Studies in Computational Intelligence 3

Bozena Kostek

Perception-Based Data Processing in Acoustics

Applications to Music Information Retrieval and Psychophysiology of Hearing

Springer

Bożena Kostek

Perception-Based Data Processing in Acoustics

Applications to Music Information Retrieval and Psychophysiology of Hearing

Springer

Professor Bożena Kostek

Multimedia Systems Department
Faculty of Electronics,
Telecommunications and Informatics
Gdansk University of Technology
ul. Narutowicza 11/12
80-952 Gdansk, Poland
and
Institute of Physiology and Pathology of Hearing
ul. Pstrowskiego 1
01-943 Warsaw, Poland
E-mail: bozenka@sound.eti.pg.gda.pl
http://sound.eti.pg.gda.pl

Library of Congress Control Number: 2005926348

ISSN print edition: 1860-949X
ISSN electronic edition: 1860-9503
ISBN-10 3-540-25729-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-25729-5 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in The Netherlands

Typesetting: by the authors and TechBooks using a Springer LaTeX macro package

Printed on acid-free paper SPIN: 11412595 89/TechBooks 5 4 3 2 1 0

Studies in Computational Intelligence, Volume 3

Editor-in-chief
Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series
can be found on our homepage:
springeronline.com

Vol. 1. Tetsuya Hoya
Artificial Mind System – Kernel Memory Approach, 2005
ISBN 3-540-26072-2

Vol. 2. Saman K. Halgamuge, Lipo Wang (Eds.)
Computational Intelligence for Modelling and Prediction, 2005
ISBN 3-540-26071-4

Vol. 3. Bożena Kostek
Perception-Based Data Processing in Acoustics, 2005
ISBN 3-540-25729-2

To my Father

FOREWORD

Recent years brought up many new techniques combining various aspects in so-called cognitive processing. This way of computing can be used with success in many areas of science and engineering by offering better analogy to human-like processing of information. Such an approach may be especially interesting in acoustics, where we deal with very inaccurate perceptions of phenomena due to the hearing sense characteristics which are highly imprecise with regard to time and spectral resolution.

This book demonstrates in which way soft computing methods (e.g. rough set-based methods) can be applied to provide flexible information processing capabilities for handling ambiguous decision making problems such as for examples musical signal processing, pattern classification, feature and rule generation. Methods of integrating rough sets, fuzzy and artificial neural networks for efficient knowledge discovery are also shown.

Fuzzy logic provides yet another tool that seems one of the best solutions for processing such inaccurate information as can be found in acoustics. In many domains building up membership functions could be problematic, however in acoustics the so-called subjective testing provides a good solution to this problem. Even, if such a testing is time consuming, it falls into the realm of human expertise, thus providing a class of perceptual membership functions. In addition, rules are quite obvious in the acoustic domain, and if not, they can be mined using other soft computing techniques, such as, for example, rough sets. In such a way it is possible to mimic human hearing perception and the way of processing perceived information in human brains.

The book addresses a number of topics such as the fundamentals of hearing and music perception, musical data representation and analysis, automatic classification of musical instrument sounds and musical phrases, musical sound separation, automatic recognition of musical styles, sound processing in hearing prostheses based on artificial neural networks, rough sets fuzzy logic principles, and others – based on cognitive approach. A review of soft computing and data mining techniques is provided, including all mentioned methods and others such as decision trees, evolutionary processing, and genetic algorithms. This book provides however a bal-

anced mixture of both theory and review of applications along with extensive bibliography.

The author has attempted to address a broad group of individuals with the common interest – music and hearing perception and sound processing based on cognitive approach. This is an important issue because of increasing specialization in science, however this book may be read by specialists from various domains and give them a comprehensive view on presented subjects.

This is a book of many new ideas and results, that will be of interest to all those interested in modern approach to imperfect data analysis. The author deserves highest appreciation for her valuable work.

Zdzisław Pawlak

PREFACE

The past can be pondered, the future must be created

E. Schillebeeckx

The emerging concept of human-centered computing or anthropomorphic approach represents a significant move towards intelligent systems, and affords a new perspective on information technology. Relationships between the human brain, mind and perception that have the potential of enhancing peoples' cognitive performance can be found in many domains, examples of which will be shown in relation to music processing and classification. On the other hand, it would be advisable to design systems capable of imitating perceptual processes that are best adapted to specific technological problems.

The objective of this monograph is to provide novel insights into perceptual mechanisms underlying the processing of sound and music in different environments. A solid understanding of these mechanisms is vital for numerous technological applications such as, for example, information retrieval from distributed musical databases. In order to investigate the cognitive mechanisms underlying music perception some soft computing methods will be used. The system proposed by the Author, based on the rough set method and fuzzy logic, provides knowledge on how humans internally represent such notions as quality and timbre and therefore it allows the human-like automatic processing of musical data. In addition, the automatically extracted knowledge on the above processes can be compared to fundamentals of hearing psychophysiology and to principles of music perception. Also other applications of hybrid decision systems to problem solving in music and acoustics will be exemplified and discussed in this book based not only on the review of some literature sources, but also on the experimental results obtained in the Multimedia System Department, Gdansk University of Technology.

