendous possibilities for people with restricted hand or finger movement. It also allows use of the PC applications where the use's hands are busy doing other things or in environments that could damage the keyboard. Even in the typical work-place voice input has the potential for increasing productivity. Touch typists can keep their hands over the standard portion of their keyboards and enter commands requiring any of the special keys by voice. They also will not have to remember the arbitrary keystrokes that make up most of those commands. People using graphical interfaces with a mouse will not have to let go of the mouse to use the keyboard or move the mouse from a work area to a command menu area*."

On August 2006 (vol. 39, no. 8), the state of the art in the domain of speech processing is some steps ahead: "*August 1990: **Speech Recognition**. Limited forms of speech recognition are available on personal workstations. Currently there is much interest in speech recognition, and performance is improving. Speech recognition has already proven useful for certain applications, such as telephone voice-response systems for selecting services or information, digit recognition for cellular phones, and data entry while walking around a railway yard or clambering over jet engine during an inspection. Nonetheless, comfortable and natural communication in a general setting (no constraints on what you say and how you say it) is beyond us for now, posing a problem too difficult to solve*." Also in the same volume: "*August 1990: **Speaker Recognition**. Speaker recognition has been applied most often as a security device to control access to buildings or information. One of the best-known examples is the Texas Instruments corporate computer center security system. Security Pacific has employed speaker verification as a security mechanism on telephone-initiated transfer of large sums of money. In addition to adding security, verification is advantageous because it reduces the turnaround time on the banking transactions. Bellcore uses speaker verification to limit remote*

access of training information to authorized field personnel. Speaker recognition also provides a mechanism to limit the remote access of a personal workstation to its owner or a set of registered users."

Nowadays, a very interesting and up-to-date analysis of the achievements in the domain and also the future trends were presented at "IEEE/ACL Workshop on Spoken Language Technology, Aruba, Dec. 11-13, 2006" (www.slt2006.org).

The evolution of speech and language technologies over the past decade has spawned a stimulating new research area: Spoken Language Technology (SLT). The important issues pointed out at this workshop could be summarized as: "*Technological advances in SLT promise to provide ubiquitous and personalized access to information, communication, and entertainment services. For example, advances in natural language understanding and large vocabulary continuous speech recognition have resulted in a new generation of automated contact center services that offer callers the flexibility to speak their request naturally using their own words as opposed to the words dictated to them by the machine. Advances in machine translation technology have resulted in speech-to-speech translation products that offer multi-party multi-lingual communication. Advances in information search and data mining are providing the means to extract intelligence information from large corpora of speech data (e.g., TV programs, call center data) to help improve business operation and search for information rapidly without having to listen to conversations*".

The research in the following areas seems to be strongly encouraged: • Spoken language understanding • Dialog management • Spoken language generation • Spoken document retrieval • Information extraction from speech • Question answering from speech • Spoken document summarization • Machine translation of spoken language • Speech data mining and search • Voice-based human computer interfaces • Spoken dialog systems, applications and standards • Multimodal processing, systems and standards • Machine learning for spoken language processing • Speech and language processing in the world wide web.

At this workshop, Hermann Ney, currently with Computer Science Department RWTH Aachen University, Germany made his "Closing Remarks: How to Continue?". The main issue he emphasized is about the interaction between speech and NLP (natural language processing) people. At a former similar workshop (*IEEE Workshop on Automatic Speech Recognition*, Arden House, Harriman, NY, 1985) the state of the art for ASR (automatic speech recognition) shows the IBM speaker-dependent isolated-word recognizer for 5,000-word vocabulary and some other groups achievements – 10-1,000 words (constrained grammar, speaker-dependent systems). The interaction and areas of contact between speech and NLP people were limited to phonetic knowledge in acoustic recognition, language models and (non-statistical) grammars.

Today, the SLT Workshop seems to be dominated by many areas of interaction between speech and NLP: language modeling, machine translation, spoken language understanding, dialogue modeling, information retrieval, question-answering, text summarization, language generation, multi-modal communication. The important conclusion is that data-driven and statistical concepts have moved from speech to virtually all NLP related tasks. The progress in the interaction speech – NLP people must lead towards improvements in ASR and MT (machine translation). These will depend on agreement on common tasks, agreement on common evaluation measures, and common databases and evaluation operations.

Another participant in Aruba SLT Workshop, David Nahamoo, Speech CTO, at "IBM Research", had an intervention on "Speech Technology Opportunities and Challenges". He pointed out the major speech application opportunities:

- In commerce: Contact Centers, Unified Communication;
- In global access: speech to speech translation, translingual multimedia mining, accessibility;
- In devices: automotive, set top boxes, mobile phones.

