Note from the Editor: Special issue on speech processing and soft computing

Amy Neustein

This special issue of the Journal is devoted to the work of twelve eminent speech scientists who apply novel soft computing methods to address some of the most difficult and persistent problems facing speech recognition systems today. I first heard these scientists discuss their innovative soft computing algorithms at the University of Salamanca where the Sixth International Conference on Soft Computing Models in Industrial and Environmental Applications, *SOCO 2011*, was held in conjunction with the 9th International Conference on Practical Applications of Agents and Multi-Agent Systems. The SOCO chair, Professor Emilio Corchado, devoted a special workshop session to Speech Processing and Soft Computing to explore how soft computing may complement conventional techniques in speech processing. The topics covered in this workshop included, but were not limited to, speech production, speech coding, speech modeling and analysis, speech recognition, speech enhancement, multichannel speech processing, text-to speech synthesis, natural language understanding and generation, and other aspects related to speech processing. Such soft computing methods encompass neural networks, Fuzzy systems, Evolutionary computation, and Swarm intelligence, as well as Bayesian networks, Chaos theory and other soft computing based approaches. Having been impressed with the research rigor and analytic insights of the workshop session speakers, I asked Dr. Corchado to help me assemble a special issue of the Journal that would allow the SOCO conference speakers to expound appreciably on their research methods and test findings in their papers submitted to this special issue. The papers chosen for publication in this special issue underwent rigorous peer review and were accepted only after authors carefully revised their papers to the satisfaction of the peer reviewers and the editor.

Volume 15, Number 1 opens with an illuminating paper authored by Canadian scientists Md Foezur Rahman Chowdhury, Sid-Ahmed Selouani and Douglas O'Shaughnessy of the University of Quebec and the University of Moncton, Shippagon Campus, respectively. Professor Chowdhury and his co-authors demonstrate an innovative frame dynamic rapid adaptation and noise compensation technique for tracking highly non-stationary noises and its application for on-line ASR. In their paper, "Bayesian on-line spectral change point detection: a soft computing approach for on-line ASR," the authors show how their algorithm that is based on a soft computing model using Bayesian on-line inference for spectral change point detection (BOSCPD) in unknown non-stationary noise environments addresses the shortcomings speech engineers routinely confront when ASR is used in real-world applications where test conditions are highly non stationary and not known *a priori*. Using this soft computing method, the Canadian researchers show that their BOSCPD when tested with the MCRA noise tracking technique for on-line rapid environmental change-learning in different non stationary noise scenarios reduces the delay in spectral change point detection significantly when compared to the baseline MCRA and its derivatives.

Next, King Saud University Professor Yosef Ajami Alotaibi investigates Artificial Neural Networks (ANN) based Automatic Speech Recognition (ASR) by using limited Arabic vocabulary corpora that consist of digits and vowels carried by specific carrier words. In his paper, Professor Alotaibi compared Hidden Markov Model (HMM) based ASR systems to two ANN based systems, namely Multilayer Perceptron (MLP) and recurrent architectures.

A. Neustein (✉)
800 Palisade Avenue, Suite: 1809, Fort Lee, NJ 07024, USA
e-mail: amy.neustein@verizon.net

Using the same corpora of digits and vowels, the author found that the ANN based recognition system achieved 99.5% correct digit recognition when compared with the HMM based recognition system which achieved 98.1% correct digit recognition. The author found the reverse with vowels carrier words where the ANN based recognition system achieved slightly higher vowel recognition accuracy, yielding a 92.13% correct vowel recognition whereas the HMM based recognition system achieved 91.6% correct vowel recognition. These findings are particularly interesting in the context of ANNs and other fuzzy systems which have become part and parcel of soft computing methods. Such meticulous comparative analyses, separating out digit from vowel recognition in the Arabic vocabulary, add at the minimum valuable new insights to the growing discourse on soft computing methods and their application to speech processing.

