SOUND **AI**

Professor, PhD Jan Larsen

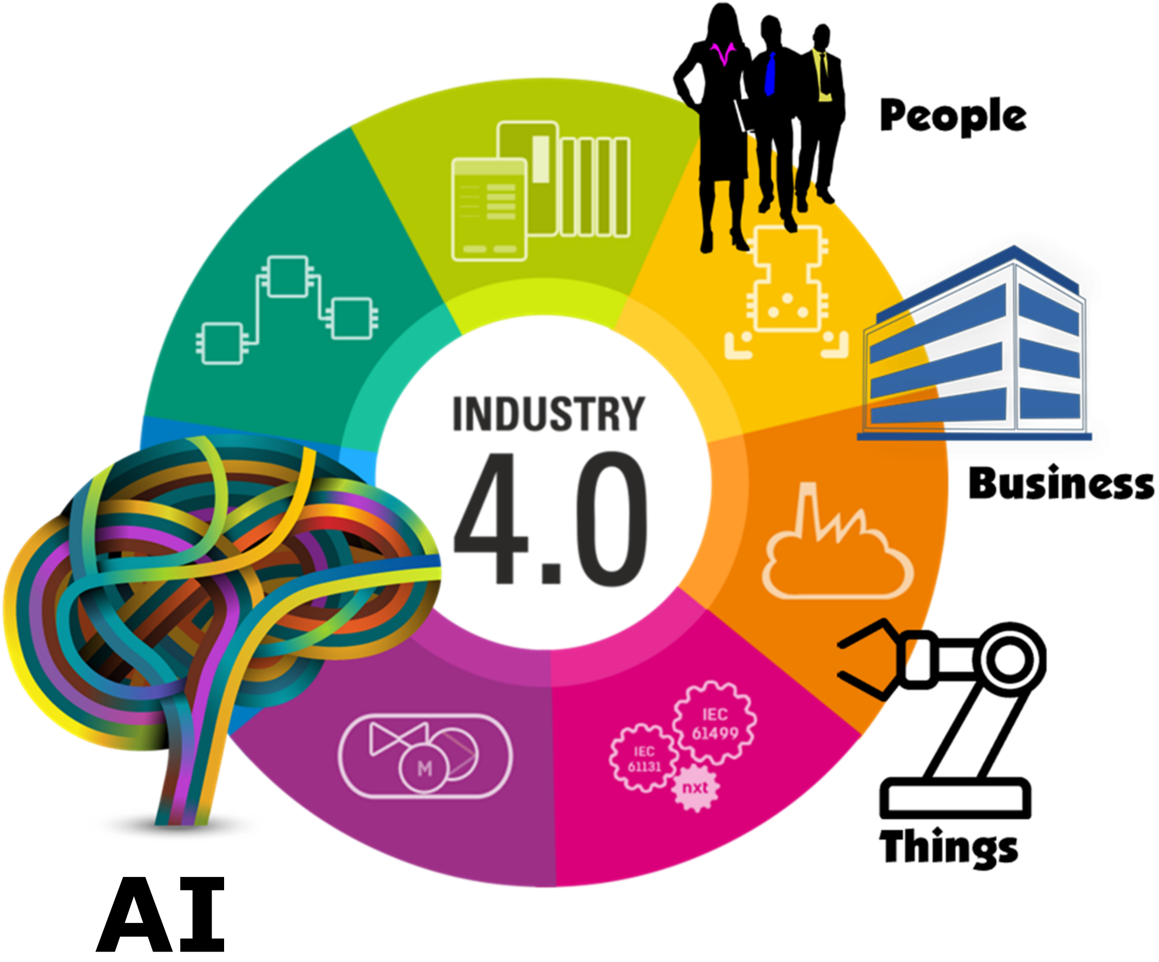
Section for Cognitive Systems
DTU Compute, Technical University of Denmark

Participation in 17 international and national collaborative research projects.
Mentoring of 2 Senior Researchers and 9 Postdocs, 34 PhD, and 90 MSc students.
>60% of projects in collaboration with private companies and stakeholders.
Danish Sound Innovation Network national network.
12 commissioned RDI projects.

My dream related to sound…

To create better quality of life by providing augmented and immersive sound experiences for people in society 4.0 using AI technology

A copy of the physical world through digitization makes it possible for cyber-physical systems to communicate and cooperate with each other and with humans in real time and perform decentralized decision-making



https://en.wikipedia.org/wiki/Industry_4.0
B. Marr: Forbes, June 20, 2016, http://www.forbes.com/sites/bernardmarr/2016/06/20/what-everyone-must-know-about-industry-4-0/#4c979f804e3b
http://www.enterrasolutions.com/2015/10/industry-4-0-facing-the-coming-revolution.html

Industry 4.0 = Civilization 4.0

It is a cognitive revolution that could be even more disruptive than earlier as it concerns not only the industry but the whole way we live our lives.

AI

Artificial Intelligence

IA

Intelligence Augmentation

research focus

CoSound

Machine learning based processing of audio data and related information, such as context, users' states, interaction, intention, and goals with the purpose of providing innovative services related to societal challenges in

Transforming big audio data into semantically interoperable data assets and knowledge: Enrichment and navigation in large sound archives such as broadcast

Experience economy and edutainment: New music services based on mood, optimization of sound systems

Healthcare: Music interventions to improve quality of life in relation to disorders such as e.g. stress, pain, and ADHD.
User-driven optimization of hearing aids.

research focus

MakeSense

Processing of sensor signals and related IoT data streams with the purpose of fostering innovative systems addressing societal challenges in

Food: Grain analysis

Security: Explosives and drug detection

Health: Blood and water analysis, intelligent drug delivery and sensing, e-health, personalized medicine

Energy: Wind mill maintenance

Environment: Exhaust gas analysis, large diesel engine predictive monitoring

Resource efficiency: Waste sorting

Digital economy: Event recommendation

SOUND IS AFFECTIVE

https://www.youtube.com/watch?v=to7uIG8KYhg

What are the mechanism? – the BRECVEM model

- **Brain stem reflexes** linked to acoustical properties, e.g. loudness
- **Evaluative conditioning** **–** association between music and emotion when they occur together
- **Emotional contagion** **–** emotion expressed in music, sad is e.g. linked low-pitches, slow, and quiet
- **Rhythmic entrainment** – movement synchronization with rhythm
- **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 analytic or aesthetic pleasure (hit-songs)
- **Musical expectancy** - balance between surprise and expectation

Ref: Juslin, P. N. and Västfäll, D. *Emotional response to music: The need to consider underlying mechanism. Behavioral and Brain Sciences, vol. 31, pp. 559–621, 2008*.
Line Gebauer & Peter Vuust, *Music interventions in Health Care, 2014.*

AI IS EFFECTIVE

What is machine learning?

Learning structures and patterns form from historical data to reliably predict outcome for new data.

Computers only do what they are programmed to do. ML infers new relations and patterns, which were not programmed. They learn and adapt to changing environment.

1. Computer systems that automatically improve through experience, or learns from data.
2. Inferential process that operate from representations that encode probabilistic dependencies among data variables capturing the likelihoods of relevant states in the world.
3. Development of fundamental statistical computational-information-theoretic laws that govern learning systems - including computers, humans, and other entities.

M. I. Jordan and T. M. Mitchell. *Machine learning: Trends, perspectives, and prospects*. Science, July 2015.
Samuel J. Gershman, Eric J. Horvitz, Joshua B. Tenenbaum. *Computational rationality: A converging paradigm for intelligence in brains, minds, and machines.* Science, July 2015.

Brief history of AI

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

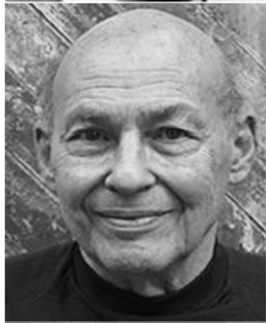
1951 Marvin Minsky’s analog neural networks (1st generation)

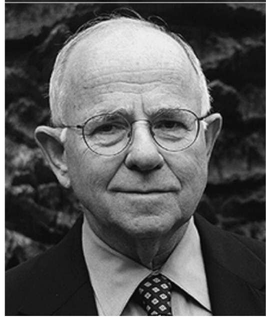
1956 Dartmouth conference: Artificial intelligence with aim of human like intelligence

1960 Bernard Widrow’s ADALINE - adatpive linear systems

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





1980's Expert systems useful in restricted domains

1980's Knowledge based systems – integration of diverse information sources

1980's The 2nd generation neural network revolution starts

Late 1980's Robotics and the role of embodiment to achieve intelligence

1990's AI and cybernetics research under new names such as machine learning, computational intelligence, evolutionary computing, neural networks, Bayesian networks, complex systems, game theory, deep neural networks (3rd generation) cognitive systems

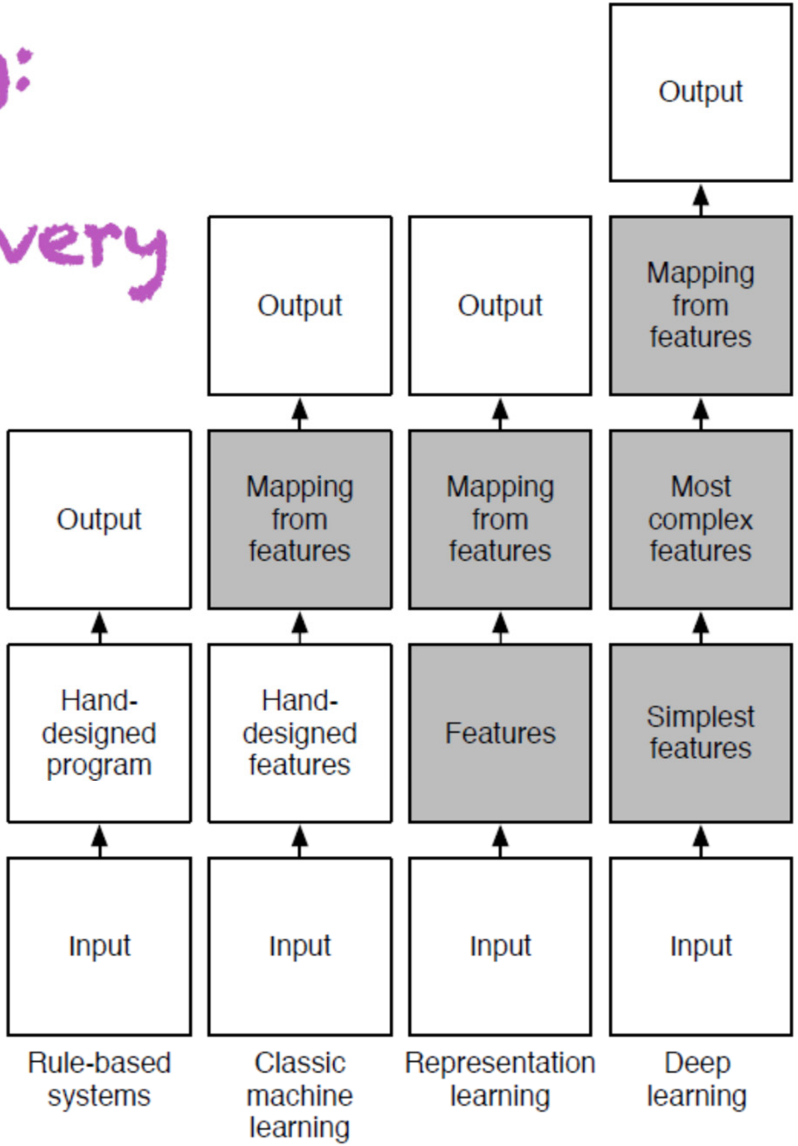
2010's deep neural networks (4rd generation) and cognitive systems, large scale data and computational frameworks, ML is commoditized

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

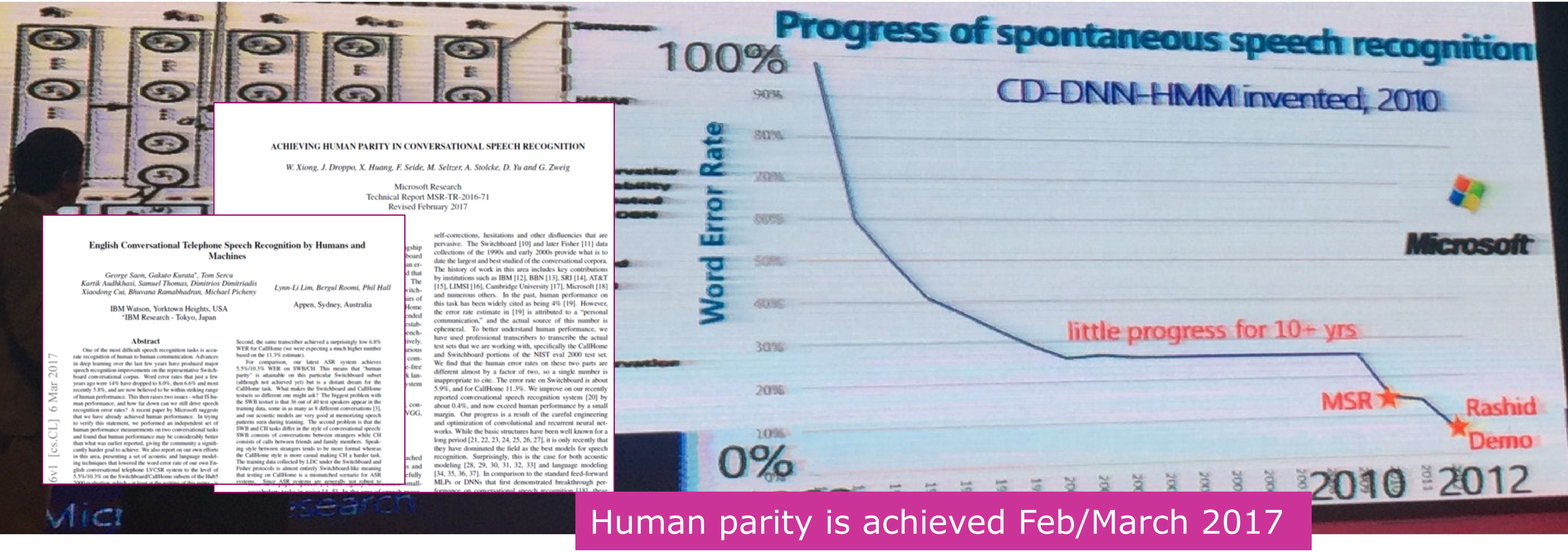
Deep Learning: Automating Feature Discovery

Geoff Hinton, Yoshua Bengio, Yann LeCun, Deep Learning Tutorial, NIPS 2015, Montreal.