In Nahamoo's opinion, speech technology innovation that matters will be in "*conversational interaction – dealing with complexity*", "*speech analytics – extracting insight/knowledge*", "*multilingual dimension – globalization*". Global language barriers due to different languages spoken by people living in different regions or even by different ethnic groups living in the same region could cause many problems: high cost for agents which need both subject matter expertise and language skills, life threatening in medical emergency, natural disaster situations, or in military applications.

One important challenge seems to be "Speech to Speech (S2S) Translation". The main issues should deal with:

- Speech recognition improvements: the need to work in noisy environments, with spontaneous, conversational speech in multiple languages and with emotional speech under stress.
- Translation has to handle output of ASR system.
- Translated text must be "speakable" for oral communication which means it is not enough to translate content adequately; output must be fluent and the need to carefully consider and tune interactions between ASR, MT and NLP.
- Cost-effective development of new languages and domains.
- Intonation translation remains a great challenge.

John Makhoul, currently with "BBN Technologies", was interested about "Information Extraction from Speech". He spoke about the amount of available resources (relatively large – many possible applications in audio indexing and broadcast monitoring), advanced extraction technologies and audio indexing systems (with all components: speech detection, speaker segmentation/tracking/identification, speech recognition, name extraction, sentence detection, machine translation).

The dimension of the "Multilingual Language Processing" was presented at Aruba SLT Workshop by Pascale Fung, from "Human Language Technology Center", Department of Electronic & Computer Engineering, Hong Kong University of Science and Technology. He offered a classification of the most spoken languages in the world: Mandarin – spoken by 1,051 million people, English – spoken by 510 million people, then Hindi – spoken by 490 million people, Spanish – spoken by 425 million people, Arabic – spoken by 255 million people, Russian – spoken by 254 million people, Portuguese – spoken by 218 million people, Bengali – spoken by 215 million people, Malay – spoken by 175 million people, French – spoken by 130 million people, etc. Meantime, the countries with highest number of Internet users are United States with 19.4 % users of the world and 69.9 % Internet penetration, China with 11.4 % users of the world and 9.4 % Internet penetration, then Japan with 8.0 % from the world users but 67.2 % Internet users, Germany with 4.7 % and 61.3 %, respectively, etc.

Pascale Fung made a comparison between multilingual systems and MT systems: machine translation systems transform source language into target language, while multilingual systems deal with the interface technology (display/input), extract information from multilingual (spoken or written) documents, understand queries in multiple languages.

Examples of multilingual systems are Google search (117 languages), Nuance products (OCR/119 languages, SLT/46 languages), spam filters, Wikipedia (12 languages).

Multilingual applications will most probably be in spoken language understanding and generation, in information retrieval (legal/medical/education), in Internet businesses, search, spam filters, user generated content.

We hope that the 4th Conference on Speech Technology and Human-Computer Dialogue "SpeD 2007" will reflect all these tendencies.

"SpeD 2007" is an international conference, under the aegis of the Romanian Academy, and in cooperation with EURASIP and IEEE, devoted primarily to the presentation of the recent trends in the field of speech technology and the human-computer dialogue. "SpeD 2007" is, in the first place, an invitational-only conference. The principal organizer of the previous editions of the conference and of SpeD 2007 has been the Faculty of Electronics, Communication and Information Technology from the University "Politehnica" of Bucharest. For the current edition, the organizers are the above-mentioned Faculty, the Institute for Computer Science of the Romanian Academy and with two other faculties, one from the University "Al. I. Cuza" Iaşi and one from the Technical University "Gh. Asachi", Iaşi.

The volume includes five sections: Speech and Speaker Recognition, Human-Computer Dialogue, Text-to-Speech Synthesis, Emotions and Expressive Speech Synthesis, Speech Analysis, Representations and Models, Natural Language Processing. This year too, we are honored to gather in this volume the papers of ones of the most representative Romanian specialists in the domain and also invited papers of our collaborators, well-known experts from Universities in Europe and U.S.A.:

- ***Speech and Speaker Recognition*** – *Jean-Paul HATON*, "Improving the Robustness of Automatic Speech Recognition"; *Laurent BESACIER, Viet-Bac LE*, "Automatic Speech Recognition for Under-Resourced Languages"; *Daniel MORARU, Guillaume GRAVIER*, "Landmark Based Large Vocabulary Continuous Speech Recognition"; *Marieta GÂTA, Gavril TODEREAN*, "Results Obtained in Speaker Recognition Using Gaussian Mixed Models"; *Doru MUNTEANU, Constantin PINTILIE, Laurențiu APOSTOL*, "Automatic Speaker Verification Experiments Using a Continuous Speech Recognizer".

