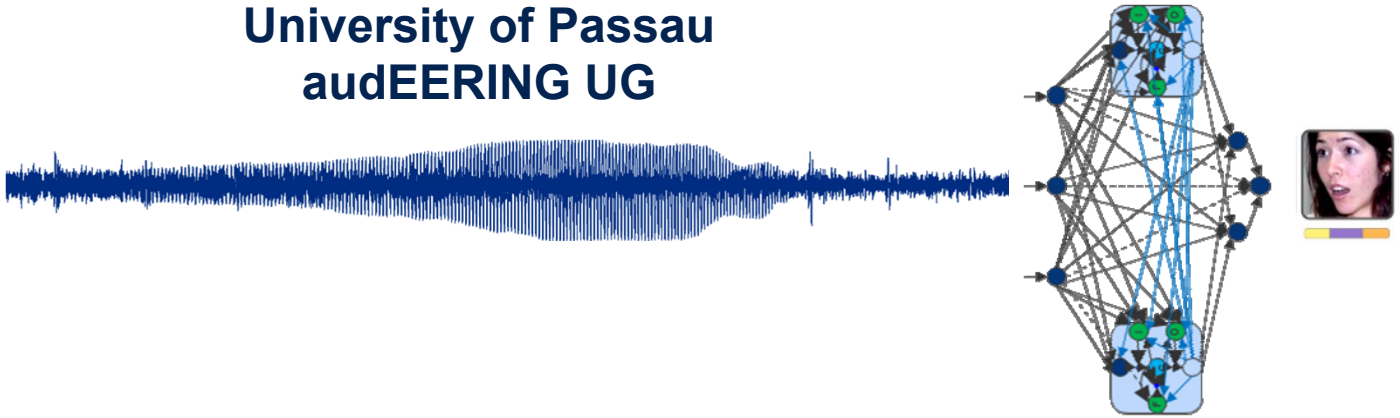
Machines' Learning on Social Behaviour Sensing

Björn W. Schuller
Machine Learning Group
Imperial College London
Chair of Complex & Intelligent Systems
University of Passau
audEERING UG



Speaker.

iHEARu

European Research Council

		# Classes	%UA/*AUC/⁺CC
2015	Nativeness	[0,1]	43.3⁺
	Parkinson's	[0,100]	54.0⁺
	Eating	7	62.7
2014	Cognitive Load	3	61.6
	Physical Load	2	71.9
2013	Social Signals	2x2	92.7*
	Conflict	2	85.9
	Emotion	12	46.1
	Autism	4	69.4
2012	Personality	5x2	70.4
	Likability	2	68.7
	Intelligibility	2	76.8
2011	Intoxication	2	72.2
	Sleepiness	2	72.5
2010	Age	4	53.6
	Gender	3	85.7
	Interest	[-1,1]	42.8⁺
2009	Emotion	5	44.0
	Negativity	2	71.2

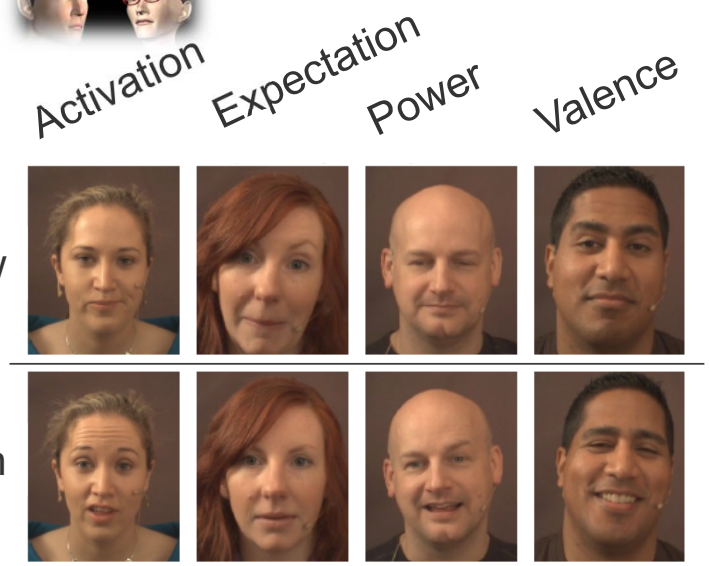
"The Computational Paralinguistics Challenge", ***IEEE Signal Processing Magazine***, *29(4): 2-6, 2012.*

Affect.

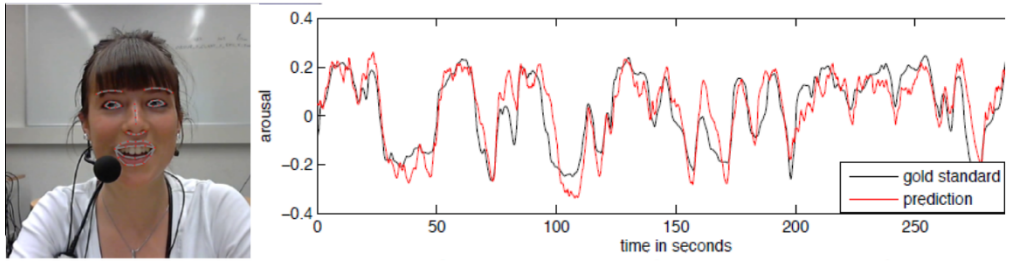
SEMAINE

ARIA VALUSPA

- **AVEC 2011 & 2012: Emotion**
  AVEC database: 50k Turns
- **AVEC 2013 & 2014: Depression**
  AVDLC database: 240 h, ~300 subjs.
  Beck Depression Inventory: 0 – 63
- **AV+EC 2015: Emotion**
  RECOLA database: 6 h, ~30 subjs.



	Activation	Expectation	Power	Valence
%Acc.	71.2	68.6	59.6	70.2
MAErr	6.3			
%CC	80.4			52.8



"LSTM-Modeling of Continuous Emotions in an Audiovisual Affect Recognition Framework", ***Image and Vision Computing****, 31(2): 153-163, 2013.* ***(best result AVEC Challenge)***

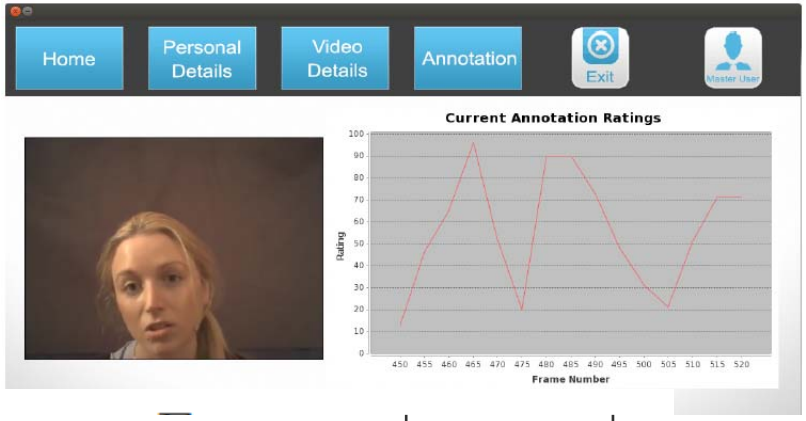
"AVEC 2014 - The Continuous Audio/Visual Emotion and Depression Recognition Challenge", ***ACM Multimedia****, 2014.*

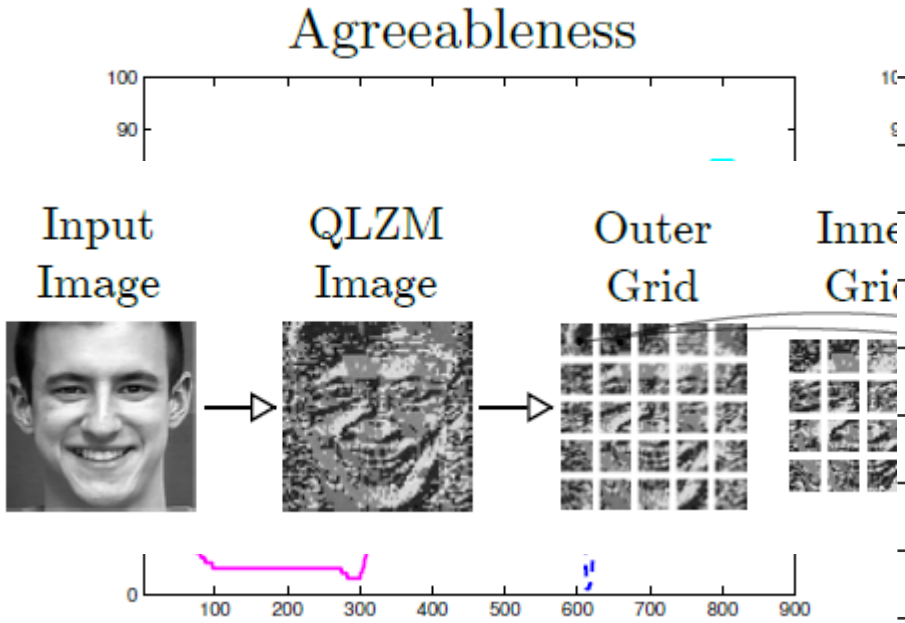
"Prediction of Asynchronous Dimensional Emotion Ratings from Audiovisual and Physiological Data", ***Pattern Recognition Letters****, 2014.*

Personality.