Deep learning is a disruptive technology



Machine learning is very successful for speech recognition and chat bots



ACHIEVING HUMAN PARITY IN CONVERSATIONAL SPEECH RECOGNITION

W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu and G. Zweig

Microsoft Research
Technical Report MSR-TR-2016-71
Revised February 2017

self-corrections, hesitations and other disfluencies that are pervasive. The Switchboard [10] and later Fisher [11] data collections of the 1990s and early 2000s provide what is to date the largest and best studied of the conversational corpora. The history of work in this area includes key contributions by institutions such as IBM [12], BBN [13], SRI [14], AT&T [15], LIMSI [16], Cambridge University [17], Microsoft [18] and numerous others. In the past, human performance on this task has been widely cited as being 4% [19]. However, the error rate estimate in [19] is attributed to a "personal communication," and the actual source of this number is ephemeral. To better understand human performance, we have used professional transcribers to transcribe the actual test sets that we are working with, specifically the CallHome and Switchboard portions of the NIST eval 2000 test set. We find that the human error rates on these two parts are different almost by a factor of two, so a single number is inappropriate to cite. The error rate on Switchboard is about 5.9%, and for CallHome 11.3%. We improve on our recently reported conversational speech recognition system [20] by about 0.4%, and now exceed human performance by a small margin. Our progress is a result of the careful engineering and optimization of convolutional and recurrent neural networks. While the basic structures have been well known for a long period [21, 22, 23, 24, 25, 26, 27], it is only recently that they have dominated the field as the best models for speech recognition. Surprisingly, this is the case for both acoustic modeling [28, 29, 30, 31, 32, 33] and language modeling [34, 35, 36, 37]. In comparison to the standard feed-forward MLPs or DNNs that first demonstrated breakthrough performance on conversational speech recognition [18], these

English Conversational Telephone Speech Recognition by Humans and Machines

George Saon, Gakuto Kurata, Tom Sercu*
Kartik Audhkhasi, Samuel Thomas, Dimitrios Dimitriadis
Xiaodong Cui, Bhuvana Ramabhadran, Michael Picheny

IBM Watson, Yorktown Heights, USA
*IBM Research - Tokyo, Japan

Lynn-Li Lim, Bergul Roomi, Phil Hall

Appen, Sydney, Australia

6v1 [cs.CL] 6 Mar 2017

Abstract

One of the most difficult speech recognition tasks is accurate recognition of human to human communication. Advances in deep learning over the last few years have produced major speech recognition improvements on the representative Switchboard conversational corpus. Word error rates that just a few years ago were 14% have dropped to 8.0%, then 6.6% and most recently 5.8%, and are now believed to be within striking range of human performance. This then raises two issues - what IS human performance, and how far down can we still drive speech recognition error rates? A recent paper by Microsoft suggests that we have already achieved human performance. In trying to verify this statement, we performed an independent set of human performance measurements on two conversational tasks and found that human performance may be considerably better than what was earlier reported, giving the community a significantly harder goal to achieve. We also report on our own efforts in this area, presenting a set of acoustic and language modeling techniques that lowered the word error rate of our own English conversational telephone LVCSR system to the level of 5.5%/10.3% on the Switchboard/CallHome subsets of the Hub5 2000 evaluation, which - at least at the writing of this paper - is

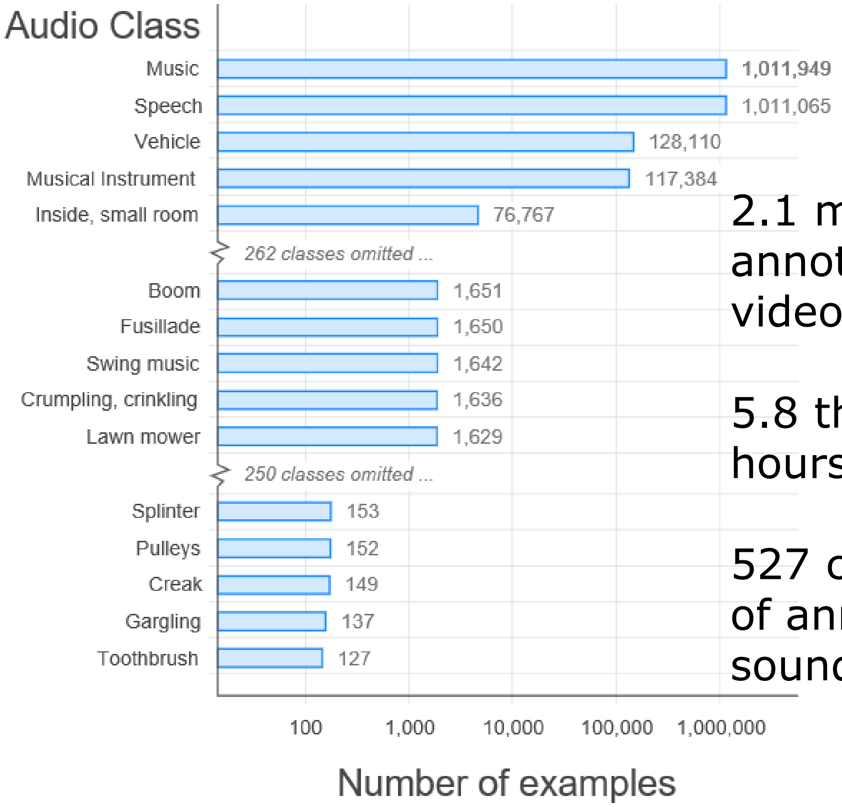
Second, the same transcriber achieved a surprisingly low 6.8% WER for CallHome (we were expecting a much higher number based on the 11.3% estimate).

For comparison, our latest ASR system achieves 5.5%/10.3% WER on SWB/CH. This means that "human parity" is attainable on this particular Switchboard subset (although not achieved yet) but is a distant dream for the CallHome task. What makes the Switchboard and CallHome testsets so different one might ask? The biggest problem with the SWB testset is that 36 out of 40 test speakers appear in the training data, some in as many as 8 different conversations [3], and our acoustic models are very good at memorizing speech patterns seen during training. The second problem is that the SWB and CH tasks differ in the style of conversational speech: SWB consists of conversations between strangers while CH consists of calls between friends and family members. Speaking style between strangers tends to be more formal whereas the CallHome style is more casual making CH a harder task. The training data collected by LDC under the Switchboard and Fisher protocols is almost entirely Switchboard-like meaning that testing on CallHome is a mismatched scenario for ASR systems. Since ASR systems are generally not robust to

Human parity is achieved Feb/March 2017

Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, and Brian Kingsbury. *Deep Neural Networks for Acoustic Modeling in Speech Recognition.* IEEE Signal Processing Magazine, 82, Nov. 2012.
George Saon, Gakuto Kurata, Tom Sercu, Kartik Audhkhasi, Samuel Thomas, Dimitrios Dimitriadis, Xiaodong Cui, Bhuvana Ramabhadran, Michael Picheny, Lynn-Li Lim, Bergul Roomi, Phil Hall. *English Conversational Telephone Speech Recognition by Humans and Machines, https://arxiv.org/abs/1703.02136, March 2017*
W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu, G. Zweig. *Achieving Human Parity in Conversational Speech Recognition, https://arxiv.org/abs/1610.05256, October 2016.*

Machine learning is very successful for audio classification



2.1 million annotated videos

5.8 thousand hours of audio

527 classes of annotated sounds

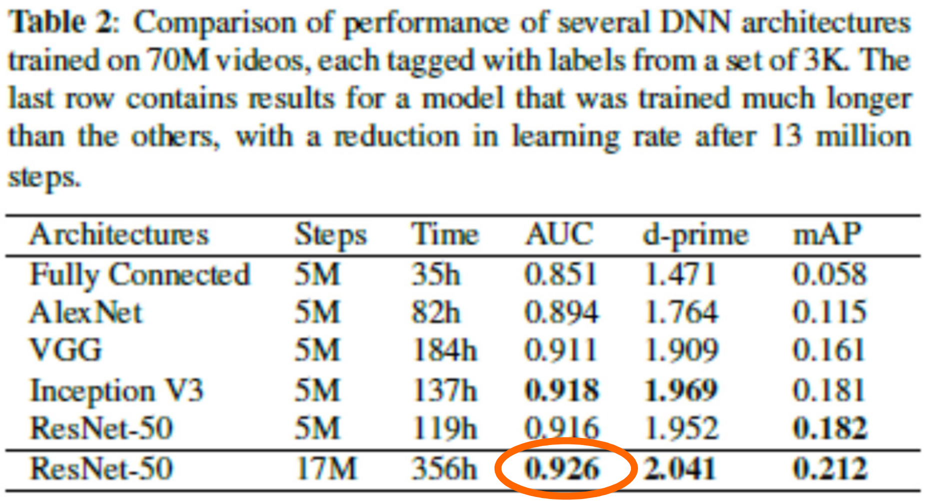


Table 2: Comparison of performance of several DNN architectures trained on 70M videos, each tagged with labels from a set of 3K. The last row contains results for a model that was trained much longer than the others, with a reduction in learning rate after 13 million steps.

Architectures	Steps	Time	AUC	d-prime	mAP
Fully Connected	5M	35h	0.851	1.471	0.058
AlexNet	5M	82h	0.894	1.764	0.115
VGG	5M	184h	0.911	1.909	0.161
Inception V3	5M	137h	**0.918**	**1.969**	0.181
ResNet-50	5M	119h	0.916	1.952	**0.182**
ResNet-50	17M	356h	**0.926**	**2.041**	**0.212**

Mean average precision mAP is low because of low class prior $<10^{-4}$.

AUC is the area under curve of true positive rate vs. false positive rate.

Jort F. Gemmeke, Daniel P. W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, Marvin Ritter. *Audio Set: An ontology and human-labeled dataset for audio events*, IEEE ICASSP 2017, New Orleans, LA, March 2017.

Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron Weiss, Kevin Wilson. *CNN Architectures for Large-Scale Audio Classification*, ICASSP 2017, New Orleans, LA, March 2017.

Machine learning is very successful for speech generation

WaveNet is a deep generative model of raw audio waveforms

WaveNets are able to generate speech which mimics any human voice and which sounds more natural than the best existing Text-to-Speech systems, reducing the gap with human performance by over 50%.

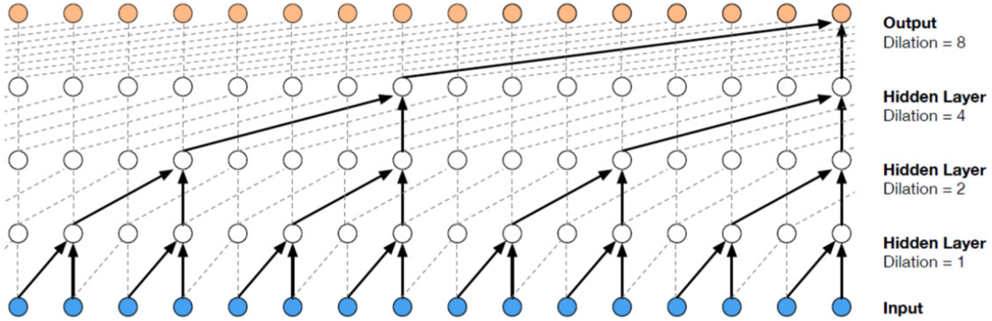
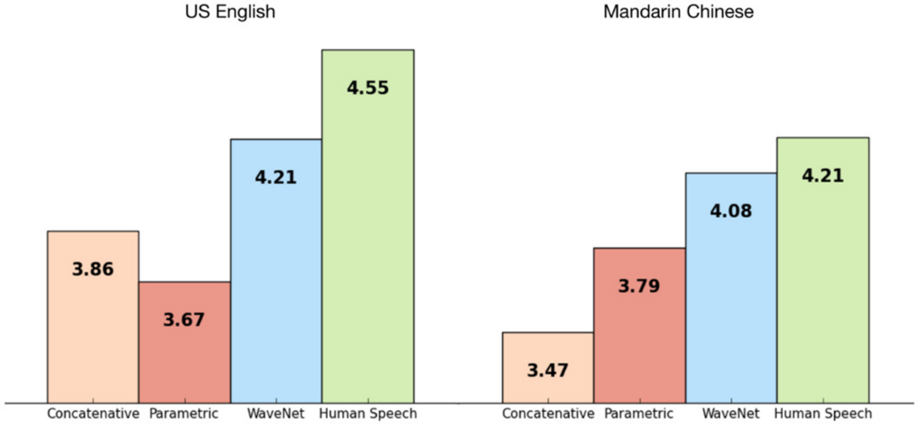


Figure 3: Visualization of a stack of *dilated* causal convolutional layers.



Äaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu. *WAWENET: A Generative Model for Raw Audio,* https://arxiv.org/pdf/1609.03499.pdf, Sept 2016*,* https://deepmind.com/blog/wavenet-generative-model-raw-audio/

BLACK BOX OF AI

Objectives

Trust

Causality

Transferability

Decomposability

Informativeness

Legal issues

European Union regulations (GDPR) on algorithmic decision-making and a “right to explanation”

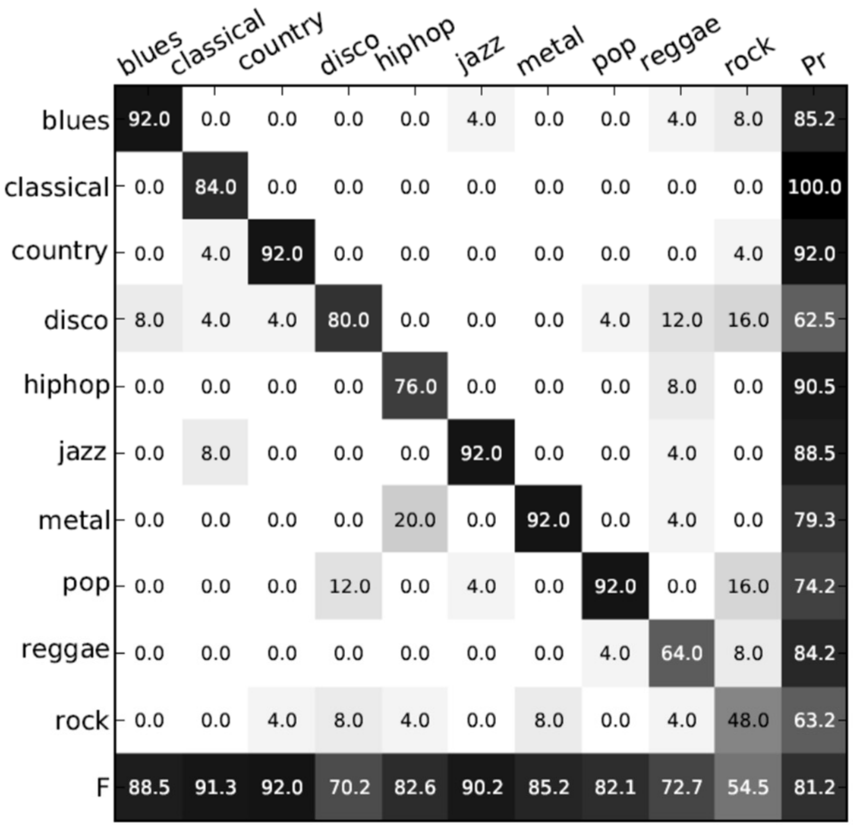
Davide Castelvecchi: http://www.nature.com/polopoly_fs/1.20731!/menu/main/topColumns/topLeftColumn/pdf/538020a.pdf, Nature, Vol. 538, 6 Oct. 2016

K.R. Müller and Wojciech Samek: *Explaining and Interpreting Deep Neural Networks, 02901 Advances Topics in Machine Learning, DTU 2017*

Z.C. Lipton: *The mythos of model interpretability*, arXiv:1606.03490, 2016.

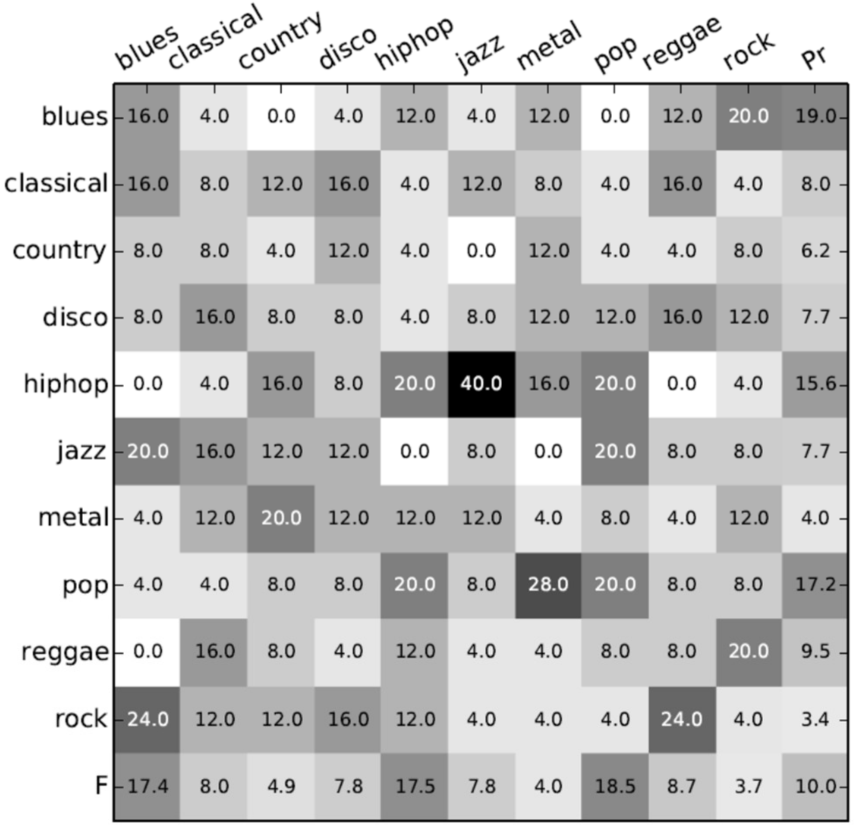
Bryce Goodman, Seth Flaxman: *European Union regulations on algorithmic decision-making and a “right to explanation”*, https://arxiv.org/pdf/1606.08813v3.pdf

Adversarial learning



	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock	Pr
blues	92.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	4.0	8.0	85.2
classical	0.0	84.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
country	0.0	4.0	92.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	92.0
disco	8.0	4.0	4.0	80.0	0.0	0.0	0.0	4.0	12.0	16.0	62.5
hiphop	0.0	0.0	0.0	0.0	76.0	0.0	0.0	0.0	8.0	0.0	90.5
jazz	0.0	8.0	0.0	0.0	0.0	92.0	0.0	0.0	4.0	0.0	88.5
metal	0.0	0.0	0.0	0.0	20.0	0.0	92.0	0.0	4.0	0.0	79.3
pop	0.0	0.0	0.0	12.0	0.0	4.0	0.0	92.0	0.0	16.0	74.2
reggae	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	64.0	8.0	84.2
rock	0.0	0.0	4.0	8.0	4.0	0.0	8.0	0.0	4.0	48.0	63.2
F	88.5	91.3	92.0	70.2	82.6	90.2	85.2	82.1	72.7	54.5	81.2

(a) Original Input



	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock	Pr
blues	16.0	4.0	0.0	4.0	12.0	4.0	12.0	0.0	12.0	20.0	19.0
classical	16.0	8.0	12.0	16.0	4.0	12.0	8.0	4.0	16.0	4.0	8.0
country	8.0	8.0	4.0	12.0	4.0	0.0	12.0	4.0	4.0	8.0	6.2
disco	8.0	16.0	8.0	8.0	4.0	8.0	12.0	12.0	16.0	12.0	7.7
hiphop	0.0	4.0	16.0	8.0	20.0	40.0	16.0	20.0	0.0	4.0	15.6
jazz	20.0	16.0	12.0	12.0	0.0	8.0	0.0	20.0	8.0	8.0	7.7
metal	4.0	12.0	20.0	12.0	12.0	12.0	4.0	8.0	4.0	12.0	4.0
pop	4.0	4.0	8.0	8.0	20.0	8.0	28.0	20.0	8.0	8.0	17.2
reggae	0.0	16.0	8.0	4.0	12.0	4.0	4.0	8.0	8.0	20.0	9.5
rock	24.0	12.0	12.0	16.0	12.0	4.0	4.0	4.0	24.0	4.0	3.4
F	17.4	8.0	4.9	7.8	17.5	7.8	4.0	18.5	8.7	3.7	10.0

(b) Input intercepted by adversary A1

Corey Kereliuk, Bob L. Sturm, Jan Larsen: Deep Learning and Music Adversaries, IEEE Transactions on Multimedia, Nov. 2015

Corey Kereliuk, Bob L. Sturm, Jan Larsen: Deep Learning, Audio Adversaries, and Music Content Analysis, 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 2015

Corey Kereliuk, Bob L. Sturm, Jan Larsen: ?El Caballo Viejo? Latin Genre Recognition with Deep Learning and Spectral Periodicity, Fifth Biennial International Conference on Mathematics and Computation in Music (MCM2015), 2015.

Adversarial learning

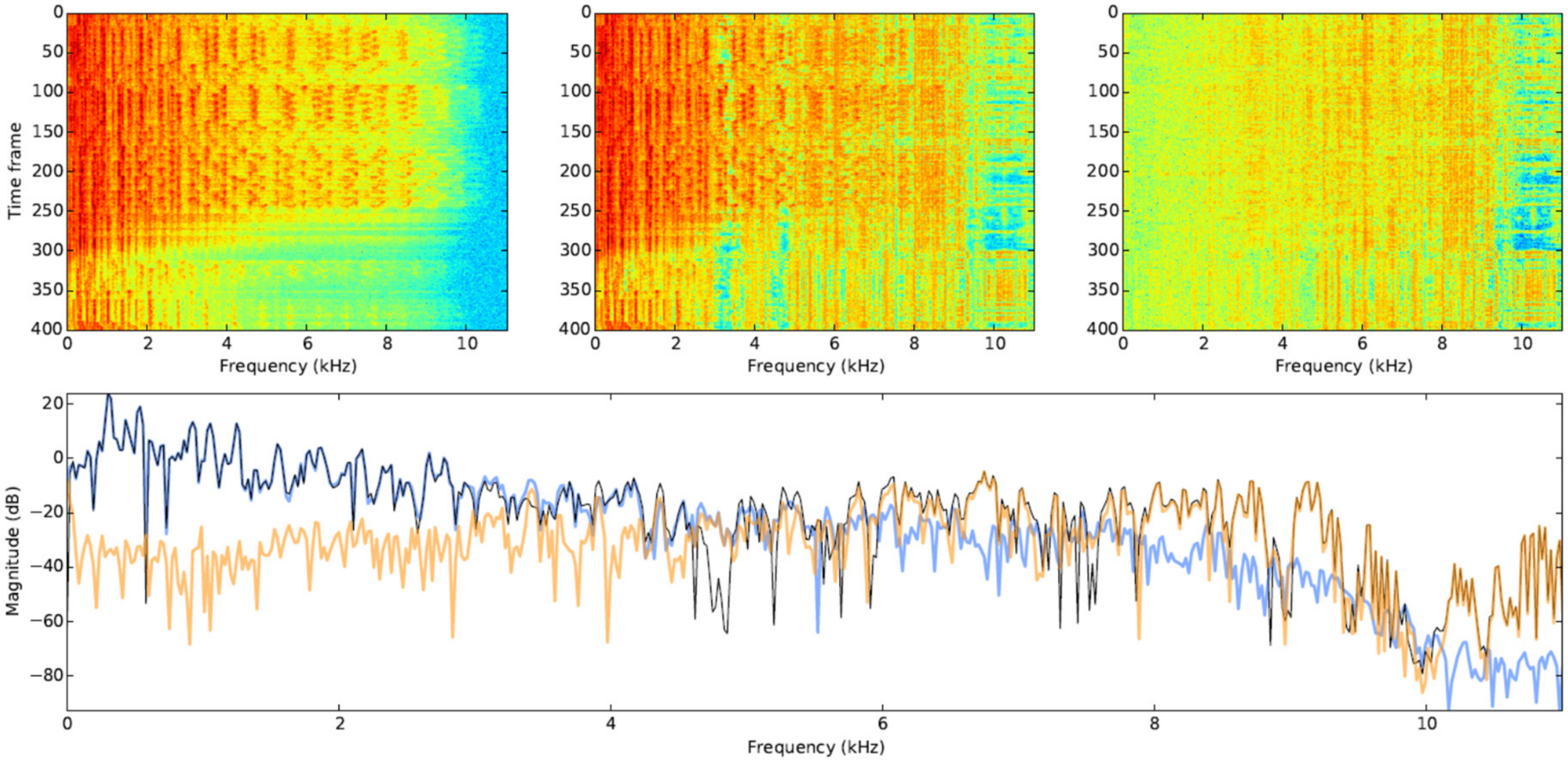


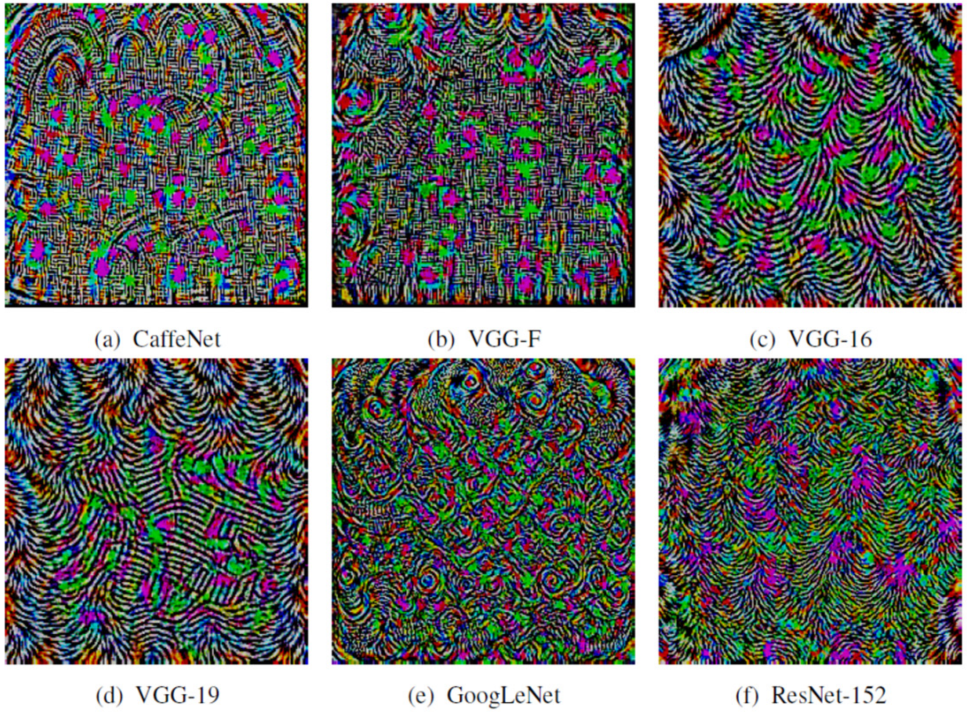
Fig. 5. Top left: spectrogram excerpt from *GTZAN* Classical “21” (Mozart, Symphony No. 39 Finale) that the DNN-based system in Fig. 2(b) classifies as *Classical*. Top middle: spectrogram of adversarial example classified as *Reggae*. Top right: spectrogram of the difference of the two. Bottom: magnitude spectrum of one frame (1024 samples) of the original (light blue), adversarial example (black), and difference (orange). Note that all excerpts in *GTZAN* have a sampling rate of 22050 Hz. The SNR = 21.1dB.

