KTH VETENSKAP OCH KONST

KTH ROYAL INSTITUTE
OF TECHNOLOGY

DT2470: Music Informatics

Bob L. Sturm (TMH)
André Holzapfel (MID)

About the projects

Project information

1. *Build and test your own music informatics algorithm, e.g.:*

- Classification (genre, mood, instrument)
- Cover song identification
- Beat tracking, tempo estimation
- Key detection
- Music/speech discrimination
- Etc.

2. *Write a brief summary of algorithm and results (~2-4 pages with results)*

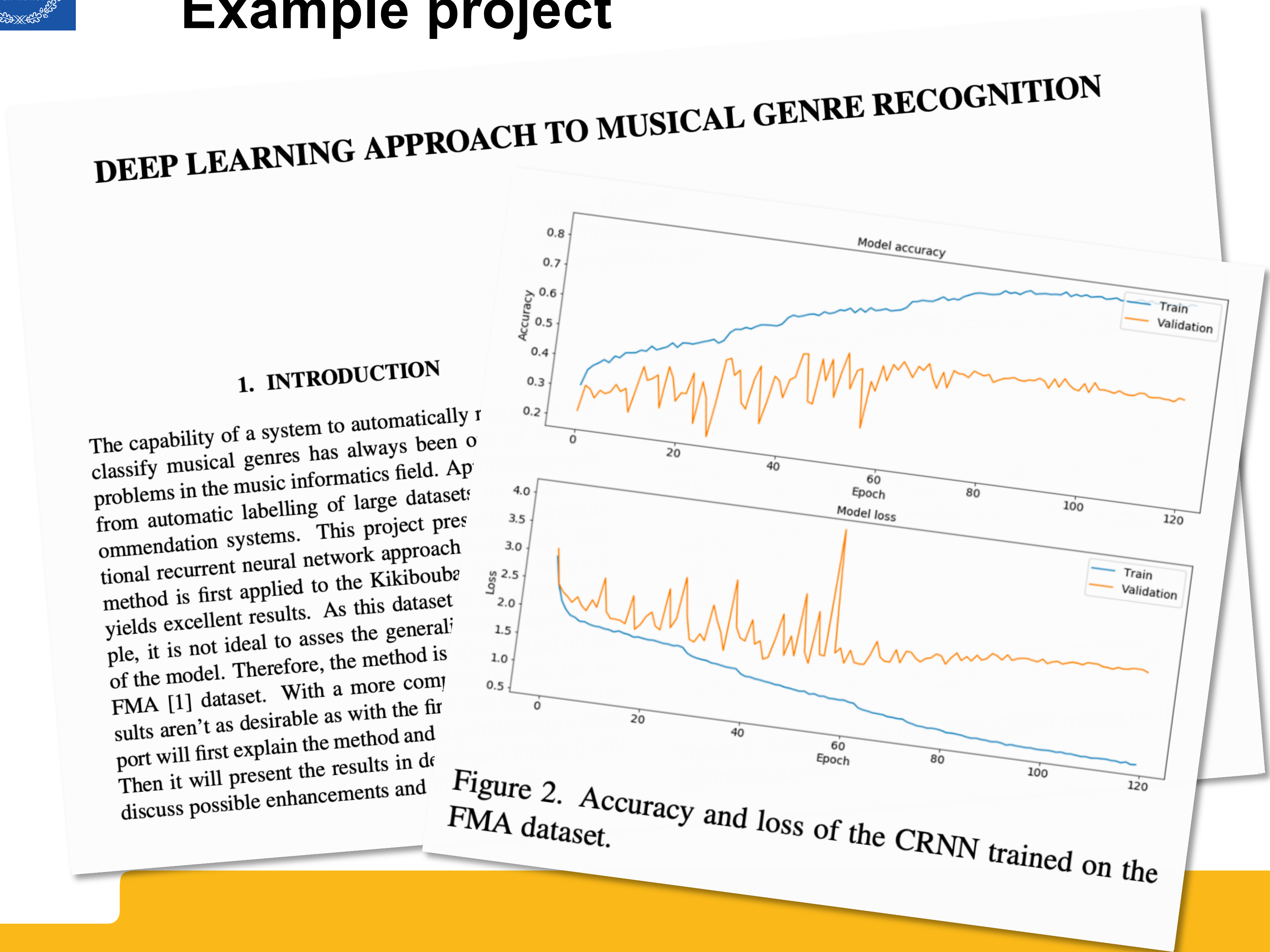
3. *Use github to host code and examples, and include link in report*

