Structure Learning in Audio

By having information about the setting a user is in, a computer is able to make decisions proactively to facilitate tasks for the user. Two approaches are taken in this thesis to achieve more information about an audio environment. One approach is that of classifying audio, and a new approach using pitch dynamics is suggested. The other approach is finding structures between the mixings of multiple sources based on an assumption of statistical independence of the sources. Three different audio classification tasks have been investigated. Audio classification into three classes, music, noise and speech, using novel features based on pitch dynamics. Within instrument classification two different harmonic models have been compared. Finally voiced/unvoiced segmentation of popular music is done based on MFCC’s and AR coefficients. The structures in the mixings of multiple sources have been investigated. A fast and computationally simple approach that compares recordings and classifies if they are from the same audio environment have been developed, and shows very high accuracy and the ability to synchronize recordings in the case of recording devices which are not connected. A more general model is proposed based on Independent Component Analysis. It is based on sequential pruning of the parameters in the mixing matrix and a version based on a fixed source distribution as well as a parameterized distribution is found. The parameterized version has the advantage of modeling both sub- and super-Gaussian source distributions allowing a much wider use of the method. All methods uses a variety of classification models and model selection algorithms which is a common theme of the thesis.

Structure learning by pruning in independent component analysis

We discuss pruning as a means of structure learning in independent component analysis (ICA). Learning the structure is attractive in both signal processing and in analysis of abstract data, where it can assist model interpretation, generalizability and reduce computation. We derive the relevant saliency expressions and compare with magnitude based pruning and Bayesian sparsification. We show in simulations that pruning is able to identify underlying structures without prior knowledge on the dimensionality of the model. We find, that for ICA, magnitude based pruning is as efficient as saliency based methods and Bayesian methods, for both small and large samples. The Bayesian information criterion (BIC) seems to outperform both AIC and test sets as tools for determining the optimal dimensionality.
Synchronization and comparison of Lifelog audio recordings

We investigate concurrent ‘Lifelog’ audio recordings to locate segments from the same environment. We compare two techniques earlier proposed for pattern recognition in extended audio recordings, namely cross-correlation and a fingerprinting technique. If successful, such alignment can be used as a preprocessing step to select and synchronize recordings before further processing. The two methods perform similarly in classification, but fingerprinting scales better with the number of recordings, while cross-correlation can offer sample resolution synchronization. We propose and investigate the benefits of combining the two. In particular we show that the combination allows sample resolution synchronization and scalability.

General information
State: Published
Organisations: Department of Informatics and Mathematical Modeling, Cognitive Systems
Authors: Nielsen, A. B. (Intern), Hansen, L. K. (Intern)
Pages: 474-479
Publication date: 2008

Host publication information
Publisher: IEEE
ISBN (Print): 978-1-4244-2375-0
Main Research Area: Technical/natural sciences
Electronic versions:
Hansen.pdf
DOIs:
10.1109/MLSP.2008.4685526

Bibliographical note
Copyright: 2008 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE
Source: orbit
Source-ID: 233580
Publication: Research - peer-review › Article in proceedings – Annual report year: 2008

VOCAL SEGMENT CLASSIFICATION IN POPULAR MUSIC

This paper explores the vocal and non-vocal music classification problem within popular songs. A newly built labeled database covering 147 popular songs is announced. It is designed for classifying signals from 1sec time windows. Features are selected for this particular task, in order to capture both the temporal correlations and the dependencies among the feature dimensions. We systematically study the performance of a set of classifiers, including linear regression, generalized linear model, Gaussian mixture model, reduced kernel orthonormalized partial least squares and (K-)means on cross-validated training and test setup. The database is divided in two different ways: with/without artist overlap between training and test sets, so as to study the so called ‘artist effect’. The performance and results are analyzed in depth: from error rates to sample-to-sample error correlation. A voting scheme is proposed to enhance the performance under certain conditions.

