Applied Informatics /
Cooperative Systems
AICOS

Technische Universität München

Towards Mapping Timbre to Emotional Affect

Niklas Klügel
<kluegel@in.tum.de>

May 30, 2013

Outline

1. Motivation
2. Issues
3. Creating the Model
4. Quantitative Evaluation

Introduction

About myself: Ph.D. Thesis (CS) about Supporting Collaborative Music Creation (in SDGs)
→ integrate people with different musical background
→ lower thresholds for engagement

Problem for users: control parameters $\mathcal{P}$ for synthesis mostly based on technical parameters (not perception/affect)

⇒ expert / technical know-how *necessary* ↯ novices & adaptive systems

Goal (long-term): Support selection and control/synthesis of *sounds* in regard to affect

Focus here is

How to get the affective emotional insight into timbre?

Introduction

What is affect?

- umbrella term, covers all evaluative/valenced states (emotion, mood, preference)
- furthermore: no difference between perspective of composer & listener

Why affect & timbre?

for singular notes , timbre contains cues that indicate affective expression independently of the presence or absence of other cues [GAMM04, EFA12, HOH+09]

Idea: create a model that has insight into the affect $\mathcal{V}$ of the timbre $\mathcal{T}$ of sounds

if we can construct mapping $\mathcal{T} \rightarrow \mathcal{V}$,
$\mathcal{V} \rightarrow \mathcal{P}$ should be feasible

Use-Cases

Some straightforward ideas. . .

If we have $\mathcal{V} \rightarrow \mathcal{P}$, the model can be applied such that:

- user(s) are given a set of sounds fitting to a preselected mood
- give set of different timbres, users can still vary timbres within bounds of similar affect

Issues

Problem I:

analytic (holistic, explicit) understanding of the affect of timbre is still missing

→ use Machine Learning (ML)

Problem II:

for ML a suitable data set is needed to construct $\mathcal{T} \rightarrow \mathcal{V}$,
e.g. sounds, each one having an affect value/class assigned

- contains a large variety of sounds with the same affect
- AND a large variety of affects;
- singular sounds (no mixture) ↯ MIR e.g. MTurk
- non specific sounds (generalization) ↯ IADS etc.

→ create own data set

Issues II

Problem III:

as will be shown, established ML methods don't perform well. . .

→ use Deep Belief Networks

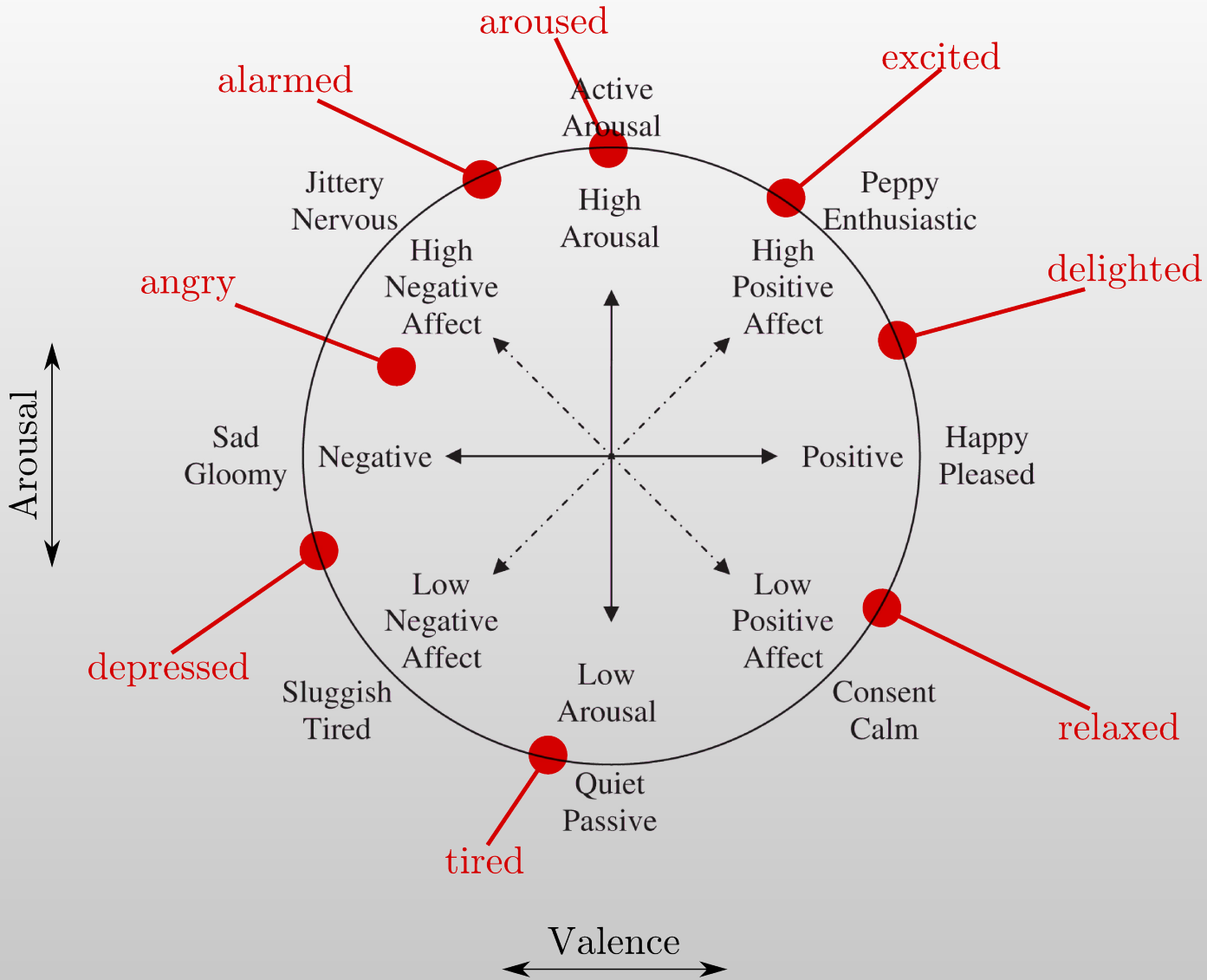
"Problem IV": ≈ definition of representation