What follows in this special issue is a fascinating two part series by an outstanding group of faculty members and researchers at the University of the Basque Country, UPV/EHU. The articles in this series, authored by Nora Barroso and her colleagues, Karmele López de Ipiña Peña, Odei Barroso, Aitzol Ezeiza, Carmen Hernández and Manuel Graña, describe the development of GorUP, a Semantic Speech Recognition System in the Basque context. Part I analyzes cross-lingual approaches oriented to under-resourced languages and Part II analyzes the development of the Language Identification system. The authors, responding to real-world constraints that make Basque an under-resourced language, adroitly use data optimization methods and Soft Computing methodologies that are oriented to complex environments in order to overcome the lack of resources needed for collecting large data samples in non noisy settings. Given that in this context three languages coexist—French, Spanish and Basque—the authors' main goal is the development of robust Automatic Speech Recognition (ASR) systems for Basque, taking into consideration that all language variability must be analyzed. In this regard, Basque speakers mix during the production of speech not only sounds but also words of the three languages which results in a strong presence of cross-lingual elements. The authors show sensitivity to the fact that Basque is an agglutinative language with a special morpho-syntactic structure inside the words that may lead to intractable vocabularies. Professor Barroso and her co-authors make the astute observation that when one considers that "the available resources for Basque ... are very few and complex to process because of the noisy environment ... the methods employed in this development (ontology-based approach or cross-lingual methodologies oriented to profit from more powerful languages) could suit the requirements of many under-resourced languages." I wholeheartedly agree and wish to point out that this soft computing approach to designing an ASR system for Basque may be seen as a template for a whole host of under-resourced languages that have been woefully neglected by designers of speech recognition systems.

As a nice coda to this two part series, Professor Barroso and her co-authors show in their subsequent paper titled, "Experiments for the selection of sub-word units in the Basque context for semantic tasks," how they meet the long term goal of their project to develop a robust ASR system in the Basque context where French, Spanish and Basque (a minority language) coexist. They show how their work which is "focused on the selection of appropriate sub-word units with under-resourced and noisy conditions" is applied to the Basque Broadcast News and to the trilingual *Infozazpi* radio which is situated in French Basque Country. The readers will see from Professor Barroso and her co-authors' discussion how they try to compensate for the paucity of language resources in the under-resourced language of Basque by applying several data optimization methodologies based on Matrix Covariance Estimation and Ontology-based approaches.

The final paper in this issue provides a fascinating look at a new framework to improve dysarthric speech recognition by using rhythm knowledge. The author, University of Moncton Professor Sid-Ahmed Selouani, and his co-authors from the University of Quebec and the School of ESPRIT in Tunisia, respectively, show how their approach builds speaker-dependent (SD) recognizers with respect to the dysarthria severity level of each speaker. The authors employ soft computing methods to assess the severity level of dysarthria by using a hybrid classifier that combines class posterior distributions and a hierarchical structure of multilayer perceptrons. They point to their use of "rhythm-based features as input parameters" given the fact that "preliminary evidence from perceptual experiments shows that rhythm troubles may be the common characteristic of various types of dysarthria." In detailing their research design Professor Sid-Ahmed and his co-authors demonstrate in their paper how speaker-dependent dysarthric speech recognition is performed using Hidden Markov Models (HMMs). Drawing from the Nemours database of American dysarthric speakers for their experiments, these authors show the relevance of rhythm metrics and the effectiveness of the proposed framework to improve the performance of dysarthric speech recognition.

In conclusion, I wish to thank the authors of Volume 15, Number 1 for their laborious efforts to perfect their papers. Their diligent efforts have insured that their critical research can be made part of the expanding body of literature on soft computing methods for speech processing, which, as we can see from the contributions mentioned above, serves some very important societal needs from helping stroke victims and others who suffer from dysarthric speech to assisting

entire communities that utilize under-resourced languages in the design of well-functioning automated speech recognition systems. Professor Emilio Corchado, who organized SOCO 2011, is equally deserving of praise for including a session on speech processing and soft computing which inspired this special issue of the Journal.