- **MAPTRAITS: Perceived Perso.**

  SEMAINE:
  44 clips, 11 subjects, 6 raters

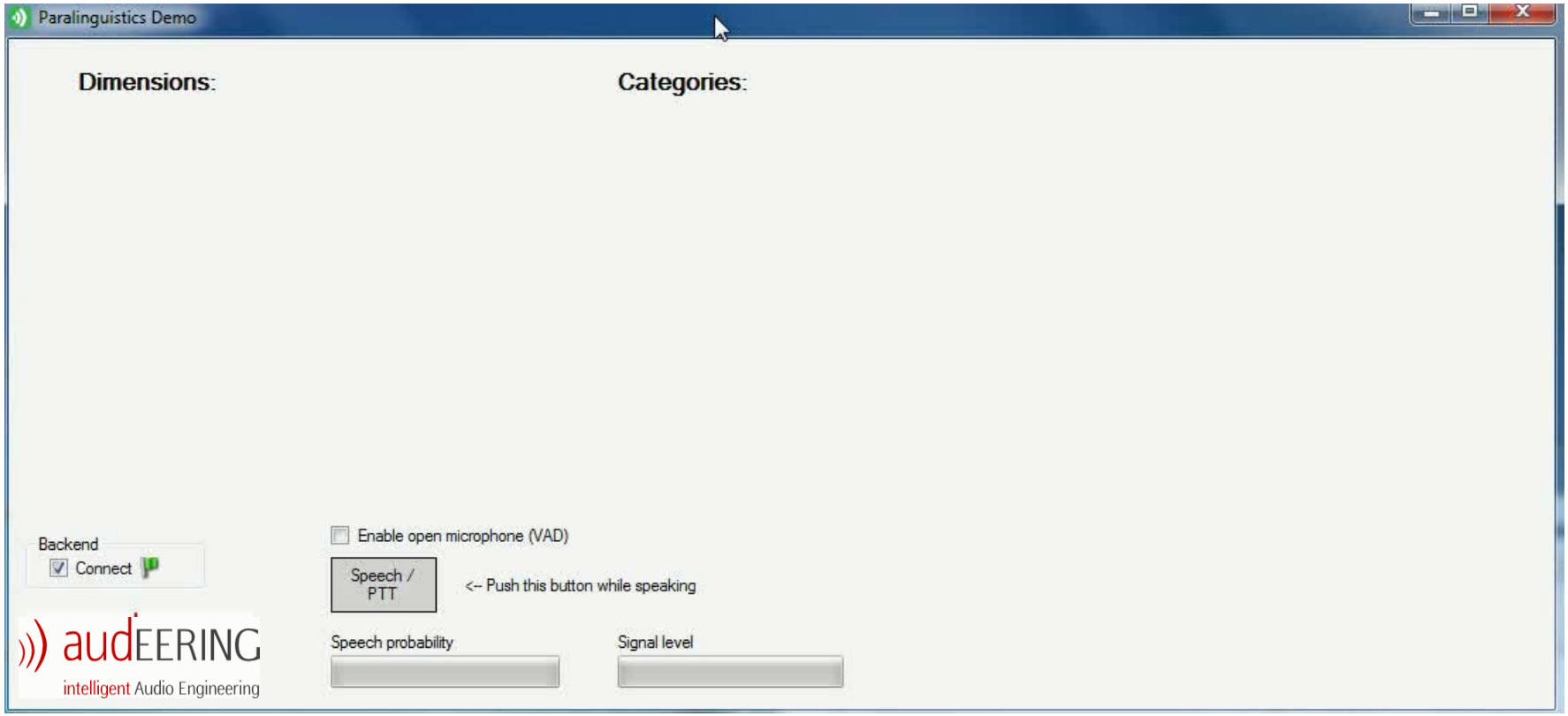




MSE	Audio	Video
Openness	.53	.39
Conscientious.	.51	.28
Extroversion	.66	.38
Agreeableness	.50	.41
Neuroticism	.77	.35
Engagement	.65	.45
Likeability	.56	.37
Facial Attractiv.	-	.34
Vocal Attractiv.	.41	-

"MAPTRAITS 2014: The First Audio/Visual Mapping Personality Traits Challenge",
***ACM ICMI**, 2014.*

Working, Yet?


Paralinguistics Demo
Dimensions:
Categories:
Backend
Connect
Enable open microphone (VAD)
Speech / PTT
<-- Push this button while speaking
Speech probability
Signal level
audEERING
intelligent Audio Engineering

Sentiment.

"You
IEE

B. S
***WIREs Data Mining and Knowledge Discovery**, 2015.*

Eye-Contact.

*"The acoustics of eye contact - Detecting visual attention from conversational audio cues", **ACM Gaze-In**, 2013.*

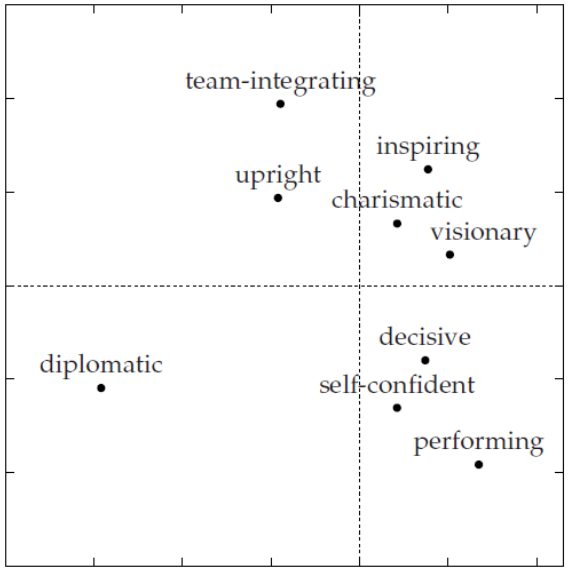
- **Eye Contact from Acoustics**

  GRAS² Corpus: 28 + 4 subjects

UA [%] / AUC	
Eye-Contact	67.4 / .732

Social Traits.

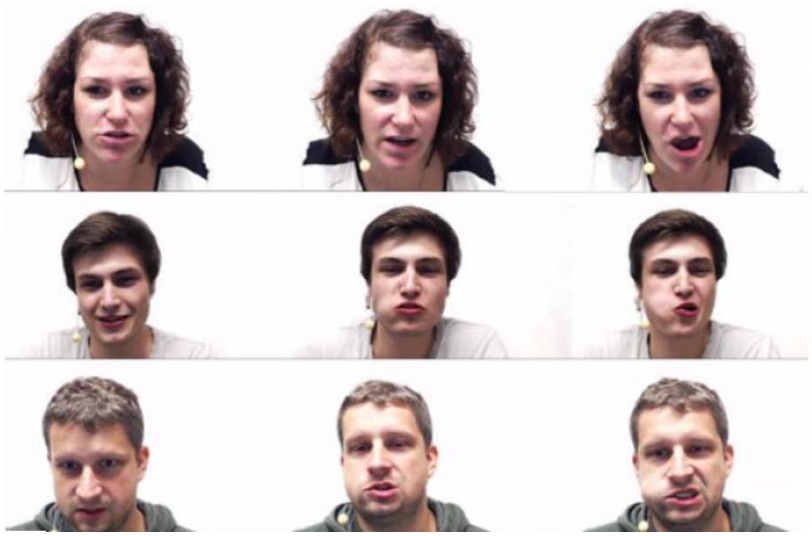
- **Perceived Leadership Traits**
  409 recordings, 10 raters
  Acoustic & Linguistic analysis



UA [%]	
Non-Malicious	61.3
Diplomatic	58.5
Achiever	72.5
Charismatic	63.2
Teamplayer	65.2

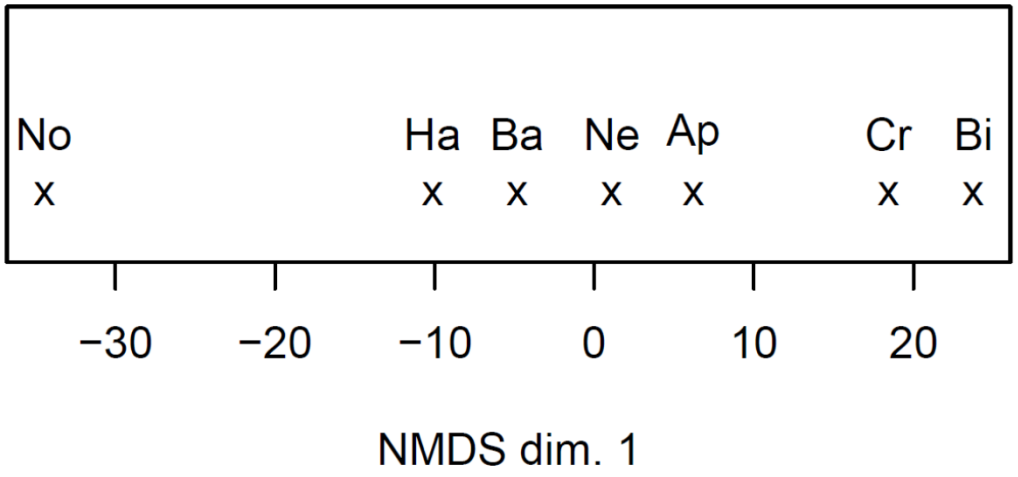
"The Voice of Leadership: Models and Performances of Automatic Analysis in On-Line Speeches",
***IEEE Transactions on Affective Computing** 3(4): 496-508, 2013.*

Eating.