what is affect and timbre?

affect	=	value in Valence/Arousal (V/A) space (emotional affective states) [Rus80] unambiguous, lin. combination possible
timbre	=	meaningful audio features [LTE08]

$\Rightarrow \mathcal{T} \rightarrow \mathcal{V}$ is a projection from high dim. audio feature space to low dim. V/A space

Valence/Arousal plot [Rus80]



Creating the data set

Source & analysis

Source: The Freesound Project (`http://freesound.org`)

user tagged online DB of audio samples

samples ~ audio features, tags ~ folksonomy to discern affect

V/A values, filtered tags using mult. dict. analyses:

- sentiment related synonyms per tag (SentiWordNet)
- remove tags not referring to affect
- give tag/synonyms a V/A value (ANEW)
- mean of remaining V/A values is used as label

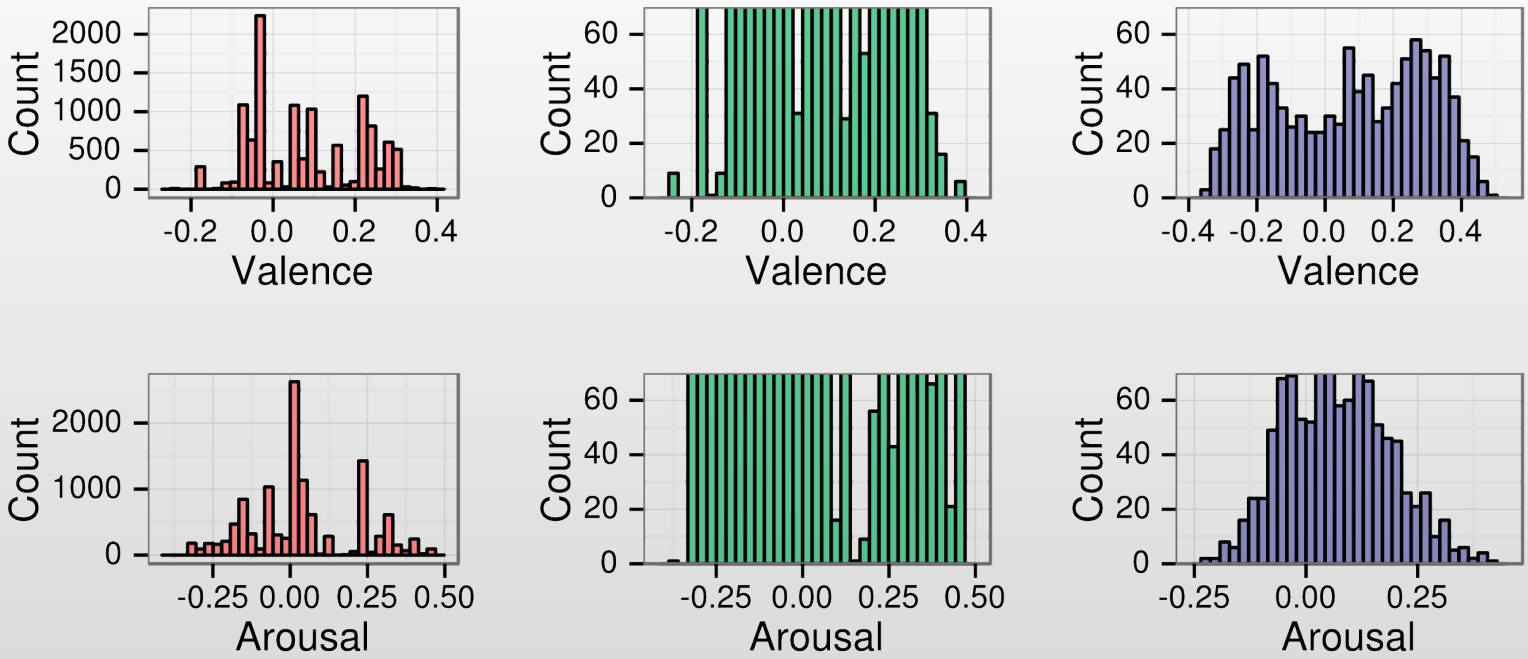
