Machine Audition:
Principles, Algorithms and Systems

Wenwu Wang
University of Surrey, UK

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE
Hershey · New York

Detailed Table of Contents

tion systems and evaluating their performance. They describe the main components of the architecture, its advantages, implementation details, and related issues.

Cocktail party problem is a classical and challenging scientific problem that is still unsolved. Many efforts have been attempted by researchers to address this problem using different techniques. This chapter provides a review on recent progresses in several areas, such as independent component analysis, computational auditory scene analysis, model-based approaches, non-negative matrix factorization, sparse representation and compressed sensing. As an example, a multistage approach is also provided for addressing the source separation issue within this problem.

The demand to function in uncontrolled listening environments has severe implications for machine audition. The natural system has addressed this demand by adapting its function flexibly to changing task demands. This chapter addresses the functional requirements of auditory systems, both natural and artificial, to be able to deal with the complexities of uncontrolled real-world input. Signal processing methods that are needed for such scenarios are also discussed.

Section 2
Audio Signal Separation, Extraction and Localization

Machine separation of moving audio sources is a challenging problem. This chapter presents a novel multimodal solution to blind source separation (BSS) of moving sources, where the visual modality is utilized to facilitate the separation of moving sources. The movement of the sources is detected by a relatively simplistic 3-D tracker based on video cameras. The tracking process is based on particle filtering which provides robust tracking performance. Positions and velocities of the sources are obtained from the 3-D tracker and if the sources are moving, real time speech enhancement and separation of the sources are obtained by using a beamforming algorithm.

activation matrix. Extensions of NMF to the multichannel case, in both instantaneous and convolutive recordings, possibly underdetermined, together with audio source separation results of a real stereo musical excerpt are also included.

Chapter 12

The authors provide a comprehensive introduction to the design of music onset detection algorithms. First, it introduces the general scheme and commonly-used time-frequency analysis for onset detection. Then, it reviews many methods for onset detection in detail, such as energy-based, phase-based, pitch-based and supervised learning methods. The chapter also includes commonly used performance measures, onset annotation software, public database, and evaluation methods.

Chapter 13

This chapter presents automatic segmentation methods using different original representations of music, corresponding to rhythm, chroma, and timbre, and by calculating a shortest path through the self-similarity calculated from each time/feature representation. Each segmentation scale quality is analyzed through the use of the mean silhouette value, which permits automatic segmentation on different time scales and gives indication on the inherent segment sizes in the music analyzed. Different methods are employed to verify the quality of the inherent segment sizes, by comparing them to the literature (grouping, chunks), by comparing them among themselves, and by measuring the strength of the inherent segment sizes.

Chapter 14

A great deal of attention has been paid recently to the automatic prediction of tags for music and audio in general. In the case of music, social tags have become an important component of ``Web 2.0'' recommender systems. In an effort to better understand the task and also to help new researchers bring their insights to bear on this problem, this chapter provides a review of the state-of-the-art methods for addressing automatic tagging of audio. It is divided in the following sections: goal, framework, audio representation, labeled data, classification, evaluation, and future directions.

Chapter 15

Non-negative matrix factorization (NMF) has been shown recently to be a useful technique for audio decomposition. However, the instantaneous NMF model has difficulty in dealing with the audio signals whose frequencies change dramatically over time, which is nevertheless in practice a case for many real signals. This chapter intends to provide a brief overview of the models and algorithms for both instantaneous and convolutive NMF, with a focus on the theoretical analysis and performance evaluation of the convolutive NMF algorithms, and their applications to audio pattern separation problems.

Section 4
Audio Cognition, Modeling and Affective Computing

This chapter investigates the modeling methods for musical cognition. The author explores possible relations between cognitive measures of musical structure and statistical signal properties that are revealed through information dynamics analysis. The addressed questions include: (1) description of music as an information source, (2) modeling of music–listener relations in terms of communication channel, (3) choice of musical features and dealing with their dependencies, (4) survey of different information measures for description of musical structure and measures of shared information between listener and the music, and (5) suggestion of new approach to characterization of listening experience in terms of different combinations of musical surface and structure expectancies.

This chapter provides a survey of research efforts in emotion recognition using different modalities: audio, visual and audio-visual combined. It also describes fifteen audio, visual and audio-visual data sets, and the types of feature that researchers have used to represent the emotional content. Several important issues, such as feature selection and reduction, emotion classification, and methods for fusing information from multiple modalities are also discussed. The chapter concludes by pointing out interesting areas in this field for future investigation.

Propagation of sound from a source to a receiver in an enclosure can be modeled as an acoustic transmission channel. Objective room acoustic parameters are routinely used to quantify properties of such channels in the design and assessment of acoustically critical spaces. This chapter discusses a number of new methods and algorithms for determining room acoustic parameters using machine audition of naturally occurring sound sources, i.e. speech and music. In particular, reverberation time, early decay time and speech transmission index can be estimated from received speech or music signals using statistical machine learning or maximum likelihood estimation in a semi-blind or blind fashion.

Chapter 19

In speech perception and recognition, many hidden phenomena are not well understood yet, including the semantic gap going from spectral time-frequency representations to the symbolic translation into phonemes and words, and the construction of morpho-syntactic and semantic structures. This chapter is intended to explore some of these facts at a simplifying level under two points of view: that of top-down analysis provided from speech perception, and the symmetric from bottom-up synthesis provided by the biological architecture of auditory pathways. It also includes an application-driven design of a neuromorphic speech processing architecture and the simulation details provided by a parallel implementation of the architecture in a supercomputer.