Cognitive Audio Information Modeling

Jan Larsen, Associate Professor PhD
Cognitive Systems Section
Dept. of Applied Mathematics and Computer Science
Technical University of Denmark
janla@dtu.dk, people.compute.dtu.dk/janla

DTU COMPUTE

Technical University of Denmark

(founded 1829; first rector H.C. Ørsted)

Ranking

Leiden *Crown Indicator* 2010

no. 1 in Scandinavia

no. 7 in Europe

DTU facts and figures (2012)

Education

7843 BSc, MSc og Beng students

incl. 627 international MSc students

1338 PhD students

627 exchange students

291 DTU students at exhange programs

Innovation

147 registered IPR

66 submitted patent applications

Research

4011 research publications

297 PhD theses

Personel

22 DVIP

1783 VIP

1148 PhD students

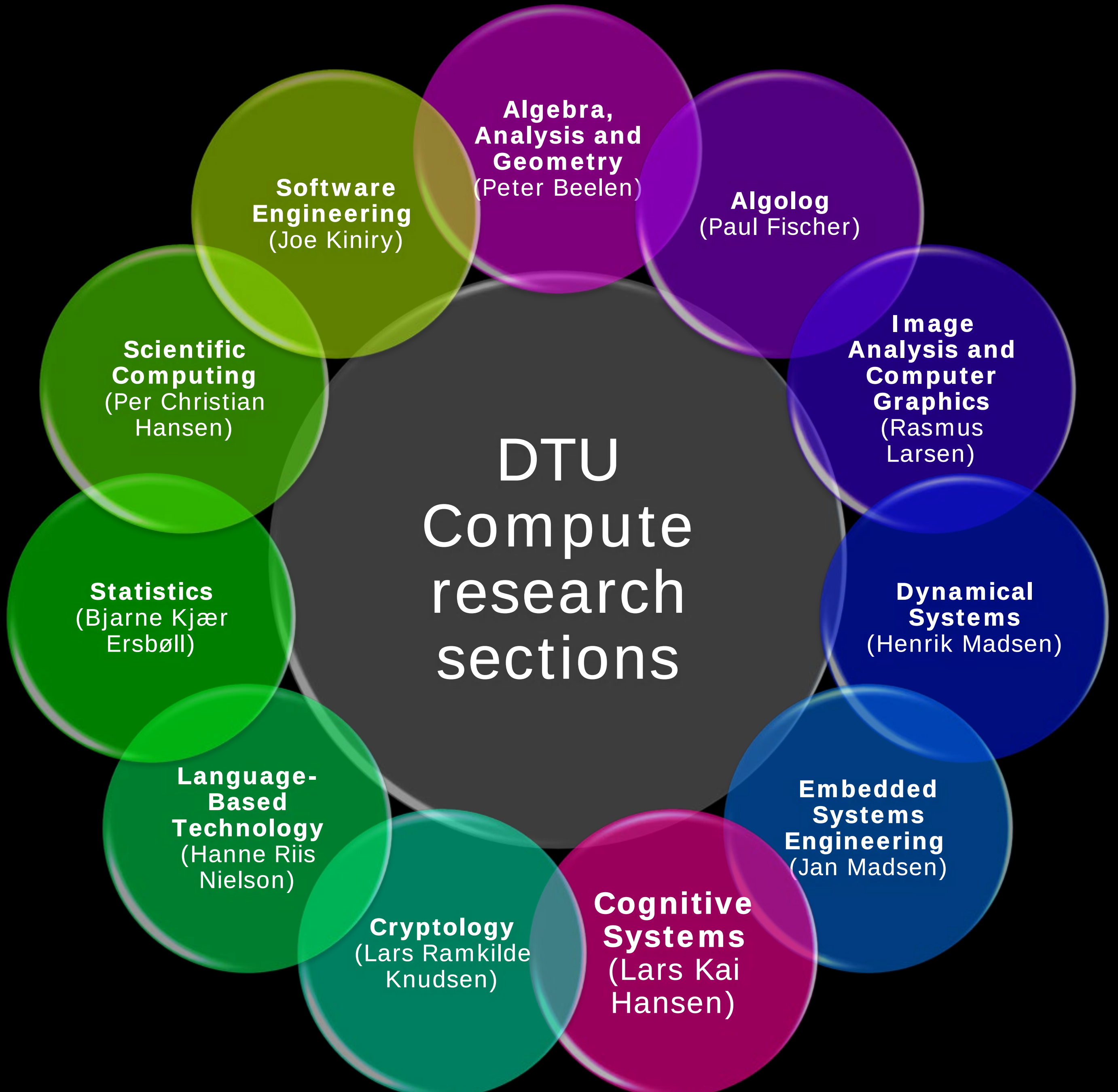
2274 TAP

Economy 7.2 bil. DKK

Public sector consultancy

Strategic contract with Danish ministries 419.9 MDKK

Buildings 482.307 m²


DTU Compute research sections
Algebra, Analysis and Geometry (Peter Beelen)
Algolog (Paul Fischer)
Image Analysis and Computer Graphics (Rasmus Larsen)
Dynamical Systems (Henrik Madsen)
Embedded Systems Engineering (Jan Madsen)
Cognitive Systems (Lars Kai Hansen)
Cryptology (Lars Ramkilde Knudsen)
Language-Based Technology (Hanne Riis Nielson)
Statistics (Bjarne Kjær Ersbøll)
Scientific Computing (Per Christian Hansen)
Software Engineering (Joe Kiniry)

Cognitive Systems Section

Why do we do it? VISION

What do we do? MISSION

machine learning

media technology

cognitive science

- 1 professor
- 7 associate prof.
- 1 assistant prof.
- 1 senior researcher
- 5 postdocs
- 17 Ph.D. students
- 5 project coordinators
- 2 programmers
- 1 admin assistant
- 10 M.Sc. students

Legacy of cognitive systems

Allan Turing

Theory of computing 1940’es

Norbert Wiener

Cybernetics

1948

processing → adaption → under-standing → cognition

information and data

people


Bjørn Sand Jensen


Jens Brehm Nielsen


Jens Madsen


Rasmus Troelsgaard


Lars Kai Hansen


Mikkel N. Schmidt


Jerónimo Arenas-García


Ling Feng

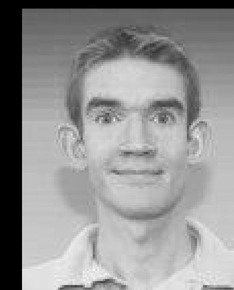

Anders Meng


Seliz Karadogan


Letizia Marchegiani


Peter Ahrendt


Michael Kai Petersen

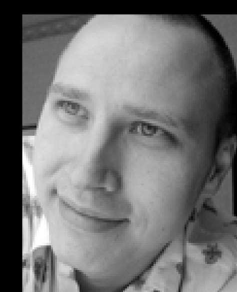

Michael Syskind Pedersen

COGNITIVE AUDIO SYSTEMS LAB

Corey Kereliuk


Lasse Lohilahti Mølgaard


Tue Lehn-Schiøler


Kaare Brandt Petersen

Mission

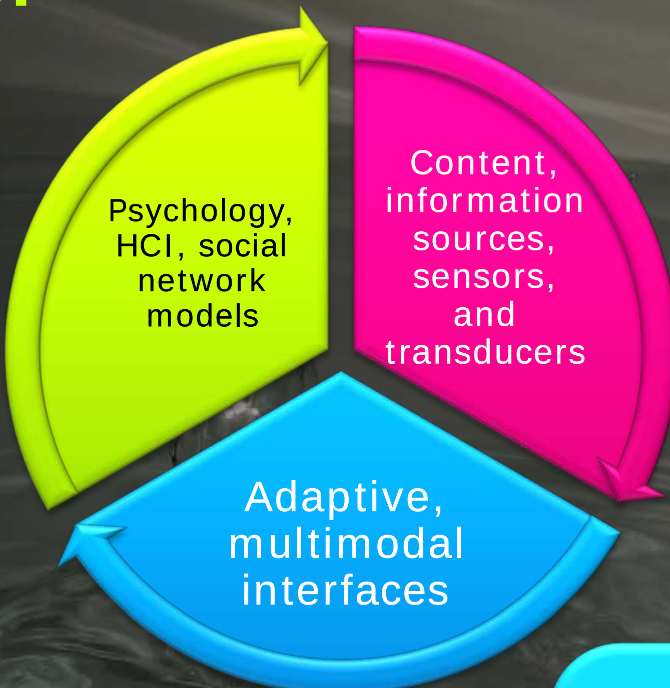
Measure, model, extract, and augment meaningful and actionable information from audio and related information, social context, psycho-physical model of the users by ubiquitous learning from data and optimizing the computational resources

Cognizant audio systems

fully informed and aware systems

Context:
who, where, what

Users in the loop:
direct and indirect



Interactive dialog with the user enables long term/continuous behavior tracking, personalization, elicitation of perceptual and affective preferences, as well as adaptation

Listen in on audio and other sensor streams to segment, identify and understand

Flexible integration with other media modalities

Mixed modality experience: Use other modalities to enhance, substitute or provide complementary information

Spectrum of research themes

- Audio segmentation
- Audio signal processing
- Genre, mood and metadata prediction
- Audio&music information retrieval
- Audio source separation
- Context based spoken document retrieval
- Elicitation of cognitive dimensions

AGENDA

- Cognitive Systems @ DTU Compute
- Introduction to cognitive systems
- Elicitation, modeling and evaluation of cognitive audio aspects
- Exercise on predicting expressed emotions in music

Literature

- Background:
  - Kenneth E. Train: Discrete Choice Methods with Simulation, Cambridge, 2nd ed., 2009. Chapters: 1,2,3.1-3.3.
  - C. E. Rasmussen & C. K. I. Williams; Gaussian Processes for Machine Learning, MIT Press, 2006, Chapters 1,2.
  - Patrik N. Juslin and Daniel Västfjäll: Emotional responses to music: The n
    Scie
- Specific
  - J. M
    Emo
    Proc
    Heid
  - Jens
    Jan
    Proc

Disclaimer: All material including documents and software is provided in accordance with the CopyDan agreement for teaching at Danish Universities. The material can only be used in connection with the PhD course and may not be redistributed or shared in any form

COGNITIVE SYSTEMS

What is it? - a vision for the future

An artificial cognitive system is the ***ultimate learning*** and thinking machine with ability to operate in ***open-ended environments*** with ***natural interaction*** with humans and other artificial cognitive systems and plays key role in the transformational society in order to achieve augmented *capabilities beyond* human and existing machines

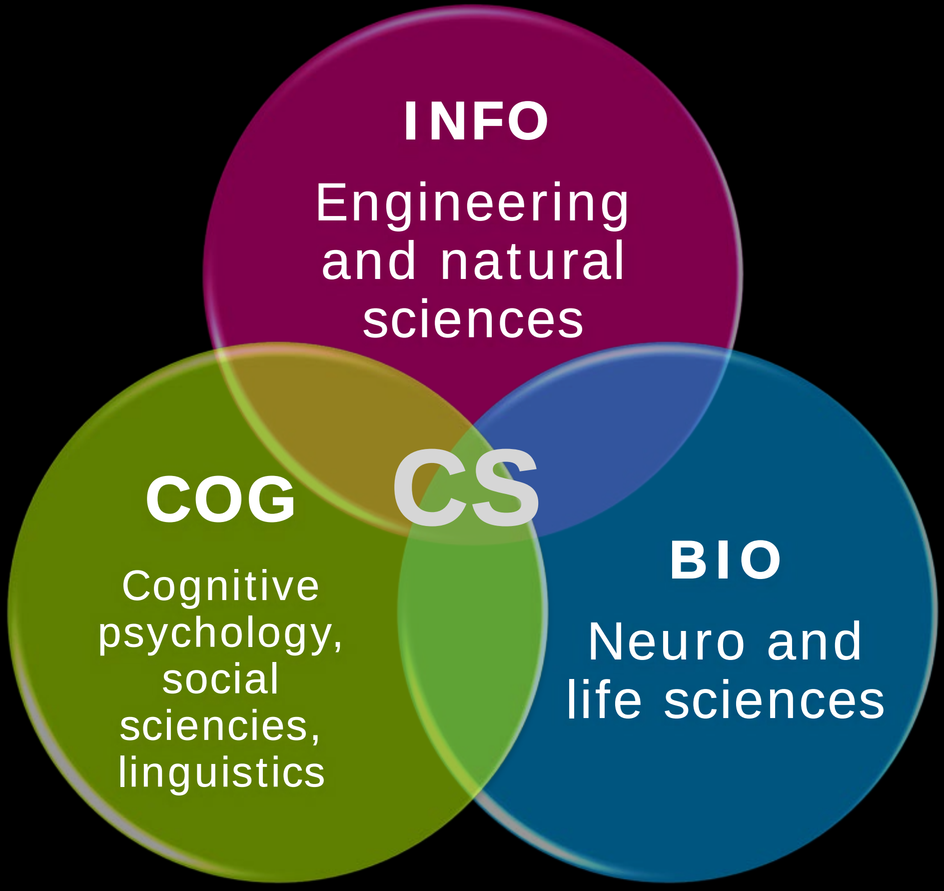
Jim Dator's definition of the transformational society: humans, and their technologies, and the environments of both, are all three merging into the same thing. Humans, as humans, are losing their monopoly on intelligence, while new forms of artificial life and artificial intelligence are emerging, eventually perhaps to supersede humanity, while the once-"natural" environments of Earth morph into entirely artificial environments that must be envisioned, designed, created and managed first by humans and then by our post-human successors.