- ***Human – Computer Dialogue*** – *Vladimir POPESCU, Jean CAELEN, Corneliu BURILEANU*, "Generic Architecture for Natural Language Generation in Spoken Human-Computer Dialogue"; *Sorin DUSAN, Sriram RAMACHANDRAN*, "Flexible Speech and Pen Interaction on Handheld Devices"; *Laurent MAZUEL, Nicolas SABOURET*, "How to Enhance Genericity in Natural Language Command Interpretation Using Introspection and Ontologies?".

- ***Text-to-Speech Synthesis; Emotion and Expressive Speech Synthesis*** –*Mihai SURMEI, Dragoş BURILEANU, Cristian NEGRESCU, Răzvan PÎRVU, Cătălin UNGUREAN, Aurelian DERVIŞ*, "Text-to-Speech Engines as Telecom Service Enablers"; *Mihai RADU, Florin POPESCU, Stefan-Adrian TOMA*, "Aspects Regarding Synthetic Speech Quality Evaluations for Military Systems"; *Ovidiu BUZA, Gavril TODEREAN, Alina NICA, Zsolt BODO*, "Original Method for Romanian Text-to-Speech Synthesis Based on Syllable Concatenation"; *Vasile APOPEI, Doina JITCĂ*, "Module for Generating the F0 Contour Using as Input a Text Structured by Prosodic Information".

- ***Speech Analysis, Representations and Models*** – *Daryle GARDNER-BONNEAU*, "Human Factors and Speech Technology: Reflections on 20 Years of User Interface Design for Speech-Enabled Applications"; *Michel VACHER, Jean-François SERIGNAT, Stéphane CHAILLOL*, "Sound Classification in a Smart Room Environment: an Approach Using GMM and HMM Methods"; *Cristian NEGRESCU, Amelia CIOBANU, Dragoş BURILEANU, Dumitru STANOMIR*, "An Improved Hybrid Time-Frequency Algorithm for Timescale Modification of Speech/Audio Signals"; *Leonardo CERNAIANU, Eugeniu OANCEA, Alexandru SERBANESCU*, "Genetic Optimization of Wavelet Packet Compression for Speech Signal"; *Horia-Nicolai TEODORESCU, Monica FERARU, Diana TRANDABĂȚ*, "Studies on the Prosody of the Romanian Language: the Emotional Prosody and the Prosody of Double-Subject Sentences".

- ***Natural Language Processing*** – *Dan TUFIŞ, Radu ION*, "Parallel Corpora, Alignment Technologies and Further Prospects in Multilingual Resources and Technology Infrastructure"; *Dan CRISTEA* et al., "The Digital Form of the Thesaurus Dictionary of the Romanian Language"; *Adriana VLAD, Adrian MITREA, Mihai MITREA*, "Orthography and Punctuation Marks in Printed Romanian Modelling: a Study Based on the Letter Statistical Structure in a Literary Corpus"; *Neculai CURTEANU, Diana TRANDABĂȚ, Mihai Alex MORUZ*, "Syntax-Prosody Interface for Romanian within Information Structure Theories"; *Ştefan DIACONESCU, Ionuț DUMITRAŞCU*, "Complex Natural Language Processing System Architecture".

ACKNOWLEDGEMENTS

I want to express our gratitude to Acad. Mihai Drăgănescu, President of the Section for Science and Technology of Information, who is our mentor for the functional electronics domain. I am very grateful to Prof. dr. eng. Ecaterina Andronescu, Rector of the University POLITEHNICA of Bucharest for her important support.

The organizers of SpeD 2007 and the Editors of this volume are grateful to the Romanian Academy – Section for Sciences and Technologies of Information and to the Iaşi-Branch of the Romanian Academy. Special thanks to the Romanian Academy Press, especially to the book editor Mihaela Marian, for competent and kind support.

We are also grateful to the panel of referees for their effort in revising the papers. We thank the members of the Scientific Committee and the members of the Organizing Committee. I must also mention the helpful collaboration with „Alexandru Ioan Cuza" University of Iaşi and the members of the Local Organizing Committee.

The "SpeD 2007" Organizing Committee acknowledges the sponsoring of the Conference by the following companies: "TEHNE" Foundation – The Center for Development and Innovation in Education, Infineon Technologies Romania, SOFTWIN Group, Freescale Semiconductor Romania, and ERICSSON Telecommunications Romania S.R.L.

I am especially grateful to Professor Horia-Nicolai Teodorescu for his valuable effort to be co-chair and "host" of our Conference. It was an honor to collaborate with Professor Teodorescu in all aspects of the Conference.