Audio features: MIRToolbox

⇒ 12k V/A labeled sounds with features ≈ **648 GiB** of data!

Creating the data set II
Remarks

- drop time-series features (memory constraint, tags describe the whole sound only) using statistical features
  → 648GiB reduced to 48MiB!
  per sound, ≈ 60 dim. time-series matrix reduced to single 386 dim. vector
- literature on genre/mood class. of *songs*: no significant singular feat./class correlation
- no key/chromatic/rhythmic features used here → less information cp. to MIR

V/A plot of collected sounds



Regarding the V/A values:

- filtered *folksonomy* is less diverse → uneven distribution of V/A values
- *spread of V/A values* similar orig. dictionary → diversity & structure of affects adequately represented

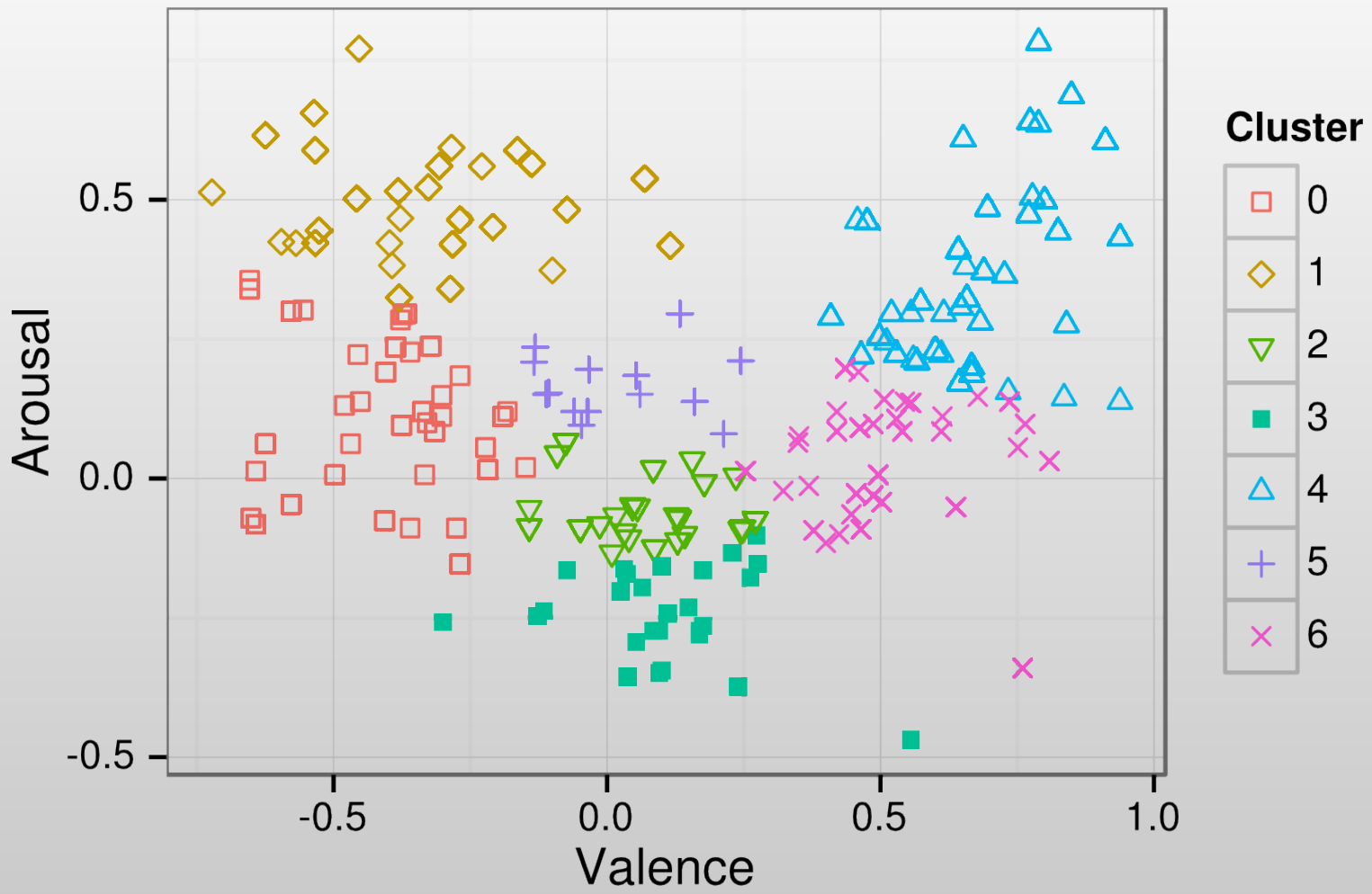
Before we go any further . . .