A vision with great implications

Ubiquitous interaction between humans and artificial cognitive systems

- Ethical (maybe new regulatory bodies)
- Cultural (inclusiveness)
- Political (regulations and policies)
- Economical (digital economy and instability)
- Social (collaboration, globalization, conflicts)
- Anthropological (transformational society)

It takes cross-disciplinary effort to create a cognitive system



Ref: EC Cognitive System Unit http://cordis.europa.eu/ist/cognition/index.html

A brief history

- Late 40's Allan Touring: theory of computation
- 1948 Claude Shannon: A Mathematical Theory of Communication
- 1948 Norbert Wiener: Cybernetics - *Control and Communication in the Animal and the Machine*
- 1950 The Touring test
- 1951 Marvin Minsky's analog neural networks
- 1956 Dartmouth conference: Artificial intelligence with aim of human like intelligence
- 1956-1974 Many small scale "toy" projects in robotics, control and game solving
- 1974 Failure of success and Minsky's criticism of perceptron, lack of computational power, combinatorial explosion, Moravec's paradox: simple tasks are not easy to solve

A brief history

- 1980's Expert systems useful in restricted domains
- 1980's Knowledge based systems – integration of diverse information sources
- 1980's The neural network revolution starts
- Late 1980's Robotics and the role of embodiment to achieve intelligence
- 1990's and onward AI research under new names such as machine learning, computational intelligence, evolutionary computing, neural networks, Bayesian networks, informatics, complex systems, game theory, **cognitive systems**

Ref: http://en.wikipedia.org/wiki/Timeline_of_artificial_intelligence
http://en.wikipedia.org/wiki/History_of_artificial_intelligence

Revitalizing old ideas through cognitive systems by means of enabling technologies

Computation

distributed and ubiquitous computing

Connectivity

internet, communication technologies and social networks

Pervasive sensing

digital, accessible information on all levels

New theories of the human brain

Neuroinformatics, brain-computer interfaces, mind reading

New business models

Free tools paid by advertisement, 99+1 principle: 99% free, 1% buys, the revolution in digital economy

The unreasonable effectiveness of data

- E. Wigner 1960: The unreasonable efffectiveness of mathematics in the natural sciences.
- Simple linear classifiers ba
  representations performs b

There is often a threshold of sufficient data

- Unsupervised learning on unlabeled data which are abundant
- The power of linking many different sources
- Semantic interpretation
  - The same meaning can be expressed in many ways – and the same expression can convey many different meanings
  - Shared cognitive and cultural contexts helps the disambiguation of meaning
  - Ontologies: a social construction among people with a common shared motive
  - Classical handcrafted ontology building is infeasible – crowd computing / crowdsourcing is possible

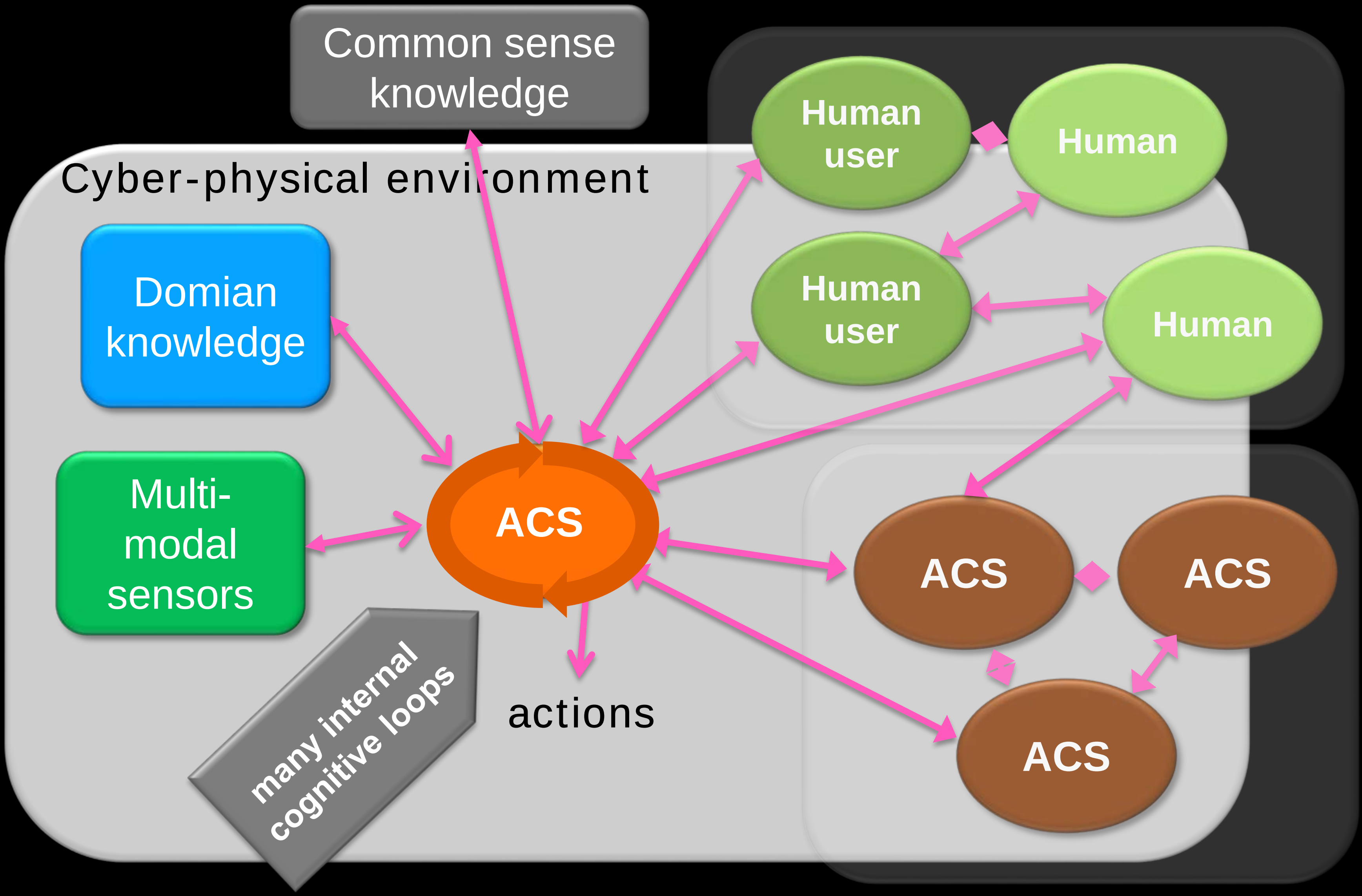
Ref: A. Halevy, P. Norvig, F. Pereira: The unreasonbale effectiveness of data,

IEEE Intelligen Systems, March/April, pp. 8-12, 2009.

A 360 degrees view of the concepts in cognitive systems

- –Why: goals
- –How: data, processing
- –What: capabilities

The cognitive system and its world



Why - goals

Disentanglement of confusing, ambiguous, conflicting and vast amounts of multimodal, multi-level data and information

Perform specific tasks

- Exploration
- Retrieval
- Search
- Physical operation and manipulation
- Information enrichment
- Making information actionable
- Navigation and control
- Decision support
- Meaning extraction
- Knowledge discovery
- Creative process modeling
- Facilitating and enhancing communication
- Narration

How – data, processing and computing

Dynamical, multi-level, integration and learning of

- heterogeneous,
- multi-modal,
- multi-representation (structured/unstructured),
- multi-quality (resolution, noise, validity)
- data, information and interaction streams

with the purpose of

- achieving relevant specific goals for a set of users,
- and ability to evaluate achievement of goals

using

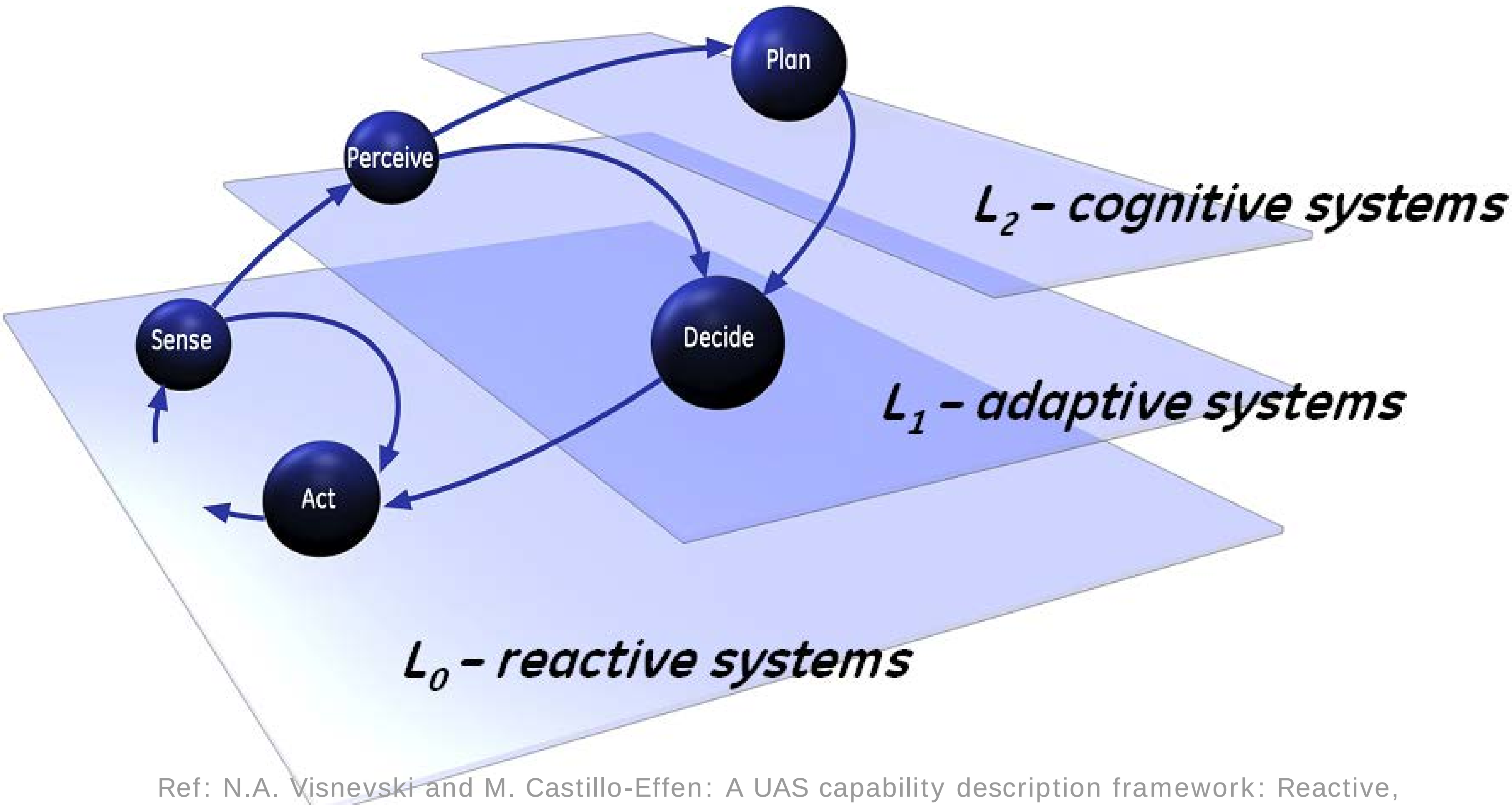
- new frameworks and architectures and
- computation (platforms, technology, swarm intelligence, grid computing, crowd computing)

Cognitive systems

How much is needed to qualify the system as being cognitive?

A tiered approach: from low to high-level capabilities

Visnevski / Castillo-Effen tiered approach



Ref: N.A. Visnevski and M. Castillo-Effen: A UAS capability description framework: Reactive, adaptive, and cognitive capabilities in robotics, 2009 IEEE Aerospace Conference, pp. 1-7, 2009.

What - capabilities

Robustness

- Perturbations and changes in the world (environment and other cognitive agents)
- Graceful degradation
- Ability to alert for incapable situations

Adaptivity

- Handling unexpected situations
- Attention
- Ability to adapt to changes at all levels: data, environment, goals
- Continuous evolution

What - capabilities

Effectiveness

- Level of autonomy
- Prediction
- Learning at all levels (interactive learning)
- Generalization
- Pro-activeness
- Multi-level planning (actions, goals)
- Simulation
- Exploration
- Self-evaluation
- Learning transfer
- Emergent behavior
- Handling of inaccuracy and deception

What - capabilities

Natural interaction

- Mediation and ontology alignment
- Handling of ambiguity, conflicts, uncertainties
- Communication
- Multi-goal achievement
- Locomotion and other physical actions

High-level emergent properties (strong AI)

- Consciousness
- Self-awareness
- Sentience (feeling)
- Empathy
- Emotion
- Intuition

Weak AI is preferred as it is easier to engineer and evaluate

A Cognitive Systems Approach to Enriched and Actionable Information from Audio Streams

COSOUND

DR

Syntonetic

Musikzonen

DTU

ChaosInsight

Royal School of Library and Information Science

Hindenburg Systems

UCL

Queen Mary University of London

B&O

Danish Council for Strategic Research Project 2012-2016

Copenhagen University

Aalborg University

State and University Library

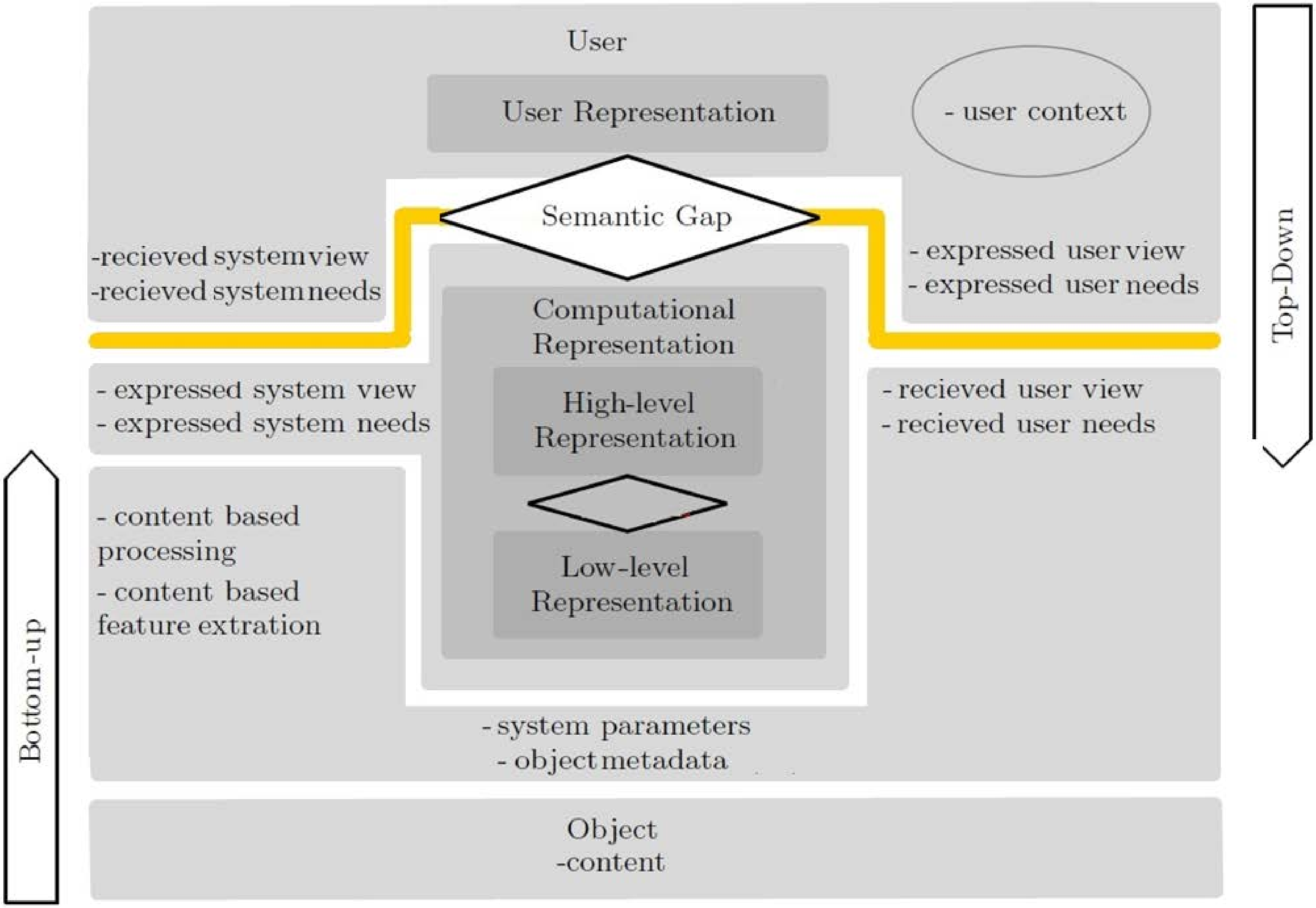
University of Glasgow

Two research tracks and overarching hypotheses

DTU

Music	Interactive enrichment
• Are emotional expressions in music essential for natural navigation and interaction as well as access to hidden but relevant music serendipity? • Is it possible to bridge the semantic gap between audio and user's semantic representations by interactive learning? • It is possible to recommend enjoyable music from the "dark" music universe using new similarities, user profiling, and interaction?	• Is it possible to effectively enrich large audio archives with additional semantic information by interactive learning and gamification, and can this lead to clarifying the importance on "Big Data as a Lens on Human Culture" and 'search tools' for the professional music/audio industry? • Is it possible to create an ontology for an audio collection, which enables the system to answer questions encoded in the ontology or can be inferred from the ontology?