Example projects

See: https://www.music-ir.org/mirex/wiki/MIREX_HOME

- **August 30th 2020**
  - 2020:Audio Fingerprinting <TC: Chung-Che Wang>
  - 2020:Audio Classification (Train/Test) Tasks <TC: Yun Hao (IMIRSEL)>, including
    - Audio US Pop Genre Classification
    - Audio Latin Genre Classification
    - Audio Music Mood Classification
    - Audio Classical Composer Identification
  - 2020:Audio K-POP Mood Classification <TC: Yun Hao (IMIRSEL)>
  - 2020:Audio K-POP Genre Classification <TC: Yun Hao (IMIRSEL)>
  - 2020:Audio Tag Classification <TC: Emre Demir>
- **September 6th 2020**
  - 2020:Audio Chord Estimation <TC: Johan Pauwels>
  - 2020:Audio Cover Song Identification <TC: Yun Hao (IMIRSEL)>
  - 2020:Audio Downbeat Estimation <TC: Mickaël Zehren>
  - 2020:Audio Key Detection <TC: Johan Pauwels>
  - 2020:Audio Melody Extraction <TC: An-Qi Huang>
  - 2020:Patterns for Prediction (offshoot of 2017:Discovery of Repeated Themes & Sections) <TC: Berit Janssen, Iris Ren, and
  - 2020:Query by Singing/Humming <TC: Makarand Velankar>
  - 2020:Multiple Fundamental Frequency Estimation & Tracking <TC: Yun Hao (IMIRSEL)>
- **September 13th 2020**
  - ***NEW!*** 2020:Singing_Transcription_from_Polyphonic_Music <TC: Jun-You Wang>
  - 2020:Lyrics Transcription (former: Automatic Lyrics-to-Audio Alignment) <TC: Georgi Dzhambazov, Daniel Stoller, Chitralekha

