An Introduction to Audio Content Analysis

IEEE Press
445 Hoes Lane
Piscataway, NJ 08854

Technical Reviewers

Roger B. Dannenberg, Carnegie Mellon University

An Introduction to Audio Content Analysis

Applications in Signal Processing and Music Informatics

Alexander Lerch
zplane.development, Berlin

IEEE

IEEE PRESS

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Lerch, Alexander.
Audio content analysis : an introduction / Alexander Lerch.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-118-26682-3 (hardback)
1. Computer sound processing. 2. Computational auditory scene analysis. 3. Content analysis (Communication)—Data processing. I. Title.
TK7881.4.L485 2012
006.4'5—dc23 2012008107

Printed in the United States of America.

10 9 8 7 6 5 4 3 2

CONTENTS IN BRIEF

CONTENTS

PREFACE

The growing amount of audio and music data on the Internet and in user databases leads to an increasing need for intelligent browsing, retrieving, and processing of this data with automated methods. *Audio content analysis*, a subfield of the research field *music information retrieval*, aims at extracting (musical and perceptual) properties directly from the audio signal to support these tasks. Knowledge of these properties allows us to improve the interaction of humans or machines with digital audio signals. It enables new ways of assessing, processing, and visualizing music.

Although analysis of audio signals covers other research areas such as automatic speech recognition, we will restrict ourselves to the analysis of music signals in the context of this book.

When preparing classes on audio content analysis with a focus on music recordings it became quickly clear that — although there is a vast and growing amount of research literature available — there exists no introductory literature. This observation led to writing this book in the hope it might assist students, engineers, and developers who have basic knowledge of digital signal processing. The focus lies on the signal processing part of audio content analysis, but wherever it may improve the understanding of either algorithmic design choices or implementation details some basic characteristics of human perception, music theory, and notation as well as machine learning will be summarized.

Chapter 2 starts by introducing some definitions and offers a short reiteration of the most important tools of digital signal processing for the analysis of audio signals. The following chapters encompass the basic four technical content categories timbre, level, pitch, and rhythm. A fifth category is reserved for purely technical and statistical signal descriptions. Chapter 3 introduces low-level or short-term features that are widely used in systems for signal analysis. A large part of the chapter deals with timbre representations of a signal, accompanied by the introduction of statistical features. The chapter concludes with a summary of approaches to feature selection and post-processing. Chapter 4 focuses on intensity-related features. It covers envelope features and simple models of human loudness perception. The extraction of pitch-related information such as the detection of fundamental frequency, harmony, key, etc. is described in Chap. 5. Chapter 6 focuses on the temporal and rhythmic aspects of the audio signal. It explains the segmentation of audio signals into musical events and covers higher level information such as the detection of tempo and meter. The remaining chapters deal with analysis systems using combinations of timbre, loudness, onset, and pitch features to derive higher level information. Chapter 7 describes the automatic synchronization of two similar audio sequences or an audio and a score sequence. Musical genre classification, one of the most prominent research fields of audio content analysis, is explained in Chap. 8. Chapter 9 is about audio fingerprinting which is probably the commercially most successful application in audio content analysis. The concluding chapter, targeting classical music, covers the analysis of music performance. It is not a core field in audio content analysis but emphasizes the differentiation between performance aspects and musical aspects of recordings and elaborates on the manual and automated analysis methods used for musicological music performance analysis. The appendices provide details and derivations of some of the most important signal processing tools as well as a short survey on available software solutions for audio content analysis.

Downloadable MATLAB files are available at: http://www.audiocontentanalysis.org.