- **Speech Under Eating & Food**

  30 subjects, 6 food types, +ASR features



R^2	
Crispness	.562

*"The Interspeech 2015 Computational Paralinguistics Challenge: Nativeness, Parkinson's & Eating Condition", **Interspeech**, 2015.*

Action Units.

ARIA VALUSPA

HORIZON 2020

- **Face Reading from Speech**

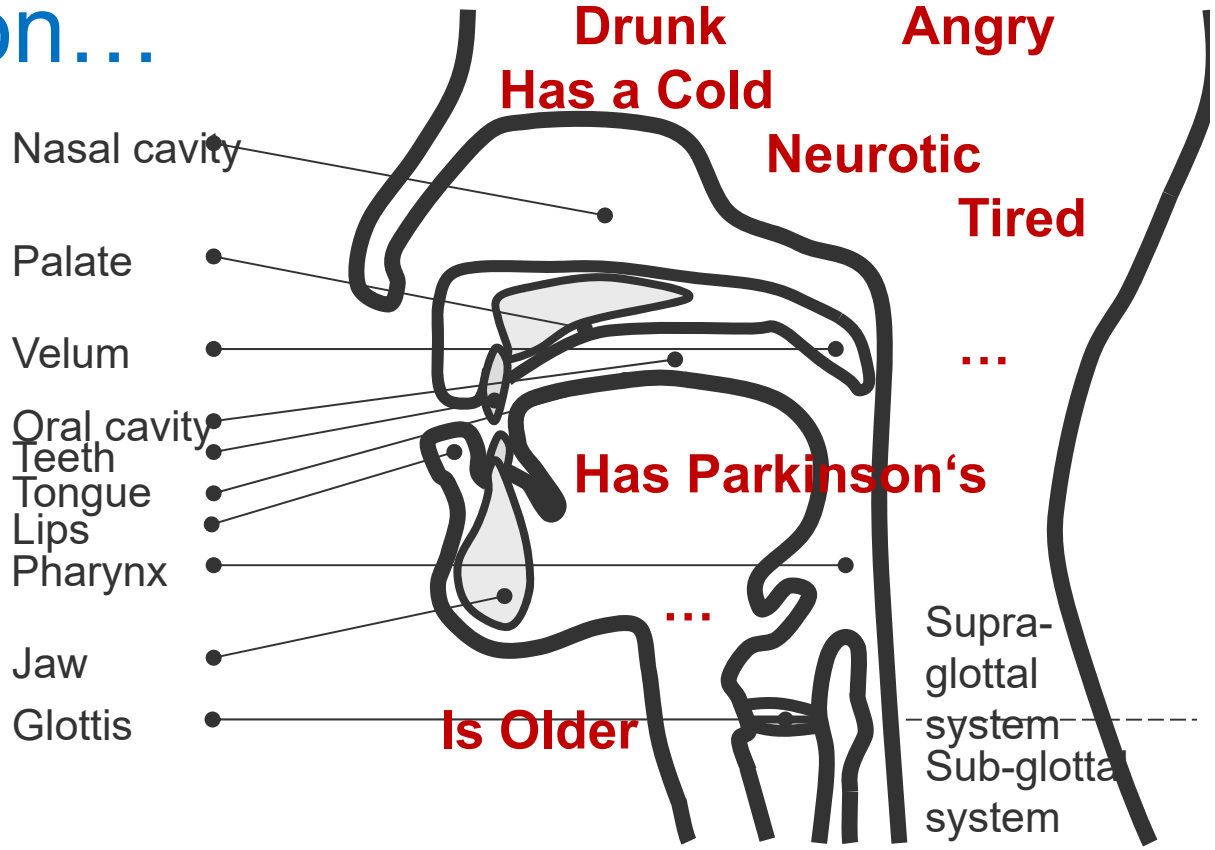
  GEMEP corpus, SVM vs DLSTMNNs, big (65) vs small (28) feature set

	ComPare				GeMAPS			
AU	**Onset**	**Apex**	**Offset**	**Occu.**	**Onset**	**Apex**	**Offset**	**Occu.**
1	61.88	64.54	59.48	67.03	62.08	**67.96**	61.18	67.62
2	64.38	67.43	62.27	69.33	64.60	**70.99**	62.98	70.90
4	61.35	66.21	63.33	64.66	61.90	66.67	63.19	**67.05**
6	64.01	65.17	63.35	63.78	64.25	**67.71**	63.17	63.11
7	62.71	56.62	58.29	54.47	**62.85**	59.74	61.46	52.72
10	61.82	60.53	56.39	60.39	**61.92**	61.03	58.02	60.34
12	59.34	59.85	56.71	58.67	58.04	**60.92**	57.77	58.72
17	60.47	64.78	62.34	64.53	61.62	64.94	63.13	**65.88**
Avg.	62.00	63.14	60.27	62.86	62.16	**65.00**	61.36	63.29

"Face Reading from Speech – Predicting Facial Action Units from Audio Cues", ***Interspeech***, *2015.*

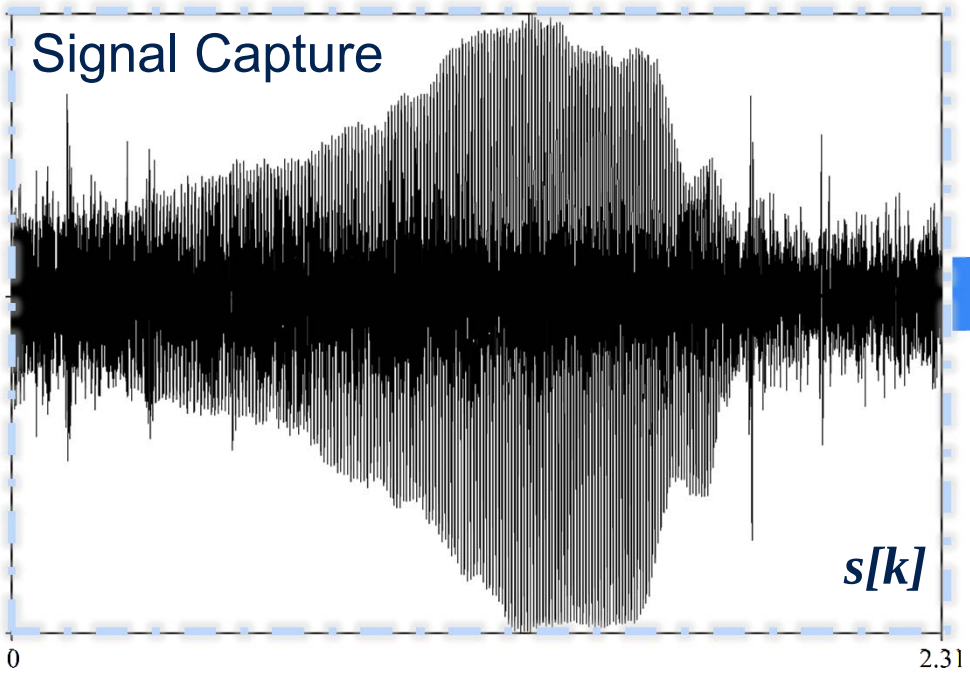
The Production…

- **Multiple-Targets**
  There is just one Vocal Production Mechanism…



% UA	Single	Multiple
Likability	59.1	(+A,G,CI) **62.2**
Neuroticism	62.9	(+G,OCEA, CI) **67.5**

Principle.



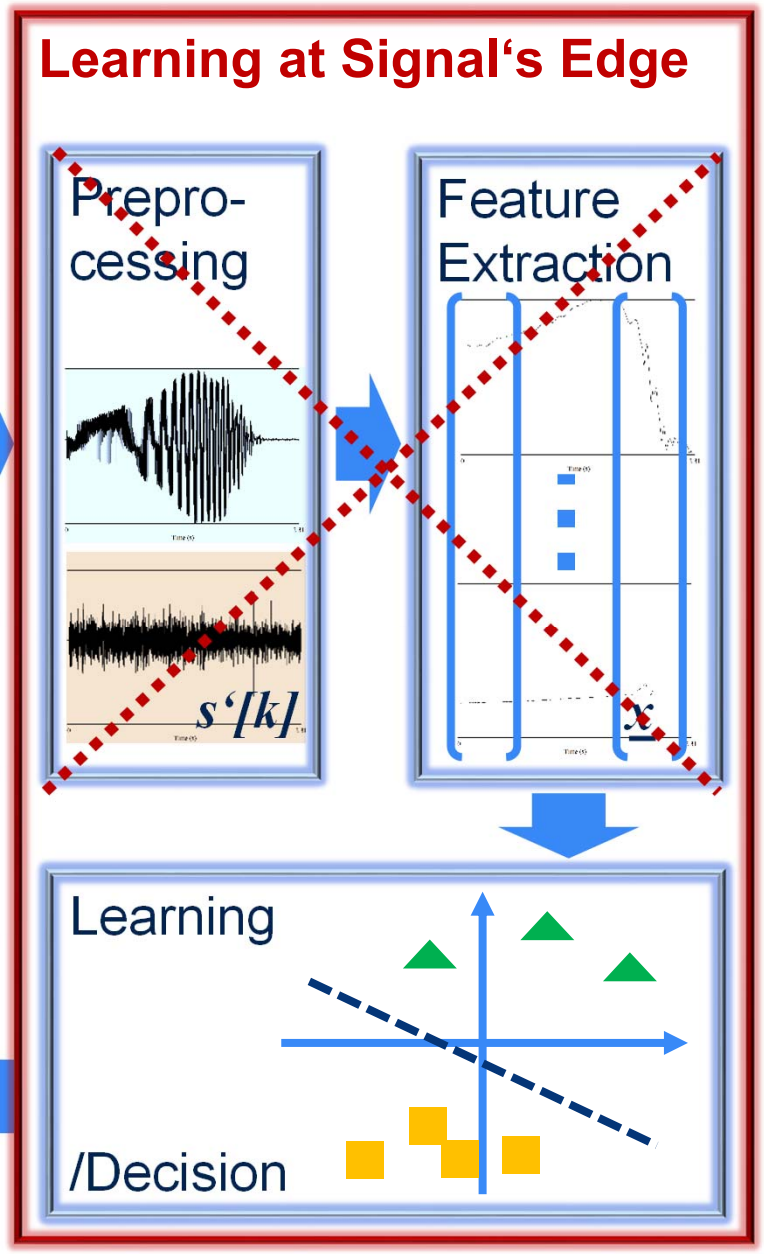
"Big Data vs. Little Labels"

