Contents

Part I Introduction

1 Intelligent Audio Analysis: A Definition 3
1.1 Intelligent Audio Analysis 3
1.2 In Real-life Conditions 4
References 6

2 Motivation, Aims, and Solutions 7
2.1 Motivation of Intelligent Audio Analysis 7
2.2 Aims of the Book 9
2.3 Solutions 10
References 10

3 Structure of the Book 13

Part II Intelligent Audio Analysis Methods

4 Chain of Audio Processing 17
References 21

5 Audio Data 23
5.1 Audio Data Requirements 23
5.2 Ground Truth and Gold Standard 24
5.3 Exemplary Databases 26
5.3.1 Example in Speech: TUM AVIC 26
5.3.2 Example in Music: NTWICM 28
5.3.3 Example in Sound: FindSounds Database 34
References 38

6 Audio Features 41
6.1 Audio Chunking 41
6.1.1 Digital Audio 41

6.1.2 Short Time Analysis . . . 43
6.1.3 Audio Activity Detection . . . 45
6.2 Audio Low Level Descriptors . . . 47
6.2.1 Speech Descriptors . . . 47
6.2.2 Music Descriptors . . . 64
6.2.3 Sound Descriptors . . . 72
6.3 Textual Descriptors . . . 73
6.3.1 Bag of Words . . . 75
6.3.2 Bag of N-grams . . . 76
6.3.3 Bag of Character N-grams . . . 77
6.3.4 On-Line Knowledge . . . 77
6.4 Supra Segmental Features . . . 83
6.5 Audio Feature Extraction: The openSMILE Toolkit . . . 83
6.5.1 openSMILE's Architecture . . . 84
6.5.2 Available Feature Extractors . . . 86
6.5.3 Performance . . . 89
6.6 Reduction and Selection of Features . . . 90
References . . . 92

7 Audio Recognition . . . 99
7.1 Audio Recognition Requirements . . . 99
7.2 Static Learning Algorithms . . . 99
7.2.1 Decision Trees . . . 99
7.2.2 Support Vectors . . . 103
7.2.3 Neural Networks . . . 110
7.3 Dynamic Learning Algorithms . . . 118
7.3.1 Estimation . . . 120
7.3.2 Hierarchical Decoding . . . 121
7.4 Ensemble Learning . . . 124
7.4.1 Bootstrapping . . . 125
7.4.2 Meta-Learning . . . 127
7.4.3 Tandem Learning . . . 129
7.5 Evaluation . . . 130
7.5.1 Partitioning and Balancing . . . 130
7.5.2 Evaluation Measures . . . 133
References . . . 136

8 Audio Source Separation . . . 139
8.1 Methodology . . . 140
8.2 Performance . . . 142
8.3 NMF Activation Features . . . 144
References . . . 146

9 Audio Enhancement and Robustness 149
9.1 Audio Signal Preprocessing 150
9.1.1 Advanced Front-End Feature Extraction 151
9.1.2 Unsupervised Spectral Subtraction 151
9.2 Feature Enhancement 153
9.2.1 Feature Normalisation 154
9.2.2 Model Based Feature Enhancement 156
9.3 Model Architectures 159
9.3.1 Conditional Random Fields 159
9.3.2 Hidden Conditional Random Fields 159
9.3.3 Audio Modelling in the Time Domain 160
References 163

Part III Intelligent Audio Analysis Applications

10 Applications in Intelligent Speech Analysis 169
10.1 Linguistics: Digits and Spelling 170
10.1.1 Automotive Digits and Spelling Database 170
10.1.2 Performance 172
10.1.3 Summary 174
10.2 Linguistics: Spontaneous Speech 176
10.2.1 The COSINE Corpus 176
10.2.2 Performance 177
10.2.3 Summary 179
10.3 Non-Linguistics: Vocalisations 180
10.3.1 Methodology 180
10.3.2 Performance 181
10.3.3 Summary 182
10.4 Paralinguistics: States and Traits 183
10.4.1 Sentiment and Opinion 185
10.4.2 Short-term States: Emotion and Interest 193
10.4.3 Long-term Traits: Age, Gender, Height 198
10.4.4 Mid-term: Intoxication and Sleepiness 204
References 210

11 Applications in Intelligent Music Analysis 225
11.1 Drum-Beat Separation 226
11.1.1 Methodology 226
11.1.2 Performance 232
11.1.3 Summary 232
11.2 Onsets 233
11.2.1 The Bello Database 235
11.2.2 Methodology 235

11.2.3 Performance . . . 238
11.2.4 Summary . . . 239
11.3 Tempo, Metre, Ballroom Dance Style . . . 239
11.3.1 BRD Database . . . 243
11.3.2 Methodology . . . 243
11.3.3 Performance . . . 251
11.3.4 Summary . . . 252
11.4 Key . . . 252
11.4.1 Key Databases . . . 254
11.4.2 Parameter Tuning . . . 255
11.4.3 Correlation-Based Analysis . . . 257
11.4.4 Data-Driven Analysis . . . 259
11.4.5 Performance . . . 260
11.4.6 Summary . . . 263
11.5 Chords . . . 264
11.5.1 ChoRD Database . . . 265
11.5.2 Methodology . . . 267
11.5.3 Performance . . . 268
11.5.4 Discussion . . . 269
11.6 Structure . . . 269
11.6.1 Methodology . . . 270
11.6.2 Performance . . . 273
11.6.3 Summary . . . 274
11.7 Mood . . . 274
11.7.1 Methodology . . . 276
11.7.2 Performance . . . 279
11.7.3 Summary . . . 282
11.8 Singer Traits: Age, Gender, Height, Race . . . 283
11.8.1 UltraStar Singer Traits Database . . . 284
11.8.2 Methodology . . . 285
11.8.3 Performance . . . 288
11.8.4 Summary . . . 289
References . . . 290

12 Applications in Intelligent Sound Analysis . . . 299
12.1 Animal Vocalisations . . . 300
12.1.1 HU-ASA Database . . . 300
12.1.2 Methodology . . . 301
12.1.3 Performance . . . 302
12.1.4 Summary . . . 304
12.2 Acoustic Events . . . 305
12.2.1 Methodology . . . 305
12.2.2 Performance . . . 306
12.2.3 Summary . . . 308

12.3 Emotion 308
12.3.1 Methodology 309
12.3.2 Performance 310
12.3.3 Summary 312
References 312

Part IV Conclusion

13 Discussion 317
13.1 Picking Up on the Goals 317
13.2 Best Practice Recommendations 320
13.3 Remaining Challenges 323
References 326

14 Vision 335
14.1 Summary of Results 335
14.2 Future Perspective 336
Reference 337

Appendix: openSMILE Standardised Feature Sets 339

Index 343

http://www.springer.com/978-3-642-36805-9

Intelligent Audio Analysis
Schuller, B.W.
2013, XXVIII, 345 p., Hardcover
ISBN: 978-3-642-36805-9