Corey Kereliuk, Bob L. Sturm, Jan Larsen: Deep Learning and Music Adversaries**,** IEEE Transactions on Multimedia, Nov. 2015

Corey Kereliuk, Bob L. Sturm, Jan Larsen: Deep Learning, Audio Adversaries, and Music Content Analysis, 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 2015

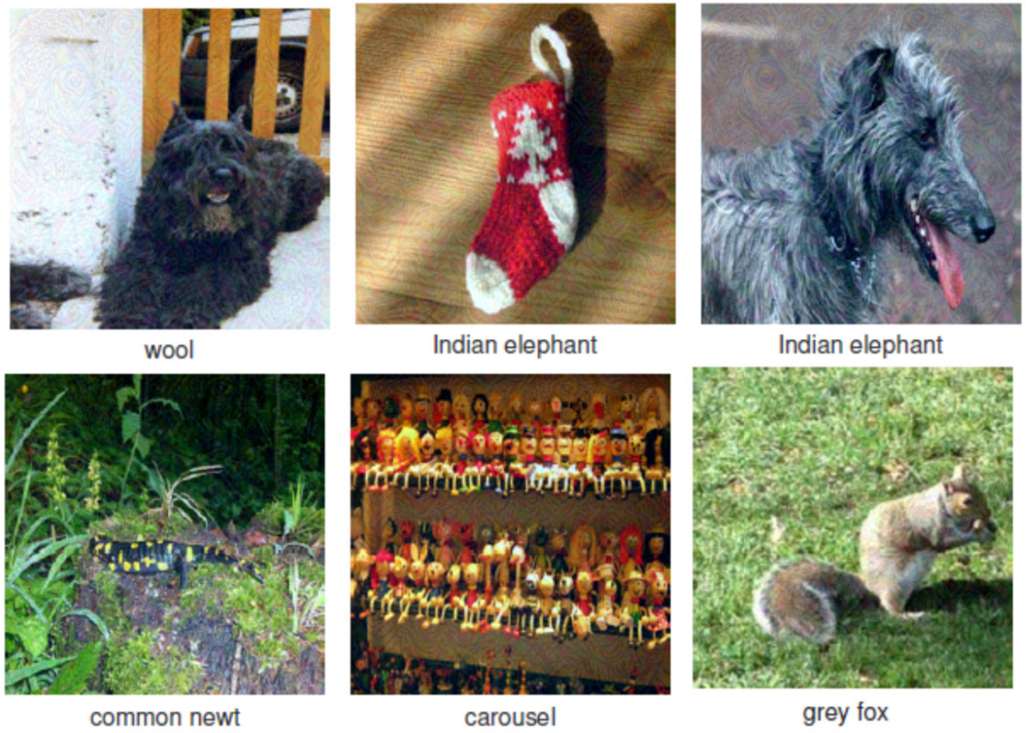
Corey Kereliuk, Bob L. Sturm, Jan Larsen: ?El Caballo Viejo? Latin Genre Recognition with Deep Learning and Spectral Periodicity, Fifth Biennial International Conference on Mathematics and Computation in Music (MCM2015), 2015.

Universal Adversarial Learning



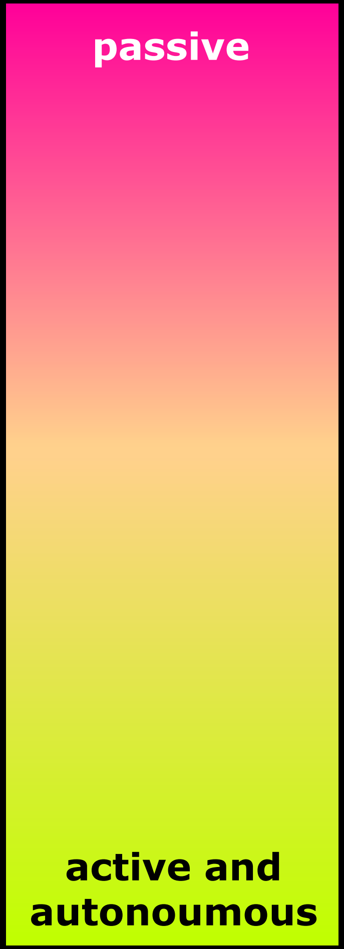
(a) CaffeNet (b) VGG-F (c) VGG-16

(d) VGG-19 (e) GoogLeNet (f) ResNet-152



Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, Pascal Frossard: Universal adversarial perturbations, arXiv:1610.08401. 2017

What defines simple and complex problems - and how do we solve them them?



passive

active and autonomous

exploration and summarization

prediction

continuous learning
reflection
pro-activeness
engagement
experimentation
creativity

Unreasonable effectiveness of

Mathematics E. Wigner, 1960

Data Halevy, Norvig, Pereira, 2009

RNNs Karpathy, 2015

Experimentation and interaction through users-in-the-loop

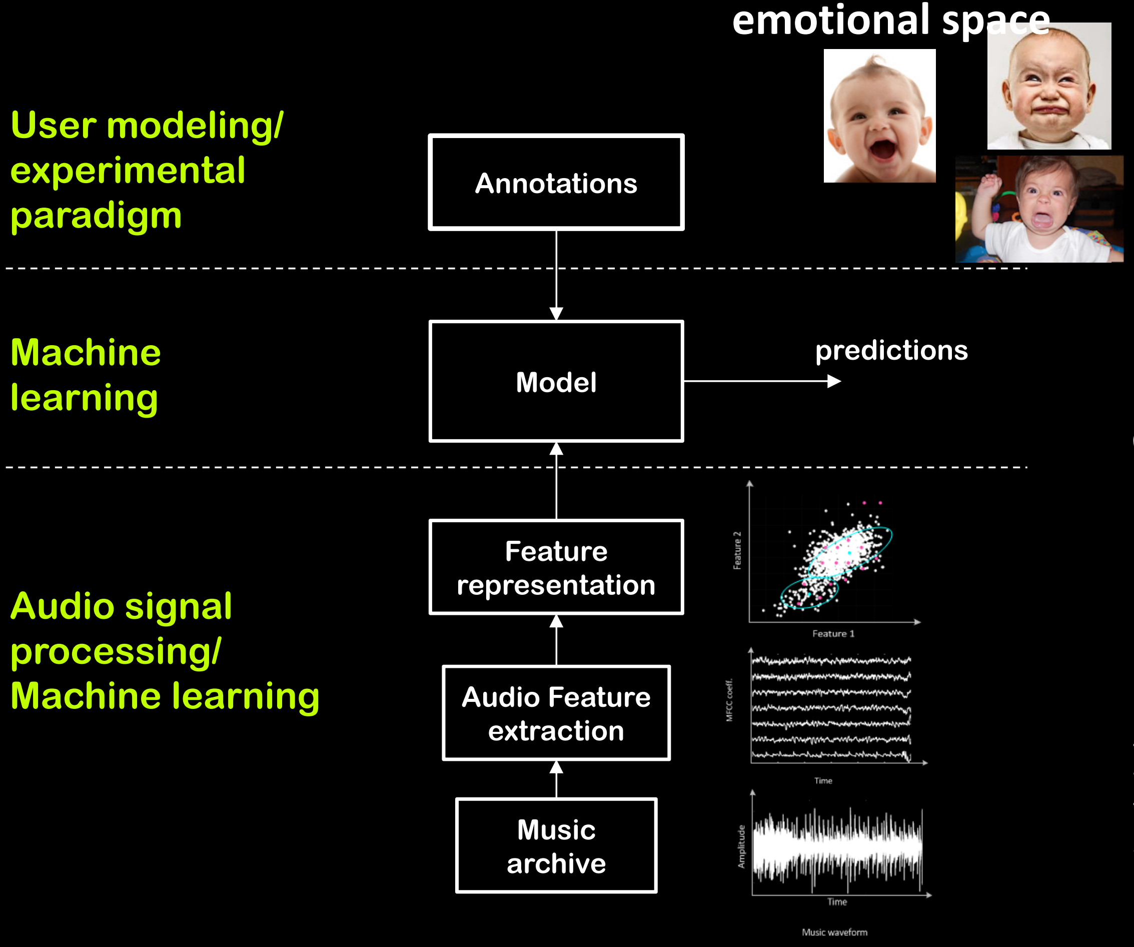
INTERACTIVE MACHINE LEARNING IN SOUND

Expressed emotions in music

- Jens Madsen, Jan Larsen. The Confidence Effect in Elicitation of Expressed Emotion in Music. To be submitted.
- Jens Madsen, Jan Larsen. Designing a Cognitive Music System. To be submitted.
- Jens Madsen, Bjø [illegible] ure Representations [illegible]
- Jens Madsen, Bjørn Sa [illegible] Music using Probabilistic Features Representations, submitted IEEE T-ASPL, 2015.
- 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 and induced emotion?
- Which scaling method should we use?
- Which role does mood play?