Efficient Weakly Supervised

<label>

(e.g., "burst")

y

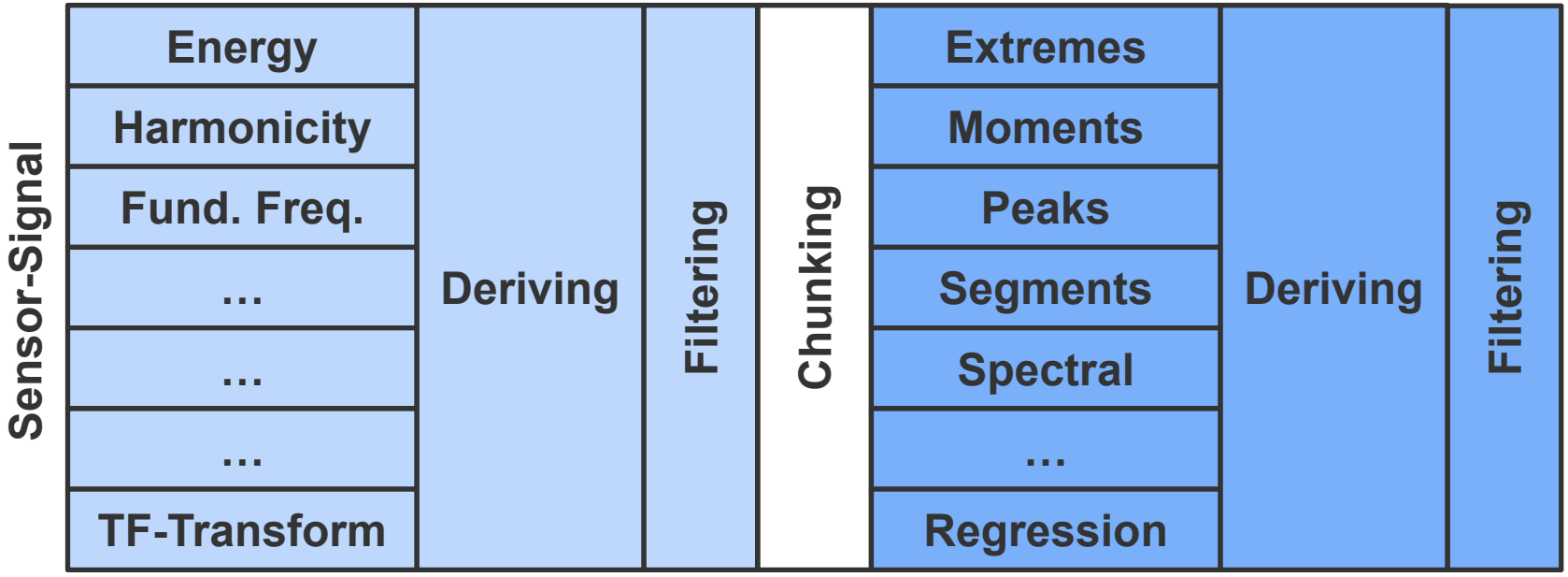


Feature Extraction.

openSMILE:) by audEERING™

- **openSMILE**
  Brute-force High-Dim. Spaces
  (Android / C++)
  Online update

#features	RTF
10k	2%
500k	**3%**



"Recent Developments in openSMILE, the Open-Source Multimedia Feature Extractor", **ACM Multimedia**, 2013.
(2nd place ACM MM Open Source Software Competition in 2010 and 2013, ~600 citations for 3 papers)

Less is More?

- **Geneva Minimalistic Set**

  GeMAPS: 18 LLDs / 62 functionals

  Pitch, Jitter, Formant 1-3, Formant 1 bandwidth

  Shimmer, Loudness, HNR

  Alpha Ratio, Hammarberg Index, Spectral Slope 0–500Hz and 500–1500 Hz, Formant 1-3 relative energy, Harmonic difference H1–H2, Harmonic difference H1–A3

  Extended: +7 LLDs / 88 functionals

  MFCC 1–4, Spectral Flux, Formant 2–3 bandwidth

"The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing", ***IEEE Transactions on Affective Computing****, 2015.*

Bag-of-X-Words

LLDs	Functionals
ZCR, ΔZCR RMS energy, ΔRMS energy F0, ΔF0 HNR, ΔHNR MFCCs 1–12, ΔMFCCs 1–12	mean, standard deviation kurtosis, skewness, linear regression (offset, slope, MSE), extremes (values, relative position, range)

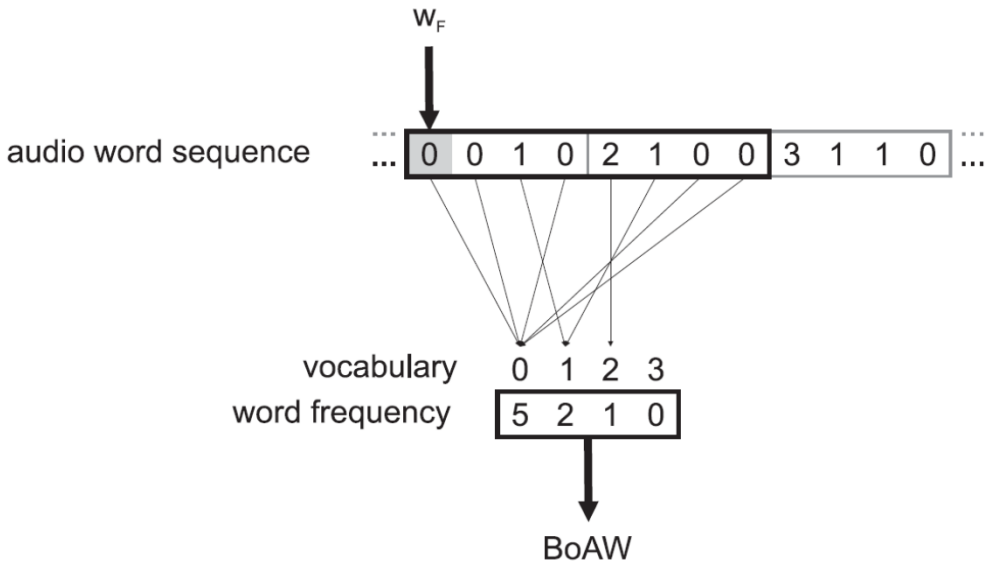
- **Audio Words**

  Base: UA/WA = 54.3/61.2%

  VQ vs SVQ

  VAM: Valence

Configuration			Recognition	
# SV	SVCS [bit]	CS [bit]	UA [%]	WA [%]
2	2	4	61.6	65.1
2	4	2	60.3	62.7
3	**2**	**5**	**64.2**	**65.6**
3	2	6	63.7	65.6
3	4	2	64.6	64.6
3	4	3	63.8	63.2
4	4	3	62.1	61.7
6	3	3	63.2	63.2
8	4	3	65.3	62.7
12	3	4	63.4	64.1



*"Detection of Negative Emotions in Speech Signals Using Bags-of-Audio-Words", **WASA**, 2015.*

Deep Learning.

$$\mathbf{y} = f_{DNN}(\mathbf{x}) = o\left(\mathbf{W}^K \sigma\left(\mathbf{W}^{K-1} \sigma(\ldots \mathbf{W}^1 \mathbf{x})\right)\right)$$

- **Deep Neural Networks**

  DNN = MLP…

  Training makes the difference:

  Build them layer-wise

  Less uninitialized parameters

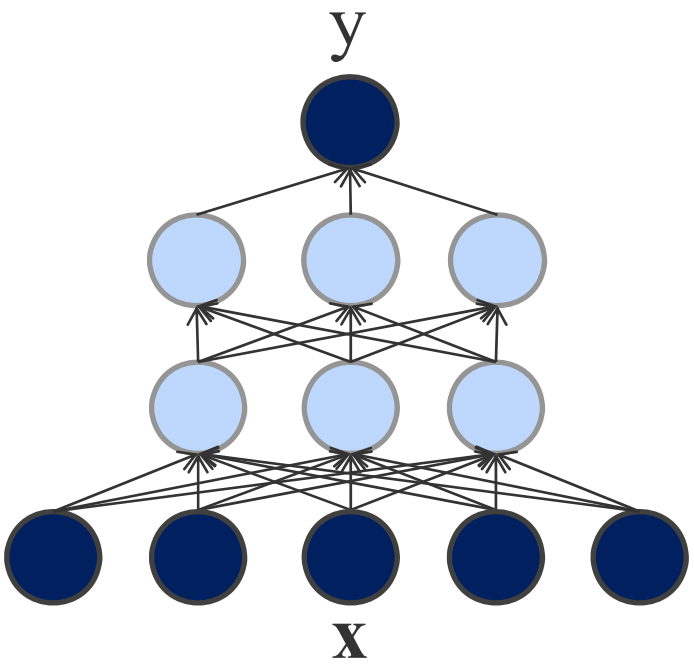
  Use “raw” features as input

  Net learns own higher-level features

  Use more training data

  On-line learning (stoch. gradient descent)

  Use ReLUs, drop-out learning, …



$$E_{Tr}(\mathbf{y}, \mathbf{y}^*) = \sum_{\mathbf{x} \in Tr} D(f_{DNN}(\mathbf{x}), \mathbf{y}^*)$$

$$w^{(i+1)} = w^{(i)} - \eta \frac{\partial E_B}{\partial w}(w^{(i)}),\ B \subset Tr$$

Deep LSTM Nets.