Framework



User
User Representation
- user context
Semantic Gap
-recieved system view
-recieved system needs
- expressed user view
- expressed user needs
Computational Representation
High-level Representation
Low-level Representation
- expressed system view
- expressed system needs
- recieved user view
- recieved user needs
- content based processing
- content based feature extration
- system parameters
- object metadata
Object
-content
Bottom-up
Top-Down

ELICITATION OF COGNITIVE ASPECTS

Goal is to efficiently and robustly to elicit, model and predict top-down aspects such as affective, perceptual and other cognitive aspects

To understand which properties of audio content in combination with context, intention/task that drives the cognitive aspect

Modelling cognitive aspects

Affection

- **Preference elicitation** refers to the problem of developing a decision support system capable of **generating recommendations to a user, thus assisting him in decision making**. It is important for such a system to model user's preferences accurately, find hidden preferences and avoid redundancy. This problem is sometimes studied as a **computational learning theory** problem (ref. Wikipedia)
- Affect refers to the experience of feeling or emotion

Modelling cognitive aspects

Perception

Perception is the organization, identification, and interpretation of sensory information in order to represent and understand the environment. All perception involves signals in the nervous system, which in turn result from physical stimulation of the sense organs. Perception is not the passive receipt of these signals, but can be shaped by learning, memory, and expectation. Perception involves these "top-down" effects as well as the "bottom-up" process of processing sensory input (ref. Wikipedia)

Research contributions 2013/2014

- Jens Brehm Nielsen, Systems for Personalization of Hearing Instruments: A Machine Learning Approach, PhD Thesis, January 2014.
- J. Madsen, B. S. Jensen, J. Larsen, Predictive Modeling of Expressed Emotions in Music using Pairwise Comparisons, *CMMR 2012 Post-Proceedings*, vol. 7900, pp. 253-277, Springer-Verlag Berlin Heidelberg, 2013
- B. S. Jensen, J. B. Nielsen, J. Larsen, *Bounded Gaussian Process Regression*, IEEE International Workshop on Machine Learning for Signal Processing, 2013
- J. B. Nielsen, B. S. Jensen, T. J. Hansen, J. Larsen, *Personalized Audio Systems - a Bayesian Approach*, 135th AES Convention, 2013
- Jens Brehm Nielsen, Jakob Nielsen: Efficient Individualization of Hearing and Processers Sound, ICASSP2013.
- Jens Brehm Nielsen, Jakob Nielsen, Jan Larsen: Perception based Personalization of Hearing Aids using Gaussian Process and Active Learning, in preparation for IEEE Trans. ASLP, 2013.

Research contributions 2012

- Bjørn Sand Jensen, Javier Saez Gallego and Jan Larsen. *A Predictive model of music preference using pairwise comparisons*. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2012.
- Jens Madsen, Bjørn Sand Jensen, Jan Larsen and Jens Brehm Nielsen. *Towards Predicting Expressed Emotion in Music from Pairwise Comparisons*, 9th Sound and Music Computing Conference, 2012.
- Jens Madsen, Jens Brehm Nielsen, Bjørn Sand Jensen and Jan Larsen. *Modeling Expressed Emotions in Music using Pairwise Comparisons*. 9th International Symposium on Computer Music Modeling and Retrieval (CMMR) 2012.
- Jens Brehm Nielsen, Bjørn Sand Jensen and Jan Larsen, *Pseudo Inputs For Pairwise Learning With Gaussian Processes*, IEEE International Workshop on Machine Learning for Signal Processing, 2012.
- S. G. Karadogan, J. Larsen, *Combining Semantic and Acoustic Features for Valence and Arousal Recognition in Speech*, Cognitive Information Processing CIP2012, IEEE Press, 2012
- Bjørn Sand Jensen, Integration of top-down and bottom-up information for audio organization and retrieval, PhD thesis, Kgs. Lyngby, Technical University of Denmark, 2012. 197 p. (IMM-PhD-2012; No. 291).
- Seliz Karadogan, Towards Cognizant Hearing Aids: Modeling of Content, Affect and Attention. PhD Thesis, Technical University of Denmark, 2012. 142 p. (IMM-PhD-2012; No. 275).

Research contributions 2011

- Bjørn Sand Jensen, Jens Brehm Nielsen, and Jan Larsen. *Efficient Preference Learning with Pairwise Continuous Observations and Gaussian Processes*, IEEE International Workshop on Machine Learning for Signal Processing, 2011.
- S. G. Karadogan, L. Marchegiani, J. Larsen, L. K. Hansen, *Top-Down Attention with Features Missing at Random*, International Workshop on Machine Learning for Signal Processing, IEEE Press, 2011
- J. B. Nielsen, B. S. Jensen, J. Larsen, *On Sparse Multi-Task Gaussian Process Priors for Music Preference Learning*, NIPS 2011 Workshop on Choice Models and Preference Learning, 2011
- L. Marchegiani, S. G. Karadogan, T. Andersen, J. Larsen, L. K. Hansen, *The Role of Top-Down Attention in the Cocktail Party: Revisiting Cherry's Experiment after Sixty Years*, The tenth International Conference on Machine Learning and Applications (ICMLA'11), 2011

Use cases

Interactive development

- Iterative system development on a budget

Performance evaluation

- Identify the best audio system among a fixed set of systems
- Audio system feature sensitivity/importance
- Evaluation and comparison of system performance

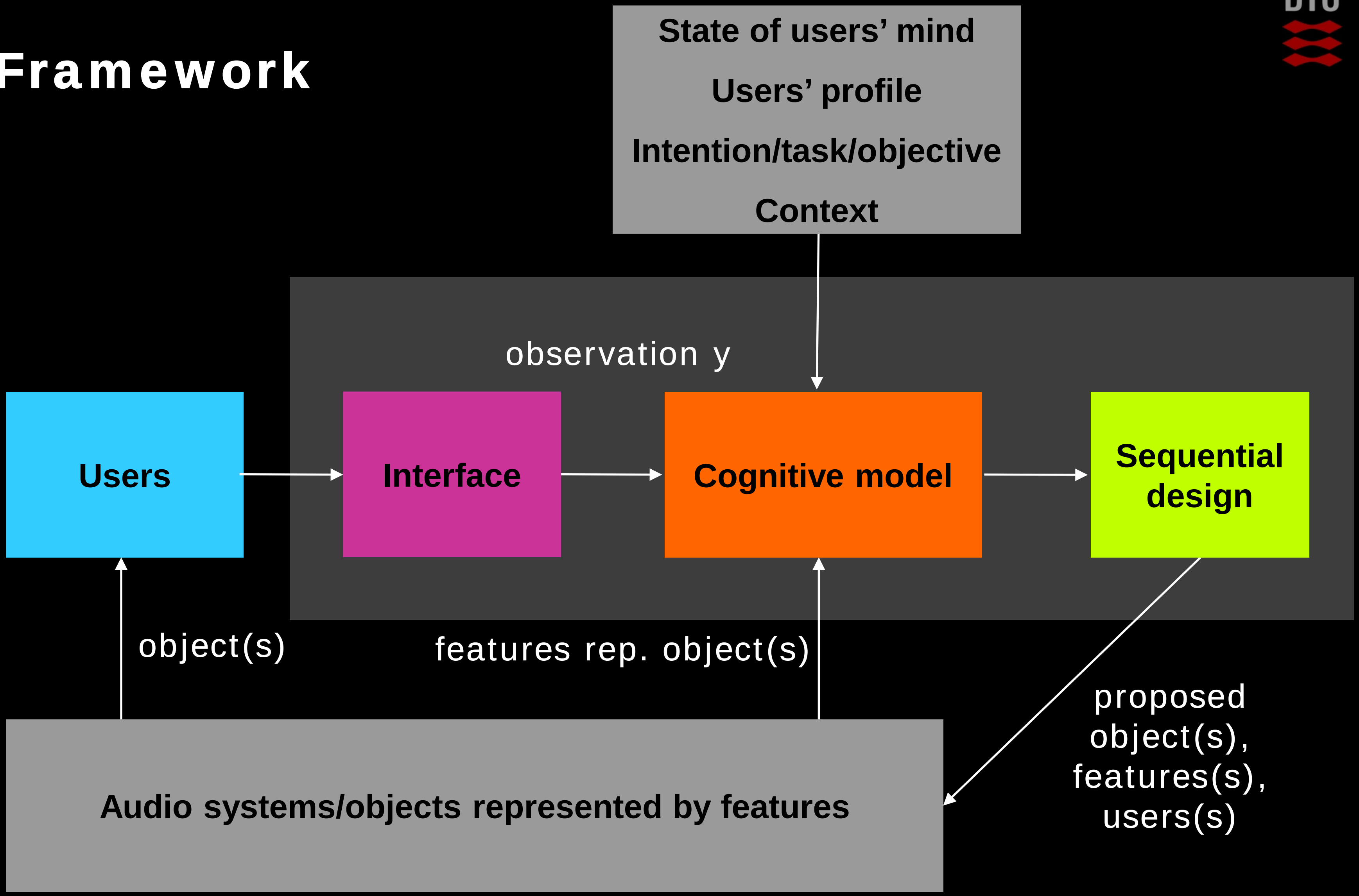
Individualization

- Personalization of audio systems

Optimization

- Predict the best *unknown* audio system from a set of evaluated audio systems
- Identify best tuning of a single audio system

Framework

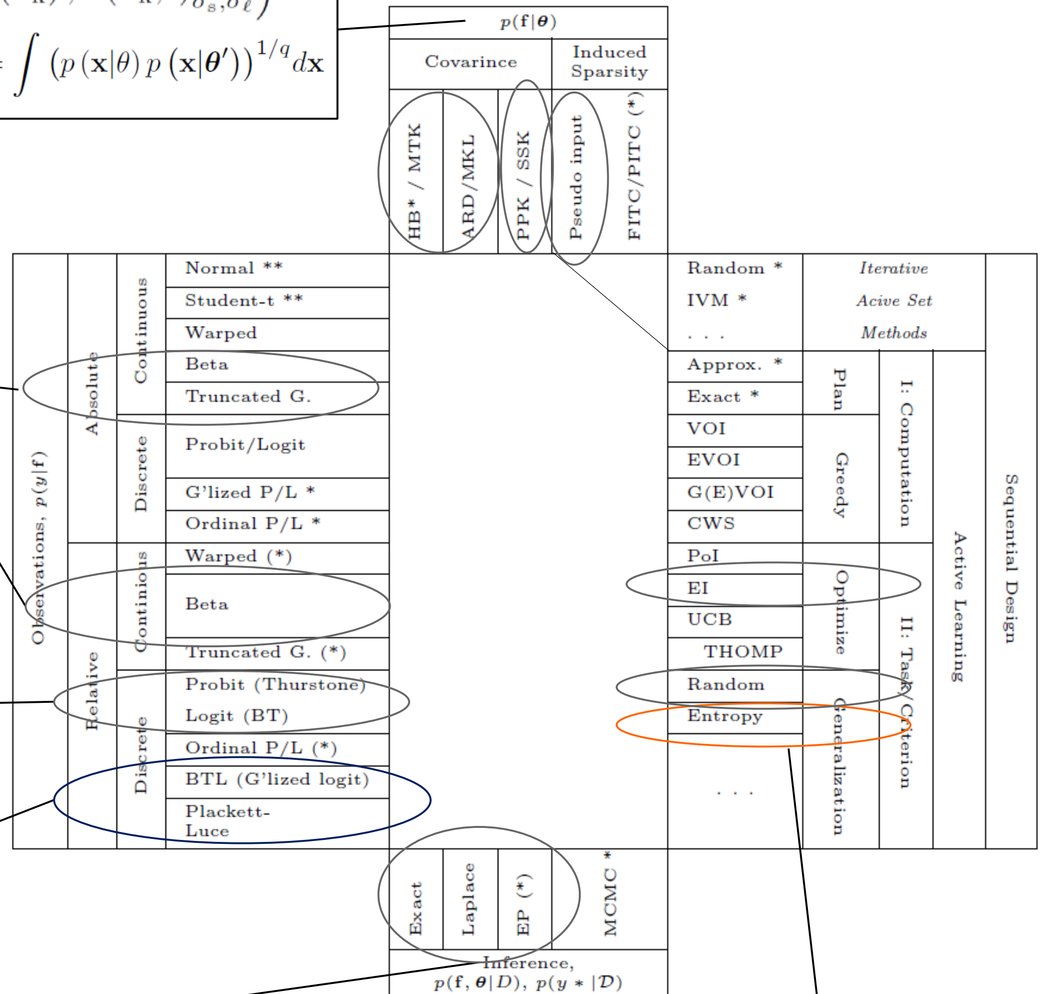


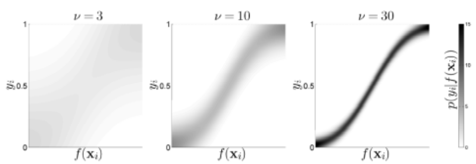
State of users' mind
Users' profile
Intention/task/objective
Context
observation y
Users
Interface
Cognitive model
Sequential design
object(s)
features rep. object(s)
proposed object(s), features(s), users(s)
Audio systems/objects represented by features