A. LERCH

Berlin
January, 2012

ACRONYMS

ACA	Audio Content Analysis
ACF	Autocorrelation Function
ADPCM	Adaptive Differential Pulse Code Modulation
AMDF	Average Magnitude Difference Function
ANN	Artificial Neural Network
AOT	Acoustic Onset Time
API	Application Programmer's Interface
BPM	Beats per Minute
CAMEL	Content-based Audio and Music Extraction Library
CASA	Computational Auditory Scene Analysis
CCF	Cross Correlation Function
CCIR	Comité Consultatif International des Radiocommunications
CD	Compact Disc
CiCF	Circular Correlation Function
CLAM	C++ Framework for Audio and Music
COG	Center of Gravity
CQT	Constant Q Transform

DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DP	Dynamic Programming
DTW	Dynamic Time Warping
EBU	European Broadcasting Union
ERB	Equivalent Rectangular Bandwidth
FEAPI	Feature Extraction Application Programmer's Interface
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FN	False Negative
FP	False Positive
FT	Fourier Transform
FWR	Full-Wave Rectification
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
HPS	Harmonic Product Spectrum
HSS	Harmonic Sum Spectrum
HTK	HMM Toolkit
HWR	Half-Wave Rectification
IBI	Inter-Beat Interval
ICA	Independent Component Analysis
IDFT	Inverse Discrete Fourier Transform
IFT	Inverse Fourier Transform
IIR	Infinite Impulse Response
IO	Input/Output
IOI	Inter-Onset Interval
ISMIR	International Society for Music Information Retrieval
ITU	International Telecommunication Union
JNDL	Just Noticeable Difference in Level
KNN	K-Nearest Neighbor

LDA	Linear Discriminant Analysis
MA	Moving Average
MFCC	Mel Frequency Cepstral Coefficient
MIDI	Musical Instrument Digital Interface
MIR	Music Information Retrieval
MIREX	Music Information Retrieval Evaluation eXchange
MPA	Music Performance Analysis
MPEG	Motion Picture Experts Group
NOT	Note Onset Time
PAT	Perceptual Attack Time
PCA	Principal Component Analysis
PDF	Probability Density Function
POT	Perceptual Onset Time
PPM	Peak Program Meter
PSD	Peak Structure Distance
RBF	Radial Basis Function
RFD	Relative Frequency Distribution
RLB	Revised Low Frequency B Curve
RMS	Root Mean Square
ROC	Receiver Operating Curve
SIMD	Single Instruction Multiple Data
SNR	Signal-to-Noise Ratio
SOM	Self-Organizing Map
STFT	Short Time Fourier Transform
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
WEKA	Waikato Environment for Knowledge Analysis
YAAFE	Yet Another Audio Feature Extractor

LIST OF SYMBOLS

A	Amplitude
a	Filter Coefficient (recursive)
$\mathcal{B}$	Number of Beats
b	Filter Coefficient (transversal)
β	Exponent
$\mathcal{C}$	Number of (Audio) Channels
$\chi(.)$	Center Clipping Function
$\boldsymbol{C}_{\mathrm{AB}}$	Cost Matrix for the Distance Matrix between Sequences A and B
$\mathfrak{C}_{\mathrm{AB}}$	Overall Cost of a Path through the Cost Matrix
$c_x(\cdot)$	Cepstrum of the Signal x
$\boldsymbol{D}_{\mathrm{AB}}$	Distance Matrix between Sequences A and B
d	Distance Measure
Δ_{Q}	Quantization Step Size
e_{P}	Prediction Error
e_{Q}	Quantization Error
$\mathfrak{e}(f)$	Equivalent Rectangular Bandwidth

η	(Correlation) Lag
e_{Tfp}	(Spectral) Prediction Error
F	F-Measure
f	Frequency in Hz
f_0	Fundamental Frequency in Hz
f_{S}	Sample Rate
f_{A4}	Tuning Frequency in Hz
$\mathcal{F}$	Number of Features
f_I	Instantaneous Frequency in Hz
$\mathfrak{F}(\cdot)$	(Discrete) Fourier Transform
G	Threshold
$\mathbf{\Gamma}$	Chord Transformation Matrix
$\gamma_{x,\mathcal{O}}$	Central Moment of Order $\mathcal{O}$ of Signal x
$H(\cdot)$	Transfer Function
$h(\cdot)$	Impulse Response
$\mathcal{H}$	Hop Size
i	Sample Index
$\mathcal{J}$	Impulse Response Length
j	Integer (Loop) Variable
$\mathcal{K}$	Block Size
k	Frequency Bin Index
κ	Percentage
$\Lambda(k,n)$	Tonalness Spectrum
λ	Weighting Factor
$\mathcal{M}$	Number of (Quantization) Steps
m	Key Index
$\mathtt{m}(f)$	Pitch (Mel)
μ_x	Arithmetic Mean of Signal x
$\mathcal{N}$	Number of Observations or Blocks
n	Block Index

