

WELCOME

to

**The 5th International Conference
on**

**Speech Technology
and**

Human-Computer Dialogue

“SpeD 2009”

Constanța, June 18-21, 2009

The “SpeD” (“Speech and Dialogue”) Conferences have already a history.

After 2000, together with Academician Mihai Drăgănescu, former President of Romanian Academy, we decided to resume the organization of a Conference in the field of spoken language technology (as an application of digital speech processing) and human-computer dialogue (the first one was organized in the 80’s years).

So, it was the birth of “SpeD” Conference, which is now at the 5th edition.

In the 2nd edition (2003), we noticed the evolution from *speech technology* to *spoken language technology*. “This had to be foreseen however from the very beginning involving the use of artificial intelligence, both for natural language processing and for acoustic- phonetic processes of the spoken language. The language technology is seen today with two great subdivisions: technologies of the written language and technologies of the spoken language. These subdivisions have to work together in order to obtain a valuable and efficient human-computer dialogue” (Mihai DRĂGĂNESCU)

In 2005, we appreciated that no dramatic changes occurred in the domain, since our past Conference; however, some trends were more and more obvious and some new fields of interest appeared to be a promise for the future.

We were able to identify a constant development of what is called “*speech interface technology*” which includes

- automatic speech recognition,
- synthetic speech,
- natural language processing.

We noticed commercial applications in computer command, consumer, data entry, speech-to-text, telephone, and voice verification. Robust speaker-independent recognition systems for command and navigation in personal computers were already available; telephone-based transaction and database inquiry systems using both speech synthesis and recognition were coming into use.

Some main directions were pointed out:

- **speech recognition software,**
- **text-to-speech, voice access interfaces,**
- **multimodality,**
- **wearable computers,**
- **portable messaging and information appliances,**
- **spontaneous speech recognition,**
- **multichannel signal processing,**
- **speech technology in biometric security.**

For the edition in 2007, we considered a very interesting and up-to-date analysis of the achievements in the domain and also the future trends that were presented at “IEEE/ACL Workshop on Spoken Language Technology, Aruba, Dec. 11-13, 2006”

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 seemed 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, one of the main issue emphasized is about the **interaction between speech and NLP (natural language processing) people.**

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.

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.

In our days edition of the Conference, “SpeD 2009”, we can notice some possible expansions of the trends already mentioned:

- **spoken language understanding,**
- **spoken language generation,**
- **machine translation of spoken language,**
- **spoken document retrieval,**
- **information extraction from speech,**
- **spoken document summarization,**
- **speech data mining and search,**
- **Web search,**
- **voice-based human computer interfaces,**
- **spoken dialogue systems,**
- **dialog management,**
- **applications and standards,**
- **multimodal processing,**
- **machine learning for spoken language processing,**
- **security, bioinformatic/genomic signals.**

The “Signal Processing Magazine” May 2008-issue is dedicated to the spoken language technology (SLT). This special issue on SLT was motivated by the first SLT workshop (we already mentioned) held in Aruba, in December 2006.

“The articles in this issue cover language and/or understanding aspects of the signal processing. This forms a stark contrast to the special section in September 2005, speech technology in human-machine communications, where half of the articles dealt with language processing while the focus was on the more traditional tasks of speech analysis and recognition with the trend towards speech understanding. This change represents a significant new development in our field of signal processing, where the conventional way of thinking about the signal as low-level, numerical-valued information is transcended to a new perspective concerning the importance of high-level, symbolic-valued informational sources (e.g., language or text) that embed their underlying contents”

Some of the achievements but also the challenges of the **spoken language technology (SLT)** are emphasized in the mentioned issue: “Technological advances in SLT promise to provide ubiquitous and personalized access to information, communication, and entertainment services.

- 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 the speech”.

The spoken document search over the Web is an important field of interest.