Framework

$$\mathbf{f}_k|\sigma_s,\sigma_\ell \sim \mathcal{GP}\left(\mathrm{m}\left(\mathbf{x}_\mathrm{k}\right),\mathrm{k}\left(\mathbf{x}_\mathrm{k},\cdot\right)_{\sigma_s,\sigma_\ell}\right)$$

$$\mathrm{k}\left(p\left(\mathbf{x}|\boldsymbol{\theta}\right),p\left(\mathbf{x}|\boldsymbol{\theta}'\right)\right)=\int\left(p\left(\mathbf{x}|\theta\right)p\left(\mathbf{x}|\boldsymbol{\theta}'\right)\right)^{1/q}d\mathbf{x}$$





$$\beta(\mathbf{f}_k)=\nu(1-\mu(\mathbf{f}_k)),\ \alpha(\mathbf{f}_k)=\nu\mu(\mathbf{f}_k)$$

$$\mu\left(\mathbf{f}_k,\sigma\right)=\Phi\left(\frac{f(v_k)-f(u_k)}{\sqrt{2}\sigma}\right)$$

$$y_k\sim\mathrm{Beta}\left(\alpha(\mathbf{f}_k),\beta(\mathbf{f}_k)\right)$$

$$p\left(y_k|\mathbf{f}_k,\sigma\right)=\Phi\left(y_k\frac{f\left(\mathbf{x}_{u_k}\right)-f\left(\mathbf{x}_{v_k}\right)}{\sqrt{2}\sigma_\mathcal{L}}\right)$$

$$p\left(\mathbf{y}_k|\mathbf{f}_k\right)=\prod_{j=1}^{C-1}\frac{e^{f\left(\mathbf{x}_{\mathbf{y}_{k(j)}}\right)}}{\sum_{i=j}^{C}e^{f\left(\mathbf{x}_{\mathbf{y}_{k(i)}}\right)}}$$

I Approximate first level posterior, $p(\mathbf{f}|\boldsymbol{\theta},\mathcal{X},\mathcal{Y})$ using Laplace or EP with $\boldsymbol{\theta}$ fixed.

II Find ML/MAP-II point-estimates of the hyperparemetrs $\hat{\boldsymbol{\theta}}$ based on marginal likelihood approximation, provided by the first level approximation.

... iterate until convergence of $\hat{\boldsymbol{\theta}}$ or the marginal likelihood / evidence.

$$EVOI\ \left(\mathcal{E}_k\right)\equiv\iint p\left(\mathbf{f}_k|\mathcal{E}_k,\mathcal{D}\right)p\left(y_k|\mathbf{f}_k,\mathcal{D}\right)\log p\left(y_k|\mathbf{f}_k,\mathcal{D}\right)dyd\mathbf{f}$$
$$-\int p\left(y_k|\mathcal{E}_k,\mathcal{D}\right)\log p\left(y_k|\mathcal{E}_k,\mathcal{D}\right)dy$$

Observations

Absolute

Relative

Continuous

Discrete (nominal/ordinal)

Multi vs. single-label

Multiple objects

Ranking

k-AFC

Triangle (odd out)

Noise models

user consistency

User modeling

individual approach

pooled approach

hierarchical approach based on:

  user features and/or user observations

Observations, $p(y\|f)$	Absolute	Continuous	Normal **
			Student-t **
			Warped
			Beta
			Truncated G.
		Discrete	Probit/Logit
			G'lized P/L *
			Ordinal P/L *
	Relative	Continious	Warped (*)
			Beta
			Truncated G. (*)
		Discrete	Probit (Thurstone) Logit (BT)
			Ordinal P/L (*)
			BTL (G'lized logit)
			Plackett- Luce

HB

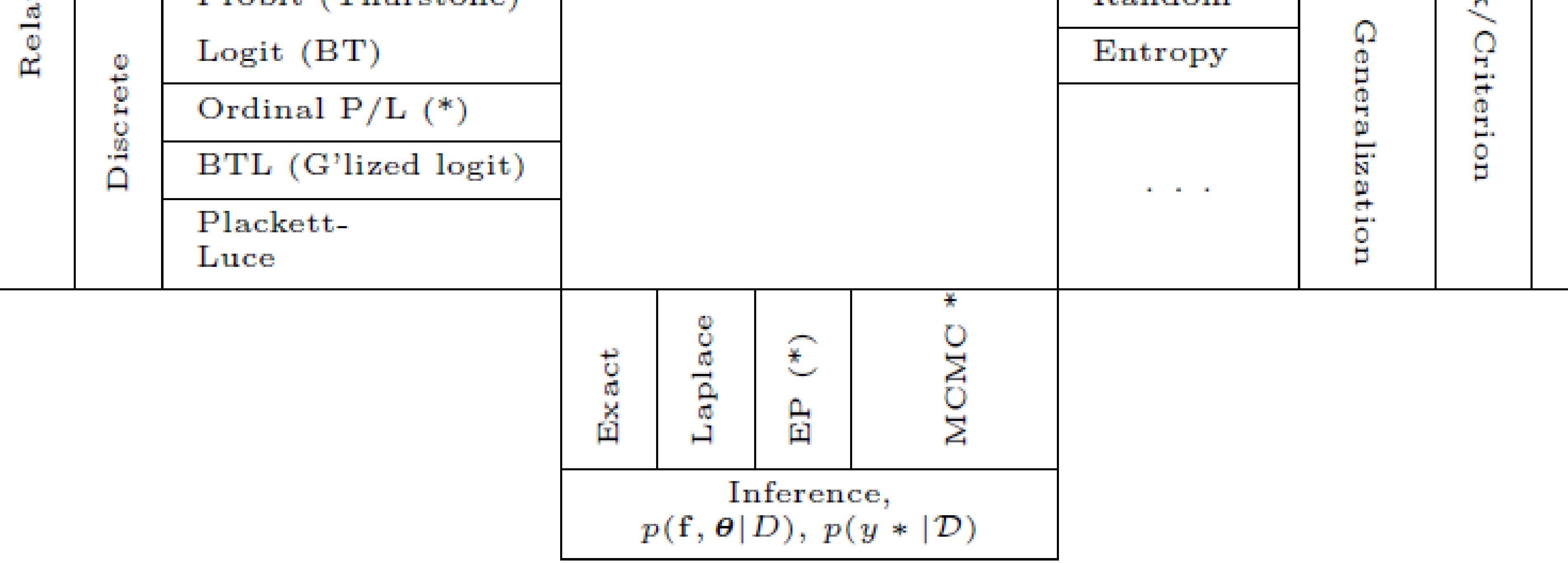
Exact

Bayesian nonlinear model

$p(\mathbf{f}\|\boldsymbol{\theta})$				
Covarince			Induced Sparsity	
HB* / MTK	ARD/MKL	PPK / SSK	Pseudo input	FITC/PITC (*)

Absolute	Continuous	Normal **
		Student-t **
		Warped
		Beta
		Truncated G.

Random *	*Iterative*	
IVM *	*Acive Set*	
. . .	*Methods*	
Approx. *	Plan	I: Co
Exact *		
VOI		



Inference (learning)

<table>
<tr><td>Random *
IVM *
. . .</td><td colspan="3">Iterative Acive Set Methods</td><td rowspan="13">Sequential Design</td></tr>
<tr><td>Approx. *</td><td rowspan="2">Plan</td><td rowspan="6">I: Computation</td><td rowspan="12">Active Learning</td></tr>
<tr><td>Exact *</td></tr>
<tr><td>VOI</td><td rowspan="4">Greedy</td></tr>
<tr><td>EVOI</td></tr>
<tr><td>G(E)VOI</td></tr>
<tr><td>CWS</td></tr>
<tr><td>PoI</td><td rowspan="4">Optimize</td><td rowspan="7">II: Task/Criterion</td></tr>
<tr><td>EI</td></tr>
<tr><td>UCB</td></tr>
<tr><td>THOMP</td></tr>
<tr><td>Random</td><td rowspan="3">Generalization</td></tr>
<tr><td>Entropy</td></tr>
<tr><td>. . .</td></tr>
</table>

Sequential design of objects, users or inputs

Fixed design:

m observations

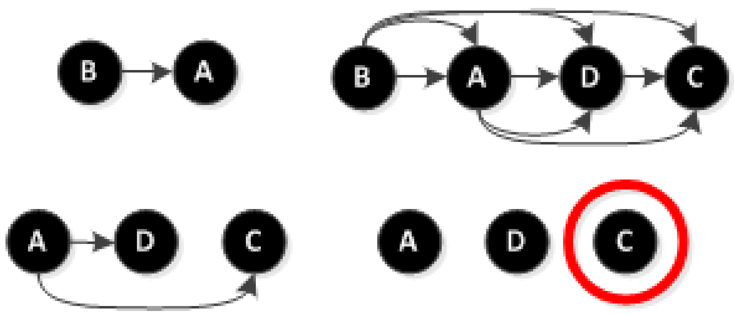
Sequential design:

αm observations

Indirect or relative scaling

- Task is comparing a set of objects and rank them in order or assign a value to the similarity between them.
- Elicitation by relative comparisons eliminates the need for absolute references and explanation - less why questions!
- Difficult to articulate experience/opinion
- Issues related to learning from limited number of objects

2AFC (Pairwise), k-AFC, ranking, odd-one out.



Similarity / Continuous (degree of preference/ confidence)

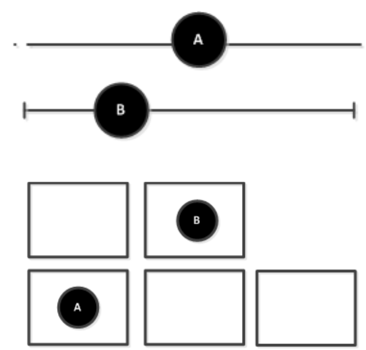
Direct or absolute scaling

- Elicitation of a specific aspect
- Learning from few objects might by complex due to perceptual and cognitive processes
- Difficult to understand/explain scale
- Difficult to consistently rate on direct scales (dimensional or categorical)
    - communication biases due to uncertainties in scales, anchors or labels
    - lack of references causes drift and inconsistencies

Infinite, ordinal, bounded, continuous scale

Categorical (classification):

Binary / multi-class



Pairwise comparison versus direct scaling

- Thurstones ”Principle of comparative judgments”
  - ”The discriminal process” – the total process of discriminating stimuli
  - Assumptions
    1. preference (utility function, or in Thurstone's terminology, *discriminal process*) for each stimulus
    2. The stimulus whose value is larger at the moment of the comparison will be preferred by the subject
    3. These unobserved preferences are normally distributed in the population
- The “phsycological scale is at best an artificial construct” (Thurstone)
- Lockhead claims that everything is relative......

G. R. Lockhead, “Absolute Judgments Are Relative: A Reinterpretation of Some Psychophysical Ideas.,” Review of General Psychology, vol. 8, no. 4, pp. 265–272, 2004.

L. L. Thurstone, “A law of comparative judgement.,” Psychological Review, vol. 34, 1927.

A. Maydeu-Olivares: ”On Thutstone’s Model For Paired Comparisons and Ranking Data”, Barcelona Univ.

Multiple aspects of users can be included

Content perception/affection

State of mind

Context

Memory/knowledge

Objective/task/intention

Modeling cognitive aspects

Is it possible to model the users representation of expressed emotion using pairwise comparisons?

Which scaling method should we use?

Is it possible to design a personalized audio system from user's preference of audio clips?

Is it possible to model, interpret and predict individual music preference based on low-level audio features and pairwise comparisons?

Expressed emotions

- Jens Madsen, Bjørn Sand Jensen, Jan Larsen and Jens Brehm Nielsen. *Towards Predicting Expressed Emotion in Music from Pairwise Comparisons*, 9th Sound and Music Computing Conference, 2012.
- Jens Madsen, Jens Brehm Nielsen, Bjørn Sand Jensen and Jan Larsen. *Modeling Expressed Emotions in Music using Pairwise Comparisons*. 9th International Symposium on Computer Music Modeling and Retrieval (CMMR) 2012.
- Madsen, J., Jensen, B.S., Larsen, J., Predictive modeling of expressed emotions in music using pairwise comparisons. M. Aramaki et al. (Eds.): CMMR 2012, LNCS 7900, pp. 253–277, 2013. Springer-Verlag Berlin Heidelberg 2013.

Is it possible to model the users representation of expressed emotion using pairwise comparisons?

Which scaling method should we use?

Internet revolutionizing the music industry

MPEG Layer 1-3 (1993-1995)

Winamp (1997)

IRC (1988), Hotline, and Usenet (1+ million users, 2003)

Napster (1999) (80+ million users)

P2P services (1999 – 2014+)

Spotify (2006) 25 million songs, (40+ million users)

iTunes (2001,2008) 37 million songs (575+ million users)

Deezer (2007), 35 million songs (5,16+ million users)

WiMP (2010),....

