

Computational Analysis of Sound Scenes and Events

Tuomas Virtanen • Mark D. Plumbley • Dan Ellis
Editors

Computational Analysis of Sound Scenes and Events

 Springer

Editors

Tuomas Virtanen
Laboratory of Signal Processing
Tampere University of Technology
Tampere, Finland

Mark D. Plumbley
Centre for Vision, Speech
and Signal Processing
University of Surrey
Surrey, UK

Dan Ellis
Google Inc.
New York, NY, USA

ISBN 978-3-319-63449-4 ISBN 978-3-319-63450-0 (eBook)
DOI 10.1007/978-3-319-63450-0

Library of Congress Control Number: 2017951198

© Springer International Publishing AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The recent progress on machine learning and signal processing has enabled the development of technologies for automatic analysis of sound scenes and events by computational means. This has attracted several research groups and companies to investigate this new field, which has potential in several applications and also has several research challenges. This book aims to present the state-of-the-art methodology in the field, to serve as a baseline material for people wishing to enter it or to learn more about it.

We would like to thank all the authors of the chapters of this book for their excellent contributions. We gave you hard times by making several requests, many of which were quite laborious to address. We would specifically like to thank those authors who agreed to help by cross-reviewing other chapters to make this book coherent. We would also like to thank the external reviewers, Joonas Nikunen, Guangpu Huang, Benjamin Elizalde, Mikko Parviainen, Konstantinos Drossos, Sharath Adavanne, Qiang Huang, and Yong Xu.

Tampere, Finland
Surrey, UK
New York, NY, USA
June 2017

Tuomas Virtanen
Mark D. Plumbley
Dan Ellis

Contents

Part I Foundations

- 1 Introduction to Sound Scene and Event Analysis** 3
Tuomas Virtanen, Mark D. Plumbley, and Dan Ellis
- 2 The Machine Learning Approach for Analysis of Sound Scenes and Events** 13
Toni Heittola, Emre Çakır, and Tuomas Virtanen
- 3 Acoustics and Psychoacoustics of Sound Scenes and Events** 41
Guillaume Lemaitre, Nicolas Grimault, and Clara Suied

Part II Core Methods

- 4 Acoustic Features for Environmental Sound Analysis** 71
Romain Serizel, Victor Bisot, Slim Essid, and Gaël Richard
- 5 Statistical Methods for Scene and Event Classification** 103
Brian McFee
- 6 Datasets and Evaluation** 147
Annamaria Mesaros, Toni Heittola, and Dan Ellis

Part III Advanced Methods

- 7 Everyday Sound Categorization** 183
Catherine Guastavino
- 8 Approaches to Complex Sound Scene Analysis** 215
Emmanouil Benetos, Dan Stowell, and Mark D. Plumbley
- 9 Multiview Approaches to Event Detection and Scene Analysis** 243
Slim Essid, Sanjeel Parekh, Ngoc Q.K. Duong, Romain Serizel, Alexey Ozerov, Fabio Antonacci, and Augusto Sarti

Part IV Applications

10 Sound Sharing and Retrieval 279
Frederic Font, Gerard Roma, and Xavier Serra

11 Computational Bioacoustic Scene Analysis 303
Dan Stowell

12 Audio Event Recognition in the Smart Home 335
Sacha Krstulović

13 Sound Analysis in Smart Cities 373
Juan Pablo Bello, Charlie Mydlarz, and Justin Salamon

Part V Perspectives

14 Future Perspective 401
Dan Ellis, Tuomas Virtanen, Mark D. Plumbley, and Bhiksha Raj

Index 417

Contributors

Fabio Antonacci Politecnico di Milano, Milano, Italy

Juan Pablo Bello Music and Audio Research Laboratory, New York University, New York, NY, USA

Emmanouil Benetos School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK

Victor Bisot LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France

Emre Çakır Tampere University of Technology, Tampere, Finland

Ngoc Q.K. Duong Technicolor, Rennes, France

Dan Ellis Google Inc, New York, NY, USA

Slim Essid LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France

Frederic Font Music Technology Group (MTG), Universitat Pompeu Fabra, Barcelona, Spain

Nicolas Grimault UMR CNRS 5292, Centre de Recherche en Neurosciences de Lyon, Université Lyon 1, Lyon Cedex, France

Catherine Guastavino School of Information Studies, McGill University, Montreal, QC, Canada

Toni Heittola Tampere University of Technology, Tampere, Finland

Sacha Krstulović Audio Analytic Ltd., Cambridge, UK

Guillaume Lemaitre STMS-Ircam-CNRS-UPMC, Paris, France

Brian McFee Center for Data Science, New York University, New York, NY, USA

Annamaria Mesaros Tampere University of Technology, Tampere, Finland

Charlie Mydlarz Center for Urban Science and Progress & Music and Audio Research Laboratory, New York University, New York, NY, USA

Alexey Ozerov Technicolor, Rennes, France

Sanjeel Parekh Technicolor, Rennes, France

Mark D. Plumbley Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey, UK

Bhiksha Raj Carnegie Mellon University, Pittsburgh, PA, USA

Gaël Richard LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France

Gerard Roma Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, Guildford, UK

Justin Salamon Center for Urban Science and Progress & Music and Audio Research Laboratory, New York University, New York, NY, USA

Augusto Sarti Politecnico di Milano, Milano, Italy

Romain Serizel Université de Lorraine, LORIA, UMR 7503, Vandœuvre-lès-Nancy, France

Xavier Serra Music Technology Group (MTG), Universitat Pompeu Fabra, Barcelona, Spain

Dan Stowell School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK

Clara Suied Institut de Recherche Biomédicale des Armées, Brétigny-sur-Orge, France

Tuomas Virtanen Laboratory of Signal Processing, Tampere University of Technology, Tampere, Finland