Example project



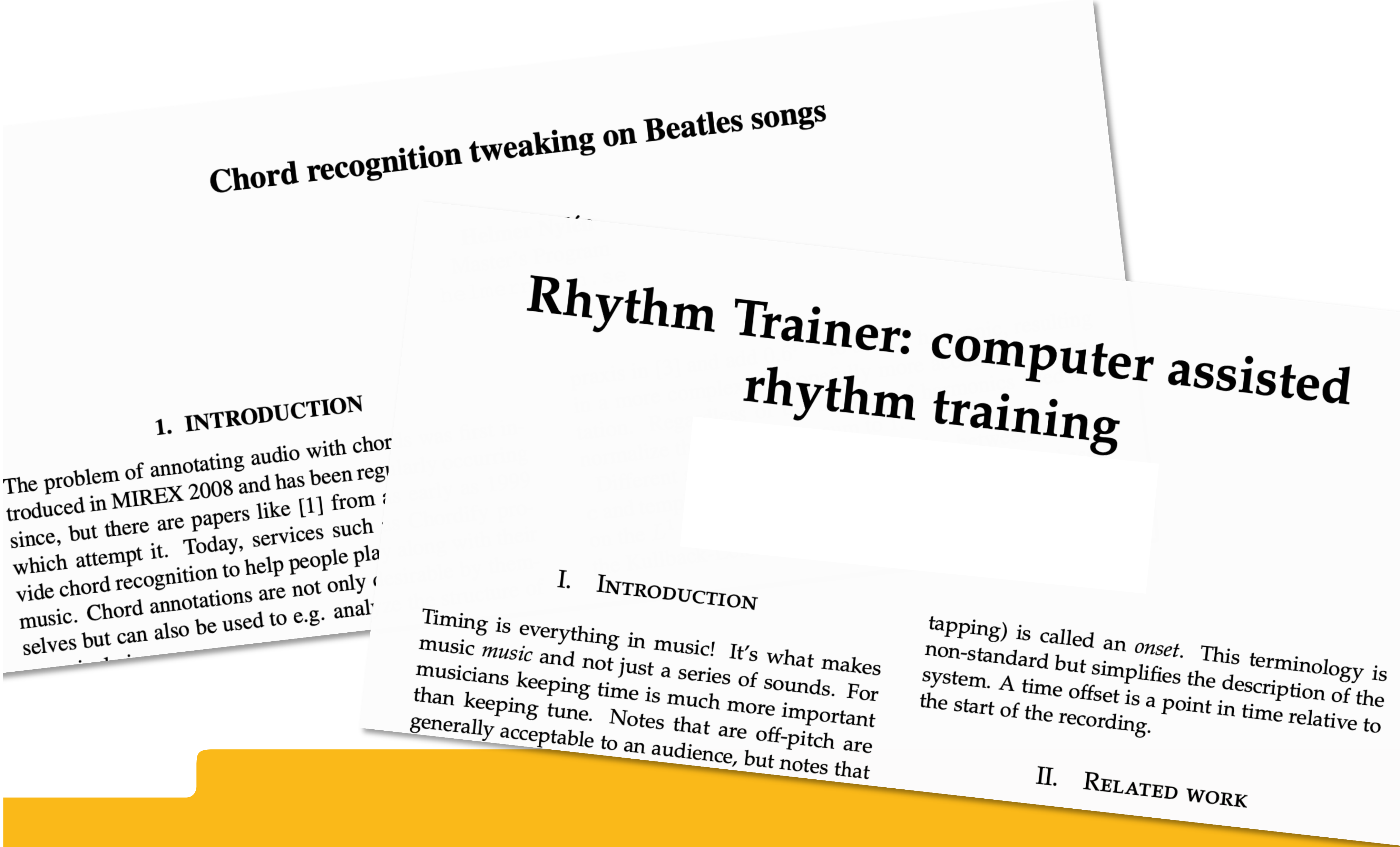
DEEP LEARNING APPROACH TO MUSICAL GENRE RECOGNITION

1. INTRODUCTION

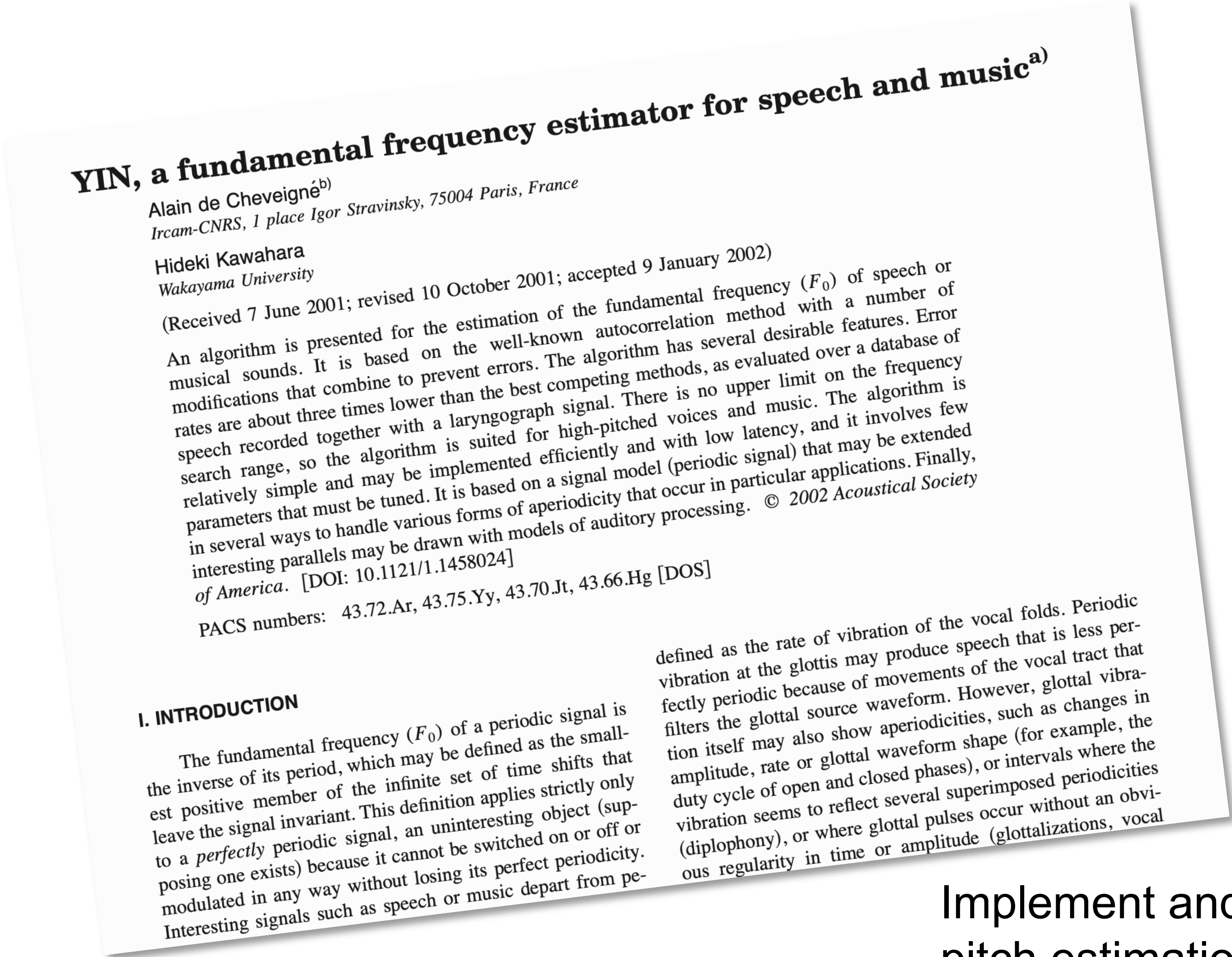
The capability of a system to automatically r
classify musical genres has always been o
problems in the music informatics field. Ap
from automatic labelling of large datasets
ommendation systems. This project pres
tional recurrent neural network approach
method is first applied to the Kikibouba
yields excellent results. As this dataset
ple, it is not ideal to asses the generali
of the model. Therefore, the method is
FMA [1] dataset. With a more com
sults aren't as desirable as with the fir
port will first explain the method and
Then it will present the results in de
discuss possible enhancements and

Figure 2. Accuracy and loss of the CRNN trained on the FMA dataset.