- **LSTM Cell**

Linear Unit
Auto weight 1
“Error Carousel”

Non-linear Gate
Input / Output / Forget

Multiplicative
Open / Shutdown

Example:ComParE:SSC

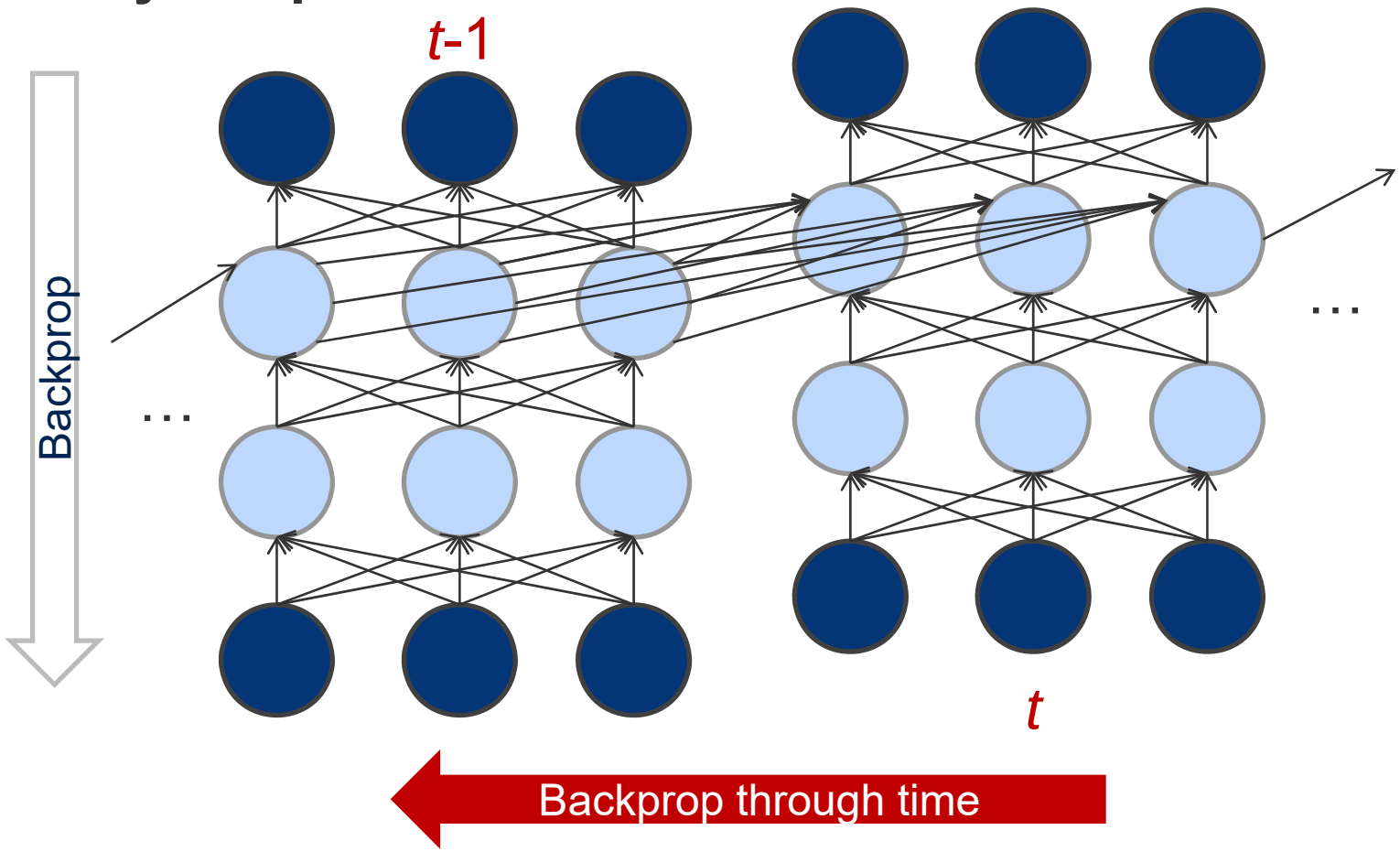
% AUC (Acc.)	Test
NN	83.3
Deep NN	92.4
LSTM	93.0
Deep LSTM	94.0

*“Social Signal Classification Using Deep BLSTM Recurrent Neural Networks”, **IEEE ICASSP**, 2014.*
(Best Result ComParE:SSC)

Deep Recurrent Nets.

$$\mathbf{y}_t = f_{RNN}(\mathbf{x}_0, \dots, \mathbf{x}_t)$$
$$= o\left(\mathbf{W}^o \sigma\left(\mathbf{W}\mathbf{x}_t + \mathbf{R}\,\sigma(\dots \mathbf{W}\mathbf{x}_0 + \mathbf{R}\mathbf{h}_0)\right)\right)$$

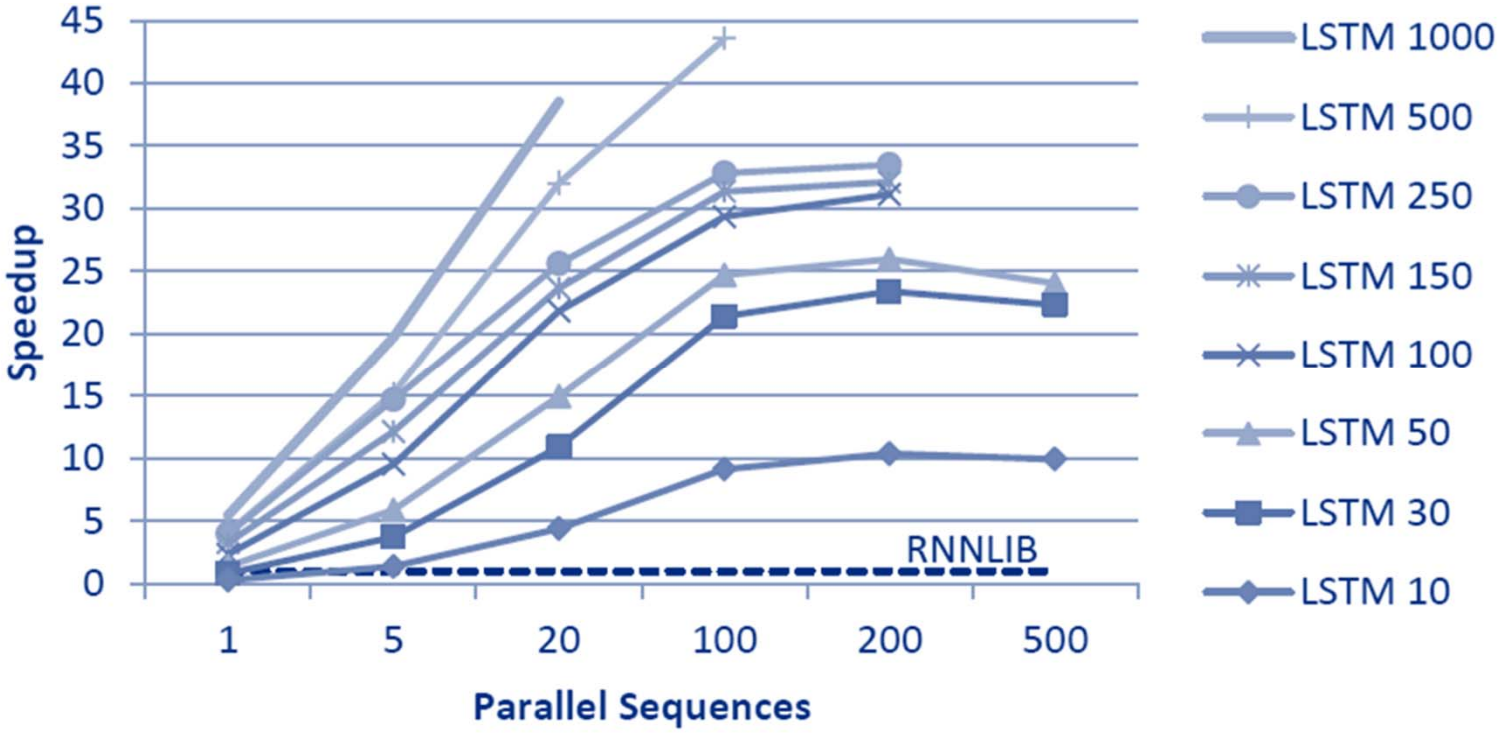
- **"Very Deep" NN**



Parallel Learning.