The aim of this book is to show examples of the implementation of computational intelligence methods in musical signal and music analysis, as well as in the classification tasks. A part of this book contains a short review of perceptual bases of hearing and music. Then methods and techniques that can be classified as computational intelligence or machine-learning are shortly introduced. The presented methods are applied in the areas considered to be most relevant to music information retrieval (MIR) and acoustics. Accordingly, methods based on such learning algorithms as neural networks, rough sets, fuzzy-logic, and genetic algorithms were conceived, implemented and tested on musical data. In addition, the above-mentioned methods were applied to the analysis of musical duets, musical phrases and audio signals. Another problem discussed within the framework of this book is the 'computing with words' concept applied to both acoustics and psychophysiology. Perception-based analysis applied to psychophysiology focuses on the evaluation of hearing impairments. Application of neural networks to the processing of beamformer signals is another issue reviewed in this book. The last application described is devoted to the problem of audio-visual correlation search. This is based on a hybrid system consisting of rough-fuzzy and evolutionary computation methods.

ACKNOWLEDGMENTS

I am indebted to Professors Z. Pawlak, H. Skarzynski and A. Skowron for their encouragement and interest in this research work.

I am very thankful to Professor J. Kacprzyk who supported the idea of this new book from the beginning.

Much gratitude I owe my distinguished Teachers of Sound Engineering and Acoustics – Professors M. Sankiewicz and G. Budzyński for introducing me to these interesting, interdisciplinary scientific domains.

I thank all my colleagues and students from the Multimedia Systems Department of the Gdansk University of Technology for the discussions and support.

I would also like to thank my husband – Prof. A. Czyżewski for his encouragement and for sharing scientific interests with me.

Finally, I would like to express my gratitude to the members of the Editorial of the Springer-Verlag, in particular Shanya Rehman, Heather King Dr. Thomas Ditzinger, who supported me in a very professional way.

January, 2005
Gdańsk, Poland

Bożena Kostek

CONTENTS

1 INTRODUCTION

Over the last decade, a series of publications has brought and established new research areas related to music, and intensified the research verging on several disciplinary boundaries, typically dealt with separately. The explosion of collaboration and competition was triggered by the Internet revolution. Research achievements published in the Internet, along with audio and video available through the Internet have made research more efficient. This creates enormous possibilities and synergy. Also standards are more easily defined and implemented. On the other hand, content search of the Internet resources must in response bring new solutions to the problem – most possibly in the form of new standards and technology. Among new emerging areas are: Music Information Retrieval (MIR), Semantic Audio Analysis (SAA), music ontology, and many others. Music Information Retrieval refers to data extraction and retrieval from musical databases found on the Internet. The MIR strategic plans were defined and re-defined many times. Strong collaboration, and at the same time strong competition, afforded solutions to many problems defined within the scope of MIR, and overcame some of the largest obstacles found in this field. In addition, these problems have been addressed by technology, thus no research plans have been immune to the demands of an increasingly competitive technology environment.

There exist several definitions on semantic audio analysis. In one of them SAA means the extraction of features from audio (live or recorded) that either have some relevance to humans (e.g. rhythm, notes, phrases) or some physical correlate (e.g. musical instruments). This may be treated as complementary to human-entered metadata. In order to differentiate between human-entered metadata and semantic data, the latter issue constitutes a form of 'technical metadata', which can accompany a recording or broadcast. Thus metadata are important elements of SAA, and should cover both the extraction of features and their semantic representation. This book will highlight examples where SAA can supplement interactions with music and audio.

Human communication includes the capability of recognition. This is particularly true of auditory communication. Information retrieval can be

Bożena Kostek: *Perception-Based Data Processing in Acoustics*, Studies in Computational Intelligence (SCI) **3**, 1–5 (2005)
www.springerlink.com

investigated with cognitive systems engineering methodologies. Music information retrieval turns out to be particularly challenging, since many problems remain unsolved to this day.

Topics that should be included within the scope of the aforementioned areas include: automatic classification of musical instrument sounds and musical phrases/styles, music representation and indexing, estimating musical similarity using both perceptual and musicological criteria, recognizing music using audio and/or semantic description, building up musical databases, evaluation of MIR systems, intellectual property rights issues, user interfaces, issues related to musical styles and genres, language modeling for music, user needs and expectations, auditory scene analysis, gesture control over musical works, and many others. Some topics contained within the notion of MIR are covered by the MPEG-7 standard, which provides description of the multimedia content in order to support better interpretation of information.

It should be stressed that solving these problems requires human assistance. Many features of multimedia content description are based on perceptual phenomena and cognition. The preparation of format description, both numerical and categorical, is done on the basis of understanding the problem area. Information retrieval systems are presupposed to give an exact match to documents involving the same cues to the user query. However, operations, which are behind the query do not always provide good responses to the user's interest. This means that retrieving multimedia content on the basis of descriptors would also require human assistance. Decision systems may produce numerous rules generated in the mining process. This necessitates the provision of the generated rules for post-processing. Another problem which needs attention is the processing of unknown, missing attribute values or incomplete data when acquiring knowledge from databases. To improve information retrieval quality, various strategies were proposed and used, such as probabilistic, clustering and intelligent retrieval. The latter technique often uses concept analysis requiring semantic calculations.

The MPEG-7 standard refers to metadata information contained in the Internet archives. This notion is often applied to the value-added information created to describe and track objects, and to allow access to those information objects. In this context descriptors that are well-defined provide means for better computing and improved users interfacing and data management. It can easily be observed that these low-level descriptors are more data- than human-oriented. This is because the idea behind this standard is to have data defined and linked in such a way as to be able to use it for more effective automatic discovery, integration, and re-use in various applications. The most ambitious task is, however, to provide seamless