Web statistics (dates of capture – end of 2006, 2007) are especially interesting:

- **domain names** - **estimated volume: 138 million;**
- **archived Web pages** - **estimated volume: 85 billion;**
- **video streams** - **estimated volume: 7 billion;**
- **blogs** - **estimated volume: 200 million;**
- **Internet users** - **estimated volume: 850 million (SpeD 2007);**
- **mobile Internet users** - **estimated volume: 35 million;**
- **broadband subscribers** - **estimated volume: 300 million;**
- **instant messaging users** - **estimated volume: 430 million;**
- **voice over IP subscribers** - **estimated volume: 34 million;**
- **blog readers** - **estimated volume: 63 million;**
- **searches per month** - **estimated volume: 14 billion.**

“Voice search is essentially the combination of automatic speech recognition (ASR) and document search. It is particularly attractive when using mobile devices for searching Web contents, such as business and consumer listings, maps and directions, media contents, shopping, etc. In voice search, a system takes a spoken query as input, retrieves multiple hypotheses using a speech recognizer, formulates a text query, queries for index and provides users the search results”

Many important difficulties related to the voice search are not yet surmounted:

- **robustness and efficiency of speech recognition,**
- **spoken language understanding/search,**
- **dialog management.**

Studies showed that the automation rate of the spoken dialog systems was 92% in a laboratory study, but only 30% automation rate in the actual field trial “due to unexpected behavior of novice spoken dialog systems users and environment noise”.

In several papers, the characteristics of a proper name recognition system are presented. The error analysis gives the surprising result that 31% of the recognition errors were noise related and 22% were pronunciation related.

Spoken language translation seems to be important target in our days: “Spoken language translation (SLT) is of great relevance in our increasingly globalized world, both from a social and economic point of view.

Prominent SLT research projects are TC-STAR in Europe and GALE in the United States:

- The European project **TC-STAR** addressed the translation of speeches recorded at the European Parliament, between Spanish and English, and of broadcast news by Voice of America, from Mandarin to English.
- The DARPA program **GALE** has been focusing instead on the translation of speech and text collected from different media, mainly from Chinese and Arabic into English”.

A summary of system-level capabilities for spoken language translation:

- *first dialog demonstration systems, 1989-1993:*

- restricted vocabulary,
 - constrained speaking style,
 - speed (2-10)xreal-time,
 - platform-workstations.

- *one-way phrasebooks, 1997-present:*

- restricted vocabulary,
 - constrained speaking style,
 - speed (1-3)xreal-time,
 - handheld devices.

- *spontaneous two-way systems, 1993-present:*

- unrestricted vocabulary,
 - spontaneous speaking style,
 - speed (1-5)xreal-time,
 - PCs/handheld devices.

- *translation of broadcast news, 2003-present:*
unrestricted vocabulary,
ready-prepared speech,
offline,
PCs/PC-clusters.

- *simultaneous translation of lectures, 2005-present:*
unrestricted vocabulary,
spontaneous speech,
real-time,
PCs/laptops.

As we already noticed in the last edition of our Conference, an important field of interest is related to **multilingual spoken language processing**: “With more than 6,900 languages in the world and the current trend of globalization, one of the most important challenges in spoken language technologies today is the need to support multiple input and output languages, especially if applications are intended for international markets, linguistically diverse user communities, and nonnative speakers. In many cases these applications have to support multiple languages simultaneously to meet the needs of a multicultural society. **Consequently, new algorithms and tools are required that support the simultaneous recognition of mixed-language input, the summarization of multilingual text and spoken documents, the generation of output in the appropriate language, or the accurate translation from one language to another”**

We hope that the 5th Conference on Speech Technology and Human-Computer Dialogue “SpeD 2009” is connected to these tendencies but **is emphasizing also the problems we still have to obtain better performances for the Romanian language.**

The Conference includes four sections:

- **Speech Recognition and Understanding and Human-Computer Dialogue,**
- **Text-to-Speech Synthesis and Real-Time Speech-Enabled Communication Applications,**
- **Natural Language Processing,**
- **Speech and Audio-Signal Processing and Speech Interface Applications.**

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 and Prof. dr. eng. Cornel Panait, Rector of Constanța Maritime University for their important support.

The organizers of SpeD 2007 are grateful to the Romanian Academy – Section for Sciences and Technologies of Information, to the Iași-Branch of the Romanian Academy, Institute for Computer Science, Iași and to the Research Institute for Artificial Intelligence.

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 the members of the Local Organizing Committee from Constanța Maritime University: dean George Căruntu, Conf. Violeta Ciucur, dr. Mihaela Hnatiuc, Alina Tudor, Alexandru Caranica.

The “SpeD 2009” 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, Microchip Romania.

I am especially grateful to Professor Horia-Nicolai TEODORESCU for his valuable effort. It was an honor to collaborate with Professor Teodorescu in all aspects of the Conference.