Example projects



Example project



YIN, a fundamental frequency estimator for speech and music[a)]

Alain de Cheveigné[b)]
Ircam-CNRS, 1 place Igor Stravinsky, 75004 Paris, France

Hideki Kawahara
Wakayama University

(Received 7 June 2001; revised 10 October 2001; accepted 9 January 2002)

An algorithm is presented for the estimation of the fundamental frequency (F_0) of speech or musical sounds. It is based on the well-known autocorrelation method with a number of modifications that combine to prevent errors. The algorithm has several desirable features. Error rates are about three times lower than the best competing methods, as evaluated over a database of speech recorded together with a laryngograph signal. There is no upper limit on the frequency search range, so the algorithm is suited for high-pitched voices and music. The algorithm is relatively simple and may be implemented efficiently and with low latency, and it involves few parameters that must be tuned. It is based on a signal model (periodic signal) that may be extended in several ways to handle various forms of aperiodicity that occur in particular applications. Finally, interesting parallels may be drawn with models of auditory processing. © 2002 *Acoustical Society of America*. [DOI: 10.1121/1.1458024]

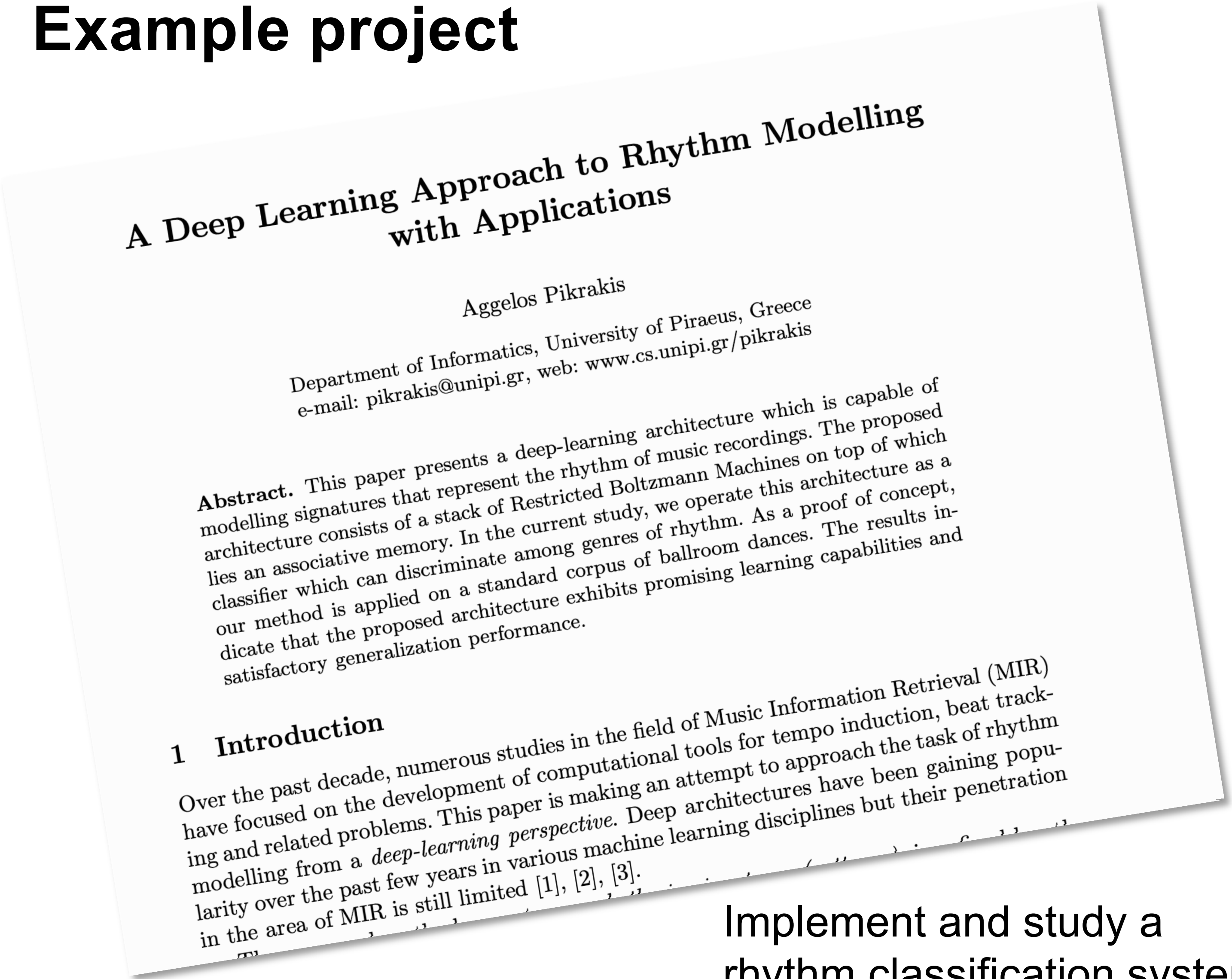
PACS numbers: 43.72.Ar, 43.75.Yy, 43.70.Jt, 43.66.Hg [DOS]

