BOOK REVIEWS

P. L. Marston

Physics Department, Washington State University, Pullman, Washington 99164

These reviews of books and other forms of information express the opinions of the individual reviewers and are not necessarily endorsed by the Editorial Board of this Journal.

Computational Auditory Scene Analysis: Principles, Algorithms and Applications

Edited by DeLiang Wang and Guy J. Brown

This excellent book consists of ten chapters by various authors on the progress that has been made in computationally analyzing sound mixtures into their component sources. This challenging problem is important in speech recognition, psychoacoustics, neuroscience, and hearing aid technology, and both introductory and advanced readers from all these fields will find many of the chapters of this book both comprehensible and enlightening.

A comprehensive and very well-written introductory chapter by the book's editors Wang and Brown whets the reader's appetite. It provides an overview of Computational Auditory Scene Analysis (CASA), reviewing background material from human auditory scene analysis (ASA) and providing a succinct summary of the relatively short history of CASA systems, their goals and their strategies. A recurrent theme at the heart of the book concerns the recognition of speech. Although human listeners are remarkably tolerant of additional sounds being present at the same time as the speech they wish to attend to, machine algorithms for speech recognition are not. Additional sounds corrupt the algorithms' inclusive input data representations and severely degrade performance. As Jon Barker says in his illuminating chapter on Robust Speech Recognition: "speech recognition remains a brittle technology." Many of the authors of the chapters in this book have been at the forefront of work that has attempted a radical solution to this problem. Taking their lead from psychoacoustic work on auditory scene analysis, in particular work covered in Albert Bregman's 1990 book of that title, they have endeavored to use simple properties of sounds, such as harmonicity, onset-time, azimuthal direction, and continuity to perform an allocation of the components in a sound mixture to their respective sources. This "bottom-up" symbolic approach initially met with limited success, at least compared to the best performance of statistical methods such as Independent Components Analysis. However, these statistical methods are themselves brittle, failing badly when the conditions for which they were designed are not met. CASA holds out the promise at least of a more robust, if partial, solution. Although, as Jon Barker points out, there are major problems in interfacing scene analysis modules with existing speech recognition methods, one of the exciting developments that appears in the book is the combination of CASA with the "missing-data" approach to speech recognition (pioneered by Martin Cooke). The issues raised at this interface are important for psychologists interested in the role that ASA plays in speech recognition in humans. Other chapters within this theme provide more technical descriptions of Feature-Based Speech Segregation by Wang and of Model-Based Scene Analysis by Daniel Ellis. These two chapters represent the two opposing approaches to the problem of separation proposed by Bregman, the first using simple properties that the sounds from a particular source share, the second, schema-based, using knowledge about specific sounds to find them in a mixture. The issues addressed here and the proposed solutions should stimulate not only those involved in speech recognition, but also psychoacousticians and other psychologists interested in auditory perception in the presence of competing sounds. The last two decades have seen a flourishing of research addressing the perception of both speech and the basic properties of complex sounds in the context of sound mixtures. There is presently a healthy interchange between this experimental work on auditory scene analysis and the computational work. For example, DeLiang Wang's concept of an ideal binary mask that reveals only those parts of a spectrogram that are dominated by a particular sound source has been a fruitful one not only in CASA but also in experiments on the recognition of speech mixed with that of other talkers. The sparse nature of harmonic sounds such as speech contributes to the success of the binary mask approach.

Another major theme of the chapters in this book is the way that a difference in azimuthal direction is exploited in scene analysis by both human listeners and computer algorithms. Some of the human data are well-reviewed by Richard Stern, Brown, and Wang who rightly raise the question of the relationship between spatial localization and scene analysis. Human listeners certainly use perceived location to help track a particular sound source over time, but how a particular source gets localized at a particular time is something of a puzzle. Do we use location to help simultaneous grouping, or do we group sounds before localizing them? Binaural information about azimuth in any one frequency channel is readily disturbed by, for example, reverberation. So although localization-based grouping works well in acoustically dead environments, as described by Albert Feng and Douglas Jones in their chapter, cues such as interaural time difference (ITD) degrade with reverberation and the grouping suffers. Perhaps as a consequence, human listeners show only weak simultaneous grouping effects with ITD, although it is the dominant cue for azimuthal localization of complex sounds, and also show some evidence for simultaneous grouping preceding localization. The varied effects of reverberation on a range of perceptual abilities and on machine speech recognition are very clearly reviewed in a chapter by Brown and Kalle Palomäki. They argue that in order to successfully combat the effects of reverberation, recognition algorithms need to exploit monaural grouping cues more and reduce their reliance on spatial differences. For example, recent developments in beamforming exploit the sparse nature of harmonic sounds such as speech to identify spectral regions that minimize the adulteration caused by reverberation. Brown and Palomäki also argue that the masking effect of late reflections can be profitably handled within a missing-data/binary mask framework.