- **CURRENNT**

  10 – 1k LSTM cells, 2k – 4Mio parameters, Distribution

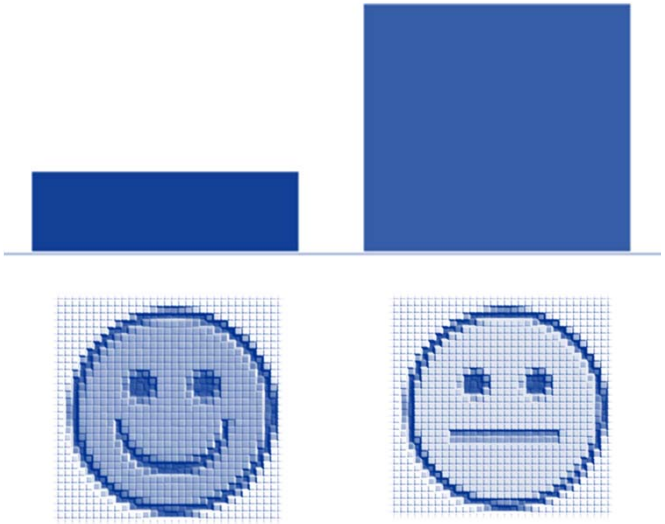


"Introducing CURRENNT - the Munich Open-Source CUDA RecurREnt Neural Network Toolkit",
***Journal of Machine Learning Research**, 2014.*

Weakly Supervised Learning.

- **Transfer Learning**
  Re-use data of related domain

- **Active Learning**
  Select “most informative” instances
  from large amount of unlabelled data
- **Semi-Supervised Learning**
  Have computer label the data
- **Cooperative Learning**
  Efficient combination of the above



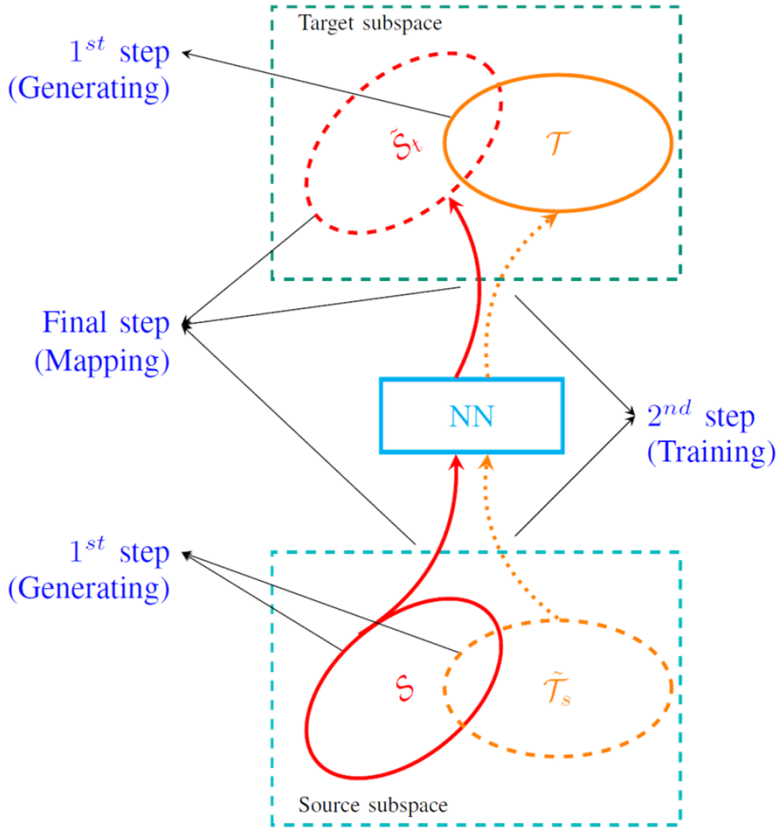
Transfer Learning.

- **Transfer Learning**

  Sparse Auto-Encoder

  target values $\overset{!}{=}$ input

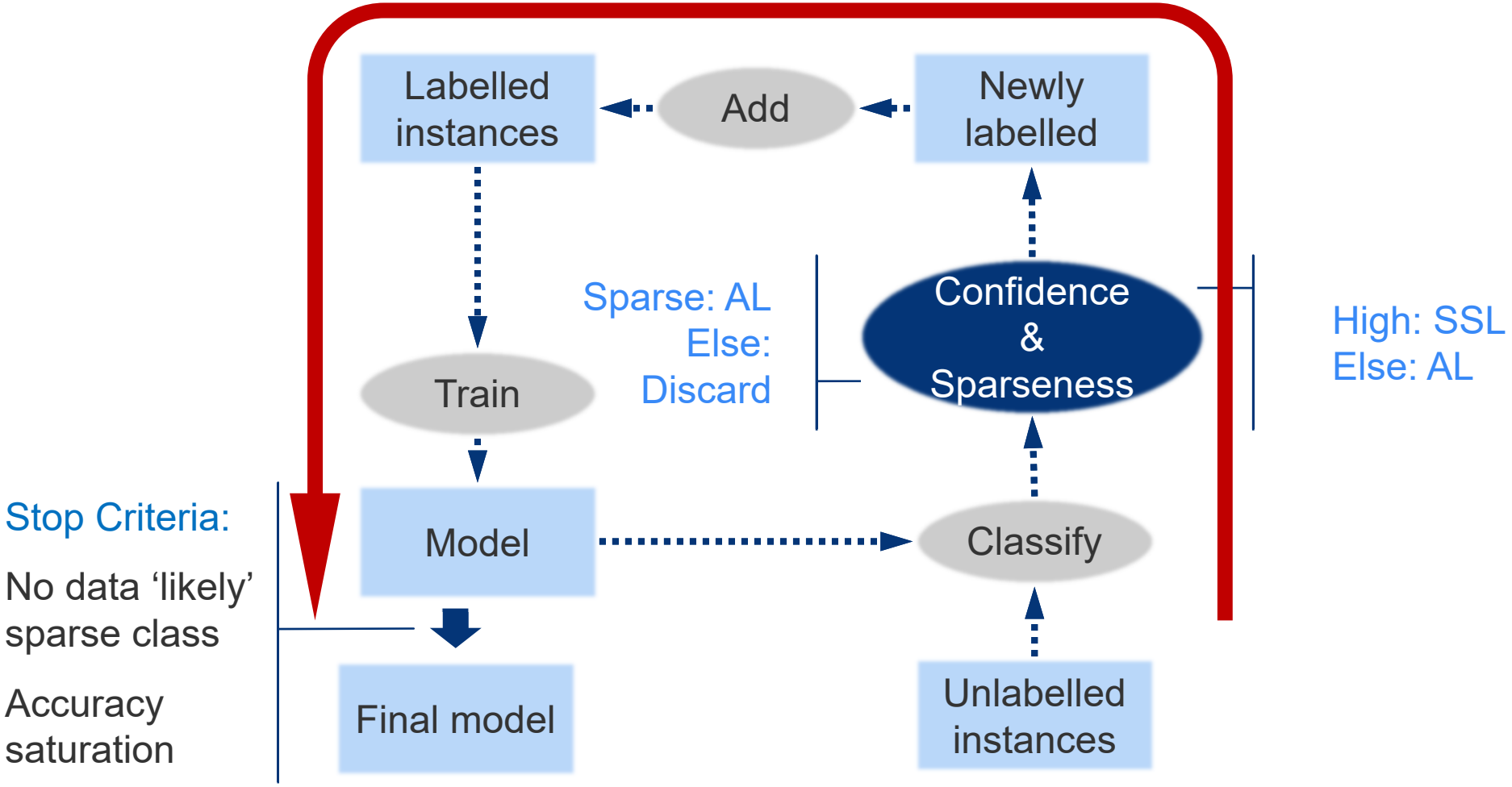
% UA / CC	Tar-get	w/o	DAE	DAE-NN
ComParE :EC	60.4	56.3	59.2	64.2
M → S: A	.82	.05	.32	-
M → S: V	.51	.06	.16	-



*"Linked Source and Target Domain Subspace Feature Transfer Learning", **ICPR**, 2014.*

*"Autoencoder-based Unsupervised Domain Adaptation for Speech Emotion Recognition", **IEEE Signal Processing Letters**, 2014.*

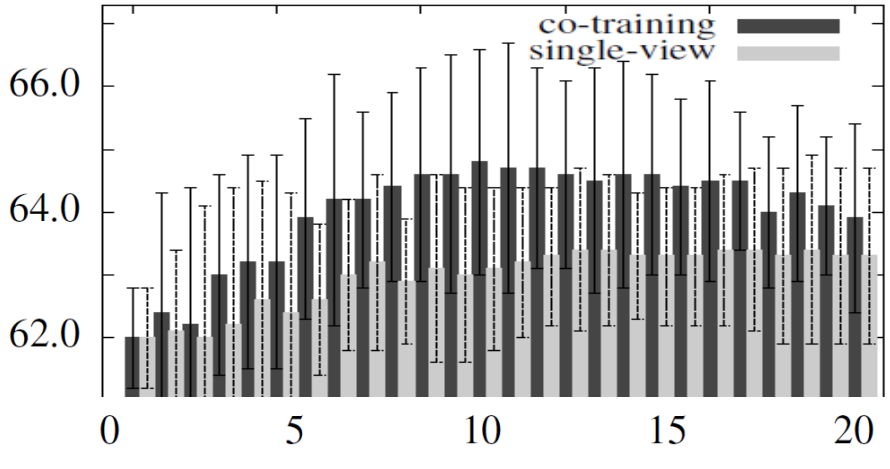
Cooperative Learning.



