An Introduction to Audio Content Analysis
Applications in Signal Processing and Music Informatics

Alexander Lerch
zplane.development, Berlin
## CONTENTS

Preface xiii  
Acronyms xv  
List of Symbols xix  

### 1 Introduction  
1.1 Audio Content 3  
1.2 A Generalized Audio Content Analysis System 4  

### 2 Fundamentals  
2.1 Audio Signals 7  
2.1.1 Periodic Signals 7  
2.1.2 Random Signals 9  
2.1.3 Sampling and Quantization 9  
2.1.4 Statistical Signal Description 13  
2.2 Signal Processing 14  
2.2.1 Convolution 14  
2.2.2 Block-Based Processing 18  
2.2.3 Fourier Transform 20  
2.2.4 Constant Q Transform 23  
2.2.5 Auditory Filterbanks 24  
2.2.6 Correlation Function 24
2.2.7  Linear Prediction 28

3  Instantaneous Features 31
3.1  Audio Pre-Processing 33
3.1.1  Down-Mixing 33
3.1.2  DC Removal 33
3.1.3  Normalization 34
3.1.4  Down-Sampling 34
3.1.5  Other Pre-Processing Options 35
3.2  Statistical Properties 35
3.2.1  Arithmetic Mean 36
3.2.2  Geometric Mean 36
3.2.3  Harmonic Mean 36
3.2.4  Generalized Mean 36
3.2.5  Centroid 37
3.2.6  Variance and Standard Deviation 37
3.2.7  Skewness 38
3.2.8  Kurtosis 39
3.2.9  Generalized Central Moments 40
3.2.10  Quantiles and Quantile Ranges 40
3.3  Spectral Shape 41
3.3.1  Spectral Rolloff 42
3.3.2  Spectral Flux 44
3.3.3  Spectral Centroid 45
3.3.4  Spectral Spread 47
3.3.5  Spectral Decrease 48
3.3.6  Spectral Slope 49
3.3.7  Mel Frequency Cepstral Coefficients 51
3.4  Signal Properties 54
3.4.1  Tonalness 54
3.4.2  Autocorrelation Coefficients 61
3.4.3  Zero Crossing Rate 62
3.5  Feature Post-Processing 63
3.5.1  Derived Features 64
3.5.2  Normalization and Mapping 65
3.5.3  Subfeatures 66
3.5.4  Feature Dimensionality Reduction 66

4  Intensity 71
4.1  Human Perception of Intensity and Loudness 71
4.2  Representation of Dynamics in Music 73
4.3  Features 73
4.3.1 Root Mean Square 73
4.3.2 Peak Envelope 76
4.3.3 Psycho-Acoustic Loudness Features 77

5 Tonal Analysis 79

5.1 Human Perception of Pitch 79
5.1.1 Pitch Scales 79
5.1.2 Chroma Perception 81

5.2 Representation of Pitch in Music 82
5.2.1 Pitch Classes and Names 82
5.2.2 Intervals 83
5.2.3 Root Note, Mode, and Key 83
5.2.4 Chords and Harmony 86
5.2.5 The Frequency of Musical Pitch 88

5.3 Fundamental Frequency Detection 91
5.3.1 Detection Accuracy 92
5.3.2 Pre-Processing 94
5.3.3 Monophonic Input Signals 97
5.3.4 Polyphonic Input Signals 103

5.4 Tuning Frequency Estimation 106

5.5 Key Detection 108
5.5.1 Pitch Chroma 108
5.5.2 Key Recognition 112

5.6 Chord Recognition 116

6 Temporal Analysis 119

6.1 Human Perception of Temporal Events 119
6.1.1 Onsets 119
6.1.2 Tempo and Meter 122
6.1.3 Rhythm 122
6.1.4 Timing 123

6.2 Representation of Temporal Events in Music 123
6.2.1 Tempo and Time Signature 123
6.2.2 Note Value 124

6.3 Onset Detection 124
6.3.1 Novelty Function 125
6.3.2 Peak Picking 127
6.3.3 Evaluation 128

6.4 Beat Histogram 133
6.4.1 Beat Histogram Features 134

6.5 Detection of Tempo and Beat Phase 135

6.6 Detection of Meter and Downbeat 136
## 7 Alignment

### 7.1 Dynamic Time Warping
- 7.1.1 Example
- 7.1.2 Common Variants
- 7.1.3 Optimizations

### 7.2 Audio-to-Audio Alignment
- 7.2.1 Ground Truth Data for Evaluation

### 7.3 Audio-to-Score Alignment
- 7.3.1 Real-Time Systems
- 7.3.2 Non-Real-Time Systems

## 8 Musical Genre, Similarity, and Mood

### 8.1 Musical Genre Classification
- 8.1.1 Musical Genre
- 8.1.2 Feature Extraction
- 8.1.3 Classification

### 8.2 Related Research Fields
- 8.2.1 Music Similarity Detection
- 8.2.2 Mood Classification
- 8.2.3 Instrument Recognition

## 9 Audio Fingerprinting

### 9.1 Fingerprint Extraction
### 9.2 Fingerprint Matching
### 9.3 Fingerprinting System: Example

## 10 Music Performance Analysis

### 10.1 Musical Communication
- 10.1.1 Score
- 10.1.2 Music Performance
- 10.1.3 Production
- 10.1.4 Recipient

### 10.2 Music Performance Analysis
- 10.2.1 Analysis Data
- 10.2.2 Research Results

## A Convolution Properties

### A.1 Identity
### A.2 Commutativity
### A.3 Associativity
### A.4 Distributivity
### A.5 Cirularity

## B Fourier Transform
CONTENTS

B.1 Properties of the Fourier Transformation 186
  B.1.1 Inverse Fourier Transform 186
  B.1.2 Superposition 186
  B.1.3 Convolution and Multiplication 186
  B.1.4 Parseval’s Theorem 187
  B.1.5 Time and Frequency Shift 188
  B.1.6 Symmetry 188
  B.1.7 Time and Frequency Scaling 189
  B.1.8 Derivatives 190

B.2 Spectrum of Example Time Domain Signals 190
  B.2.1 Delta Function 190
  B.2.2 Constant 191
  B.2.3 Cosine 191
  B.2.4 Rectangular Window 191
  B.2.5 Delta Pulse 191

B.3 Transformation of Sampled Time Signals 192

B.4 Short Time Fourier Transform of Continuous Signals 192
  B.4.1 Window Functions 193

B.5 Discrete Fourier Transform 195
  B.5.1 Window Functions 196
  B.5.2 Fast Fourier Transform 197

C Principal Component Analysis 199
  C.1 Computation of the Transformation Matrix 200
  C.2 Interpretation of the Transformation Matrix 200

D Software for Audio Analysis 201
  D.1 Software Frameworks and Applications 202
    D.1.1 Marsyas 202
    D.1.2 CLAM 202
    D.1.3 jMIR 203
    D.1.4 CoMIRVA 203
    D.1.5 Sonic Visualiser 203
  D.2 Software Libraries and Toolboxes 204
    D.2.1 Feature Extraction 204
    D.2.2 Plugin Interfaces 205
    D.2.3 Other Software 206

References 207

Index 243