I. INTRODUCTION

The fundamental frequency (F_0) of a periodic signal is the inverse of its period, which may be defined as the smallest positive member of the infinite set of time shifts that leave the signal invariant. This definition applies strictly only to a *perfectly* periodic signal, an uninteresting object (supposing one exists) because it cannot be switched on or off or modulated in any way without losing its perfect periodicity. Interesting signals such as speech or music depart from periodicity... defined as the rate of vibration of the vocal folds. Periodic vibration at the glottis may produce speech that is less perfectly periodic because of movements of the vocal tract that filters the glottal source waveform. However, glottal vibration itself may also show aperiodicities, such as changes in amplitude, rate or glottal waveform shape (for example, the duty cycle of open and closed phases), or intervals where the vibration seems to reflect several superimposed periodicities (diplophony), or where glottal pulses occur without an obvious regularity in time or amplitude (glottalizations, vocal

Implement and study a pitch estimation algorithm
- Write a lab/tutorial.

Example project



A Deep Learning Approach to Rhythm Modelling with Applications

Aggelos Pikrakis

Department of Informatics, University of Piraeus, Greece
e-mail: pikrakis@unipi.gr, web: www.cs.unipi.gr/pikrakis

Abstract. This paper presents a deep-learning architecture which is capable of modelling signatures that represent the rhythm of music recordings. The proposed architecture consists of a stack of Restricted Boltzmann Machines on top of which lies an associative memory. In the current study, we operate this architecture as a classifier which can discriminate among genres of rhythm. As a proof of concept, our method is applied on a standard corpus of ballroom dances. The results indicate that the proposed architecture exhibits promising learning capabilities and satisfactory generalization performance.