"Cooperative Learning and its Application to Emotion Recognition from Speech,"
***IEEE Transactions on Audio, Speech and Language Processing 23(1):115-126**, 2015.*

Cooperative Learning.

- **Example: ComParE:EC**
  1) Active Learning (AL)
     +
  2) Semi-Supervised
     =
  3) Cooperative Learning
     cross-view (xv)
     multi-view (mv)
  4) Dynamic
     further 79.2% reduction



Sign. levels	PL	AL	coAL	svCL	xvCL
PL					
AL	■				
coAL	■	/			
svCL	▒	/	■		
xvCL	■	■	■	■	
mvCL	■	▒	■	▒	\

p>.05 p<.05 p<.01 p<.001

"Cooperative Learning and its Application to Emotion Recognition from Speech,"
***IEEE Transactions on Audio, Speech and Language Processing 23(1):115-126**, 2015.*

iHEARu-PLAY

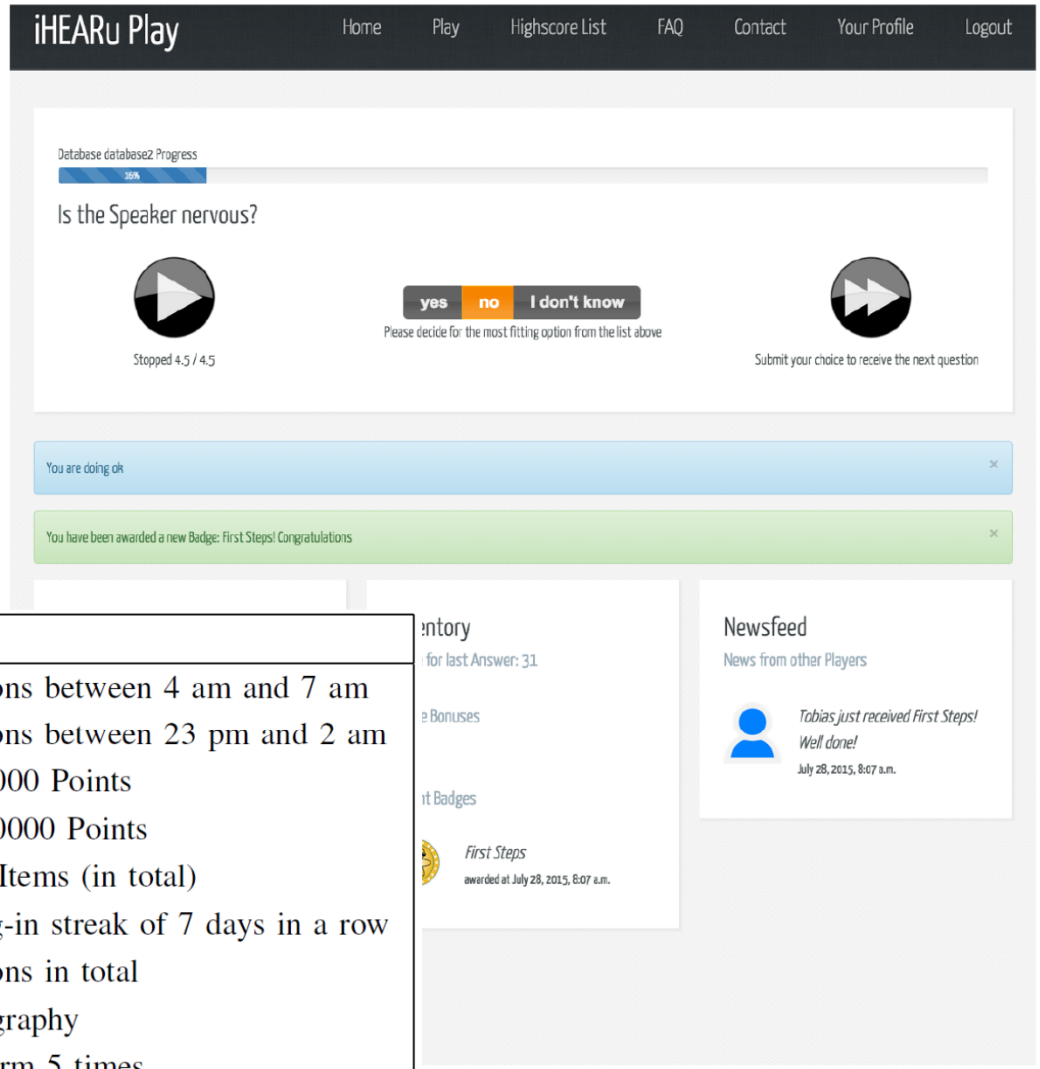
- **Playful Sourcing**

  iHEARu-PLAY:

  Gamified annotation

  Competing w/ others

  Gratification…



Badge Name	Conditions
Early Bird	Answer 100 questions between 4 am and 7 am
Night Owl	Answer 100 questions between 23 pm and 2 am
Expert	Reach a score of 5000 Points
Master	Reach a score of 20000 Points
Powerman	Collect 100 Bonus Items (in total)
Regular Customer	Have a constant log-in streak of 7 days in a row
Way to go	Answer 100 questions in total
Autobiographer	Fill out own bibliography
Chatterbox (hidden)	Used the contact form 5 times

*"iHEARu-PLAY: Introducing a game for crowdsourced data collection for affective computing", **WASA**, 2015.*

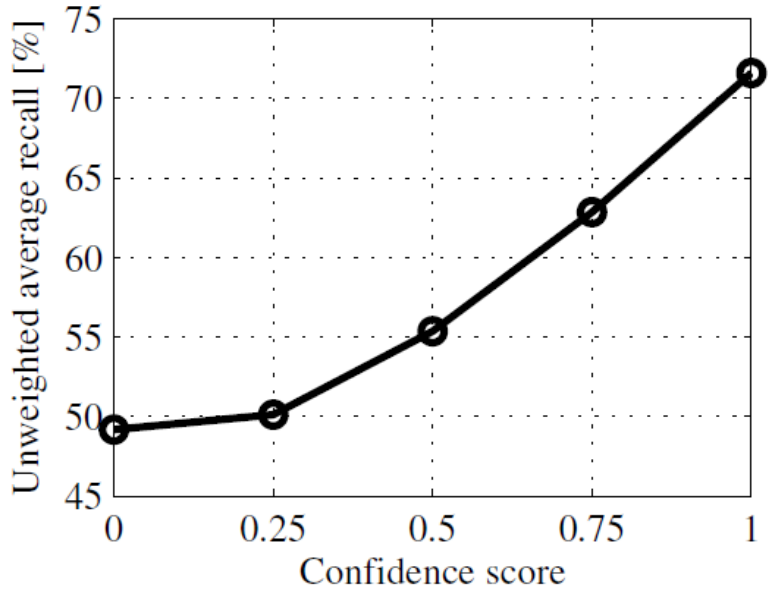
Cooperative Learning.