Music Emotion Modeling



emotional space

User modeling/ experimental paradigm

Annotations

Machine learning

Model

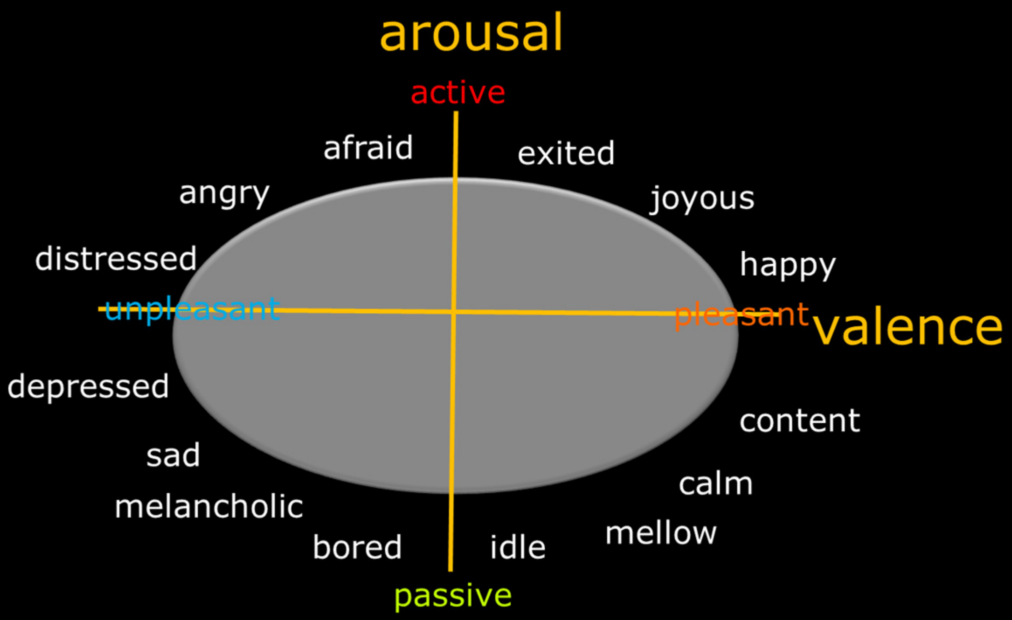
predictions

Audio signal processing/ Machine learning

Feature representation

Audio Feature extraction

Music archive



arousal

active

afraid

exited

angry

joyous

distressed

happy

unpleasant

pleasant

valence

depressed

content

sad

calm

melancholic

mellow

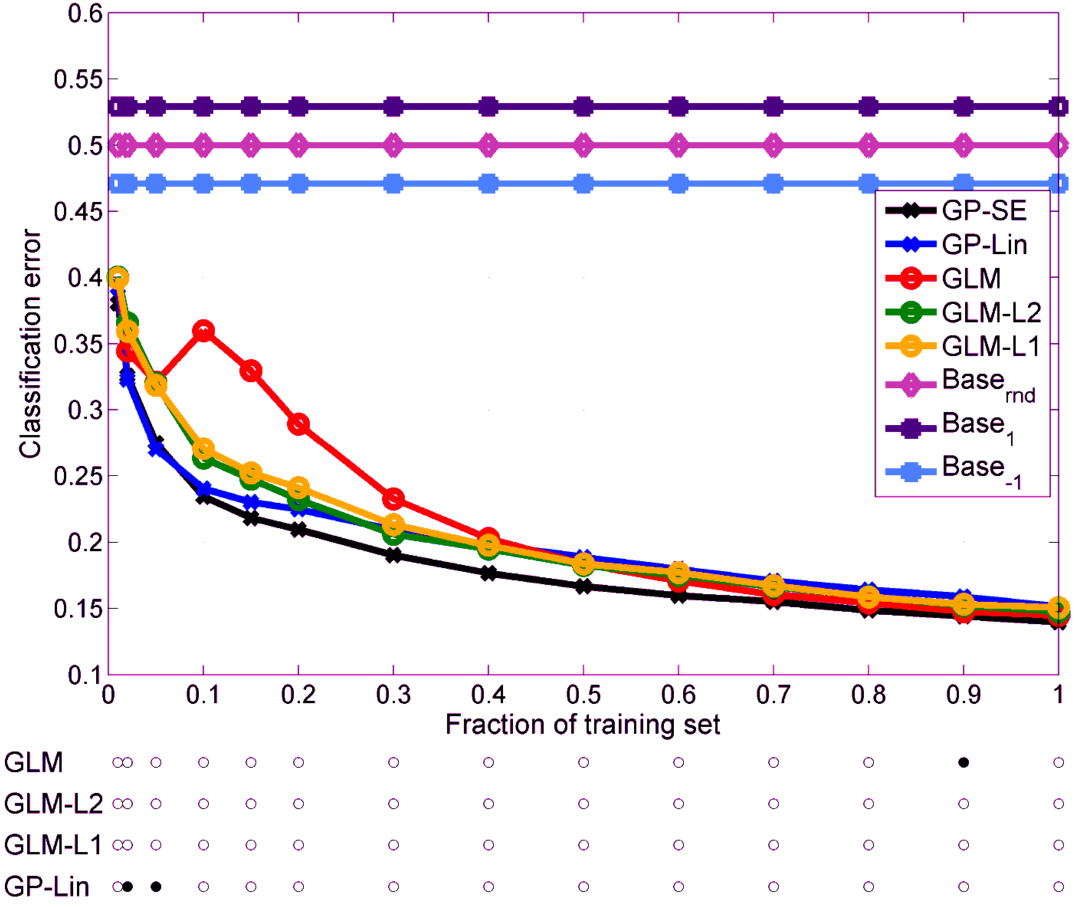
bored

idle

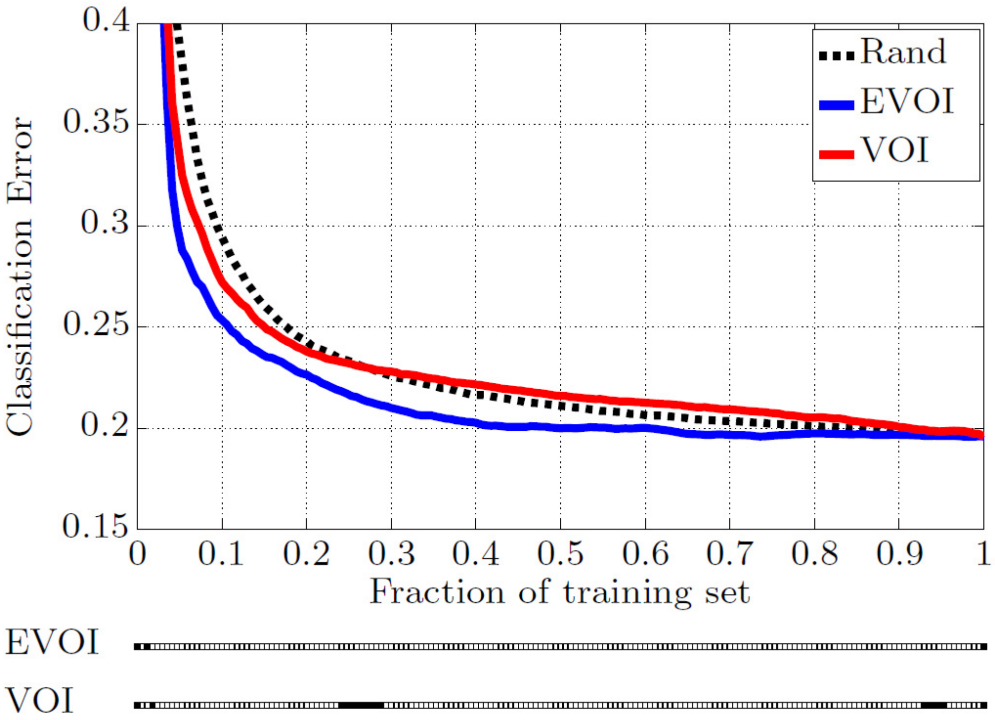
passive

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

Learning curve modeling arousal shows nonlinear modelling is best



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

The power of human data

Why - Humans as a measurement device

How - Humans in the loop

Who - Humans in the loop

Why - Humans as a measurement device

- With the purpose of individualization and dynamical response.
- With the purpose of group studies and population models.
- For eliciting perceptual, affective, and cognitive aspects.
- For acquiring other aspects e.g. behavioral and physical.
- For quality measurement and control.
- For obtaining shared cognitive and cultural information and contexts that helps disambiguation of meaning.

How - Humans in the loop

- Direct measurement of physiological, cognitive and behavior states from physical devices.
- Indirect measurements from self-reports, experiments using direct, indirect and related scaling methods of objective or subjective information.

Whether data are Experimental or Observational plays an important role!

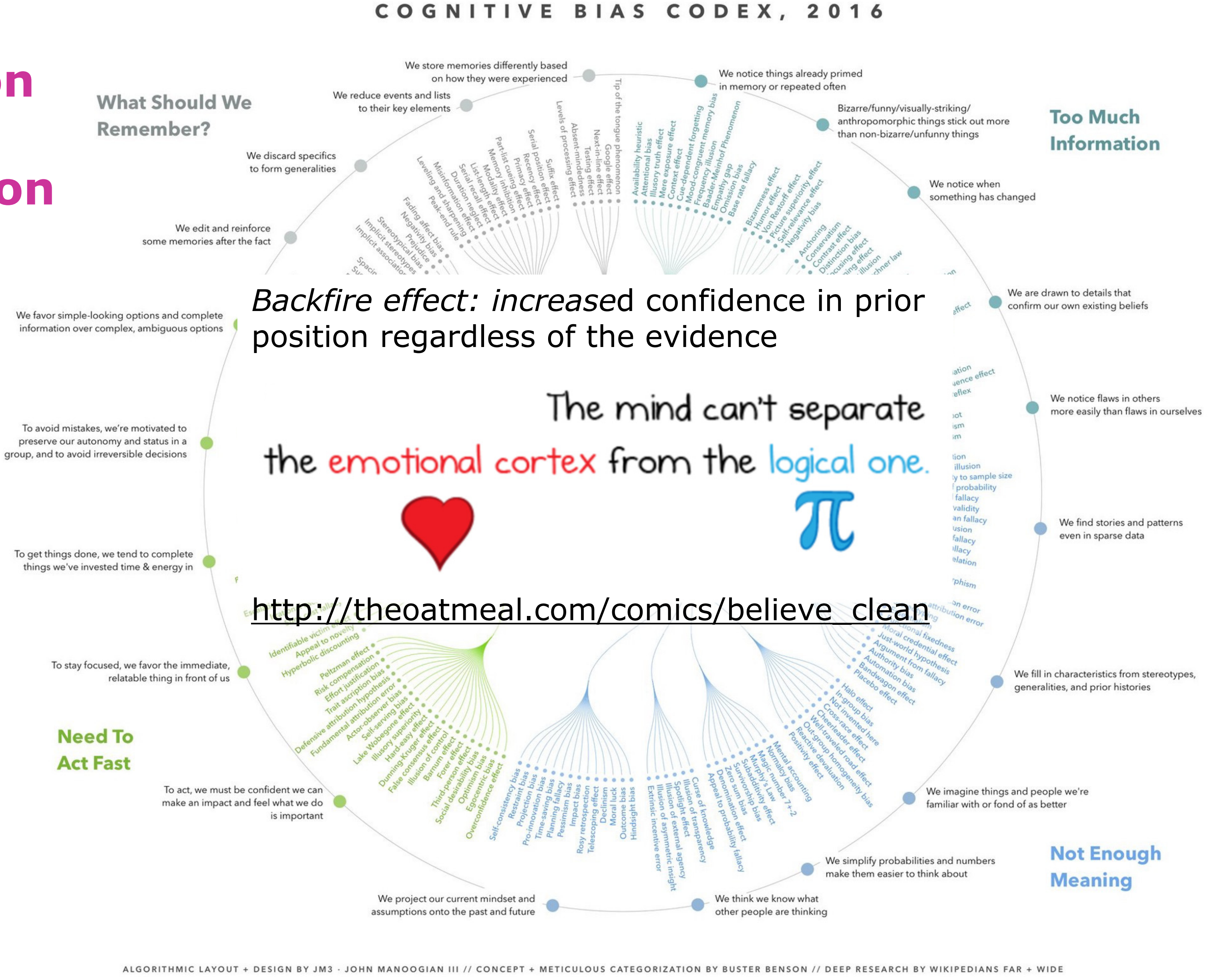
Who - Humans in the loop

- End-user
- Experimenter
- Developer
- Expert user
- Collaborative, transfer learning for crowds of humans

Challenge: Robust adaptive learning and optimization from interaction with inconsistent, biased and often inattentive users

– Modeling and/or knowledge of many aspects of the state of person(s) and the environment
– Modeling and representing uncertainty
    - concerning the “objective” (incl. needs, intentions, level of engagement)
    - concerning the interaction/answers/measurements from the subjects
– Support for varying complexity of a multi-aspect objective function
– Adaptive/online elicitation and learning of the objective function

Human interaction with information


COGNITIVE BIAS CODEX, 2016
What Should We Remember?
Too Much Information
Need To Act Fast
Not Enough Meaning
We store memories differently based on how they were experienced
We reduce events and lists to their key elements
We discard specifics to form generalities
We edit and reinforce some memories after the fact
We notice things already primed in memory or repeated often
Bizarre/funny/visually-striking/anthropomorphic things stick out more than non-bizarre/unfunny things
We notice when something has changed
We are drawn to details that confirm our own existing beliefs
We notice flaws in others more easily than flaws in ourselves
We find stories and patterns even in sparse data
We fill in characteristics from stereotypes, generalities, and prior histories
We imagine things and people we're familiar with or fond of as better
We simplify probabilities and numbers make them easier to think about
We think we know what other people are thinking
We project our current mindset and assumptions onto the past and future
To act, we must be confident we can make an impact and feel what we do is important
To stay focused, we favor the immediate, relatable thing in front of us
To get things done, we tend to complete things we've invested time & energy in
To avoid mistakes, we're motivated to preserve our autonomy and status in a group, and to avoid irreversible decisions
We favor simple-looking options and complete information over complex, ambiguous options
Tip of the tongue phenomenon
Google effect
Next-in-line effect
Testing effect
Absent-mindedness
Levels of processing effect
Suffix effect
Serial position effect
Recency effect
Primacy effect
Part-list cueing effect
Memory inhibition
Modality effect
List-length effect
Serial recall effect
Duration neglect
Misinformation effect
Leveling and sharpening
Peak-end rule
Fading affect bias
Negativity bias
Prejudice
Stereotypical bias
Implicit stereotypes
Implicit association
Availability heuristic
Attentional bias
Illusory truth effect
Mere exposure effect
Context effect
Cue-dependent forgetting
Mood-congruent memory bias
Frequency illusion
Baader-Meinhof Phenomenon
Empathy gap
Omission bias
Base rate fallacy
Bizarreness effect
Humor effect
Von Restorff effect
Picture superiority effect
Self-relevance effect
Negativity bias
Anchoring
Conservatism
Contrast effect
Distinction bias
Focusing effect
Weber-Fechner law
Identifiable victim effect
Appeal to novelty
Hyperbolic discounting
Peltzman effect
Risk compensation
Effort justification
Trait ascription bias
Defensive attribution hypothesis
Fundamental attribution error
Actor-observer bias
Self-serving bias
Lake Wobegone effect
Illusory superiority
Hard-easy effect
Dunning-Kruger effect
False consensus effect
Illusion of control
Barnum effect
Forer effect
Third-person effect
Social desirability bias
Optimism bias
Egocentric bias
Overconfidence effect
Self-consistency bias
Restraint bias
Projection bias
Pro-innovation bias
Time-saving bias
Planning fallacy
Pessimism bias
Impact bias
Rosy retrospection
Telescoping effect
Declinism
Moral luck
Outcome bias
Hindsight bias
Extrinsic incentive error
Illusion of asymmetric insight
Illusion of external agency
Spotlight effect
Illusion of transparency
Curse of knowledge
Appeal to probability fallacy
Denomination effect
Zero sum bias
Survivorship bias
Subadditivity effect
Magic number 7+-2
Murphy's Law
Normalcy bias
Mental accounting
Positivity effect
Reactive devaluation
Out-group homogeneity bias
Well-traveled road effect
Cheerleader effect
Cross-race effect
Not invented here
In-group bias
Halo effect
Placebo effect
Bandwagon effect
Automation bias
Authority bias
Argument from fallacy
Just-world hypothesis
Moral credential effect
Functional fixedness
ALGORITHMIC LAYOUT + DESIGN BY JM3 · JOHN MANOOGIAN III // CONCEPT + METICULOUS CATEGORIZATION BY BUSTER BENSON // DEEP RESEARCH BY WIKIPEDIANS FAR + WIDE

*Backfire effect: increase*d confidence in prior position regardless of the evidence

The mind can't separate
the emotional cortex from the logical one.
π

http://theoatmeal.com/comics/believe_clean

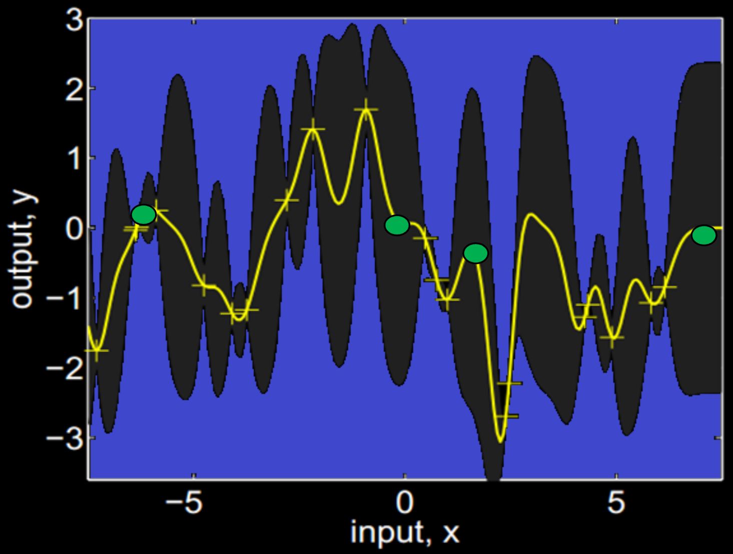
Interactive Learning / Sequential Experimental Design

Generalization objective

Eliciting and learning the entire model / objective function.
Expected change in relative entropy is derived from the posterior and predictive distribution.

