"Trends" Expert Overview Sessions Revived at ICASSP 2011

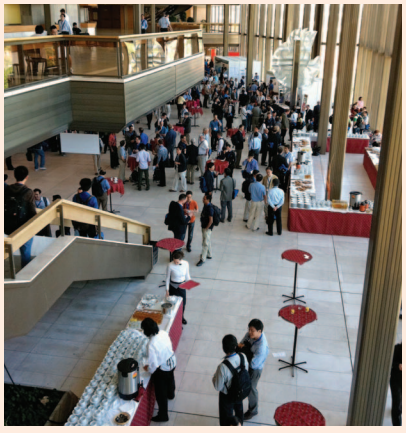
INTRODUCTION

The breadth of topics in signal processing continues to grow, propelled by advances in digital signal processing theory as well as in new application domains. Researchers are therefore bombarded by new information in their areas of interest, but it is rather difficult to achieve a perspective of trends and advances in collateral areas. At the 2011 IEEE ICASSP in Prague, the program committee revived the idea of expert overview sessions called "Trends," which are summary talks authored by the IEEE Signal Processing Society Technical Committees and presented by their keynote experts. As drivers of the technology areas, the Technical Committees have the obvious expertise to gauge the advances and to put in perspective the directions of future research, also due to their vantage point of reviewers of papers.

The expert overview sessions were very well attended during the conference, which is an indication of the usefulness of the initiative. The ICASSP 2011 program committee thought that these sessions should be captured in writing and made available to the members of the Society. Therefore we contacted the editors of *IEEE Signal Processing Magazine,* who welcomed the initiative. Starting with this issue, readers will have an opportunity to access the summaries of the Trends sessions authored by the Technical Committees. Enjoy reading the summaries, and let us know your thoughts!

Alle-Jan van der Veen and Jose C. Principe

—Note: Additional multimedia resources for these sessions can be found at http://www.signalprocessingsociety.org/publications/periodicals/spm/columns-resources/.



The "Trends" expert overview sessions were well attended at ICASSP 2011 in Prague.

Digital Object Identifier 10.1109/MSP.2011.941984
Date of publication: 22 August 2011

Audio and Acoustic Signal Processing

Malcolm Slaney and
Patrick A. Naylor

Many of the technology trends we see today in audio and acoustic signal processing (AASP) can be traced back to visionary ideas of those we now call luminaries. Key figures in our community include Colin Cherry (1914–1979) for defining the cocktail party problem; Manfred Schroeder (1926–2009) for contributions to understanding of acoustics, hearing, music, and speech coding; and many other researchers equally important but too numerous to list.

The AASP area has seen strong growth in research activity in recent years. Drivers for such growth include rapidly increasing customer expectations for quality, mobility, and functionality for audio in communications and entertainment scenarios. Many consumers have come to expect that there is really nothing that cannot be done with audio, whether it's music or speech, providing your battery holds out. This "demand" is coupled with the "supply" of low-cost, high-performance audio processing capability.

The number of AASP submissions to ICASSP grew again this year by 25% with papers in speech enhancement, source separation, array processing, and music analysis being the most numerous. Similar trends can be seen for *IEEE Transactions on Audio, Speech, and Language Processing*.

(continued on page 150)

Digital Object Identifier 10.1109/MSP.2011.941985
Date of publication: 22 August 2011

MICROPHONE ARRAYS

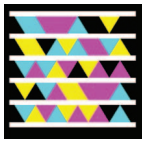
Two long-standing aspects of microphone array signal processing continue to challenge our community: 1) source localization and tracking and 2) source separation. In a new approach to aid robustness to room acoustic effects, environment-aware acoustic sensing uses microphone arrays to infer the physical geometry of the acoustic environment to compute, for example, information on the location of reflective surfaces in a room. A welcome and bold trend towards analyses of alternative geometries for microphone arrays is also clear. Research into distributed acoustic sensing as well as optimal geometries for directivity of fixed arrays has generated significant outcomes.