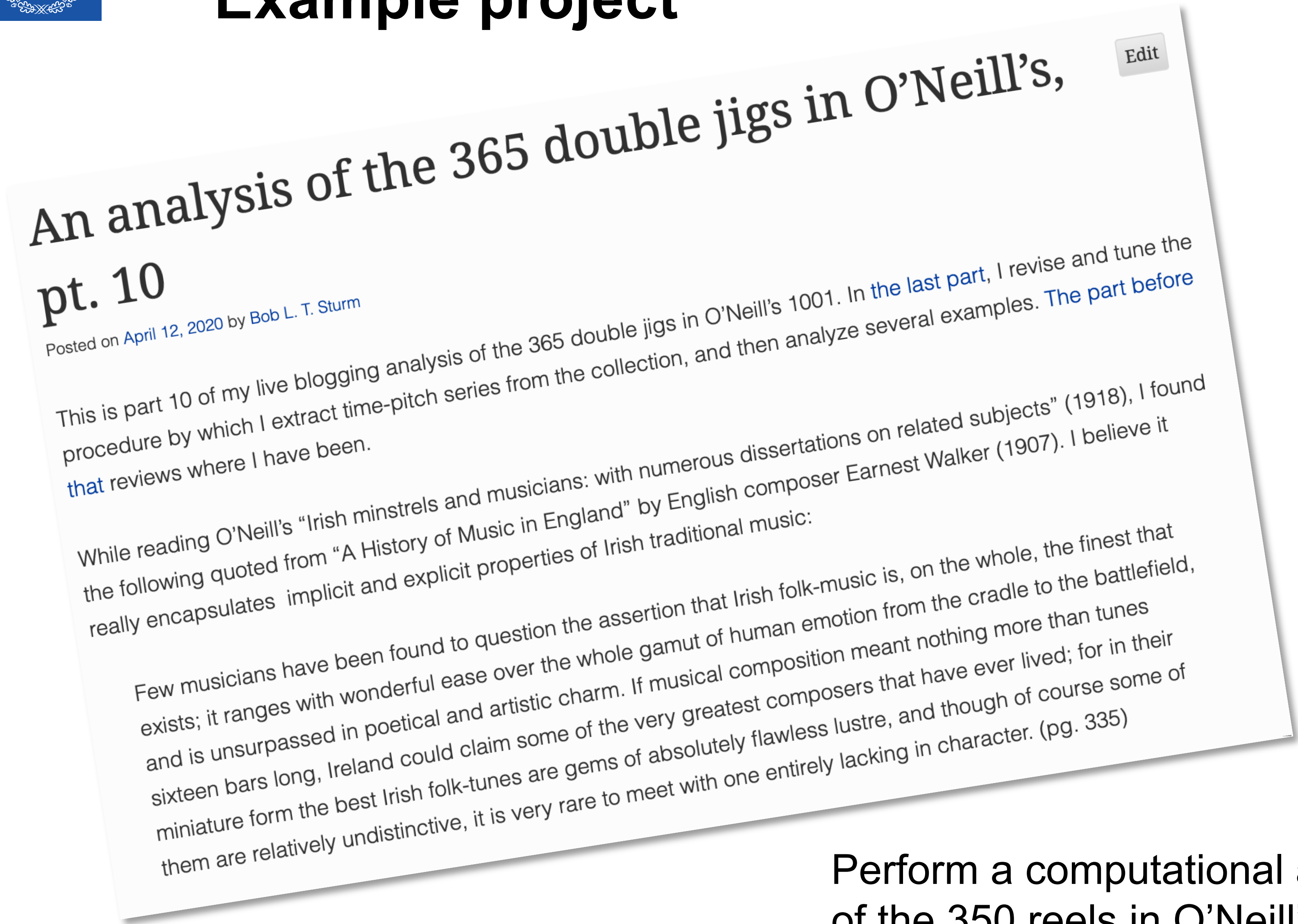
1 Introduction

Over the past decade, numerous studies in the field of Music Information Retrieval (MIR) have focused on the development of computational tools for tempo induction, beat tracking and related problems. This paper is making an attempt to approach the task of rhythm modelling from a *deep-learning perspective*. Deep architectures have been gaining popularity over the past few years in various machine learning disciplines but their penetration in the area of MIR is still limited [1], [2], [3].

Implement and study a rhythm classification system trained with data augmentation.

- Write a lab/tutorial.
- Email bobs@kth.se

Example project



Edit

An analysis of the 365 double jigs in O'Neill's, pt. 10

Posted on April 12, 2020 by Bob L. T. Sturm

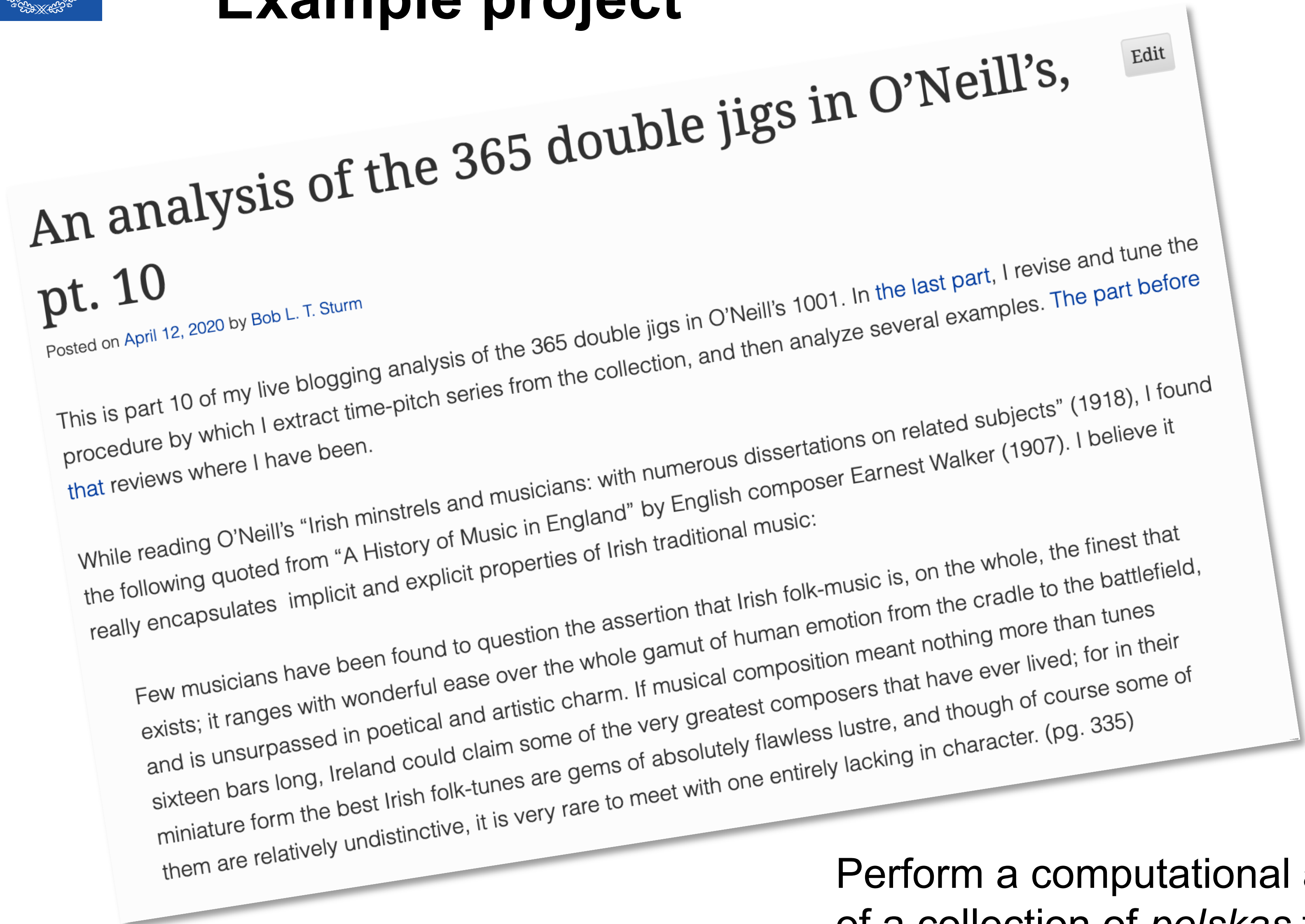
This is part 10 of my live blogging analysis of the 365 double jigs in O'Neill's 1001. In the last part, I revise and tune the procedure by which I extract time-pitch series from the collection, and then analyze several examples. The part before that reviews where I have been.