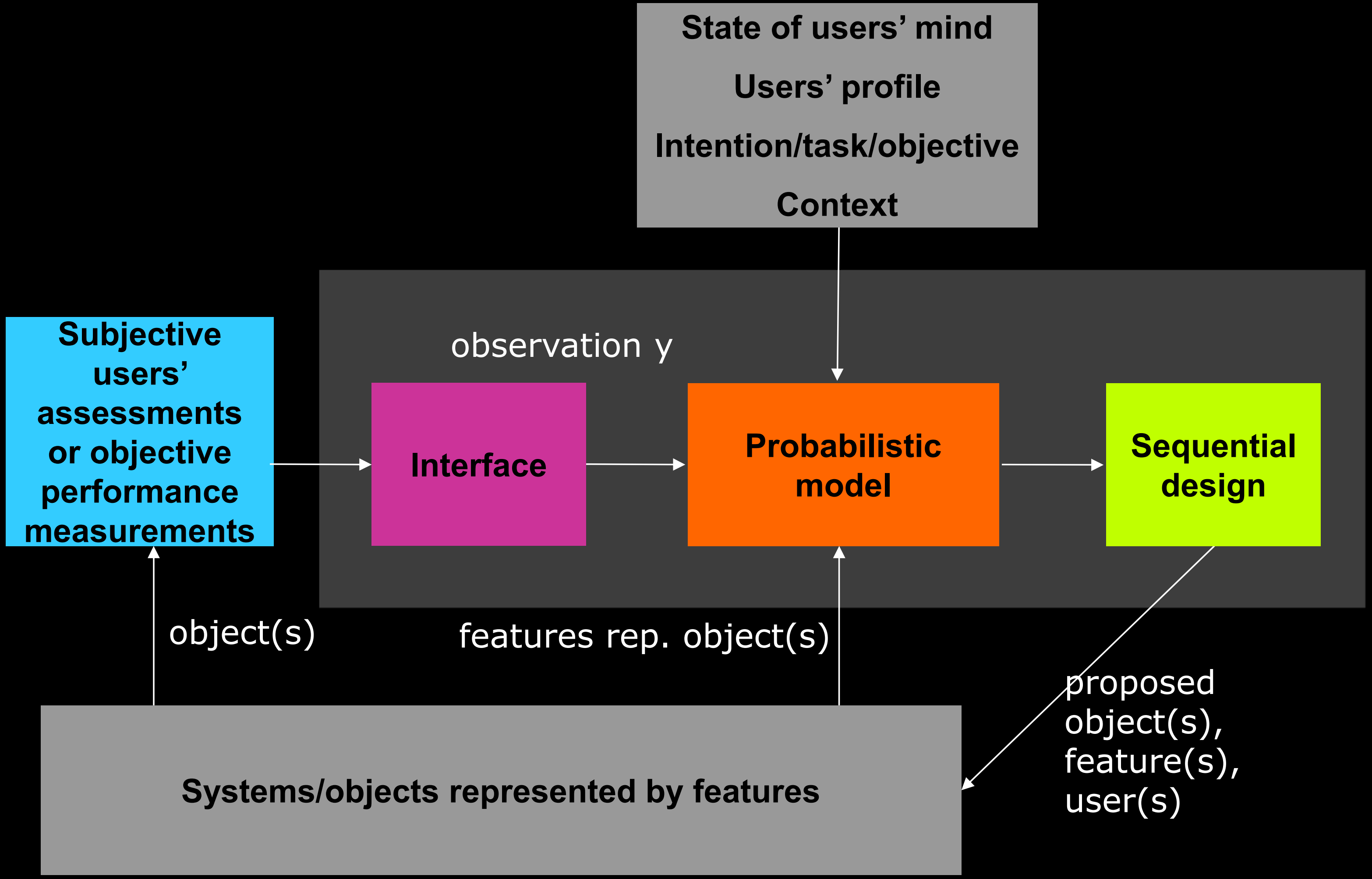
Optimization objective

Learning and identifying optimum
The Expected Improvement of a new candidate sample (green points) is derived from the predictive distribution.

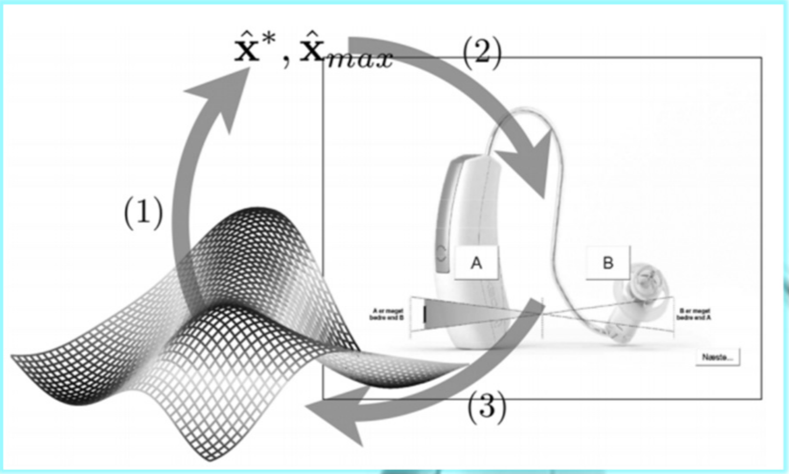


Which of the four green parameters settings/products/interface, x, should the user assess (rate/annotate/see/hear), or where do we need tp evaluate objective performance measurements

General framework



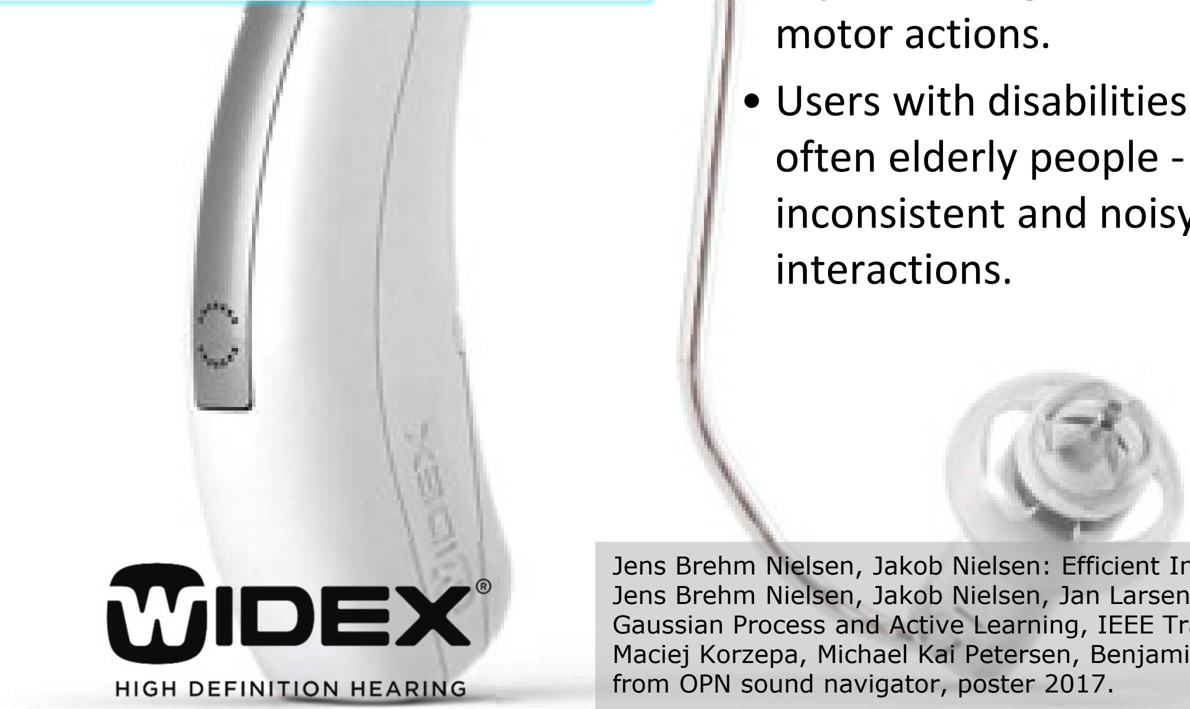
Optimization of hearing aids using Bayesian optimization



- Highly personalization needs.
- Dynamic environment and use with different needs.
- Latent, convoluted object functions which are difficult to express though verbal and motor actions.
- Users with disabilities – and often elderly people - provide inconsistent and noisy interactions.