. . . a ***preliminary*** test:

- cluster V/A values (create 7 V/A classes), EM based
- run established ML algorithms on data set

Question: can we simply map *features* to *cluster memberships* using established ML algorithms?

V/A plot of collected sounds, clusters are used for *preliminary* testing



Preliminary Test

ML algorithm shootout, clustered V/A values

Classifier	368 feat. **error %**	25 feat.(relieff) **error %**	4 feat.(relieff) **error %**
Naïve Bayes	70.61	67.82	74.75
Bayes Net	65.73	65.85	71.84
J48 DT	62.78	62.40	74.10
RandomForest	**55.59**	**52.59**	71.09
LibSVM	76.37	76.20	**70.97**

Résumé:

- Simple/shallow classifiers may not generalize enough

  → use something more powerful (DBN)

- Problem is really about finding a good projection

  → use different error measure

Deep Belief Networks

Short introduction

Deep belief nets are probabilistic generative models that are composed of multiple layers of stochastic, latent variables.

- DBN vs. NN: in practise, *several* hidden layers + final layer for classification/regression
- Complexity theory: deep architecture more efficient than shallow arch. (e.g. SVM) for modeling complex problems

→ more layers of latent variables ≈ more layers of abstraction of input

Final Method

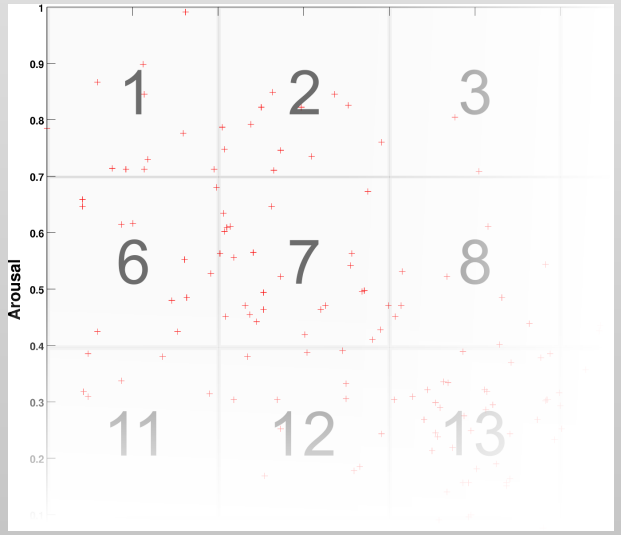
Learning method: Deep Belief Network

Coordinate “based” class-membership:

- V/A values are discretized on a grid (cheating a little)
- each sound has class membership according to grid tile

