Signal Processing Methods for Music Transcription
Contents

Preface .................................................................................. ix

List of Contributors ............................................................. xi

Part I Foundations

1 Introduction to Music Transcription
Anssi Klapuri ......................................................................... 3
1.1 Terminology and Concepts ............................................... 7
1.2 Perspectives on Music Transcription ................................. 11
1.3 Outline .......................................................................... 17

2 An Introduction to Statistical Signal Processing and Spectrum Estimation
Manuel Davy ......................................................................... 21
2.1 Frequency, Time-Frequency, and Cepstral Representations ........................................ 21
2.2 Basic Statistical Methods .................................................. 28
2.3 Bayesian Statistical Methods ............................................. 39
2.4 Pattern Recognition Methods .......................................... 52

3 Sparse Adaptive Representations for Musical Signals
Laurent Daudet, Bruno Torrésani ........................................... 65
3.1 Introduction .................................................................... 65
3.2 Parametric Representations ............................................. 68
3.3 Waveform Representations ............................................. 70
3.4 Conclusion ..................................................................... 97
<table>
<thead>
<tr>
<th>Part II Rhythm and Timbre Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Beat Tracking and Musical Metre Analysis</td>
</tr>
<tr>
<td>Stephen Hainsworth</td>
</tr>
<tr>
<td>4.1 Introduction</td>
</tr>
<tr>
<td>4.2 Summary of Beat-Tracking Approaches</td>
</tr>
<tr>
<td>4.3 Musical Background to Rhythmic Structure</td>
</tr>
<tr>
<td>4.4 Onset Detection</td>
</tr>
<tr>
<td>4.5 Rule-Based Approaches</td>
</tr>
<tr>
<td>4.6 Autocorrelation Methods</td>
</tr>
<tr>
<td>4.7 Oscillating Filter Approaches</td>
</tr>
<tr>
<td>4.8 Histogramming Methods</td>
</tr>
<tr>
<td>4.9 Multiple Agent Approaches</td>
</tr>
<tr>
<td>4.10 Probabilistic Models</td>
</tr>
<tr>
<td>4.11 Comparison of Algorithms</td>
</tr>
<tr>
<td>4.12 Conclusions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5 Unpitched Percussion Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derry FitzGerald, Jouni Paulus</td>
</tr>
<tr>
<td>5.1 Introduction</td>
</tr>
<tr>
<td>5.2 Pattern Recognition Approaches</td>
</tr>
<tr>
<td>5.3 Separation-Based Approaches</td>
</tr>
<tr>
<td>5.4 Musicological Modelling</td>
</tr>
<tr>
<td>5.5 Conclusions</td>
</tr>
<tr>
<td>5.6 Acknowledgements</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6 Automatic Classification of Pitched Musical Instrument Sounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfecto Herrera-Boyer, Anssi Klapuri, and Manuel Davy</td>
</tr>
<tr>
<td>6.1 Introduction</td>
</tr>
<tr>
<td>6.2 Methodology</td>
</tr>
<tr>
<td>6.3 Features and Their Selection</td>
</tr>
<tr>
<td>6.4 Classification Techniques</td>
</tr>
<tr>
<td>6.5 Classification of Isolated Sounds</td>
</tr>
<tr>
<td>6.6 Classification of Sounds from Music Files</td>
</tr>
<tr>
<td>6.7 Conclusions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part III Multiple Fundamental Frequency Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 Multiple Fundamental Frequency Estimation Based on Generative Models</td>
</tr>
<tr>
<td>Manuel Davy</td>
</tr>
<tr>
<td>7.1 Noisy Sum-of-Sines Models</td>
</tr>
<tr>
<td>7.2 Off-line Approaches</td>
</tr>
</tbody>
</table>
7.3 On-Line Approaches .............................................. 217
7.4 Other On-Line Bayesian Approaches ......................... 225
7.5 Conclusions ...................................................... 227

8 Auditory Model-Based Methods for Multiple Fundamental Frequency Estimation
Anssi Klapuri ......................................................... 229
8.1 Introduction ..................................................... 229
8.2 Musical Sounds and F0 Estimation .............................. 231
8.3 Pitch Perception Models ........................................ 234
8.4 Using an Auditory Model as a Front End ...................... 244
8.5 Computational Multiple F0 Estimation Methods .............. 248
8.6 Conclusions ...................................................... 264

9 Unsupervised Learning Methods for Source Separation in Monaural Music Signals
Tuomas Virtanen ...................................................... 267
9.1 Introduction ..................................................... 267
9.2 Signal Model .................................................... 268
9.3 Independent Component Analysis ............................... 274
9.4 Sparse Coding .................................................. 278
9.5 Non-Negative Matrix Factorization ............................. 282
9.6 Prior Information about Sources ............................... 284
9.7 Further Processing of the Components ......................... 286
9.8 Time-Varying Components .................................... 289
9.9 Evaluation of the Separation Quality .......................... 294
9.10 Summary and Discussion ..................................... 295

Part IV Entire Systems, Acoustic and Musicological Modelling

10 Auditory Scene Analysis in Music Signals
Kunio Kashino ......................................................... 299
10.1 Introduction .................................................... 299
10.2 Strategy for Music Scene Analysis .............................. 304
10.3 Probabilistic Models for Music Scene Analysis .............. 313
10.4 Conclusion: From Grouping to Generative Estimation ........ 324

11 Music Scene Description
Masataka Goto ......................................................... 327
11.1 Introduction .................................................... 327
11.2 Estimating Melody and Bass Lines ............................ 330
11.3 Estimating Beat Structure ..................................... 341
11.4 Estimating Drums ............................................. 342
11.5 Estimating Chorus Sections and Repeated Sections .......... 342
Contents

