Editorial

IN THE 56 years since the inception of electronic computation, which we mark from the 1946 dedication of ENIAC, computer systems have shrunk by a factor of 1 million in power consumption, 100 million in size, and 300 million in weight. The dimensions of the familiar keyboard and mouse interfaces now dwarf the computation mechanism itself and its power supply. While we can take our computers everywhere, their limited control mechanisms prevent us from realizing their full power. For example, access to a powerful information processing system while driving is enormously appealing (for navigation, travel information, etc.), providing you don't have to risk your life by taking your eyes off the road. The same goes for all mobile information-access activities where the eyes and hands are otherwise engaged.

The value of a portable computer that could be controlled by voice, and other means as well, is clear. The problem with this idea is that if people can take these systems everywhere, they will do so, and will want them to continue to work! Thus, mobile speech recognition systems face special challenges of noise robustness, low power consumption, endurance to shock, and other physical extremes.

With these opportunities and requirements both in mind, it is our pleasure to present this collection of papers, addressing automatic speech recognition (ASR) for mobile and portable devices. In doing so, we aim to advance the possibilities and prospects for the useful application of our field. Assembling the issue has been an enlightening and enjoyable experience for us; we hope that you will find equal benefit in reading it.

We solicited and included papers on a broad range of subtopics relevant to our theme: our call for papers explicitly cited signal processing, noise robustness, acoustic and language modeling, search algorithms, physical, electronic and software design, low-cost and high-end implementations, on-board, remote or in-network recognition, and scientific and engineering issues from fundamentals to complete applications. We also expressed interest in receiving papers on topics not on this list.

Our aim in doing so was not just to broaden the issue's appeal, though of course we want a wide variety of people to pick it up and read it. But we also believe that the challenges facing the field remain daunting, and that many different perspectives are needed to address them.

The reaction to our call was quite positive: 27 manuscripts were submitted, covering a wide range of topics. Our first words of thanks go to these potential authors, and also to the nearly 70 reviewers who helped us make the final selections.

We selected 11 papers for publication. The first two papers deal with information retrieval in a mobile environment: Chang *et al.* address recognition and retrieval accuracy in the mobile environment, with special attention to microphone acoustics and cellphone transmission channel issues; they show that with high-quality microphones, retrieval precision can come close to that for perfect text input. Dharanipragada and Roukos also target information retrieval, presenting a new algorithm for open-vocabulary wordspotting.

The third and fourth papers, respectively, by Deligne *et al.*, and Varga *et al.*, address local speech recognition, wherein the recognition is performed entirely on the portable or mobile device, such as a cellphone or PDA. Both papers provide architectural overviews and discuss the resource-constraint and noise-robustness issues that are special to mobile devices.

The succeeding four papers explore the distributed speech recognition (DSR) architecture, wherein the recognition computation step is divided between a coding or front-end stage, which takes place in the local terminal or client, and the balance, which takes place in a remote server. The first two DSR papers (Bernard and Alwan, Boulis *et al.*) explore issues such as recognition in the face of transmission channel errors and packet loss. Kim *et al.* investigate a speech enhancement method that operates on the network side of a wireless communications system. The final paper in this group (Deng *et al.*) also covers robustness issues, as well as spoken language understanding, but now in the framework of a complete multimodal user interface.

The final three papers cover a variety of topics. Bessette *et al.* describe the AMR-WB speech codec standard, recently adopted by the ITU-T for wideband speech coding. Servetti and De Martin discuss a method for reduced power consumption through encryption. Finally, Wang *et al.* describe the design of a special-purpose chip for speech recognition and coding.

It is worth reflecting on what is absent from this collection. Here are a few titles and abstracts of potential papers that we would have liked to receive, but did not:

"Worst-Case, Expected-Case, and Best-Case Modalities of ASR Usage": In this paper, we assess the current state of the art in automatic speech recognition for portable and mobile devices. We consider accuracy, latency, noise robustness, speaker-independence, vocabulary size, resource requirements and power supply, and project the capabilities of the technology ten years into the future, under pessimistic, neutral, and optimistic assumptions. We identify areas of likely progress, and areas where few ideas have been put forward. Based on these projections, we speculate on new applications for the technology, with special attention to the obstacles that must be overcome to make them effective and economical.