While reading O'Neill's "Irish minstrels and musicians: with numerous dissertations on related subjects" (1918), I found the following quoted from "A History of Music in England" by English composer Earnest Walker (1907). I believe it really encapsulates implicit and explicit properties of Irish traditional music:

> Few musicians have been found to question the assertion that Irish folk-music is, on the whole, the finest that exists; it ranges with wonderful ease over the whole gamut of human emotion from the cradle to the battlefield, and is unsurpassed in poetical and artistic charm. If musical composition meant nothing more than tunes sixteen bars long, Ireland could claim some of the very greatest composers that have ever lived; for in their miniature form the best Irish folk-tunes are gems of absolutely flawless lustre, and though of course some of them are relatively undistinctive, it is very rare to meet with one entirely lacking in character. (pg. 335)

Perform a computational analysis of the 350 reels in O'Neill's "1001".
See: https://highnoongmt.wordpress.com
Email bobs@kth.se

Example project



Edit

An analysis of the 365 double jigs in O'Neill's, pt. 10

Posted on April 12, 2020 by Bob L. T. Sturm

This is part 10 of my live blogging analysis of the 365 double jigs in O'Neill's 1001. In the last part, I revise and tune the procedure by which I extract time-pitch series from the collection, and then analyze several examples. The part before that reviews where I have been.

While reading O'Neill's "Irish minstrels and musicians: with numerous dissertations on related subjects" (1918), I found the following quoted from "A History of Music in England" by English composer Earnest Walker (1907). I believe it really encapsulates implicit and explicit properties of Irish traditional music:

> Few musicians have been found to question the assertion that Irish folk-music is, on the whole, the finest that exists; it ranges with wonderful ease over the whole gamut of human emotion from the cradle to the battlefield, and is unsurpassed in poetical and artistic charm. If musical composition meant nothing more than tunes sixteen bars long, Ireland could claim some of the very greatest composers that have ever lived; for in their miniature form the best Irish folk-tunes are gems of absolutely flawless lustre, and though of course some of them are relatively undistinctive, it is very rare to meet with one entirely lacking in character. (pg. 335)

Perform a computational analysis of a collection of *polskas* from www.folkwiki.se
Email bobs@kth.se

Another project: Cover song detection

1. Use the Second Hand Song dataset.
2. Explore its genre and geographical distribution.
3. Use some songs from various countries (not in the dataset) and see how many Beatles songs people of the world seem to cover.