General information
State: Published
Organisations: Department of Informatics and Mathematical Modeling
Authors: Feng, L. (Intern), Nielsen, A. B. (Intern), Hansen, L. K. (Intern)
Pages: 121-126
Publication date: 2008

Host publication information
Title of host publication: Proceedings of the Ninth International Conference on Music Information Retrieval (ISMIR2008)
Publisher: ISMIR
ISBN (Print): 978-0-615-24849-3
Main Research Area: Technical/natural sciences
Electronic versions:
Pop music database, Music retrieval, vocal segment classification
On the relevance of spectral features for instrument classification

Automatic knowledge extraction from music signals is a key component for most music organization and music information retrieval systems. In this paper, we consider the problem of instrument modelling and instrument classification from the rough audio data. Existing systems for automatic instrument classification operate normally on a relatively large number of features, from which those related to the spectrum of the audio signal are particularly relevant. In this paper, we confront two different models about the spectral characterization of musical instruments. The first assumes a constant envelope of the spectrum (i.e., independent from the pitch), whereas the second assumes a constant relation among the amplitude of the harmonics. The first model is related to the Mel frequency cepstrum coefficients (MFCCs), while the second leads to what we will refer to as harmonic representation (HR). Experiments on a large database of real instrument recordings show that the first model offers a more satisfactory characterization, and therefore MFCCs should be preferred to HR for instrument modelling/classification.

General information
State: Published
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling
Authors: Nielsen, A. B. (Intern), Sigurdsson, S. (Intern), Hansen, L. K. (Intern), Arenas-Garcia, J. (Ekstern)
Pages: 485-488
Publication date: 2007

Host publication information
Title of host publication: IEEE International Conference on Acoustics, Speech and Signal Processing, 2007
Volume: 2
Place of publication: Hawaii
Publisher: IEEE
ISBN (Print): 1-4244-0728-1
Main Research Area: Technical/natural sciences
Electronic versions:
Nielsen.pdf
Paper.pdf
DOIs:
10.1109/ICASSP.2007.366278

Bibliographical note
Copyright: 2007 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE
Source: orbit
Source-ID: 209397
Publication: Research - peer-review › Article in proceedings – Annual report year: 2007

On The Relevance of Spectral Features for Instrument Classification

General information
State: Published
Organisations: Cognitive Systems, Department of Informatics and Mathematical Modeling
Authors: Nielsen, A. B. (Intern)
Publication date: 2007

Publication information
Original language: English
Publisher: Honolulu, Hawaii
Main Research Area: Technical/natural sciences
Electronic versions:
imm5492.pdf
Pitch Based Sound Classification

A sound classification model is presented that can classify signals into music, noise and speech. The model extracts the pitch of the signal using the harmonic product spectrum. Based on the pitch estimate and a pitch error measure, features are created and used in a probabilistic model with soft-max output function. Both linear and quadratic inputs are used. The model is trained on 2 hours of sound and tested on publicly available data. A test classification error below 0.05 with 1 s classification windows is achieved. Furthermore it is shown that linear input performs as well as a quadratic, and that even though classification gets marginally better, not much is achieved by increasing the window size beyond 1 s.

General information
State: Published
Organisations: Department of Informatics and Mathematical Modeling, Cognitive Systems
Authors: Nielsen, A. B. (Intern), Hansen, L. K. (Intern), Kjems, U. (Ekstern)
Publication date: 2006

Host publication information
Volume: 3
Publisher: IEEE
Main Research Area: Technical/natural sciences
Electronic versions: Nielsen2.pdf
DOIs: 10.1109/ICASSP.2006.1660772

Bibliographical note
Copyright: 2006 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE

Projects:
Signalbehandling for proactive pervasive computing
Department of Informatics and Mathematical Modeling
Period: 01/10/2005 → 05/05/2009
Phd Student:
Nielsen, Andreas Brinch (Intern)
Main Supervisor:
Hansen, Lars Kai (Intern)
Examiner:
Larsen, Jan (Ekstern)
Jensen, Søren Holdt (Intern)
Oja, Erkki (Ekstern)

Financing sources
Source: Internal funding (public)
Name of research programme: DTU-lønnet stipendie
Project: PhD