A final theme of the book is the application of CASA to problems in music processing. Relevant to this theme as well as to speech is a careful and thorough chapter by Alain de Cheveigné on techniques for the estimation of the pitches of multiple simultaneous sounds; it points out the common principles underlying an apparently wide diversity of methods. Masataka Goto's chapter on the Analysis of Musical Audio Signals primarily reviews in some technical detail his approach to extracting melody and bass lines, beat structure and larger sectional structure from music.

The book's closing chapter by Brown and Wang covers Neural and Perceptual Modelling. It provides a useful review of the use of neural oscillator models to solve the "binding problem" posed by auditory scene analysis, and also covers autocorrelation approaches to vowel separation. It recognizes the somewhat surprising dearth of models for top-down or schema-based grouping and the need to integrate the two approaches, echoing Barker's comments on the challenge of integrating low-level grouping in CASA with model-based recognition algorithms.

In summary, the editors should be congratulated on assembling a well-integrated set of chapters that provide a comprehensive, comprehensible, and astute summary of the state of CASA and its relationship to human auditory scene analysis. As a bonus, the book has an associated website (www.casabook.org) which includes audio demonstrations and software downloads.

CHRIS DARWIN
Department of Psychology
University of Sussex
Brighton, BN1 9QG, U.K.

Wave Scattering by Small Bodies of Arbitrary Shapes

Alexander G. Ramm

World Scientific Publishing, Hackensack, 2005. xviii+293 pp. $66 (hardcover), ISBN: 9812561862.

As a reader of the *Journal of the Acoustical Society of America*, what would you expect this book to contain? The title suggests a coherent survey of the literature on the scattering of sound waves by obstacles that are small compared to the wavelength, the domain of "low frequency scattering." If you want such a survey, consult the fine book with this title by G. Dassios and R. E. Kleinman (Oxford University Press, 2000). Professor Ramm's book is concerned with the same topic, but its aims are quite different. Thus, the "book is based mostly on the author's papers and results" (p. ix). On the back cover, we read "Applications of these results to ultrasound mammography and electrical engineering are considered," but this reviewer could not find such applications in the book.

So, what is in the book? There are 12 chapters: 1. Basic problems (11 pages); 2. Iterative processes for solving Fredholm's integral equations for static problems (11); 3. Calculating electric capacitance (17); 4. Numerical examples (9); 5. Calculating polarizability tensors (15); 6. Iterative methods: Mathematical results (23); 7. Wave scattering by small bodies (31); 8. Fredholm alternative and a characterization of Fredholm operators (12); 9. Boundary-value problems in rough domains (26); 10. Low frequency asymptotics (30); 11. Finding small inhomogeneities from scattering data (7); and 12. Modified Rayleigh conjecture (19). There is a 53-page appendix on quadrature rules for multi-dimensional integrals and a bibliography of 165 items (75 of which are authored or co-authored by Ramm). Chapters 8 and 9 seem to have little to do with the book's title.

The first seven chapters are a re-typed version of Ramm's earlier book, "Iterative Methods for Calculating Static Fields and Wave Scattering by Small Bodies" (Springer, New York, 1982). Everything else is taken almost exactly from his published papers. Indeed, the last five chapters and the appendix all contain statements such as "the presentation follows closely" (p. 138) followed by a citation to Ramm's papers; closer inspection (conveniently, many of his papers are available from his website) suggests that the word "closely" should be replaced by "exactly"! One minor virtue of re-packaging a number of papers is that each chapter is self-contained. On the other hand, there is little cross-referencing, no attempt at synthesis of an overview of the subject, and pointless repetition (compare p. 202 and p. 210). In addition, some quantities have several notations. For example, chapter 10 opens with "In this chapter the exterior domain $D' = D_e$ is denoted by Ω": apparently, the author could not be bothered to edit his text so as to use a common notation throughout.

In summary, this is a disappointing book. It contains nothing that is unavailable elsewhere. I cannot recommend it.

P. A. MARTIN
Department of Mathematical and Computer Sciences
Colorado School of Mines
Golden, Colorado 80401-1887