Navigating and finding new music

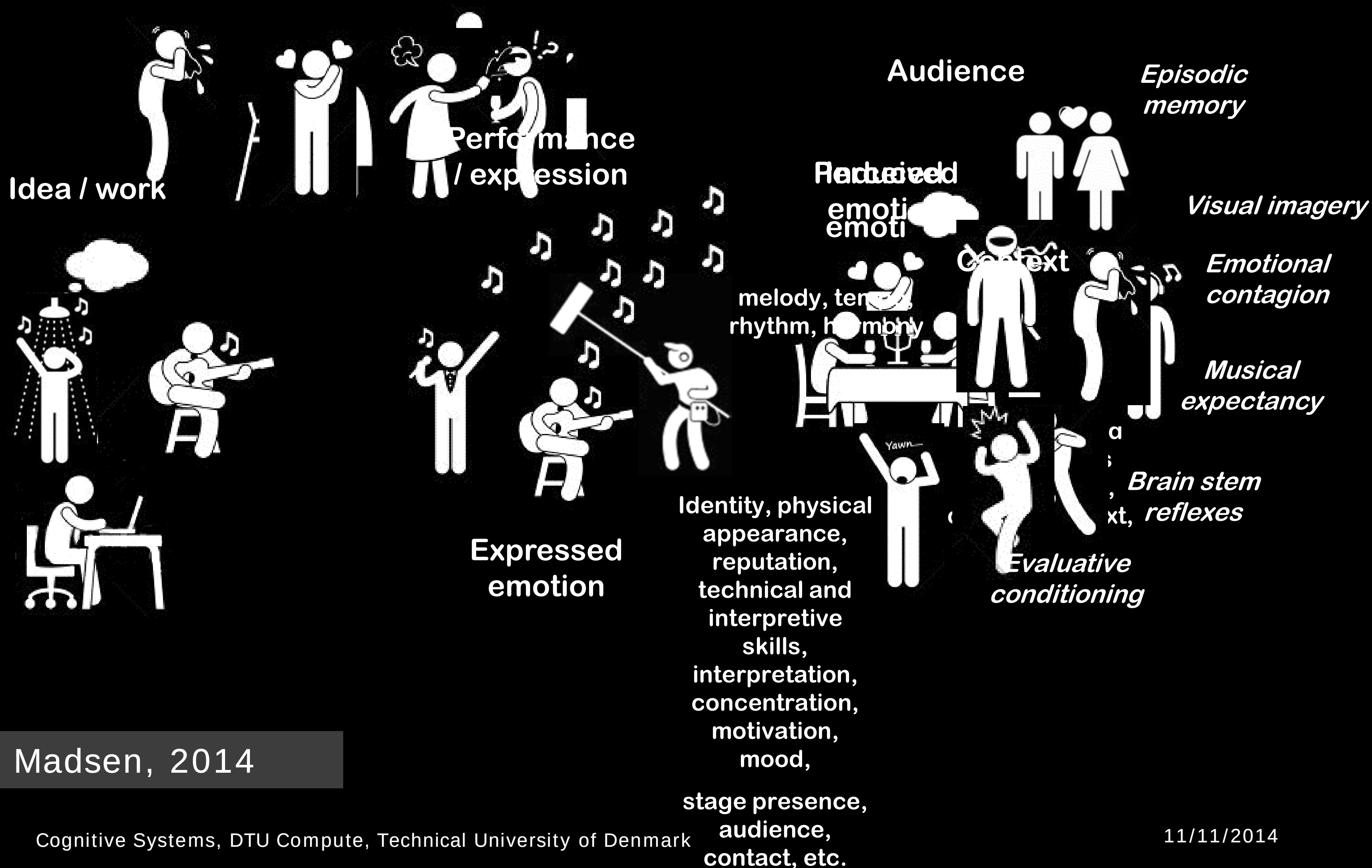
- How do we navigate in music archives? (navigation)
  - Search by artist name, genre, similar artist, etc.
  - Own listening history
  - Friends listening history

- How do we find new music? (recommendation)
  - Passive: Radio stations,
  - Semi-active: playlists, Last.fm, 8tracks, stereomo
  - Active: Pandora,

Using emotions to navigate in music archives

- Give me some happy music!
- Find me some sad jazz from the 1960 with trumpet!

Musical experience



Jens Madsen, 2014

What can we model?

- Induced emotion, can we model what makes us happy?

- We model the expressed/perceived emotion in music!

User profile

musical experience

familiarity

current motivation

mood

learned associations
conditioning

cultural context

nationality

Influences of
induced emotions

Episodic memory

Brain stem reflexes

Visual imagery

Evaluative conditioning

Emotional contagion

Musical expectancy

Mechanisms

- **Brain stem reflexes** linked to acoustiscal properties, e.g. loudness
- **Evaluative conditioning** – association between music and emotion when they occur together
- **Emotional contagion** – emotion expressed in music, sad is linked low-pitches, slow, and low
- **Visual images** – creation of visual images
- **Episodic memories** – e.g. strong emotion when you hear a melody linked to an episode
- **Cognitive appraisal** - mental analysis of music an creation of aesthetic pleasure (hit-songs)
- **Musical expectancy** - balance between surprise and expectation

Ref: Music interventions in Health Care, Line Gebauer & Peter Vuust, Danish Sound, 2014

Patrik N. Juslin and Daniel Västfjäll: Emotional responses to music: The need to consider underlying mechanisms, Behavaioral and Brain Sciences, vol. 31, pp. 559–621, 2008

Modelling expressed emotion in music

- **Too many tracks to annotate!**
  - 26 mio tracks = 148 years playtime

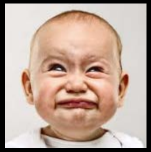
- **Automatic music emotion prediction**
  - Method of quantifying and representing the emotions expressed in music. (experimental paradigm, model of emotions, etc.)
  - How to represent the audio (feature extraction, representation)
  - Methods to predict annotations, evaluations, rankings, ratings etc. (machine learning)

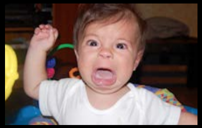
Music Emotion Modeling - overview

User modeling/

experimental paradigm

Annotations





Machine learning

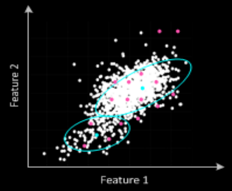
Model

predictions

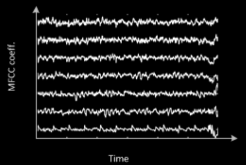
Audio signal processing/

Machine learning

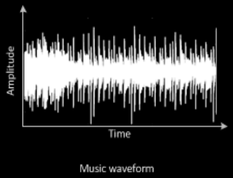
Feature representation



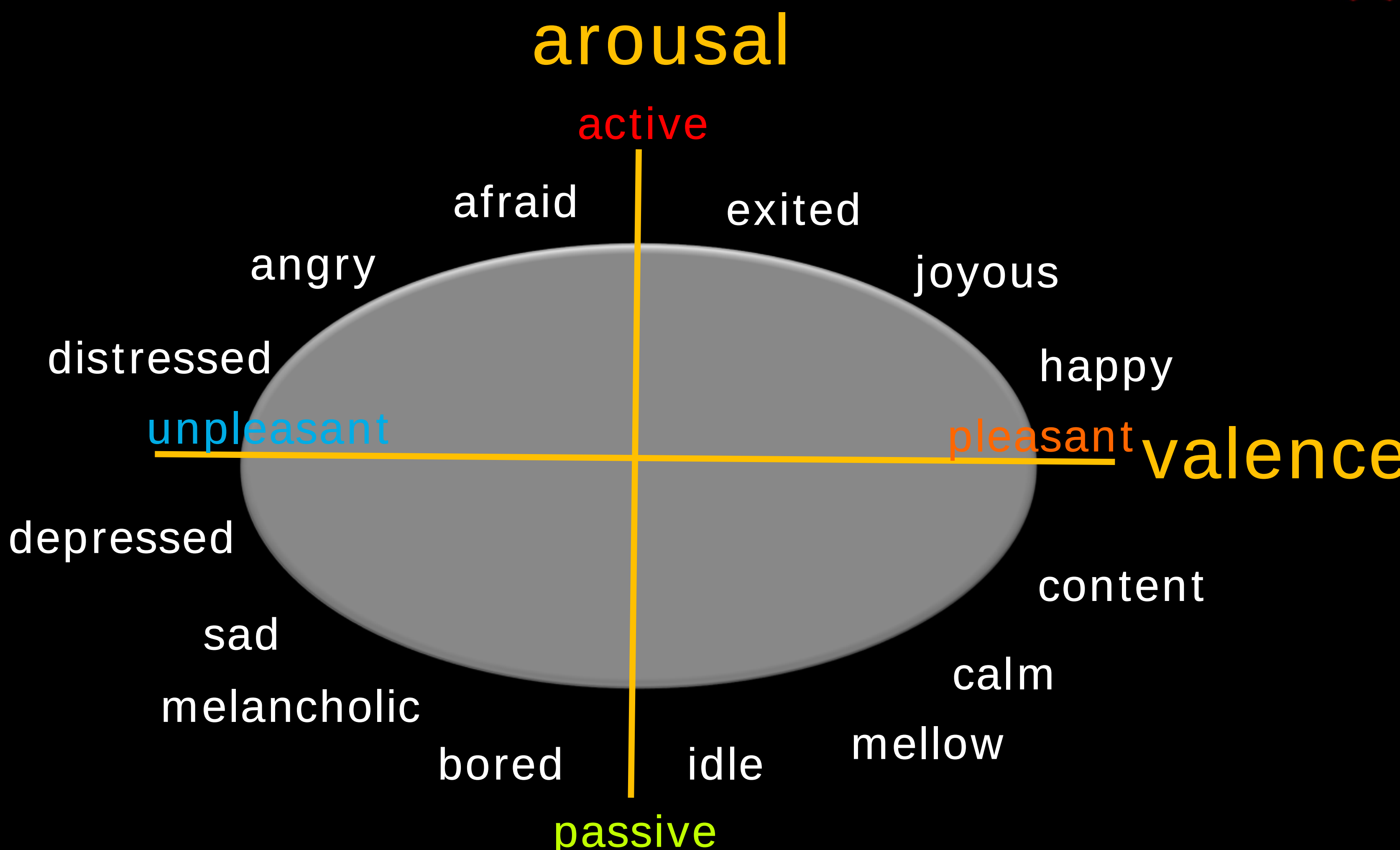
Audio Feature extraction



Music archive



Emotional spaces

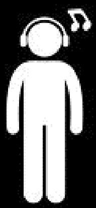


J. A. Russel: "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, 39(6):1161, 1980

J. A. Russel, M. Lewicka, and T. Niit, "A Cross-Cultural Study of a Circumplex Model of Affect," *Journal of Personality and Social Psychology*, vol. 57, pp. 848-856, 1989

Using relative measures of emotion elicitation

- **Arousal:** Which sound clip was the most exciting, active, awake?
- **Valence:** Which sound clip was the most positive, glad, happy?



Excerpt A

Excerpt B

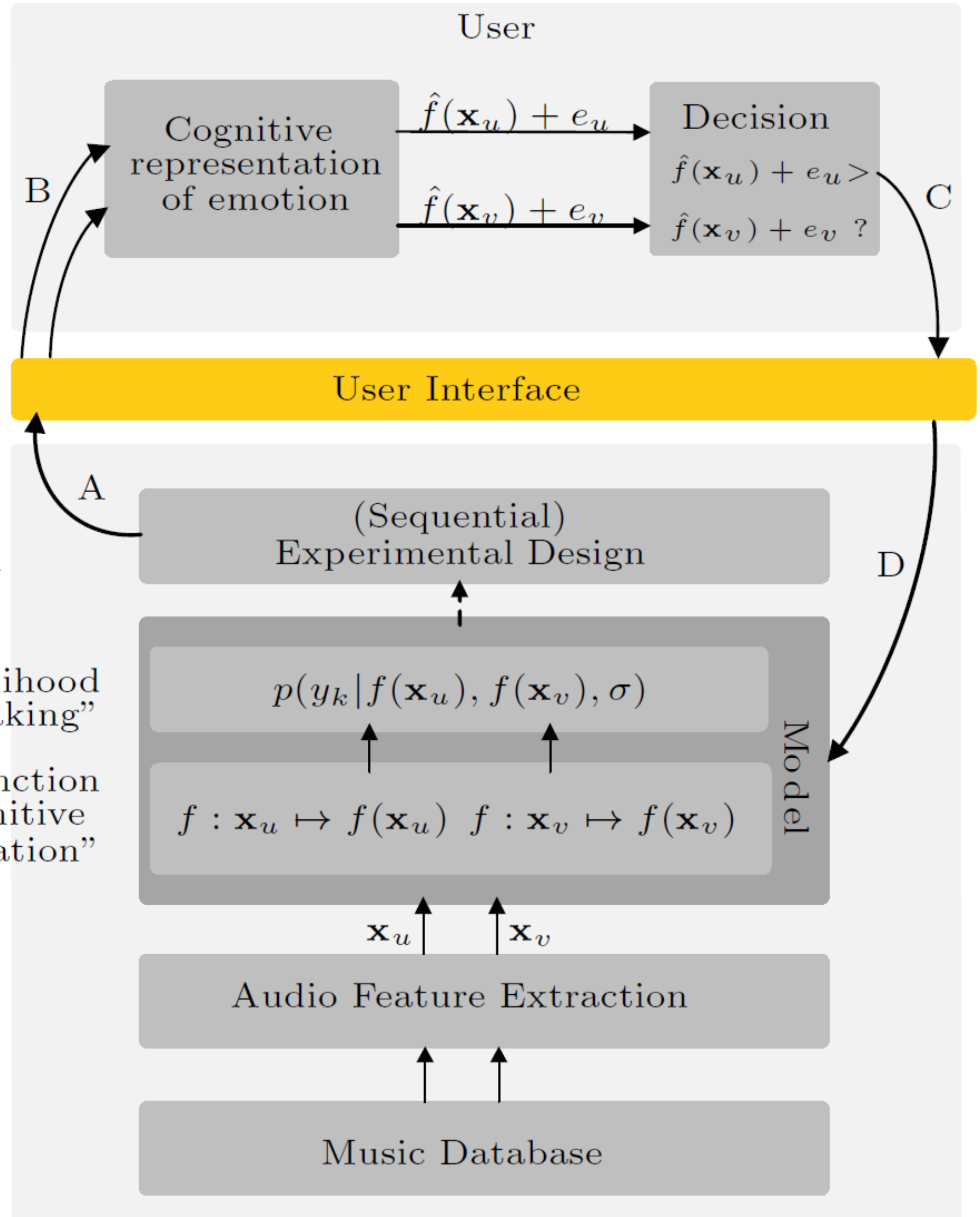
$$\mathcal{X} = \left\{ \left(\mathbf{x}_{u_m}, \mathbf{x}_{v_m} \right) | m = 1{:}M \ \wedge \ \mathbf{x}_u \mathbf{x}_{,v} \in \mathbb{R}^D \right\}$$

$$\mathcal{Y} = \{ y_m | m = 1{:}M \ \wedge \ y \in \{-1,1\} \}$$

$$\mathcal{D} = \left\{ \left(y_m, \mathbf{x}_{u_m}, \mathbf{x}_{v_m} \right) | m = 1{:}M \wedge y \in \{-1,1\} \wedge \ \mathbf{x}_u \mathbf{x}_{,v} \in \mathbb{R}^D \right\}$$

B: Present two different excerpts (and instructions)

C: The user selects the excerpt which is either the most 'happy' or 'excited.'



A: The experimental design selects an comparison, ε_k, with two music excerpts u_k and v_k, for presentation

Likelihood "Decision Making"

Latent Function "Model of the Cognitive Representation"

D: Update model based on the users reponse to comparison k .

Modelling and evalutation

- How many pairwise comparisons did we predict corectly?
- How do we rank excerpts on the dimensions of valence and arousal?

