



Editorial

Introduction to the special issue on the recognition and organization of real-world sound

Classification and recognition of acoustic signals—particularly speech—has reached a level of sophistication enabling many practical applications. The best recognition systems, however, make strong assumptions about the nature of the signal and any possible interference or noise. At the same time, there is a growing body of research concerned with a more general class of audio signals, and dealing with them within a more flexible framework. From noise-robust speech recognition, through automatic transcription of recorded music, to various forms of source separation, this research acknowledges that getting information from sound is more than a simple pattern recognition problem—not least because even defining what information is desired is a challenge, and in almost every case the information must be extracted at many overlapping levels.

The eight papers in this Special Issue represent a spectrum of work unified by the common theme of handling real-world sounds, with all its implicit complications. They cover a range of different processing strategies and a range of different acoustic analysis tasks which jointly capture both the variety and the commonality of problems raised by real-world sounds.

One of the main themes tackled in this collection of papers is that of missing data. Many studies of robust ASR using missing data processing have used a modified recognizer to handle partial evidence in the spectral domain and have not been

able to exploit the well-known benefits of cepstral features. In their paper, Raj et al. demonstrate two algorithms which reconstruct complete spectral vectors from partial information, allowing the use of cepstral features and unmodified recognisers. The paper by Palomaki et al. applies missing data processing to the problem of robust ASR in the presence of both additive noise and reverberation through the use of a binaural auditory model with precedence-effect processing.

A fundamental problem in missing-data based ASR is the need to determine which spectrographic regions are reliable and which are noise-corrupted. Seltzer et al. present a Bayesian classification approach to determining the reliability or otherwise of such features. Barker et al.¹ show how decisions about whether spectrographic regions belong to the target or background source can be left to the decoding stage. Their paper introduces a probabilistic framework for decoding one source in the presence of other sources.

Two papers in this collection describe the processing of dense acoustic mixtures using only one

¹ The Publisher regrets that the following article: “Decoding speech in the presence of other sources” by J.P. Barker, M.P. Cooke and D.P.W. Ellis is not published in this Special Issue on the Recognition and Organization of Real-World Sound. The article is published in Volume 44, Issue 1, October 2004 and can be viewed by visiting the following link: <http://dx.doi.org/10.1016/j.specom.2004.05.002>.

or two microphones. This is a challenging domain because of the high degree of spectral and temporal overlap of sources, which causes even those algorithms developed for several concurrent sources to break down. Goto's paper tackles the problem of estimating melody and bass lines in polyphonic music signals, while Okuno et al. explore the effect of increasing the number and types of sensors (monaural, binaural, visual) on recognition rates for mixtures of three simultaneous talkers.

How well do listeners identify real-world sounds or their properties? Most studies with real-world sounds to date have used musical or speech signals, but two papers in this Special Issue deal with this question using very different stimuli. Collier compares the ability of novices and experts in classifying sonar signals as being of biological or man-made origin. Houben et al. studied the perception of the size and speed of rolling balls. These studies raise important questions about our ability

to extract source properties from acoustic signals, the balance between general purpose and source-specific processing, and the related issue of innate versus learned abilities.

In summary, this collection of papers represents a snapshot of progress in a diverse range of applications, unified by the fundamental challenge of making sense of everyday sounds.

M.P. Cooke

*Department of Computer Science
University of Sheffield Regent Court
211 Portobello Street, Sheffield, S1 4DP, UK
Tel.: +44 114 222 1800; fax: +44 114 222 1810
E-mail address: m.cooke@dcs.shef.ac.uk*

D.P.W. Ellis

*Department of Electrical Engineering
Columbia University
500 W. 120th Steet, New York
NY 10027, USA*