Inaugural Editorial:

Riding the Tidal Wave of Human-Centric Information Processing— Innovate, Outreach, Collaborate, Connect, Expand, and Win

This editorial marks the start of my three-year term as Editor-in-Chief (EiC) of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING (T-ASL), right after the completion of my three-year term as EiC of the *IEEE Signal Processing Magazine* (SPM). To begin, I would like to thank Helen Meng for passing me a very strong T-ASL after three years of her outstanding work. I also thank Helen for the valuable help during the EiC transition over the past few months.

In the past three years, working as EiC of SPM I had a unique and exciting journey to the heart of our Signal Processing Society (SPS). I also had an opportunity to make the progress in our ASL technology widely informed to the broader SP community (e.g., [1], [2]). With waves of innovation introduced, SPM clinches the prestigious recognition as the most cited publication in the most recent consecutive years among all 247 journals in the field worldwide, including all 147 IEEE publications. Some of the innovations, such as translations for global reach and multimedia content access via Tag, are equally applicable to our T-ASL, which we can readily implement after major obstacles have been cleared by the effort during the past two years. Yet with new waves of innovation for our T-ASL under planning in coming years, we are in a position to broaden the intellectual influence and technical scope of our field in an unprecedented way, which I will briefly outline below.

One special advantage of our ASL field is its intimate connection, more than any other technical fields within our SP Society, to human-centric information processing and to "artificial intelligence" which is coming of age. Indeed, the tidal wave of human-centric computing is upon us in various forms, including natural user interface between human and machines/devices, and there is a strong need to improve related fundamental technologies in all aspects. Together with the new technological trends in mobility and in social computing, closely tied to audio, speech, and human language processing, we have a unique opportunity to make our publication the best among the best. Luckily, this opportunity has been offered in the expanded field of interest in our society [3], [4] a few years ago in anticipation of the human-centric computing tidal wave which is arriving. Our new innovation will focus on forging close alliance with related research fields and communities to take advantage of our unique position in the forefront of human-centric information processing. One of our ongoing efforts pertains to a potential collaboration with ACM, promising to expand the technical scope of our T-ASL in human language technology (both spoken and written) and to connect with a new huge intellectual pool directly relevant to our SPS's expanded field of interest. Another example of collaboration that already took place and has more to come is with prominent researchers in modern machine learning, which I believe will have a vast role to play in advancing all areas of our audio, speech, and human language technology. The fruit of that collaboration which I helped forge is the "deep learning" special issue you are reading now. New ways and more of such collaboration will help elevate the status of our T-ASL to new heights, and we need your help in this effort.

In relation to how to run our T-ASL more effectively, one topic under discussion and planning is to organize the editorial team (e.g., new appointments of area editors or associate EiCs) more in line with the technical areas of the Transactions. This will give three or four individual but interrelated areas of Audio, Speech, Spoken Language, and Written Human Language, all unifying under a single journal that provides synergy. The synergy of audio, speech, and language is natural not only because there is sufficient overlap in methodology, but more importantly our field has a well-established hierarchy that moves from acoustics/audio to speech to language, as a manifestation of moving from low-level signal/audio processing to higher-level, human-centric information processing as I advocated some time ago while managing SPM [3]. Further, modern research and commercial systems developed in our field (e.g., robust speech translation, spoken dialogue systems, etc.) have all components of audio, speech, and language processing, and many researchers and students are carrying out research crossing all these areas as a unity.

Just as SPM is a vehicle for SP community building, I envision that our T-ASL also be our ASL's community building device. To this end, we will further strengthen our ties with IEEE Audio/Acoustics Technical Committee (TC) and Speech/Language TC as well as with other technical organizations and get help from each other more. We will also strive to make closer connections to the conferences and workshops both in our and close to our fields. We want the best researchers to write for our T-ASL. In collaboration with relevant conference boards and organizing/review committees, we are considering setting up special recognition for outstanding conference papers with rigorous selection processes. We are also considering a more efficient T-ASL review process for these papers (e.g., leveraging the award selection reviews) so that the exceptional authors of these outstanding conference papers will have more incentive to devote their precious time in writing for our T-ASL. The desirable effect of his effort is that while it gives their work archival value, at the same time it also helps improve the quality of our T-ASL. Further, beginning from ICASSP-2013 (taking advantage of my role of General Chair), we plan to increase the conference paper limit from the historical four pages to five including one additional reference page, and make ICASSP papers more connected with our T-ASL publications. Such con-

nection makes sense because conference papers are meant to be timely with quick development to report while Transactions publications are structured to be utmost comprehensive and be the gold standard in the field to refer.