11.6 Evaluation Issues ................................................................. 355
11.7 Applications of Music Scene Description ............................... 355
11.8 Conclusion .................................................................. 358

12 Singing Transcription

Matti Ryyänen † ......................................................................... 361
12.1 Introduction ................................................................. 361
12.2 Singing Signals ............................................................. 364
12.3 Feature Extraction .......................................................... 368
12.4 Converting Features into Note Sequences .............................. 375
12.5 Summary and Discussion .................................................. 390

References ........................................................................ 391

Index .................................................................................. 429
Signal processing techniques, and information technology in general, have undergone several scientific advances which permit us to address the very complex problem of automatic music transcription (AMT). During the last ten years, the interest in AMT has increased rapidly, and the time has come for a book-length overview of this subject.

The purpose of this book is to present signal processing algorithms dedicated to the various aspects of music transcription. AMT is a multifaceted problem, comprising several subtasks: rhythm analysis, multiple fundamental frequency analysis, sound source separation, musical instrument classification, and integration of all these into entire systems. AMT is, in addition, deeply rooted in fundamental signal processing, which this book also covers. As the field is quite wide, we have focused mainly on signal processing methods and Western polyphonic music. An extensive presentation of the work in musicology and music perception is beyond the scope of this book.

This book is mainly intended for researchers and graduate students in signal processing, computer science, acoustics, and music. We hope that the book will make the field easier to approach, providing a good starting point for newcomers, but also a comprehensive reference source for those already working in the field. The book is also suitable for use as a textbook for advanced courses in music signal processing. The chapters are mostly self-contained, and readers may want to read them in any order or jump from one to another at will. Whenever an element from another chapter is needed, an explicit reference is made to the relevant chapter. Chapters 1 and 2 provide some background of AMT and signal processing for the entire book, respectively. Otherwise, only a basic knowledge of signal processing is assumed.

Editing a book is a great deal of work. This volume was made possible by those who provided us support and help. We would like to thank Vaishali Damle and Ana Bozicevic at Springer for their help and support, and for their quick replies to our e-mails. Also thanks to Teemu Karjalainen for his practical assistance with \LaTeX.
Preface

Early versions of individual chapters were reviewed by the following people, whose valuable comments and suggestions are gratefully acknowledged:
- Michael Casey, Goldsmiths College, University of London, UK
- A. Taylan Cemgil, University of Cambridge, UK
- Alain de Cheveigné, Ecole Normale Supérieure, Paris, France
- Simon Dixon, Austrian Research Institute for Artificial Intelligence, Vienna
- Dan Ellis, Columbia University, New York
- Olivier Gillet, Télécom-Paris (ENST), Paris, France
- Aki Härnä, Philips Research Laboratories, Eindhoven, The Netherlands
- Marc Leman, Ghent University, Belgium
- Emanuele Pollastri, Erazero S.r.l., Milan, Italy

Thanks go also to the chapter authors, many of whom proofread another, related chapter in the book and provided helpful comments.

Tampere, Finland
Lille, France
December 2005

Anssi Klapuri
Manuel Davy
List of Contributors

Laurent Daudet  
Laboratoire d'Acoustique Musicale  
11 rue de Lourmel  
75015 Paris, France  
daudet@lam.jussieu.fr

Manuel Davy  
LAGIS/CNRS  
BP 48, Cité Scientifique  
59651 Villeneuve d'Ascq Cedex  
France  
Manuel.Davy@ec-lille.fr

Derry FitzGerald  
Cork Institute of Technology  
Rossa Avenue  
Bishopstown  
Cork, Ireland  
derry.fitzgerald@cit.ie

Masataka Goto  
National Institute of Advanced Industrial Science and Technology (AIST)  
1-1-1 Umezono, Tsukuba  
Ibaraki 305-8568, Japan  
m.goto@aist.go.jp

Stephen Hainsworth  
Tillinghast-Towers Perrin  
71 High Holborn  
London WC1V 6TH, UK  
swh21@cantab.net

Perfecto Herrera-Boyer  
Institut Universitari de l'Audiovisual  
Universitat Pompeu Fabra  
Pg. Circumval-lació 8  
08003 Barcelona, Spain  
pherrera@iua.upf.es

Kunio Kashino  有NTT Communication Science Laboratories  数Nippon Telegraph and Telephone Corporation 3-1  
Morinosato-Wakamiya  
Atsugi, 243-0198, Japan  
kunio@eye.brl.ntt.co.jp

Anssi Klapuri  
Institute of Signal Processing  
Tampere University of Technology  
Korkeakoulunkatu 1  
33720 Tampere, Finland  
Anssi.Klapuri@tut.fi

Jouni Paulus  
Institute of Signal Processing  
Tampere University of Technology  
Korkeakoulunkatu 1  
33720 Tampere, Finland  
Jouni.Paulus@tut.fi
Bruno Torrésani  
Laboratoire d’Analyse  
Topologie et Probabilités  
CMI, Université de Provence  
39 rue F. Joliot-Curie  
13453 Marseille cedex 13, France  
Bruno.Torresani@cmi.univ-mrs.fr

Matti Ryynänen  
Institute of Signal Processing  
Tampere University of Technology  
Korkeakoulunkatu 1  
33720 Tampere, Finland  
Matti.Ryynanen@tut.fi

Tuomas Virtanen  
Institute of Signal Processing  
Tampere University of Technology  
Korkeakoulunkatu 1  
33720 Tampere, Finland  
Tuomas.Virtanen@tut.fi