Distance-based error measure:

- baseline: random points on grid
- distance-based error measure: sum of dist. obs./pred. versus mean dist. baseline



Final result & résumé

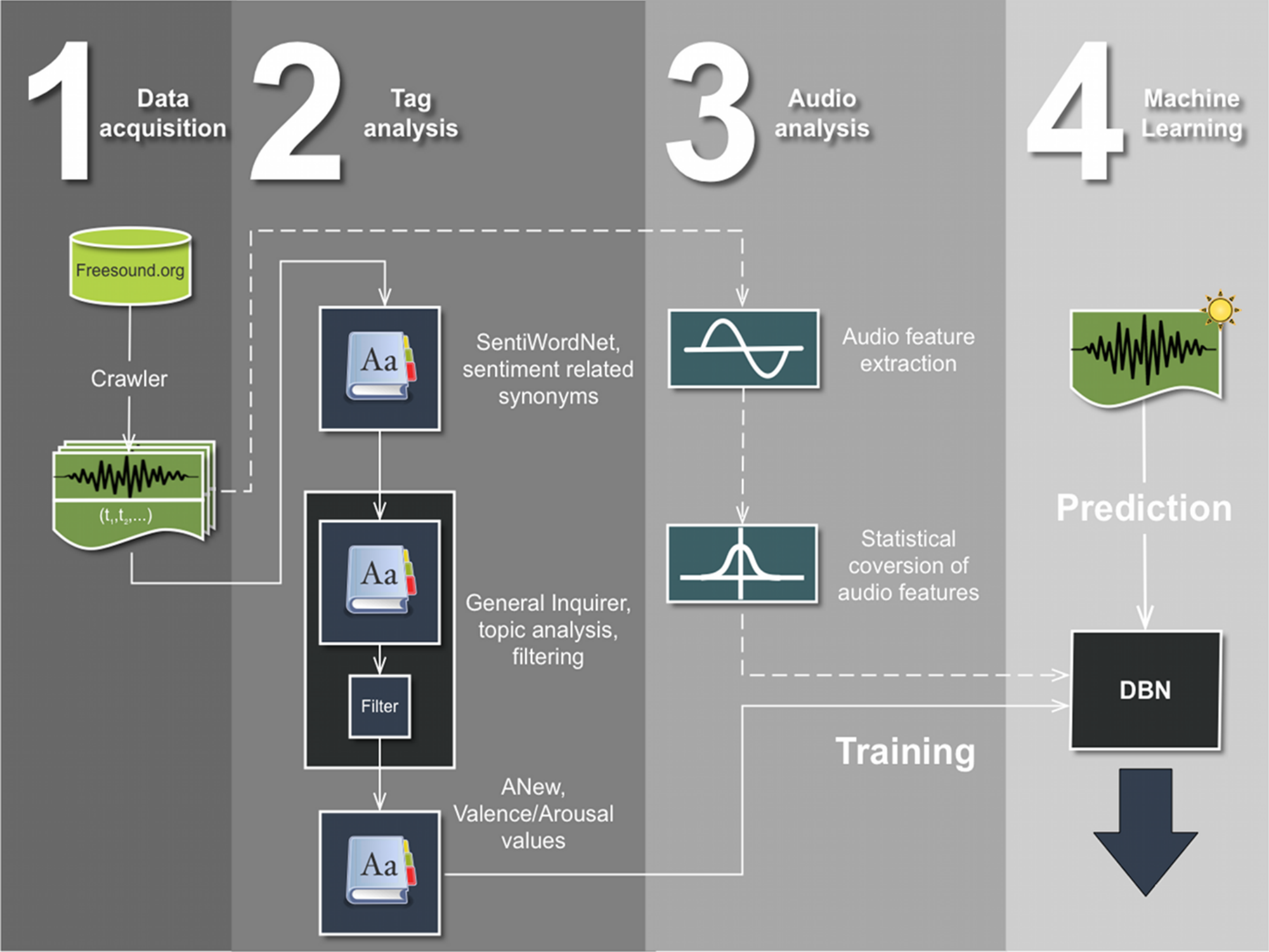
31% **error** with a 124×124 grid, 158 features
(10-fold cross validation, stratified)

- calculation for all grid sizes 25-145 took 90 days (2 GPUs)!
- the results are satisfactory such that the model is applicable to practical use-cases
- data set will be available @ `http://bit.ly/15Mi9WP`