SOURCE SEPARATION AND SIGNAL ENHANCEMENT

A solution to the cocktail party problem requires separation and enhancement of multiple talkers in reverberant and noisy scenarios. Dereverberation continues to be a significant research thread. Spatial and statistical blind source separation (BSS) methods coupled with clustering approaches have led to impressive results even in underdetermined cases with significant reverberation. Key achievements include methods to solve the inherent permutation problem, improved efficiency of iterative optimization schemes, and the integration of BSS techniques with the sensor network paradigm. It is interesting to observe both a broadening of the use of multichannel methods as well as an ongoing interest in single channel source separation—the practical value of such algorithms still being dominant in several important applications—leading to new single-channel approaches that exploit design criteria previously applied in the multichannel case.

MUSIC INFORMATION RETRIEVAL

Video



In the last few years, music analysis and applications have become much more important at ICASSP. One reason for this is a matter of marketing—the Editors Information Classification Scheme (EDICS) for ICASSP was changed to give music analysis an explicit home. It doesn't hurt that digital music has become a large and commercially relevant industry. But other reasons are more technical.

Most importantly, music is a good sandbox for researchers. All of us have at least modest collections of data at home. Music presents a problem that all of us care about and the complexity of the task varies from simple tone complexes to full orchestrations, to combinations of music and speech. This led people to work on all sorts of problems such as chord and key recognition, tagging music with labels such as genre or emotional content, and sound separation and transcription.

Second, large collections of research data are now available. Copyright laws prevent the large-scale distribution of music data, like those distributions that have made speech recognition or video classification research so successful. Instead, the music information retrieval evaluation exchange (MIREX) project brings researcher's algorithms to a large data collection housed at the University of Illinois (http://music-ir.org/mirex). Each year, one or two dozen different competitions are held in all areas of music analysis. This competition has become the standard benchmark for music analysis. Some of the best papers also come to ICASSP. More recently, a "Million-Song Database" has been published. Instead of distributing music, the music is analyzed and only the features are distributed. We expect many new ideas to develop based on this data set.

SPARSE REPRESENTATIONS

An interesting trend during the last decade is the increasing interest in sparse and overcomplete signal representations. Malcolm Slaney, an author of this summary, has to admit to feeling bemusement upon hearing the first work on matching pursuit—it seemed like a lot of work for not much information. Since then, compressive sampling has made the idea much more appealing. In our domain, the value of sparsity has been demonstrated in several different ways including perception, statistical models, and regularization.

It is worth noting that the auditory cortex evolved into the very first sparse representation. We are still trying to understand how this overcomplete representation helps us perceive the world. Other researchers have used deep belief networks to model the statistical codependencies between different parts of a signal. Finally, L1 regularization is a means to encourage a sparse representation and thus a simple "explanation" for the signal.

We suspect that explaining the benefit of sparsity is much like the blind men and the elephant. One explanation we find intriguing is the possibility that an overcomplete basis set, and thus a sparse explanation of the signal, promotes a highly specific explanation of the data. While a signal model based on an orthogonal basis representation can approximate any signal in the spanned subspace, whether it is seen in real life or not, a sparse overcomplete representation is very specific to the training data. Thus, sparsity allows nonlinear data manifolds to be modeled and then ignore data that is not seen in training.

CONCLUSIONS

Yes, many research questions new and old remain unanswered. Many topics, however, are presented to deliver new applications and new products leveraging AASP research. Thinking purely in terms of volume, the mobile devices market for audio technology is undoubtedly huge. New initiatives for the living room also present great potential for both communications and entertainment.

AUTHORS

Malcolm Slaney (malcolm@ieee.org) is a principal scientist at Yahoo! Research.

Patrick A. Naylor (p.naylor@imperial.ac.uk) is a reader in speech and audio signal processing at Imperial College London and a (consulting) professor at Stanford University.

Slides