"A System-Level Analysis of the Technology and Economics of Distributed Speech Recognition": This comprehensive study addresses the engineering of local, distributed, or centralized provisioning of speech recognition functionality. The present and anticipated computing, memory, and communication bandwidth requirements of ASR systems and applications are analyzed from the standpoint of silicon technology, electromagnetic spectrum availability, network reliability, battery life, and

Digital Object Identifier 10.1109/TSA.2002.807734

the political and economic sovereignty of telecommunication service providers. We draw conclusions regarding the appropriate economic and functional match for each architecture, with emphasis on network management requirements, performance tradeoffs, and security.

"Principles of Engineering Error-Robust Automatic Speech Recognition Systems": In virtually every other field of engineering, the limits of the underlying materials or methods—such as the strengths and failure modes of construction materials—are systematically assessed. These assessments then feed the design of the desired artifact (e.g., a bridge), which is built to meet the specified limits of reliability, endurance, cost, and so on. This mode of analysis is glaringly absent from the design of products and services for automatic speech recognition. In this paper, we attempt to develop such an analysis. We both explore its application to speech recognition, and identify the challenges that prevent us from carrying the approach to its logical conclusion.

These imaginary papers would be broader in scope, and possibly less mathematical and algorithmic in their orientation, than the papers that appear here. However, we believe that serious study of any of these topics, by technical experts, would be of enormous value in advancing applications of ASR. We invite our readers to take up these challenges, and look forward to reading discussions of these and related issues in future numbers of these TRANSACTIONS.

As a final word, we thank Fred Juang, José M. F. Moura, and Kathy Jackson, for essential guidance and assistance in assembling this issue.

HARRY PRINTZ
Agile TV Corporation
333 Ravenswood Ave.
Menlo Park, CA 94025 USA
printz@agile.tv

ISABEL TRANCOSO
INESC-ID/IST
R. Alves Redol, 9
Lisboa, 1000-029 Portugal
Isabel.Trancoso@inesc-id.pt

Harry Printz received the B.A. and M.A. degrees in physics from Harvard University, Cambridge, MA, in 1978, the B.A. degree in mathematics and philosophy from the University of Oxford, Oxford, U.K., in 1980, and the Ph.D. degree in computer science from Carnegie Mellon University, Pittsburgh, PA, in 1991.

From 1991 to 1993, he was with the DEC Paris Research Center, Rueil-Malmaison, France. From 1993 to 2001 he was with the IBM T. J. Watson Research Center, where he worked on machine translation and speech recognition, and managed research for the IBM embedded speech recognition product. He is presently Vice President of Speech and Language Technology at AgileTV Corporation, Menlo Park, CA.

Isabel Trancoso received the Licenciado, Mestre, Doutor, and Agregado degrees in electrical and computer engineering from the Instituto Superior Técnico, Lisbon, Portugal, in 1979, 1984, 1987, and 2002, respectively.

She has been a Lecturer at this University since 1979, where she currently teaches speech processing courses and coordinates the EEC course. She is also a Senior Researcher at INESC ID Lisbon, having launched the Speech Processing Group, now restructured as Spoken Language Systems Lab, in 1990. Her first research topic was medium-to-low bit rate speech coding. From October 1984 to June 1985, she worked on this topic at AT&T Bell Laboratories, Murray Hill, NJ. Her current scope is much broader, encompassing many areas in speech recognition and synthesis, with a special emphasis on tools and resources for the Portuguese language.

Dr. Trancoso was a member of the International Speech Communication Association (ISCA) Board (1993–1998). She is a member of the IEEE Speech Technical Committee (since 1999), and the Permanent Council for the organization of the International Conferences on Spoken Language Processing (since 1998).