Opn

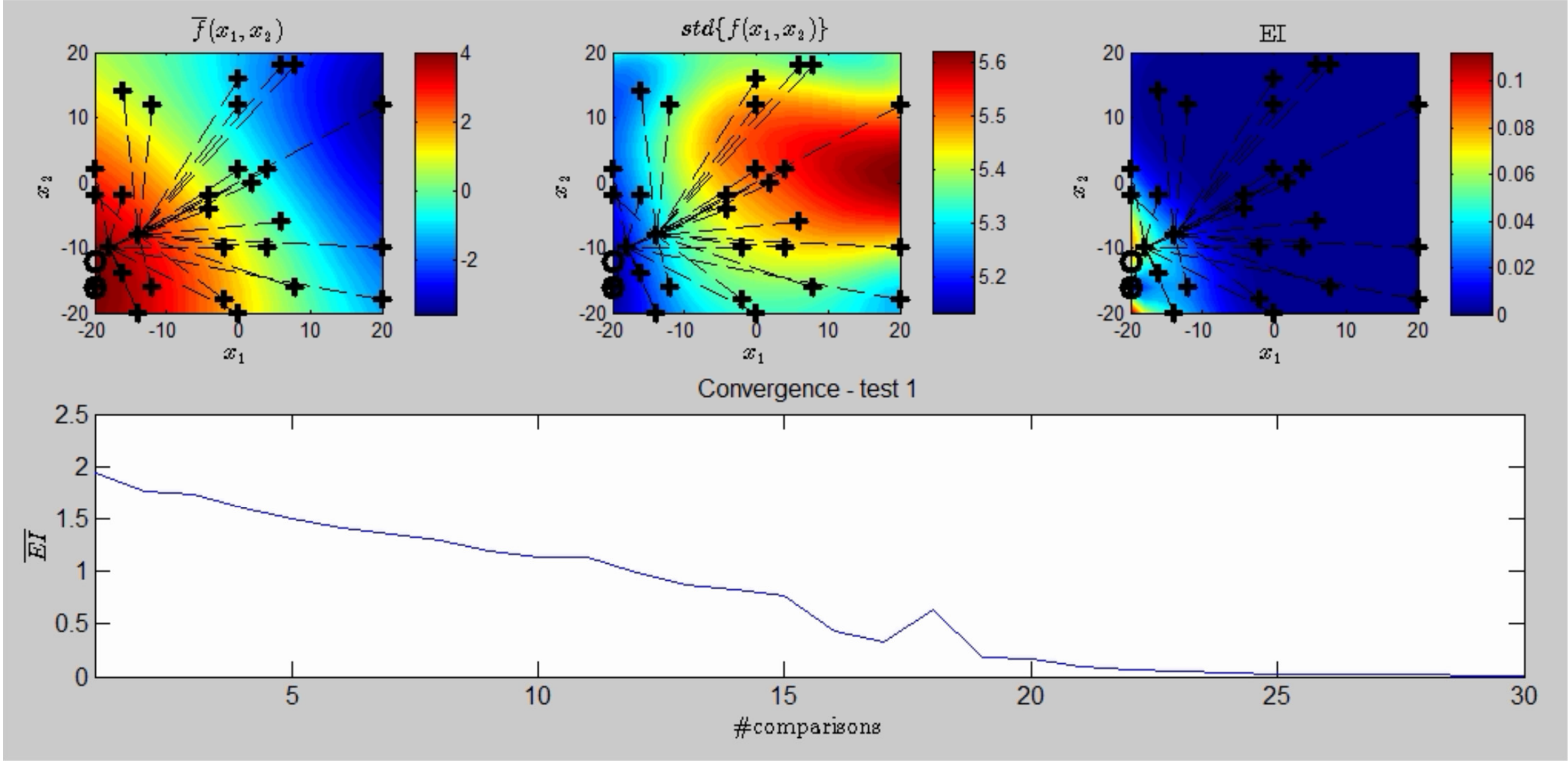
oticon
PEOPLE FIRST



WIDEX®
HIGH DEFINITION HEARING

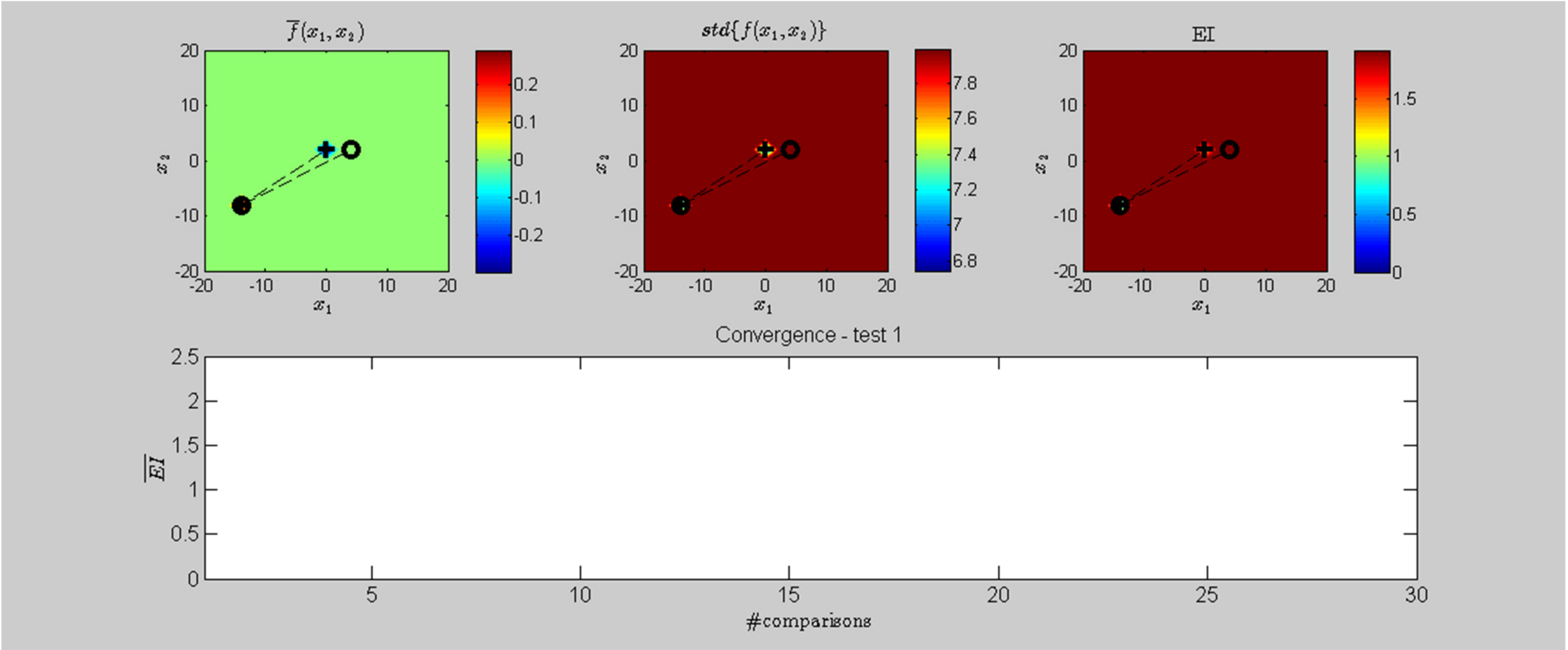
Jens Brehm Nielsen, Jakob Nielsen: Efficient Individualization of Hearing and Processers Sound, ICASSP2013.
Jens Brehm Nielsen, Jakob Nielsen, Jan Larsen: Perception based Personalization of Hearing Aids using Gaussian Process and Active Learning, IEEE Trans. ASLP, vol. 23, no. 1, pp. 162 – 173, Jan 2015.
Maciej Korzepa, Michael Kai Petersen, Benjamin Johansen, Jan Larsen, Jakob Eg Larsen: Learning soundscapes from OPN sound navigator, poster 2017.

Pairwise (2AFC) personalization of HA

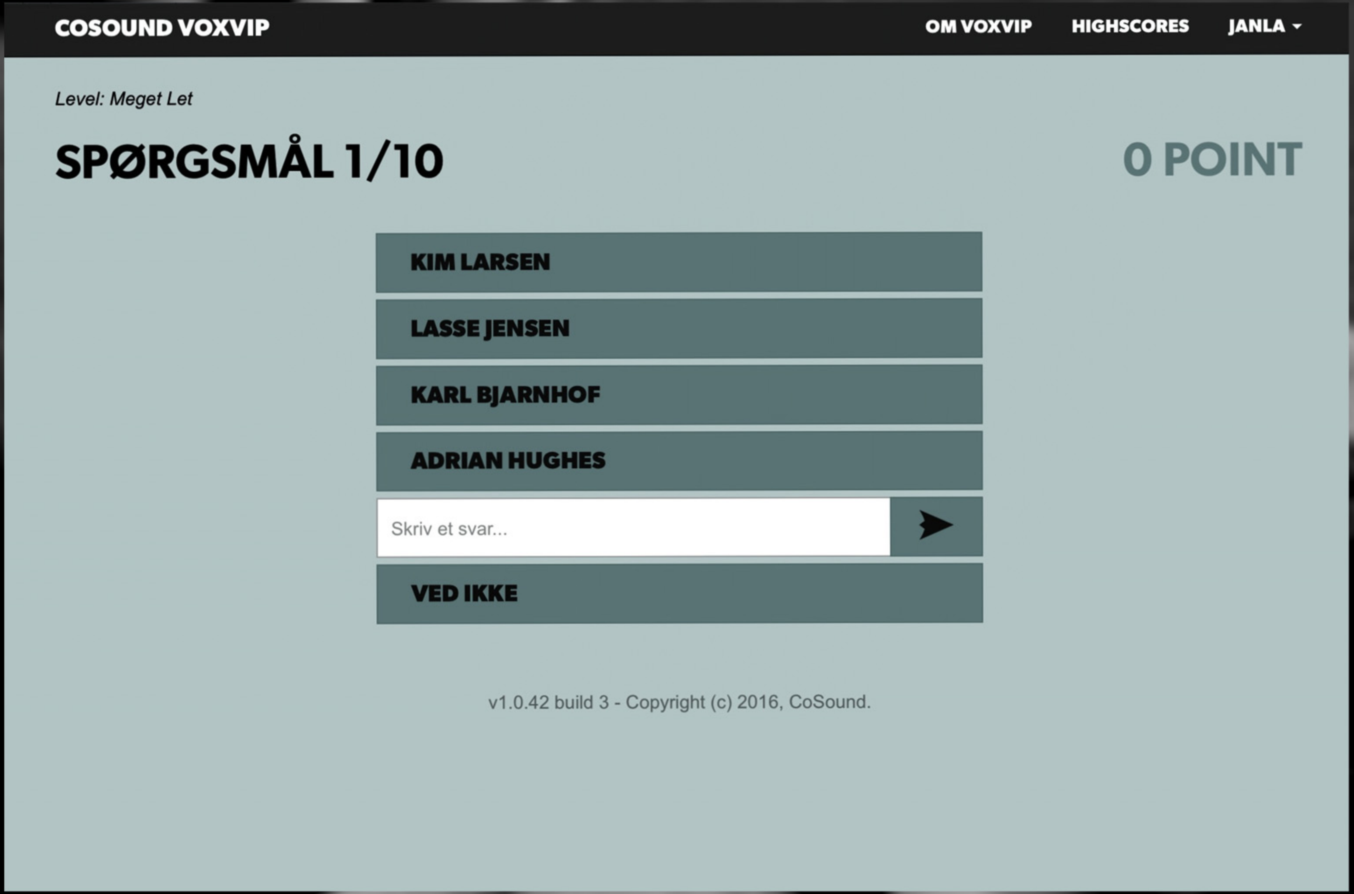


Hearing Aids Personalization

A real interactive optimization sequence in 30 iterations



VOXVIP - smart crowdsourcing of the DR radio archive



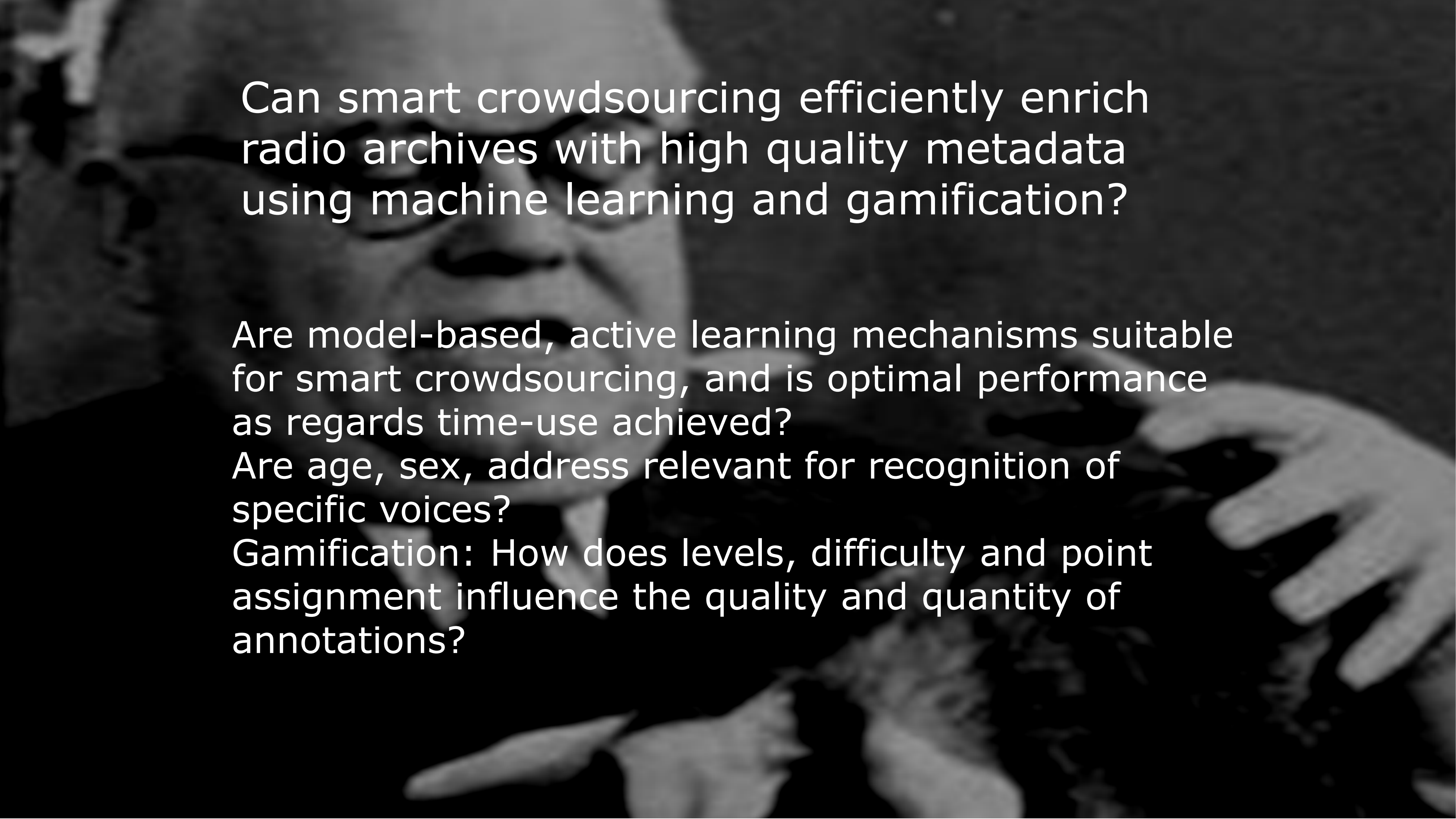
Level: Meget Let

SPØRGSMÅL 1/10

0 POINT

v1.0.42 build 3 - Copyright (c) 2016, CoSound.

voxvip.cosound.dk



Can smart crowdsourcing efficiently enrich radio archives with high quality metadata using machine learning and gamification?

Are model-based, active learning mechanisms suitable for smart crowdsourcing, and is optimal performance as regards time-use achieved?

Are age, sex, address relevant for recognition of specific voices?

Gamification: How does levels, difficulty and point assignment influence the quality and quantity of annotations?

What is meta information?

Infinite number of aspects provides information about the individual clip/object or similarity between such objects

Objective information

- Who is speaking
- What is the topic discussed?
- Which objects are present in the clip?

Subjective information

- Which emotions are expressed in the clip?
- What is the sound quality?
- Which clip is preferred?

How can meta information be created?