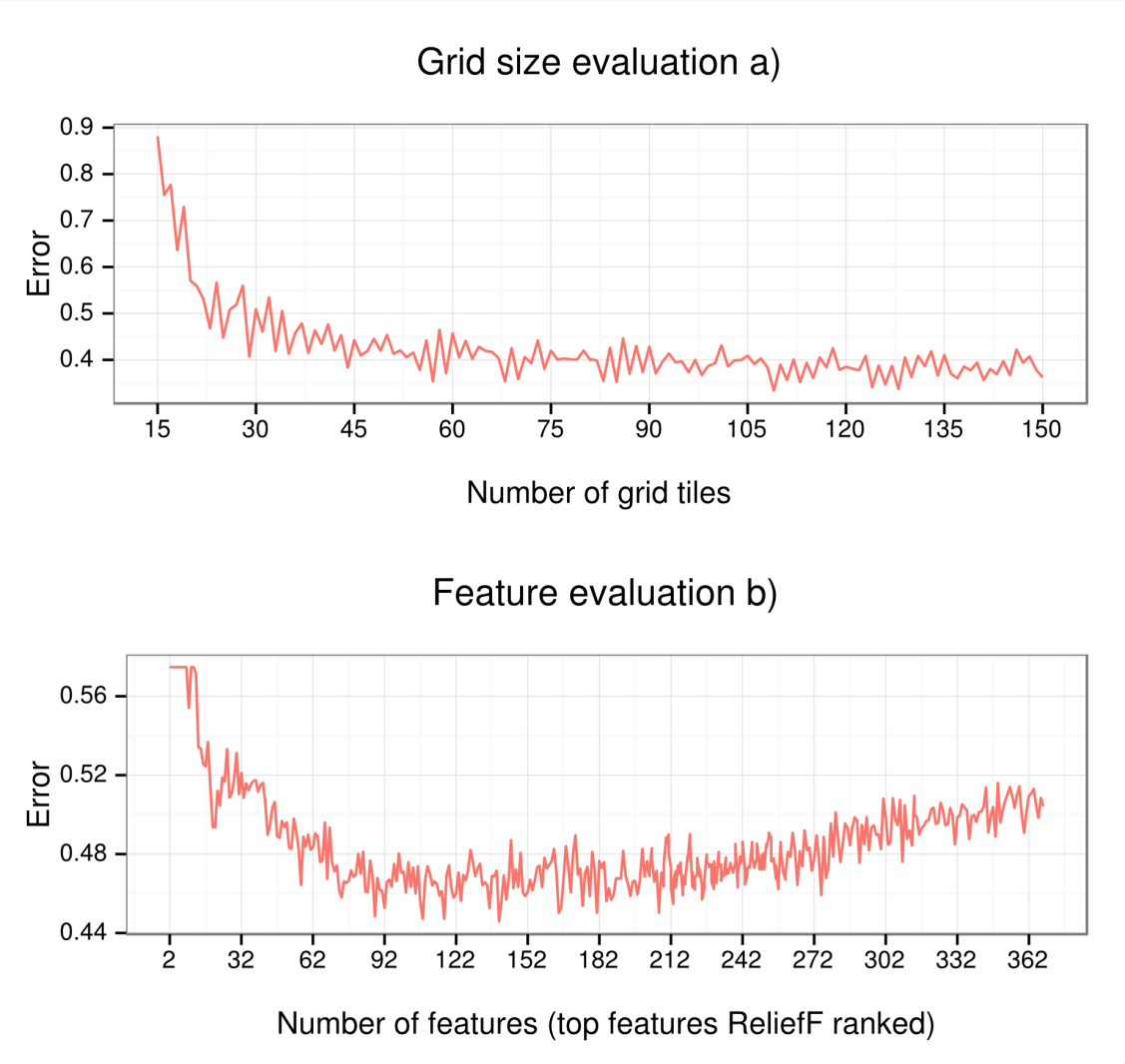
Current affairs & future:

- integration in synthesis environment
- prototypical implementation in collaborative sound design application
- qualitative evaluation