$\mathcal{O}$	Order (e.g., Filter Order)
o_{r}	Block Overlap Ratio
ω	Angular Velocity ($\omega = 2\pi f$) in radians per second
O	Number of Onsets
P	Precision
$\boldsymbol{p}$	Alignment Path
Φ_X	Phase Spectrum of the Signal x
$\varphi(\cdot)$	Gaussian Function
$\mathfrak{p}$	(MIDI) Pitch
$\boldsymbol{\nu}$	Pitch Chroma Vector/Key Profile
P_x	Power of the Signal x
$p_x(x)$	Probability Density Function of the Signal x
$\boldsymbol{\psi}(\cdot)$	Chord Probability Vector
$\mathcal{Q}$	Quality Factor (Mid-Frequency divided by Bandwidth)
q	Evaluation Metric
$Q_x(\cdot)$	Quantile Boundary
R	Recall
$r_{xy}(\cdot)$	Correlation Function between the Signals x and y
r	Radius
$\boldsymbol{R}$	Covariance Matrix
σ_x	Standard Deviation of Signal x
σ_x^2	Variance of Signal x
SNR	Signal-to-Noise Ratio
T	Time Period in s
t	Time in s
T_0	Time Period of the Fundamental Frequency in s
$\mathcal{T}$	Number of (Chord) Templates
$\mathfrak{T}$	Tempo in BPM
$\boldsymbol{T}$	(PCA) Transformation Matrix
T_{S}	Sample Period in s
$\boldsymbol{V}$	Feature Matrix with dimensions $\mathcal{F} \times \mathcal{N}$
v_{ACF}^{η}	ηth Autocorrelation Coefficient

v_{C}	Centroid
$\mathcal{V}$	Feature Set
v_{K}	Kurtosis
v^j_{MFCC}	jth MFCC
v_{Peak}	Peak Envelope
v_{PPM}	Peak Program Meter
v_{RMS}	RMS
v_{SC}	Spectral Centroid
v_{SD}	Spectral Decrease
v_{SF}	Spectral Flux
v_{SK}	Spectral Kurtosis
v_{Sk}	Skewness
v_{SR}	Spectral Rolloff
v_{SS}	Spectral Spread
v_{SSk}	Spectral Skewness
v_{SSl}	Spectral Slope
v_{Ta}	ACF Maximum
v_{Tf}	Spectral Flatness
v_{Tfp}	Spectral Predictivity
v_{Tp}	Predictivity Ratio
v_{Tpr}	Tonal Power Ratio
v_{Tsc}	Spectral Crest Factor
v_{ZC}	Zero Crossing Rate
w	Word Length in Bit
w_{AB}	Window Function with Alternative Blackman Shape
w_{B}	Window Function with Blackman Shape
w_{BH}	Window Function with Blackman-Harris Shape
w_{C}	Window Function with Cosine Shape
w_{H}	Window Function with von-Hann Shape
w_{Hm}	Window Function with Hamming Shape
w_{R}	Window Function with Rectangular Shape
w_{T}	Window Function with Bartlett Shape
$X(\cdot)$	Fourier Representation of the Signal x
$\mathfrak{x}(f)$	Normed Frequency Position on the Cochlea
$X^*(\cdot)$	Conjugate-Complex Spectrum of the Signal x
$\mathfrak{z}(f)$	Critical Band Rate (Bark)