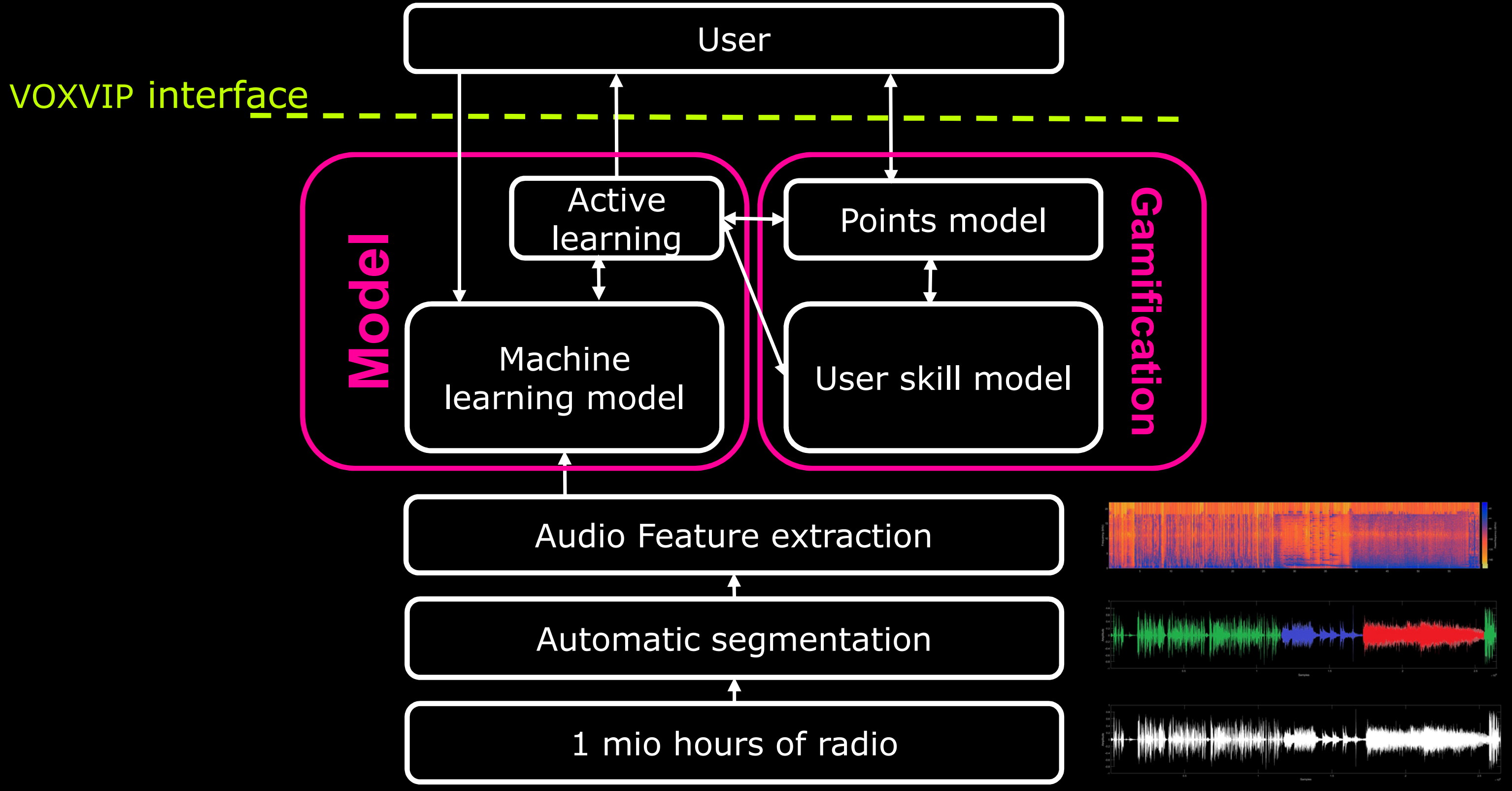
Lack of specific annotations requires prior knowledge

Manual annotation is limited or impossible due to the size of the archive, human resources, or annotators qualifications.

Semi-automatic machine learning can be used to predict information in the enture archive based on limited number of annotations.

Smart crowdsourcing exploits machine learning to predict information in the entire archive based on ‘crowd annotators’ annotations. The individual clip is selected based on uncertain information about the label, the annotators’ qualifications and engagement based on active learning mechanisms.

VOXVIP model



User

VOXVIP interface

Model

Active learning

Machine learning model

Gamification

Points model

User skill model

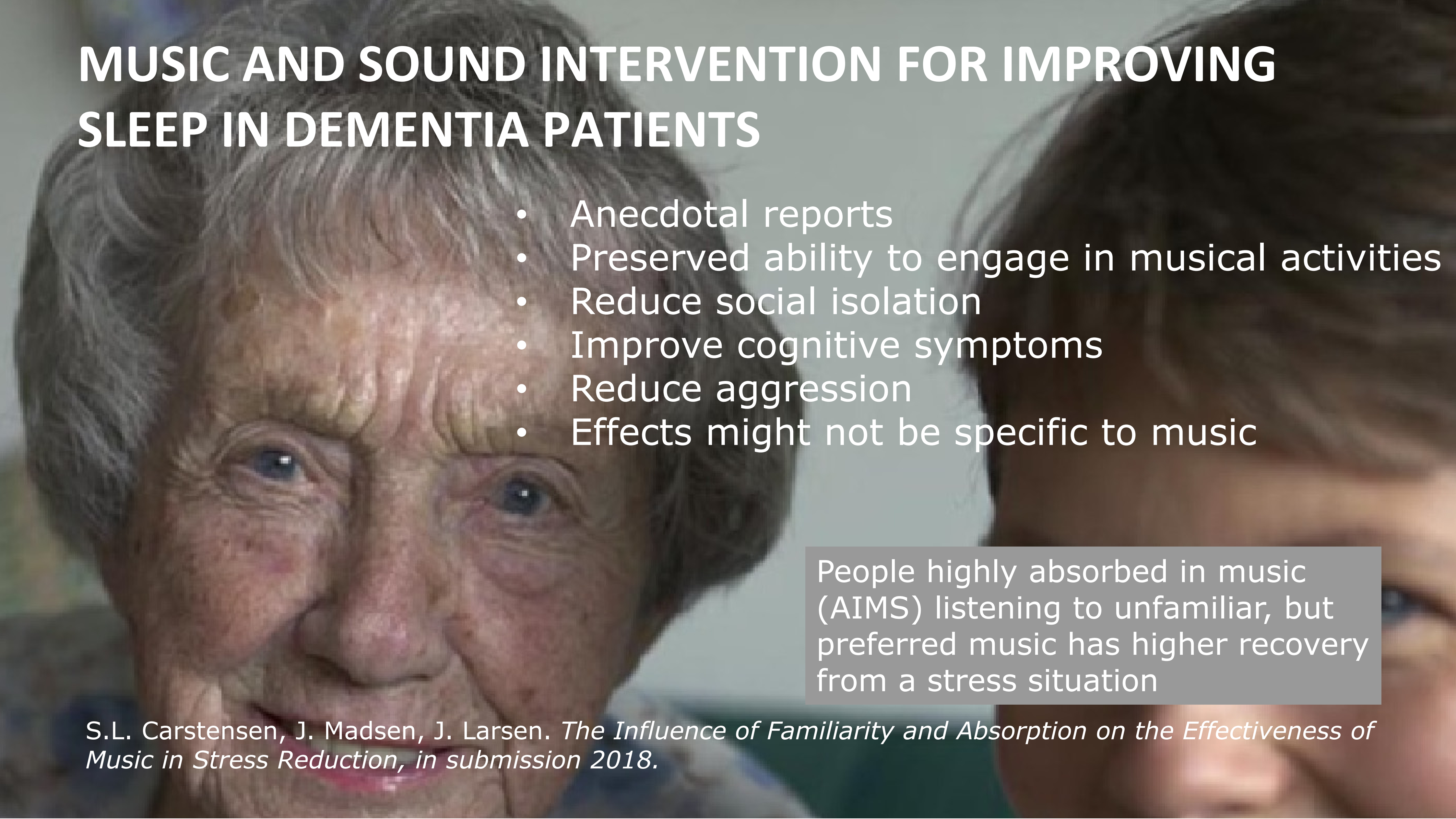
Audio Feature extraction

Automatic segmentation

1 mio hours of radio

ALIVE INSIDE

by Michael Rossato Bennett, 2014
www.youtube.com/watch?v=5FWn4JB2YLU



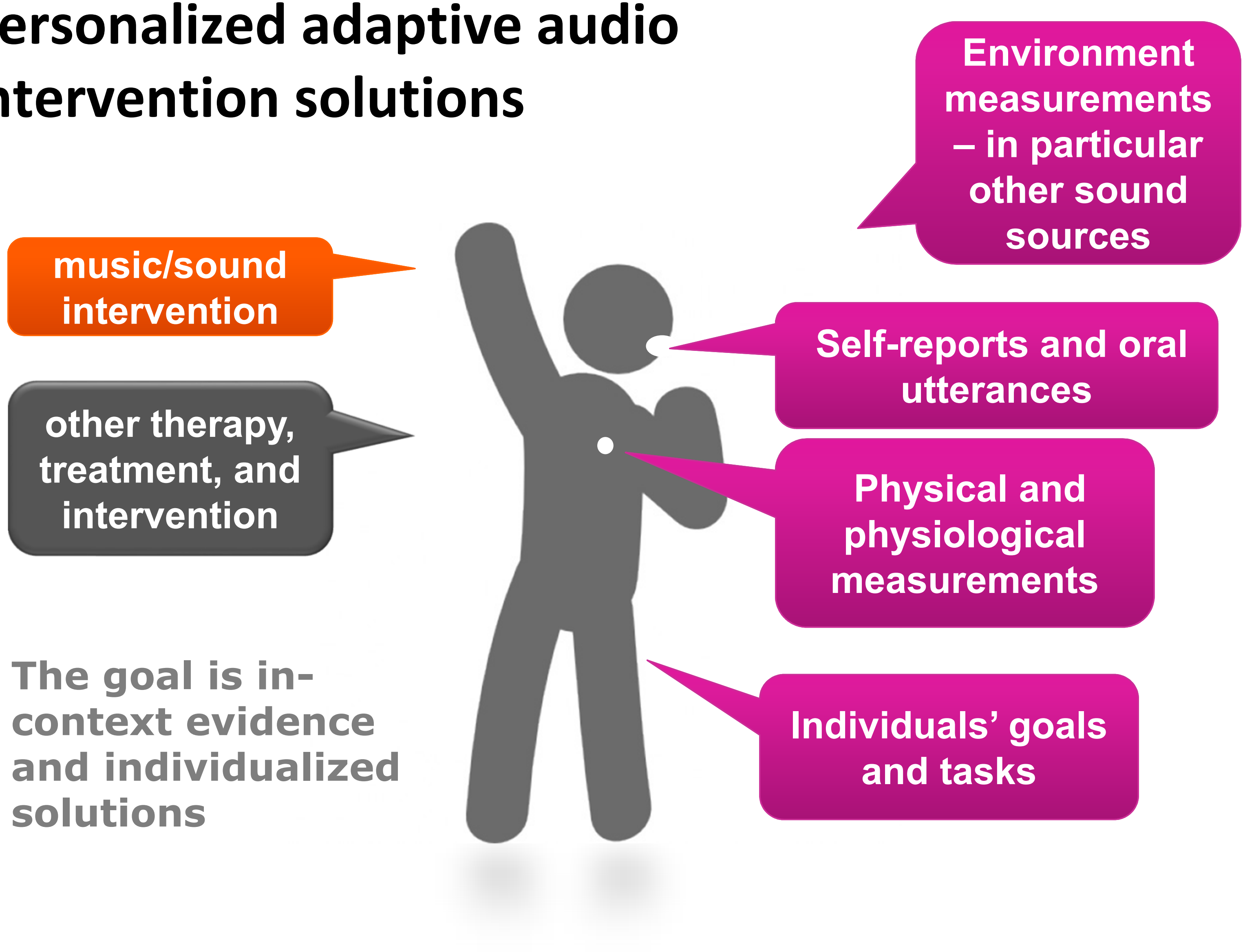
MUSIC AND SOUND INTERVENTION FOR IMPROVING SLEEP IN DEMENTIA PATIENTS

- Anecdotal reports
- Preserved ability to engage in musical activities
- Reduce social isolation
- Improve cognitive symptoms
- Reduce aggression
- Effects might not be specific to music

People highly absorbed in music (AIMS) listening to unfamiliar, but preferred music has higher recovery from a stress situation

S.L. Carstensen, J. Madsen, J. Larsen. *The Influence of Familiarity and Absorption on the Effectiveness of Music in Stress Reduction, in submission 2018.*

Personalized adaptive audio intervention solutions



music/sound intervention

other therapy, treatment, and intervention

Environment measurements – in particular other sound sources

Self-reports and oral utterances

Physical and physiological measurements

Individuals’ goals and tasks

The goal is in-context evidence and individualized solutions

FUTURE

Cognizant audio systems

fully informed and aware systems

Context:

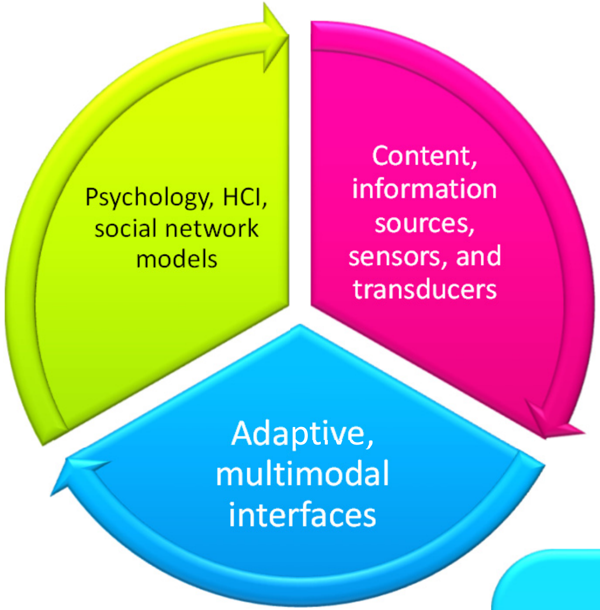
who, where, what

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

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


Psychology, HCI, social network models
Content, information sources, sensors, and transducers
Adaptive, multimodal interfaces

Flexible integration with other media modalities

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

TRANSFORMATION

Digital
Interactive - humans in the loop
Human centric
Cognitive
AI drives IA
Multimodal - IoT and hearables

THE WAYS AHEAD

- Need for possibility to include co-creation and production.
- Need for more data across domains and situations.
- Need for systems and platforms that enables experimentation and direct user interaction.
- Need for better AI and machine learning methodology that can provides robust, interpretable, interactive learning from few examples.