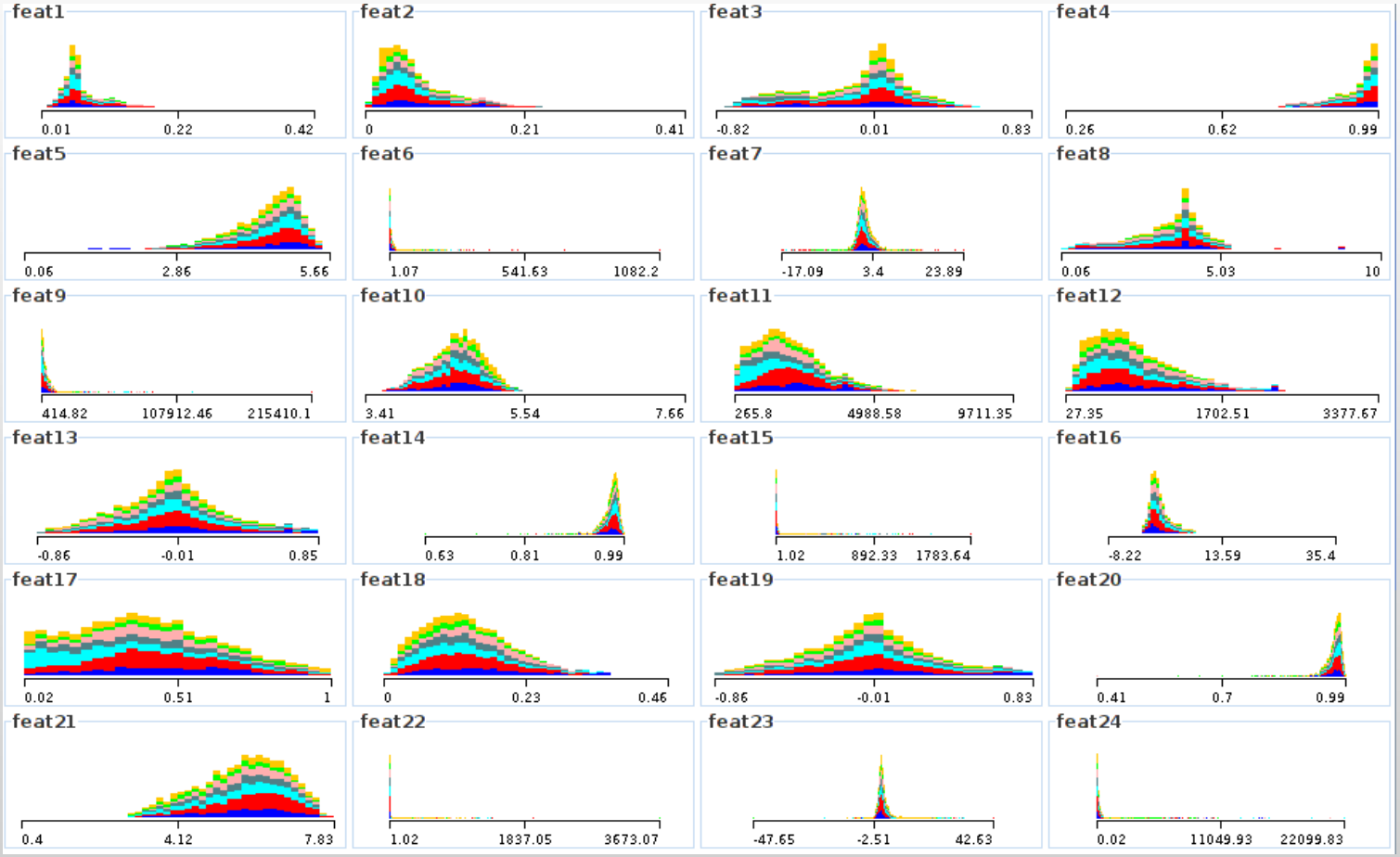
Thanks!


1
Data acquisition
2
Tag analysis
3
Audio analysis
4
Machine Learning
Freesound.org
Crawler
(t1,t2,...)
SentiWordNet, sentiment related synonyms
General Inquirer, topic analysis, filtering
Filter
ANew, Valence/Arousal values
Audio feature extraction
Statistical coversion of audio features
Prediction
DBN
Training