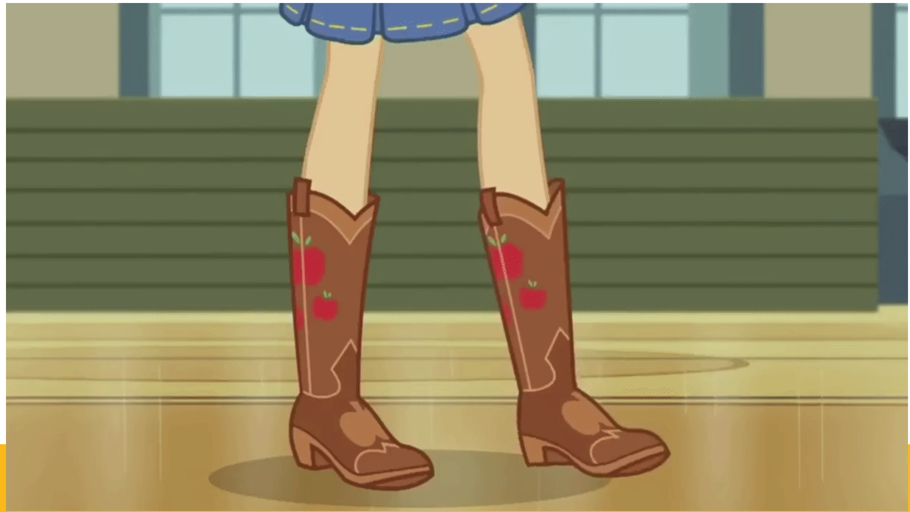
https://youtu.be/F5ky5CIljL8

Another: Time precision of onset detection

1. use MadMom onset detector and a one based on spectral flux
2. determine their accuracies with a dataset of Jazz drums and bass signals annotated with high time accuracy
   - See https://www.audiocontentanalysis.org/data-sets/
3. Can retraining MadMom increase time accuracy?

Another: “Track the Tap”

1. Use audio recordings of musicians tapping the beat to folk dance music while they play, e.g., Swedish or Irish.
2. Analyze these tapping signals with downbeat trackers to explore the consistency of the musicians tapping
3. Is the human foot a reliable metronome?



Grading

1. Laboratory work, 3.0 credits, Grading scale: P, F
2. **Project (implementation and evaluation), 3.0 credits, Grading scale: A–F**
3. **Written report and presentation, 1.5 credits, Grading scale: A–F**

Intended Learning Outcomes

After passing the course, the student should be able to:

- explain how music can be represented in reality and in the computer:

E: can **describe** a few representations (e.g., sampled audio, MIDI, feature vectors, piano roll, ...), and a few applications (e.g., search and retrieval, recommendation, transcription ...)

C: also, can **compare and contrast** representations and **explain** applications

A: also, can **extend** representations and applications

Assessed in Labs and Project Report

Intended Learning Outcomes

After passing the course, the student should be able to:

- account for how feature extraction works and explain why it is needed:

E: can **explain** windowing and computing statistics of windowed data

C: also, can **illustrate** procedure

A: also, can **choose** appropriate parameter settings for a given music informatics application

Assessed in Labs and Project Report

Intended Learning Outcomes

After passing the course, the student should be able to:

- summarise and explain which distinctive features can be extracted from a music signal, based on time, frequency and time-frequency:

E: can **name** some features in each domain

C: also, can **describe** how the features are computed

A: also, can **choose** appropriate features for a given music informatics application

Assessed in Labs and Project Report

Intended Learning Outcomes

After passing the course, the student should be able to:

- use existing software libraries for feature extraction and interpret distinctive features that have been extracted from a music signal:

E: can **extract** features and display them for a given signal

C: also, can **explain** and **interpret** these features

A: also, can **integrate** these features into a music informatics tool

Assessed in Labs and Project

Intended Learning Outcomes

After passing the course, the student should be able to:

- recommend methods for comparing and modelling of music data:

E: can **name** some modeling approaches in music informatics

C: also, can **explain** and **illustrate** these methods

A: also, can **compare and contrast** these methods

Assessed in Project Report

Intended Learning Outcomes

After passing the course, the student should be able to:

- design and implement own methods for modelling of music data:

E: can **describe** a modeling approach as an algorithm

C: also, can **implement** a method in computer code

A: also, can **explain** how such methods can be evaluated

Assessed in Labs and Project

Intended Learning Outcomes

After passing the course, the student should be able to:

- describe how information on different abstraction levels can be extracted from music data (acoustic as well as symbolic) and be used in many applications (e.g., search, retrieval, synthesis):

E: can **describe** the music informatics pipeline

C: also, can **identify** where limitations are encountered

A: also, can **choose** appropriate evaluation methodologies

Assessed in Labs and Project Report

Intended Learning Outcomes

After passing the course, the student should be able to:

- design algorithms for handling and modelling of music data as well as evaluate their performance:

E: **use** existing software libraries to extract a few basic features, model them, and evaluate them using basic approaches

C: also **use** existing software libraries to extract a variety of relevant features, model them with relevant methods, and evaluate them using many approaches

A: also **improve** the performance of developed models

Assessed in Labs and Project

Intended Learning Outcomes

After passing the course, the student should be able to:

- be able to appreciate the latest technology in music informatics and build on it:

E: **name** a few companies that use music informatics

C: also **explain** what aspects of music informatics might be employed at said companies

A: also **elaborate** upon novel applications of music informatics

Assessed in Project Report

Project Report

An excellent project report will clearly answer the following:

- What are you doing?
- Why are you doing that?
- Who has done something similar?
- How are you doing that?
- How well did you do that?
- How could that be improved?

What to do next

1. Think about problems!
2. Send a mail with proposed problems to get approval:

bobs@kth.se

holzap@kth.se

Important dates:

Seminar in week 43 to present work in progress

Final report is due end week 44