I encourage our ASL researchers to outreach to other related areas of signal processing and beyond [4]—SP theory and methods, array processing, applied math, computer science, computational linguistics, information retrieval (music, speech, text), computational phonetics and phonology, auditory science, multimodal human–machine interface, human and machine learning, artificial intelligence, etc.—and bring back useful concepts, techniques, and methodology to enrich our ASL field. As an example, an attempt to outreach to neural information processing and machine learning via co-organizing the NIPS 2009 Workshop on Deep Learning for Speech Recognition and Related Applications (http://nips.cc/Conferences/2009/Program/event.php?ID=1512) brings some very powerful techniques to speech processing. This outreach effort has become the genesis of the current T-ASL special issue in your hand. I envision human-centric computing as machine operation of human intelligence, and thus needing all kinds of multidisciplinary research. Our ASL field is an integral but partial constituent of the whole picture, focusing mainly on the auditory/sound sensory part via ears and a specialized part of cognitive brain operation via speech/language understanding. To fill in the missing sensory parts, such as visual and touch, and the remaining components of cognitive brain function, we need to reach out to and connect with relevant research such as multimodal sensing and machine learning. This will not only enrich our own ASL field but also help create greater societal impact from our ASL technology.

As I learned during my EiC tenure for SPM, one of the open secrets of running an impactful journal is to identify hot topics, and encourage the best qualified researchers to contribute as authors on these topics. We will welcome a diversity of authors from other research fields with common research methodology to ours and with a keen interest in our very challenging ASL applications. Here in this issue, we introduce our readers to one hot and emerging topic—deep learning—with one such "outsider" author as the pioneer in this area and with stellar research

achievements [5]. In coming years, our T-ASL will continue to welcome articles on novel techniques developed by researchers not traditionally associated with ASL but that are proven effective for ASL applications. We will also balance the articles with respect to their scientific value versus their contributions to fundamental technology and to applications. In addition, we will focus more on the novelty of the work submitted than on reporting incremental improvement.

The past three years have been remarkable for me working as EiC for SPM, and I look forward to continuing Helen's success and devoting more (rather than remaining) energy to T-ASL for the next three years expecting more excitement to come. As Helen said in her farewell editorial, Mari Ostendorf is assuming her new leadership role as the Vice-President of Publications, and has already done a lot of collaborative work with ACM for our T-ASL. I agree with Helen that we have a dream team, and there is no reason not to take our T-ASL to new heights and to win honors for our beloved field.

Let us ride on the new technological tidal wave, and work together to boost the value and status of T-ASL with attentive ears, dedicated mind/brain, and passionate heart. I invite you to publish your best work here, and volunteer your best service when called upon. Thank you all in advance and let's begin a new journey.

<div align="right">

Li Deng, *Editor-in-Chief*
Microsoft Research
Redmond, WA 98052 USA
deng@microsoft.com

</div>

REFERENCES

[1] J. Baker *et al.*, "Research developments and directions in speech recognition and understanding," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 75–80, May 2009.

[2] J. Baker *et al.*, "Updated MINDS report on speech recognition and understanding," *IEEE Signal Process. Mag.*, vol. 26, no. 4, pp. 78–85, Jul. 2009.

[3] L. Deng, "Expanding the scope of signal processing," *IEEE Signal Process. Mag.*, vol. 25, no. 3, pp. 2–4, May 2008.

[4] L. Deng, "Cross-pollination in signal processing technical areas," *IEEE Signal Process. Mag.*, vol. 26, no. 6, pp. 2–4, Nov. 2009.

[5] G. Hinton *et al.*, "Acoustic modeling using deep belief networks (Special Section)," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 1, Jan. 2012.

Li Deng (M'88–SM'91–F'04) joined the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 1989 as an Assistant Professor, where he became a Full Professor in 1996. In 1999, he joined Microsoft Research, Redmond, WA, where he is currently a Principal Researcher. Since 2000, he has also been an Affiliate Full Professor in the Department of Electrical Engineering, University of Washington, Seattle. Prior to Microsoft Research, he also worked or taught at the Massachusetts Institute of Technology (Cambridge, MA), ATR Interpreting Telecommunications Research Laboratories (Kyoto, Japan), Hong Kong University of Science and Technology, and Nortel (Canada). In the general areas of audio/speech/language processing, neural information processing, digital communication, and machine learning, he has published over 300 refereed papers and 3 books. He has been granted over 60 patents. He is a Fellow of the Acoustical Society of America and ISCA, and is ISCA's Distinguished Lecturer. He has received awards/honors bestowed by the IEEE, ISCA, ASA, Microsoft, and other organizations. He served on the Board of Governors of the IEEE Signal Processing Society (2008-2010). He served as Editor-in-Chief for the *IEEE Signal Processing Magazine* (2009-2011), which, according to the Thomson Reuters Journal Citation Report released June 2010 and 2011, ranks first in both years among all IEEE publications and all publications within the Electrical and Electronics Engineering Category worldwide in terms of its impact factor.