Grid size vs. validation error



audio feature histogram (complete dataset) vs. cluster membership



DBN, architecture: #feat, 10·#feat, 10·#feat, #gridcells

feature vector

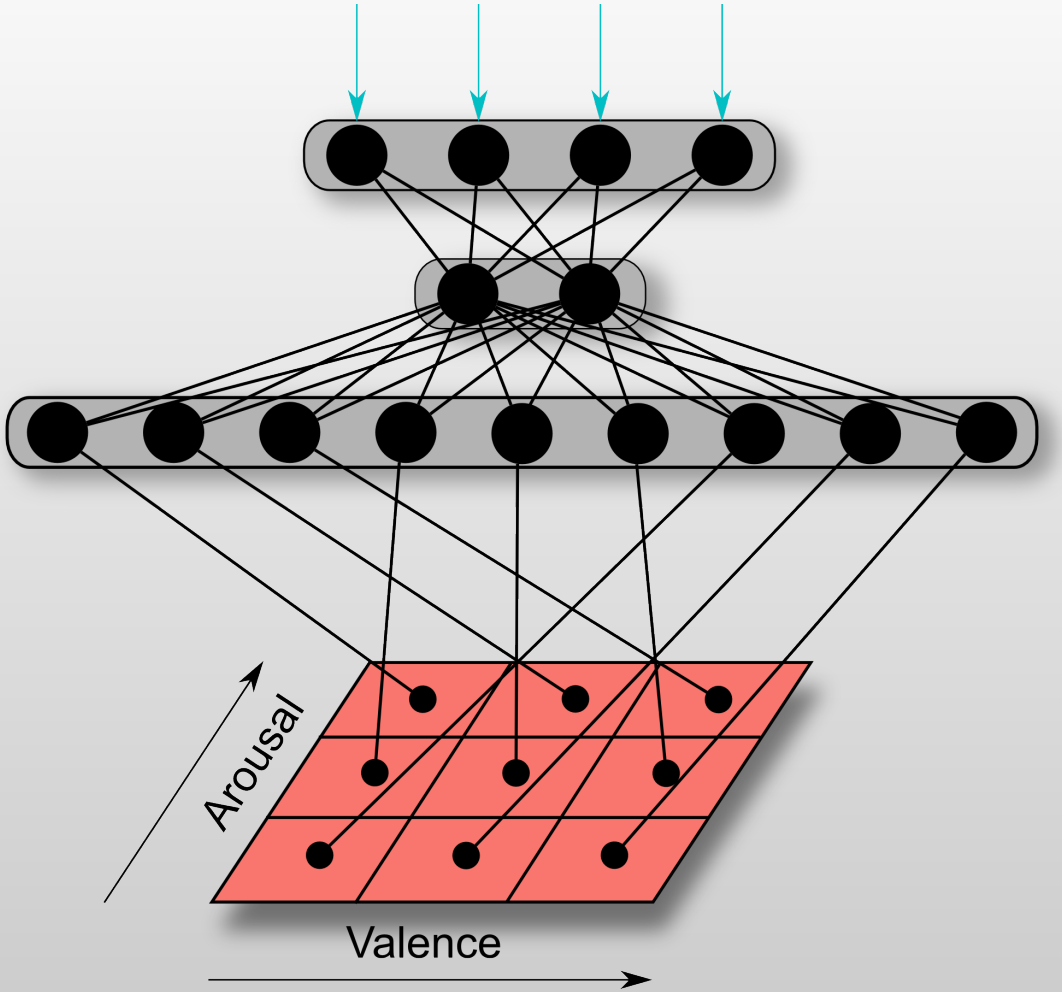
RBM_1

RBM_2

RBM_3

Logistic Regression

Valence/Arousal projection (discretized)



Dataset, V/A related stats

# unique V/A coord in dictionary	813
# unique V/A coord in data set	203
# V/A val per sample	1.32
unique V/A tags	445
# sentiment related syn.	405 (diff concepts)

2.2% of V/A tags make up 44.5% of complete data set

T Eerola, R Ferrer, and V Alluri.
Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds.
Music Perception: An Interdisciplinary Journal, 30(1):49–70, 2012.

Katja N Goydke, Eckart Altenmüller, Jürn Möller, and Thomas F Münte.
Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity.
Brain Research, 21(3):351–359, 2004.

Julia C Hailstone, Rohani Omar, Susie M D Henley, Chris Frost, Michael G Kenward, and Jason D Warren.
It's not what you play, it's how you play it: timbre affects perception of emotion in music.
Quarterly journal of experimental psychology (2006), 62(11):2141–55, November 2009.

Olivier Lartillot, Petri Toiviainen, and Tuomas Eerola.
A Matlab Toolbox for Music Information Retrieval.
Data Analysis Machine Learning and Applications, 35(M):261–268, 2008.

James A Russell.
A circumplex model of affect.
Journal of Personality and Social Psychology, 39(6):1161–1178, 1980.