Model: nonlinear logistic regression using Bayes learning and Gaussian processes

$$\sigma_l, \sigma_f, \sigma_n \sim \mathcal{U}(-\infty, \infty) \;\; / Gamma(\eta, \rho)$$

$$\mathrm{k}(\mathbf{x}, \mathbf{x}')_{\sigma_l, \sigma_f} = \frac{1}{\sigma_f^2} \exp\left(-\frac{1}{2\sigma_l^2}(\mathbf{x} - \mathbf{x}')^2\right), \mathrm{m}(\mathbf{x}) = \mathbf{0}$$

$$\mathbf{f} | \mathcal{X}, \sigma_l, \sigma_f \sim \mathcal{G}\mathrm{P}\left(\mathrm{m}(\mathbf{x}), \mathrm{k}(\mathbf{x}, \cdot)_{\sigma_l, \sigma_f}\right)$$

$$\pi_m | \mathbf{f}, \mathbf{x}_{u_m}, \mathbf{x}_{v_m} = \Phi\left(\frac{f_{u_m} - f_{v_m}}{\sigma_n^2}\right) \quad \forall m = 1{:}\, M$$

$$y_m | \pi_m \sim Bernoulli(\pi_m) \quad \forall m = 1{:}\, M$$

Obervations

$$\mathcal{X} = \{x_i | i = 1, ..., n\} \quad x_i \in \mathbb{R}^d$$

binary where $y_k = d_k, d_k \in \{-1, 1\}$
continuous and bounded where $y_k = \pi_k, \pi_k \in]0, 1[$

$$\mathcal{D} = \{(y_k; u_k, v_k) | k = 1, ..., m\}$$

Likelihood in binary case

$$p\left(\mathcal{Y} = y_k | f_k\left(u_k\right), f\left(v_k\right)\right)$$

$$p(y_k|\mathbf{f}_k) \qquad \mathbf{f}_k = \left[f\left(u_k\right), f\left(v_k\right)\right]^{\top}$$

$$\mathcal{L}_{bin} \equiv p\left(d_k|\mathbf{f}_k\right) = \Phi\left(d_k \frac{f\left(v_k\right) - f\left(u_k\right)}{\sqrt{2}\sigma}\right)$$

$\Phi(x)$ is the cumulative Gaussian

$d_k, d_k \in \{-1, 1\}$

GP preference function prior

Posterior we want to infer

Likelihood

GP function prior

$$p\left(\mathbf{f}|\mathcal{D}\right) = \frac{p\left(\mathcal{D}|\mathbf{f}\right)p(\mathbf{f})}{p\left(\mathcal{D}\right)}$$

No analytical form, hence, approximate inferece. We use Laplace approximation

Predicting preference

$$p\left(y_t|\mathcal{D}\right) = \int p\left(y_t|\mathbf{f}_t, \mathcal{D}\right) p\left(\mathbf{f}_t|\mathcal{D}\right) d\mathbf{f}_t$$

Non-Gaussian shape but for Probit likelihood analytical expression

Gaussian when using Laplace approximation

Formal definition of a GP

A function $f(\boldsymbol{x})$ can be sought of as an *infinitely* long vector

A Gaussian process is a collection of random variables where every finite number has a Gaussian distribution

N function values

$$\boldsymbol{f} = [f_1, \cdots f_N] \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

GP

$$f(\boldsymbol{x}) \sim \mathcal{GP}\left(m(\boldsymbol{x}), \boldsymbol{\Sigma}(\boldsymbol{x}, \boldsymbol{x}')\right)$$

How do we handle infinitely long vectors?

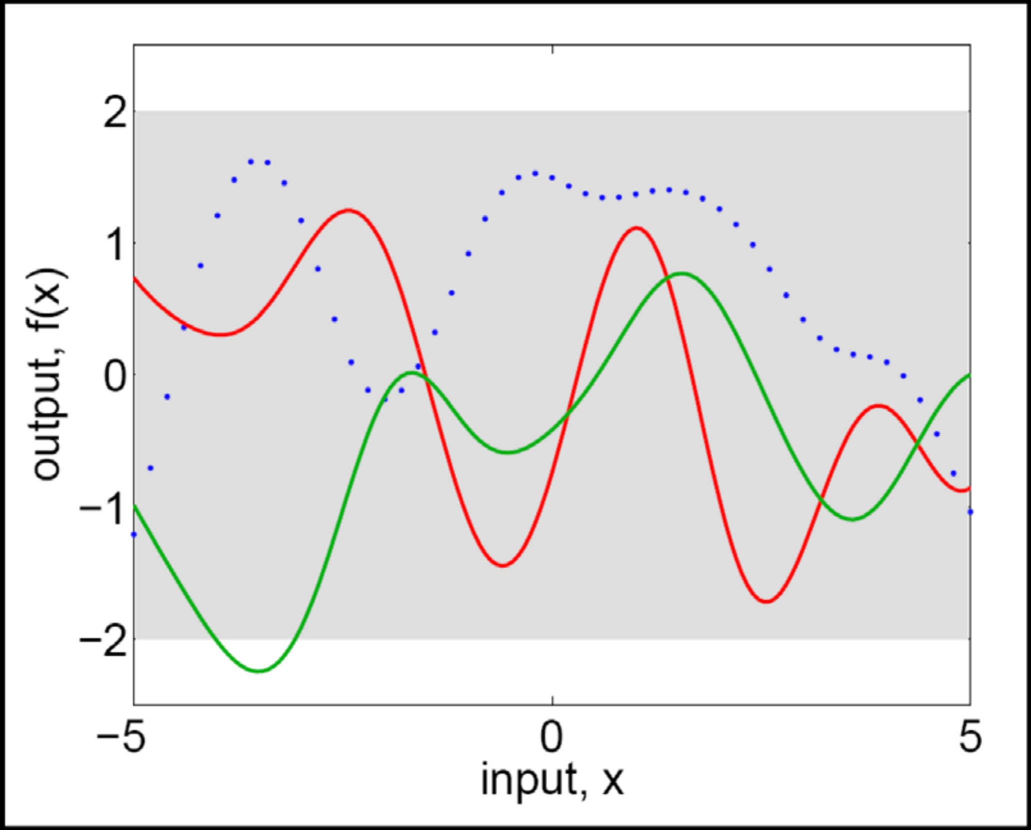
Marginalization property

Any finite sample has a fixed distribution

$$[\boldsymbol{f}_1, \boldsymbol{f}_2] \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right)$$

$$p(\boldsymbol{f}_1) = \int p(\boldsymbol{f}_1, \boldsymbol{f}_2)\, d\boldsymbol{f}_2 = \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{11}}\right)$$

Example of GP function priors



squared exponential covariance function $\exp(-|\boldsymbol{x} - \boldsymbol{x}'|^2/2)$

GP regression

Model

$$y = f(\boldsymbol{x}) + \epsilon, \quad f(\boldsymbol{x}) \sim \mathcal{GP}(\mathbf{0}, k(\boldsymbol{x}, \boldsymbol{x}')), \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

Data set, $\mathcal{D}$

$$\boldsymbol{X} = [(\boldsymbol{x}^\top(1); \cdots; \boldsymbol{x}^\top(N)], \quad N \times d \text{ matrix}$$

$$\boldsymbol{y} = [y(1), \cdots, y(N)]^\top, \quad N \times 1 \text{ column vector}$$

Predictive distribution

$$p(y^*|\boldsymbol{x}^*, \mathcal{D}) = \int p(y^*|\boldsymbol{x}^*, f)p(f|\mathcal{D})\, df$$

GP regression

$$\begin{bmatrix} \boldsymbol{y} \\ y^* \end{bmatrix} = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{K} & \boldsymbol{k} \\ \boldsymbol{k} & k \end{bmatrix}\right)$$

$$\begin{aligned} \boldsymbol{K} &= \{k(\boldsymbol{x}(i), \boldsymbol{x}(j))\} \\ \boldsymbol{k} &= \{k(\boldsymbol{x}(i), \boldsymbol{x}^*)\} \\ k &= k(\boldsymbol{x}^*, \boldsymbol{x}^*) \end{aligned}$$

Conditional Gaussian

$$p(y^*|\boldsymbol{y}) = \mathcal{N}\left(\boldsymbol{k}^\top(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{y}\,,\ \sigma^2 + k - \boldsymbol{k}^\top(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{k}\right)$$

Active learning by value of information VOI

$$S\left(\mathbf{f}_*|\varepsilon_*, \mathcal{E}_a, \mathcal{Y}_a, \boldsymbol{\theta}\right) = \frac{1}{2}\log\left(\left(2\cdot\pi\cdot e\right)^D|\mathbf{K}^*|\right)$$

$$\arg\max_{\varepsilon_*\in\mathcal{E}_c} S\left(\mathbf{f}_*|\varepsilon_*, \mathcal{E}_a, \mathcal{Y}_a, \boldsymbol{\theta}\right)$$

E. Bonilla, S. Guo, and S. Sanner, "Gaussian Process preference elicitation," in Advances in Neural Information Processing Systems 23, J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, Eds., pp. 262–270. 2010.

Active learning by expected value of information EVOI

$$\Delta S(\mathbf{f}) = S(\mathbf{f}|y_*, \varepsilon_*, \mathcal{X}_a, \mathcal{Y}_a, \boldsymbol{\theta}) - S(\mathbf{f}|\mathcal{X}_a, \mathcal{Y}_a, \boldsymbol{\theta})$$

$$\mathrm{EVOI}(\varepsilon_*) = \sum_{y\in\{-1,1\}} p(y_*|\varepsilon_*, \mathcal{X}_a, \mathcal{Y}_a, \boldsymbol{\theta})\, \Delta S(\mathbf{f}|y_*, \varepsilon_*, \mathcal{X}_a, \mathcal{Y}_a, \boldsymbol{\theta}) \tag{4}$$

$$= \sum_{y\in\{-1,1\}} \int p(y_*|\mathbf{f}_*, \mathcal{X}_a, \mathcal{Y}_a, \boldsymbol{\theta})\, p(\mathbf{f}_*|\varepsilon_*, \mathcal{X}_a, \mathcal{Y}_a, \boldsymbol{\theta}) \log p(y_*|\mathbf{f}_*, \mathcal{X}_a, \mathcal{Y}_a, \boldsymbol{\theta})\, d\mathbf{f}_*$$

$$- \sum_{y\in\{-1,1\}} p(y_*|\varepsilon_*, \mathcal{X}_a, \mathcal{Y}_a, \boldsymbol{\theta}) \log p(y_*|\varepsilon_*, \mathcal{X}_a, \mathcal{Y}_a, \boldsymbol{\theta})$$

$$\arg\max_{\varepsilon_*\in\mathcal{E}_c} \mathrm{EVOI}(\varepsilon_*)$$

Houlsby, N., Hernandez-Lobato, J.M., Huszar, F., Ghahramani, Z.: Collaborative Gaussian processes for preference learning. In: Bartlett, P., Pereira, F., Burges, C., Bottou, L., Weinberger, K. (eds.) Advances in Neural Information Processing Systems, vol. 25, pp. 2105–2113 (2012)

J. Madsen, B. S. Jensen, J. Larsen, Predictive Modeling of Expressed Emotions in Music using Pairwise Comparisons, CMMR 2012 Post-Proceedings, vol. 7900, pp. 253-277, Springer-Verlag Berlin Heidelberg, 2013

Experimental setup

IMM dataset

- **20 excerpts of 15 second** length were chosen to be evenly distributed in the AV space using a linear regression model and subjective evaluation.
- **13 participants** each evaluated all **190 unique pairwise comparisons.**

YANG dataset

- **1240 excerpts of 30 second** length evaluated on the dimension of valence
- Multiple participants evalulate 7952 pairwise comparisons

Audio representation

Echonest features

YAAFE (Yet-Another-Audio-Feature-Extraction) Toolbox

MA toolbox (Pampalk)

MIR toolbox

CM toolbox

Features

Feature	Description	Dimension(s)
Mel-frequency cepstral coefficients (MFCCs)[1]	The discrete cosine transform of the log-transformed short-time power spectrum on the logarithmic mel-scale.	20
Envelope (En)	Statistics computed on the distribution of the extracted temporal envelope.	7
Chromagram CENS, CRP [23]	The short-time energy spectrum is computed and summed appropriately to form each pitch class. Furthermore statistical derivatives are computed to discard timbre-related information.	12 12 12
Sonogram (Sono)	Short-time spectrum filtered using an outer-ear model and scaled using the critical-band rate scale. An inner-ear model is applied to compute cochlea spectral masking.	23
Pulse clarity [16]	Ease of the perception by listeners of the underlying rhythmic or metrical pulsation in music.	7
Loudness [22]	Loudness is the energy in each critical band.	24
Spectral descriptors (sd) [22] (sd2) [17]	Short-time spectrum is described by statistical measures e.g., flux, roll-off, slope, variation, etc.	9 15

Pulse clarity [16]	Ease of the perception by listeners of the underlying rhythmic or metrical pulsation in music.	7
Loudness [22]	Loudness is the energy in each critical band.	24
Spectral descriptors (sd) [22] (sd2) [17]	Short-time spectrum is described by statistical measures e.g., flux, roll-off, slope, variation, etc.	9 15
Mode, key, key strength [17]	Major vs. Minor, tonal centroid and tonal clarity.	10
Tempo [17]	The tempo is estimated by detecting periodicities on the onset detection curve.	2
Fluctuation Pattern [17]	Models the perceived fluctuation of amplitude-modulated tones.	15
Pitch [23]	Audio signal decomposed into 88 frequency bands with center frequencies corresponding to the pitches A0 to C8 using an elliptic multirate filterbank.	88
Roughness [17]	Roughness or dissonance, averaging the dissonance between all possible pairs of peaks in the spectrum.	2
Spectral Crest factor [22]	Spectral crest factor per log-spaced band of 1/4 octave.	23
Echonest *Timbre*	Proprietary features to describe timbre.	12
Echonest *Pitch* [17]	Proprietary chroma-like features.	12

Performance predicting arousal using different audio features

Training size	5%	7%	10%	20%	40%	60%	80%	100%
MFCC	0.3402	0.2860	0.2455	0.2243	0.2092	0.2030	0.1990	0.1949
Envelope	0.4110*	0.4032	0.3911	0.3745	0.3183	0.2847	0.2780	0.2761
Chroma	0.3598	0.3460	0.3227	0.2832	0.2510	0.2403	0.2360	0.2346
CENS	0.3942	0.3735	0.3422	0.2994	0.2760	0.2676	0.2640	0.2621
CRP	0.4475	0.4336	0.4115	0.3581	0.2997	0.2790	0.2735	0.2729
Sonogram	0.3325	0.2824	0.2476	0.2244	0.2118	0.2061	0.2033	0.2026
Pulse clarity	0.4620	0.4129	0.3698	0.3281	0.2964	0.2831	0.2767*	0.2725
Loudness	0.3261	0.2708	**0.2334**	**0.2118**	**0.1996**	**0.1944**	**0.1907**	**0.1862**
Spec. disc.	0.2909	0.2684	0.2476	0.2261	0.2033	0.1948	0.1931	0.1951
Spec. disc. 2	0.3566	0.3223	0.2928	0.2593	0.2313	0.2212	0.2172	0.2138
Key	0.5078	0.4557	0.4059	0.3450	0.3073*	0.2959	0.2926	0.2953
Tempo	0.4416	0.4286	0.4159	0.3804	0.3270	0.3043	0.2953	0.2955
Fluctuations	0.4750	0.4247	0.3688	0.3117	0.2835	0.2731	0.2672	0.2644*
Pitch	0.3173	0.2950	0.2668	0.2453	0.2301	0.2254	0.2230	0.2202
Roughness	**0.2541**	**0.2444**	0.2367	0.2304	0.2236	0.2190	0.2168	0.2170
Spectral crest	0.4645	0.4165	0.3717	0.3285	0.2979	0.2866*	0.2828	0.2838
Echo. timbre	0.3726	0.3203	0.2797	0.2524	0.2366	0.2292	0.2258	0.2219
Echo. pitch	0.3776	0.3264	0.2822	0.2492	0.2249	0.2151	0.2089	0.2059
$Base_{low}$	0.4122	0.3954	0.3956	0.3517	0.3087	0.2879	0.2768	0.2702

J. Madsen, B. S. Jensen, J. Larsen, Predictive Modeling of Expressed Emotions in Music using Pairwise Comparisons, CMMR 2012 Post-Proceedings, vol. 7900, pp. 253-277, Springer-Verlag Berlin Heidelberg, 2013

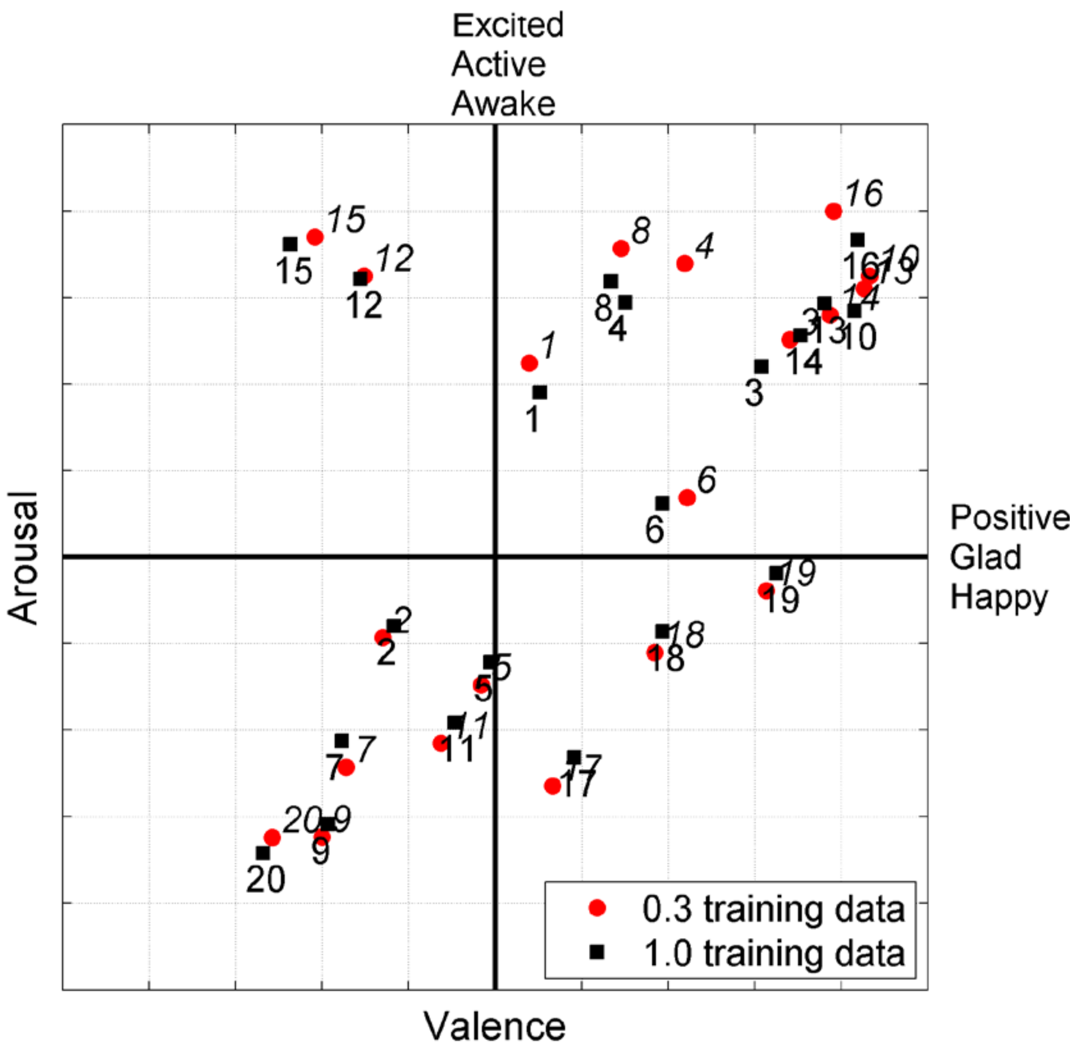
Performance predicting valence using different audio features

Training size	5%	7%	10%	20%	40%	60%	80%	100%
MFCC	0.4904	0.4354	0.3726	0.3143	0.2856	0.2770	0.2719	0.2650
Envelope	**0.3733**	**0.3545**	0.3336	0.3104	0.2920	0.2842	0.2810	0.2755
Chroma	0.4114*	0.3966*	0.3740	0.3262	0.2862	0.2748	0.2695	0.2658
CENS	0.4353	0.4139	0.3881	0.3471	0.3065	0.2948	0.2901*	0.2824
CRP	0.4466	0.4310	0.4111	0.3656	0.3066	0.2925	0.2876	0.2826
Sonogram	0.4954	0.4360	0.3749	0.3163	0.2884	0.2787	0.2747	0.2704
Pulse clarity	0.4866	0.4357	0.3856	0.3336	0.3026	0.2930	0.2879	0.2810
Loudness	0.4898	0.4310	0.3684	0.3117	0.2854	0.2768	0.2712	0.2664
Spec. disc.	0.4443	0.4151	0.3753	0.3263	0.2939	0.2857	0.2827	0.2794
Spec. disc. 2	0.4516	0.4084	0.3668	0.3209	0.2916	0.2830	0.2781	0.2751
Key	0.5303	0.4752	0.4104	0.3370	0.2998	0.2918	0.2879	0.2830*
Tempo	0.4440	0.4244	0.3956	0.3559*	0.3158	0.2985	0.2933	0.2883
Fluctuations	0.4015	0.3584	**0.3141**	**0.2730**	**0.2507**	**0.2433**	**0.2386**	**0.2340**
Pitch	0.4022	0.3844	0.3602	0.3204	0.2926	0.2831	0.2786	0.2737
Roughness	0.4078	0.3974	0.3783	0.3313	0.2832	0.2695	0.2660	0.2605
Spec. crest	0.4829	0.4289	0.3764	0.3227	0.2994	0.2942	0.2933	0.2923
Echo. timbre	0.4859	0.4297	0.3692	0.3127	0.2859	0.2767	0.2732	0.2672
Echo. pitch	0.5244	0.4643	0.3991*	0.3275	0.2942	0.2841	0.2790	0.2743
$Base_{low}$	0.4096	0.3951	0.3987	0.3552	0.3184	0.2969	0.2893	0.2850

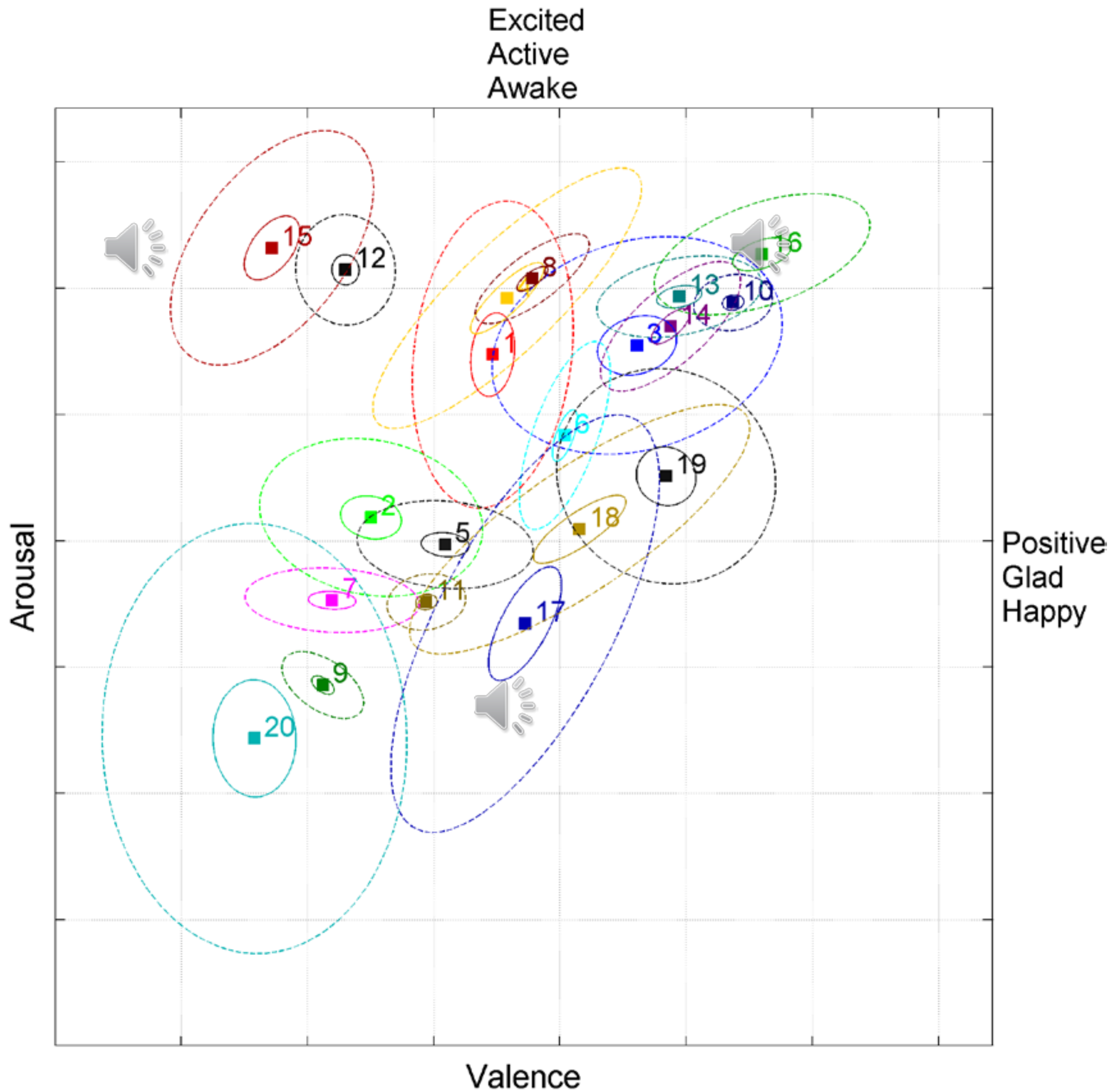
Madsen, J., Nielsen, J.B., Jensen, B.S., Larsen, J.: Modeling expressed emotions in music using pairwise comparisons. In: 9th International Symposium on Computer Music Modeling and Retrieval (CMMR) Music and Emotions. (June 2012)

Vizualization in AV-space

- No. Song name
- 1 311 - T and p combo
- 2 A-Ha - Living a boys adventure
- 3 Abba – That's me
- 4 ACDC - What do you do for money hone
- 5 Aaliyah - The one I gave my heart to
- 6 Aerosmith - Mother popcorn
- 7 Alanis Morissette - These r the thoughts
- 8 Alice Cooper – I'm your gun
- 9 Alice in Chains - Killer is me
- 10 Aretha Franklin - A change
- 11 Moby – Everloving
- 12 Rammstein - Feuer frei
- 13 Santana - Maria caracoles
- 14 Stevie Wonder - Another star
- 15 Tool - Hooker with a pen..
- 16 Toto - We made it
- 17 Tricky - Your name
- 18 U2 - Babyface
- 19 UB40 - Version girl
- 20 ZZ top - Hot blue and righteous



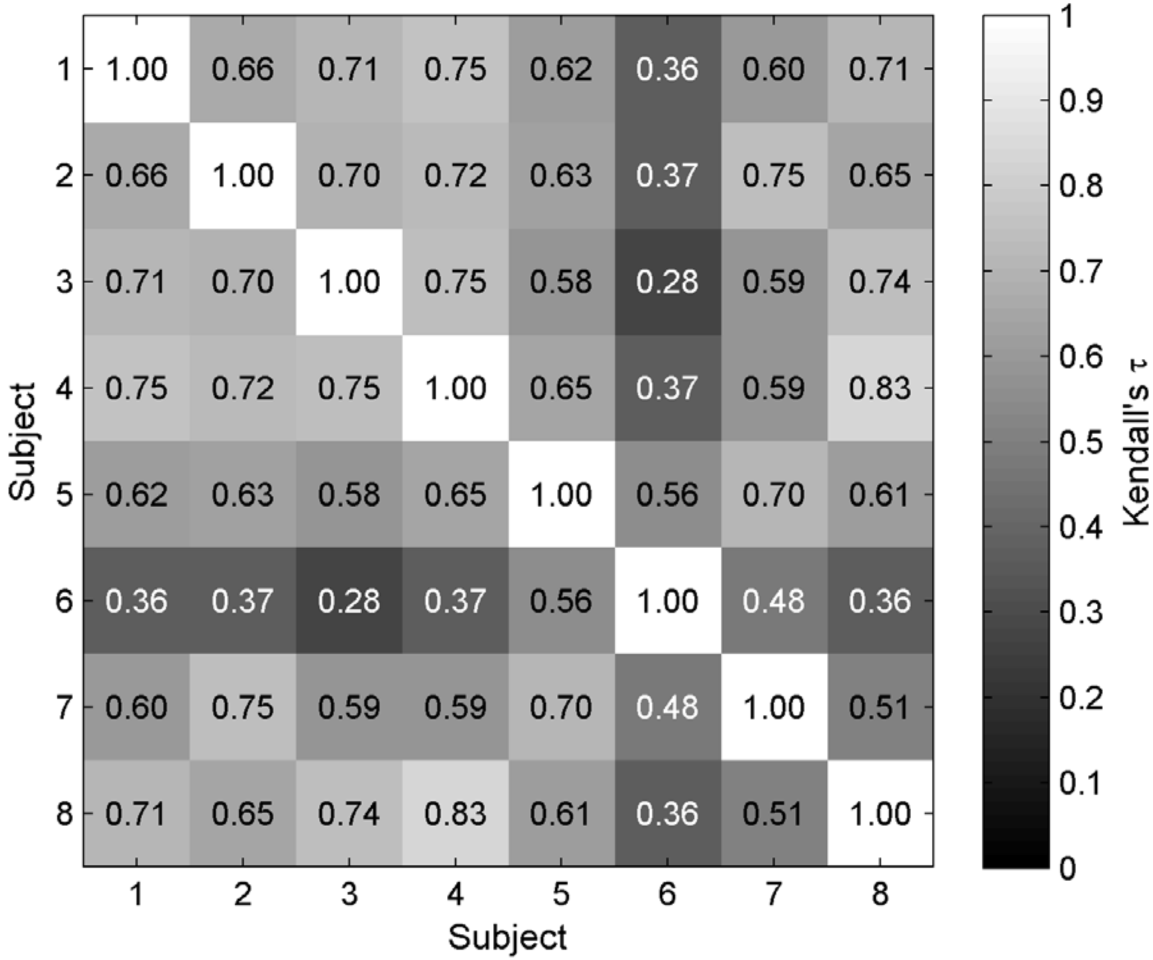
Is ranking of music subject dependent?



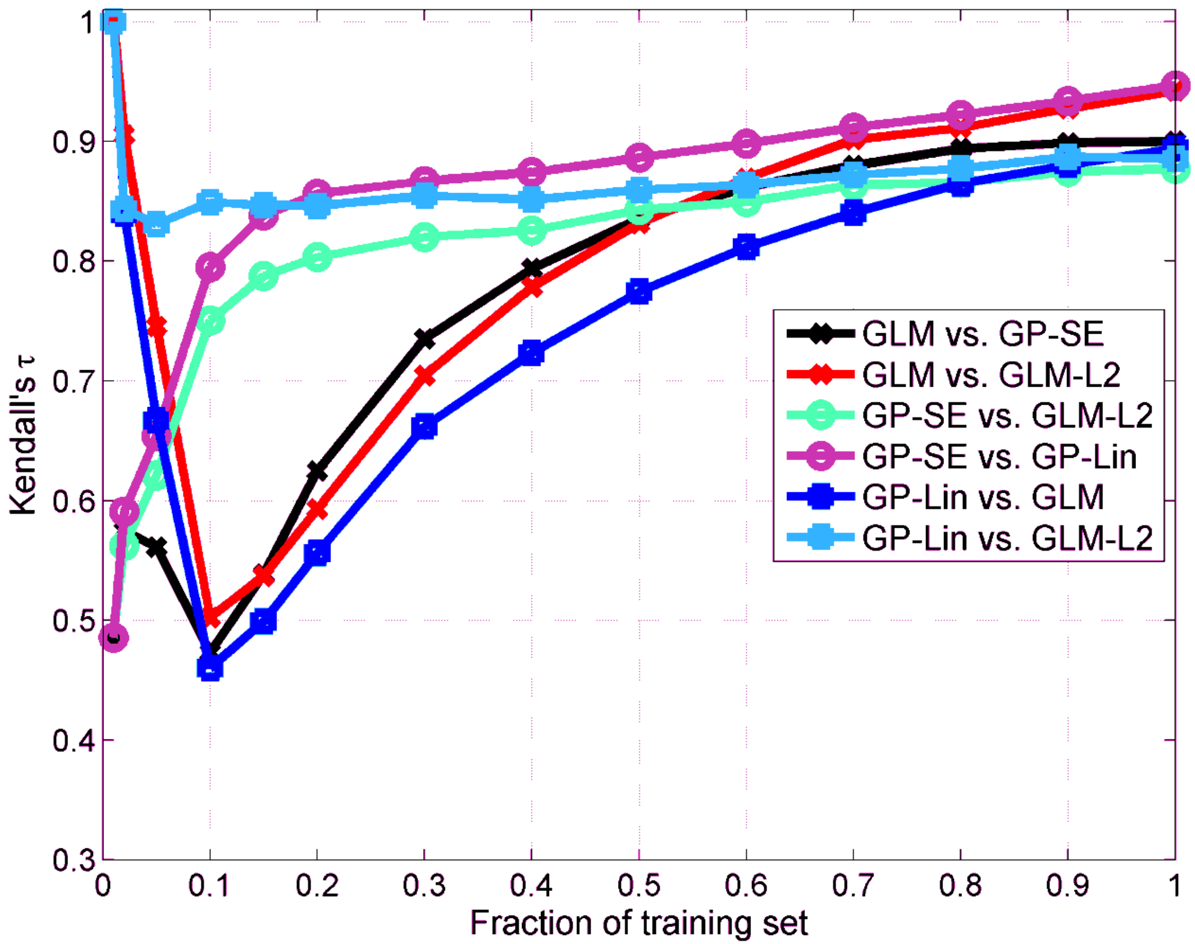
Valence / Arousal Space for GP model

Madsen, J., Jensen, B.S., Larsen, J., Nielsen, J.B.: Towards predicting expressed emotion in music from pairwise comparisons. In: 9th Sound and Music Computing Conference (SMC) Illusions. (July 2012)

Subjective difference in ranking (Arousal)

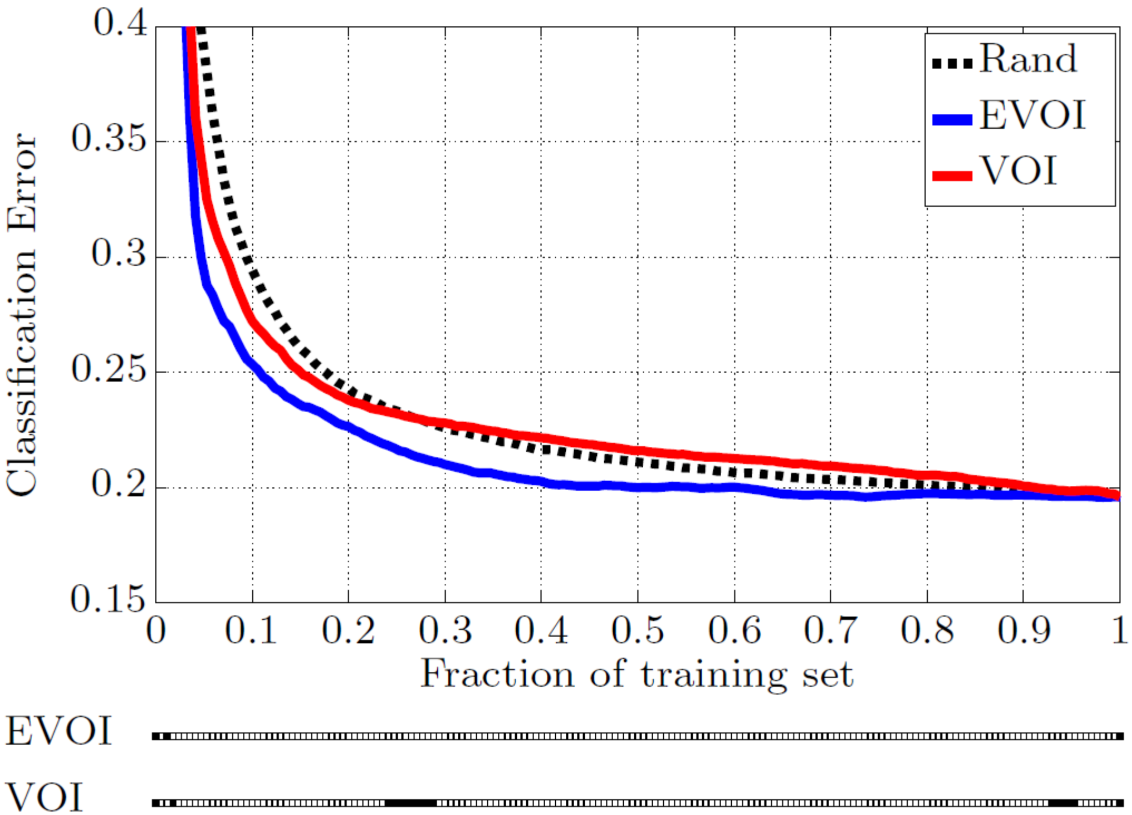


Are rankings dependent on model choice? Ranking difference (Arousal)



Madsen, J., Jensen, B.S., Larsen, J., Nielsen, J.B.: Towards predicting expressed emotion in music from pairwise compari

How many pairwise comparisons do we need to model emotions?



Using active learning

15% for valence

9% for arousal

Madsen, J., Jensen, B.S., Larsen, J., Predictive modeling of expressed emotions in music using pairwise comparisons. M. Aramaki et al. (Eds.): CMMR 2012, LNCS 7900, pp. 253–277, 2013. Springer-Verlag Berlin Heidelberg 2013

Main conclusion on eliciting emotions

- Models produce similar results using a learning curve
- Models produce different rankings specially when using a fraction of comparisons
- Large individual differences between the ranking of music expressed in music on dimensions of Valence and Arousal
- Promising error rates for both arousal and valence using as little as 30% of the training set corresponding to 2.5 comparisons per excerpt.
- Pairwise comparisons (2AFC) can scale when using active learning.

METADATA PREDICTION


WINAMP
File
Play
Options
View
Help
Detach Visualizer
Alternative Rock
Blues
Classical
Country
Dance
Folk
Jazz
Opera & Vocal
Pop
R&B
Rap & Hip-Hop
Rock
Prev
Next
Random
0:39
MUDDY WATERS - GOT MY MOJO WORKING (3:00)

AUDIO SOURCE SEPARATION

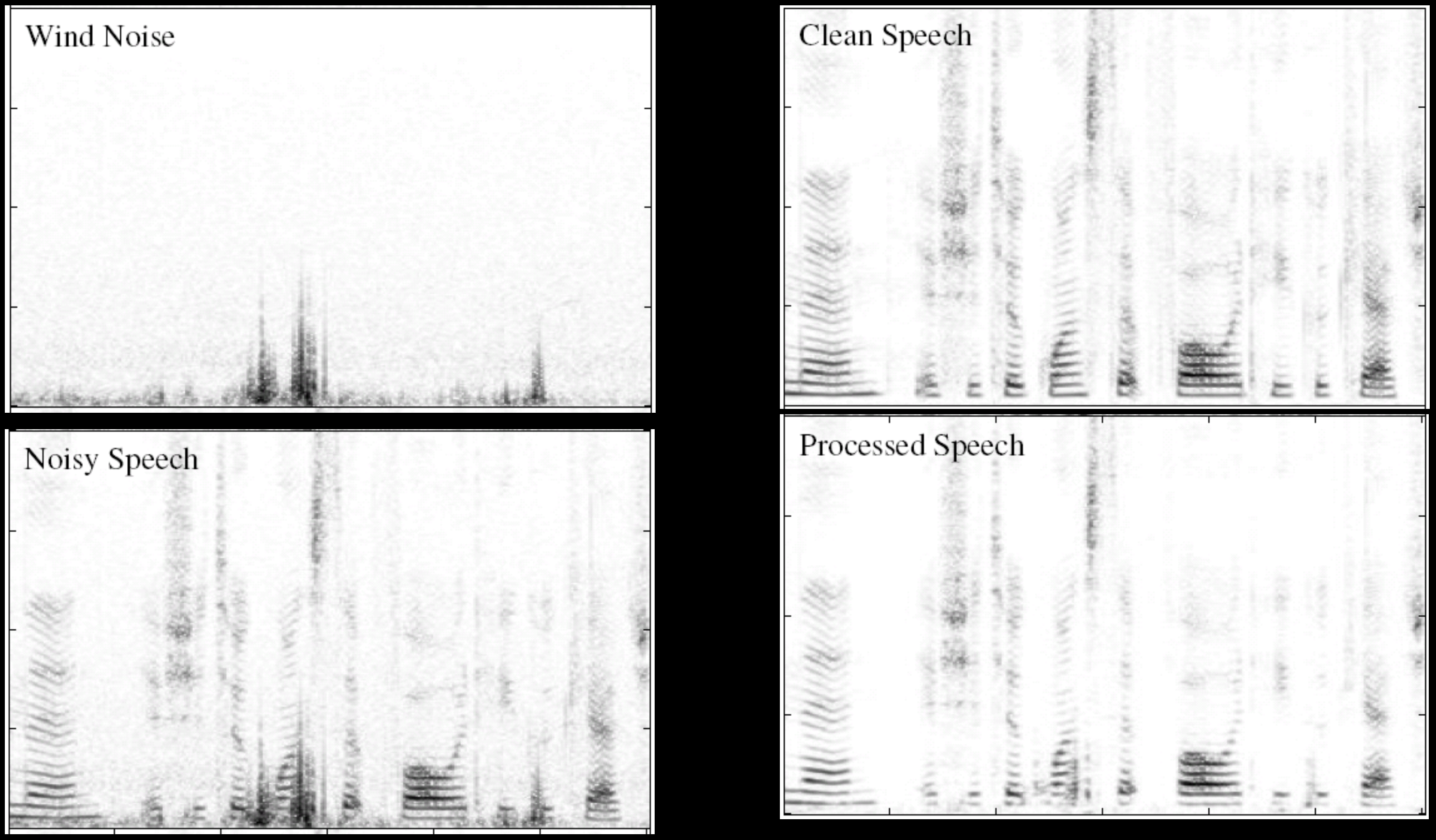
Audio separation

- A possible front end component e.g. the music search framework
- Noise reduction
- Music transcription
- Instrument detection and separation
- Vocalist identification

Semi-supervised learning methods

Pedersen, M. S., Larsen, J., Kjems, U., Parra, L. C., *A Survey of Convolutive Blind Source Separation Methods*, Springer Handbook of Speech, Springer Press, 2007

Wind noise reduction



M.N Schmidt, J. Larsen, F.T. Hsiao: Wind noise reduction using non-negative sparse coding, 2007.

Single channel separation: Sparse NMF decomposition

- Code-book (dictionary) of noise spectra is learned
- Can be interpreted as an advanced spectral subtraction technique

original

cleaned

alternative method (qualcom)

EXERCISE

- Modeling of 2AFC mood data using probit model with Gaussian Process
- Three covariance functions
  - Delta (ranking of data from 2AFC observations)
  - Linear
  - Squared Exponential
- Audio features
  - MFCC
  - Chroma
  - Loudness
- Inference and predictions
  - Laplace + MAP-II
  - 2D plots of AV predictions for individual users
- Active Learning mechanisms
  - Random
  - Entropy change (EVOI)