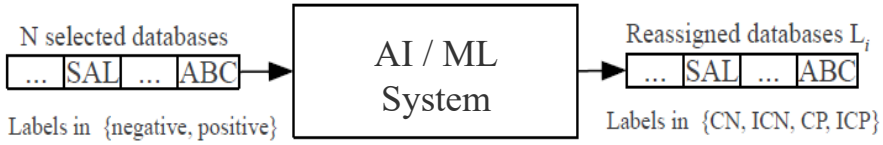
- **Confidence Measure**

  Learning Recogniser Behaviour

  Adaptation to target domain:
  Semi-supervised learning

Example: ComParE:EC



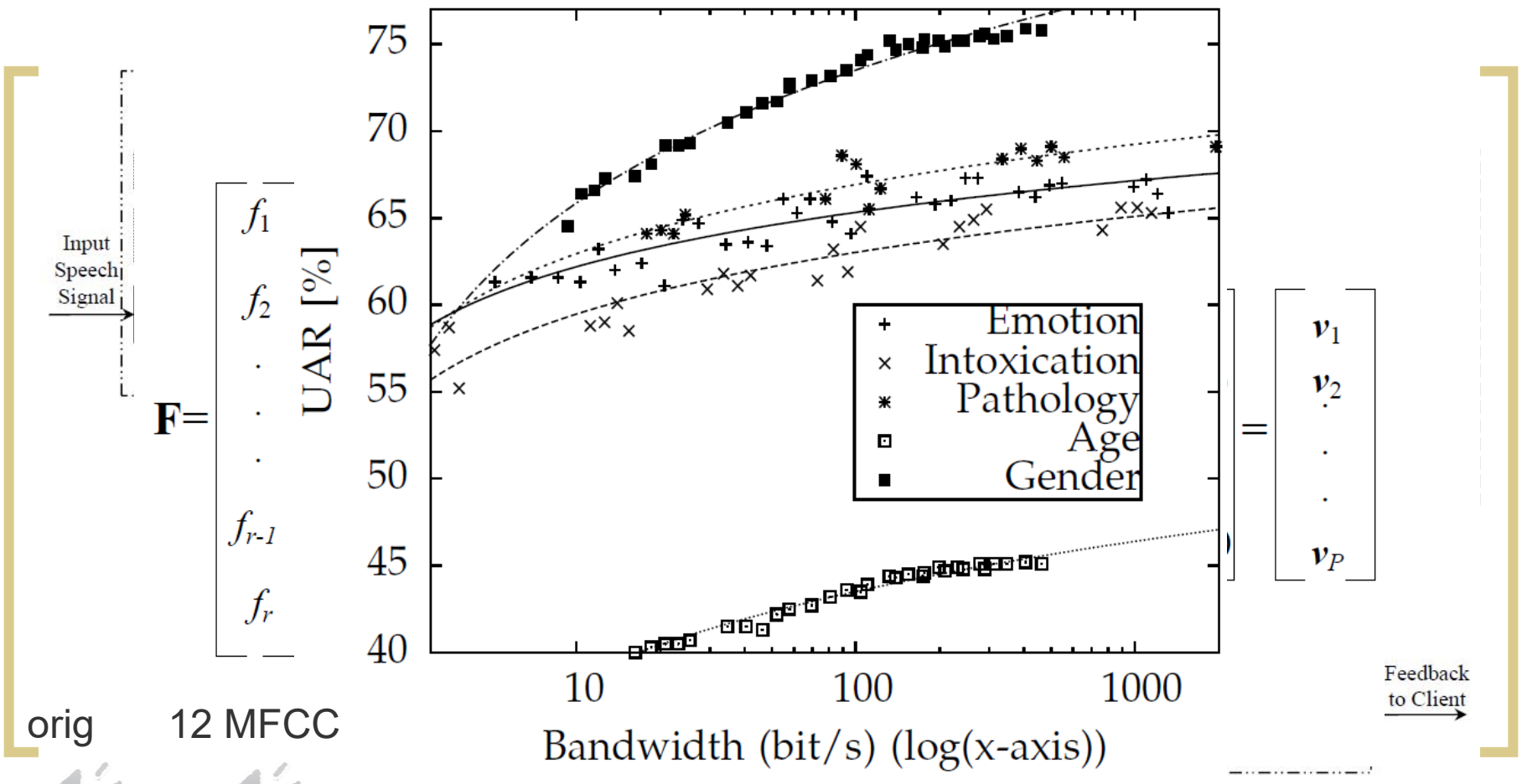


	Predicted -	Predicted +
Actual -	CN	ICN
Actual +	ICP	CP

*"Confidence Measures in SER Based on Semi-supervised Learning", **ISCA Interspeech**, 2012.*

Distribution.

*"Distributing Recognition in Computational Paralinguistics," **IEEE Transactions on Affective Computing** 5:406–417, 2014.*



orig 12 MFCC

Embedding.

- **Mobile Platform**

  Samsung Galaxy S3 (Android)

  vs Intel Core i3 2.1 GHz Laptop

Set	# feats.	RTF (S3)	RTF (i3)
GeMAPS Extended	88	2.63	0.11
Interspeech Emotion Ch.	384	0.43	0.04
Interspeech ComParE	6373	2.81	0.12

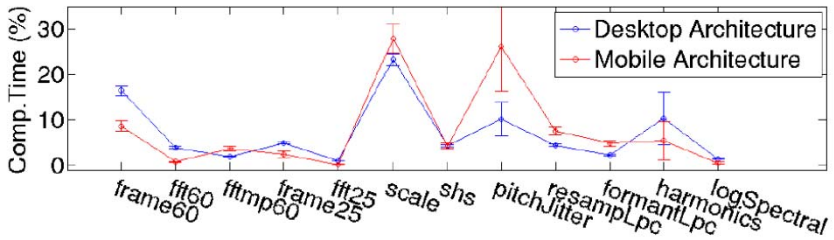


Fig. 1. GeMAPS extended feature set [19]

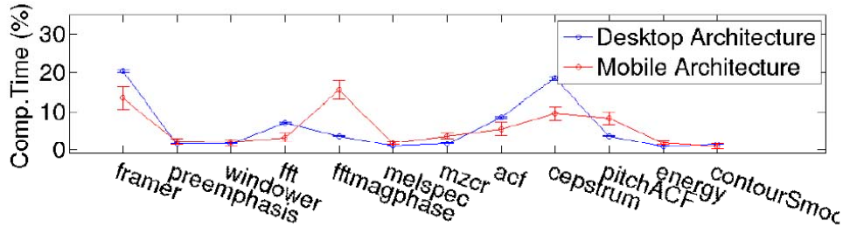


Fig. 2. Interspeech 2009 Emotion Challenge feature set [20]

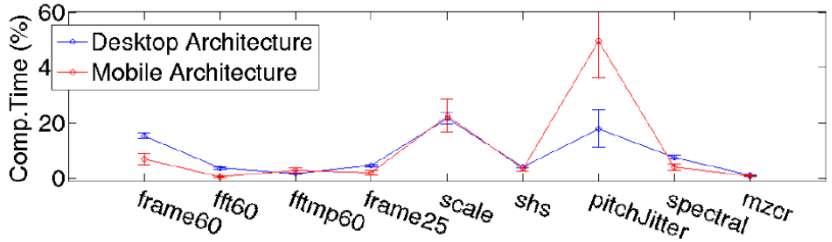


Fig. 3. Interspeech ComParE feature set [21]

Learning Hidden Representations.

- **Semi-NMF**

$$\boldsymbol{X}^{\pm} \approx \boldsymbol{Z}^{\pm}\boldsymbol{H}^{+}$$

Matrix factorization, uncovers meaningful features

Clustering interpretation:

Z: cluster centroids

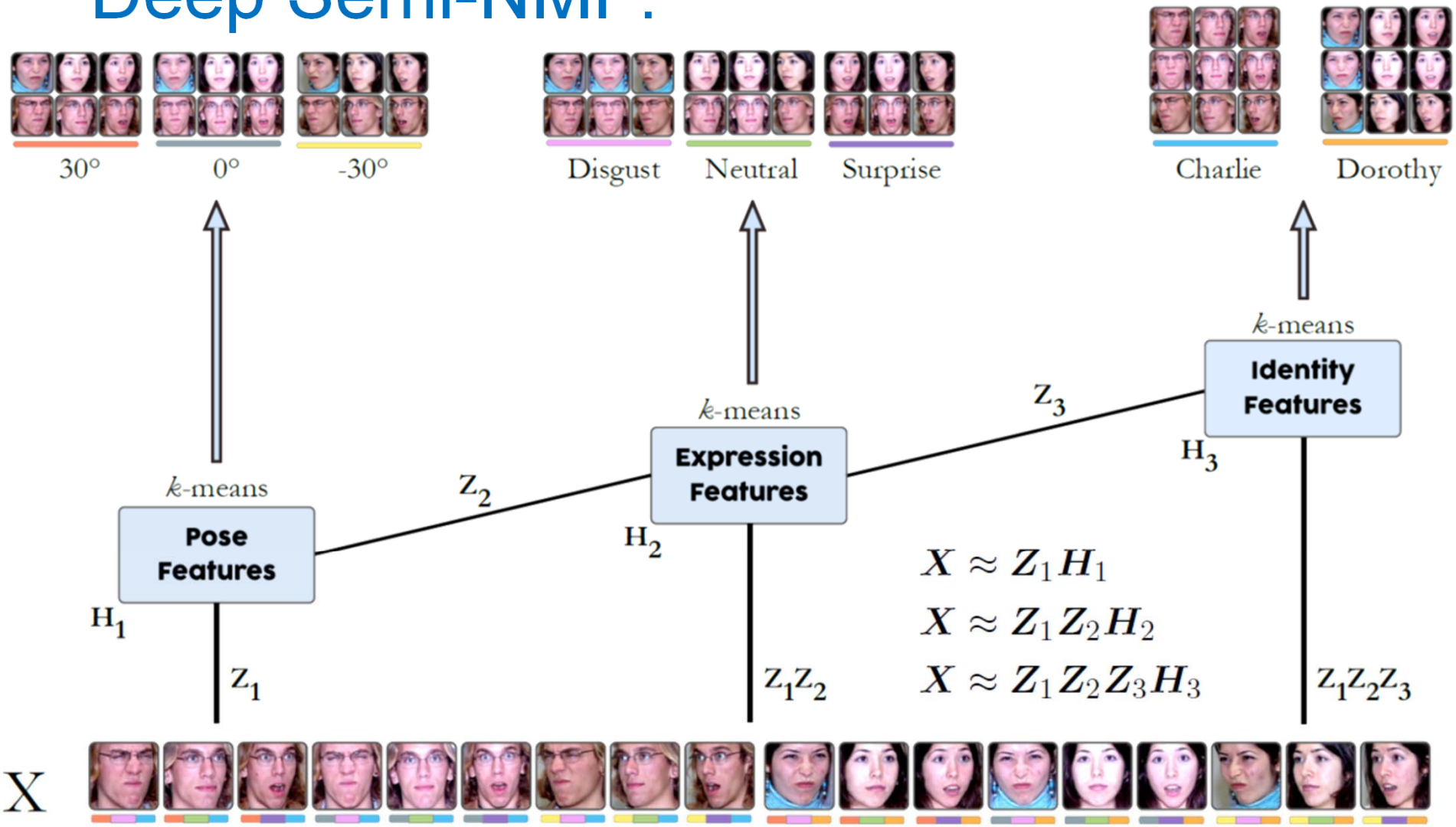
H: soft membership for every data point $h_{ik} = \begin{cases} 1 \text{ if } x_i \text{ belongs to cluster} \\ 0 \text{ otherwise} \end{cases}$

Soft version of k-means clustering – equivalent: iff H is orthogonal with:

$$C_{k-\text{means}} = \sum_{i=1}^{n}\sum_{j=1}^{k} h_{ki}\|\boldsymbol{x}_i - \boldsymbol{z}_k\|^2 = \|\boldsymbol{X} - \boldsymbol{Z}\boldsymbol{H}\|_F^2$$

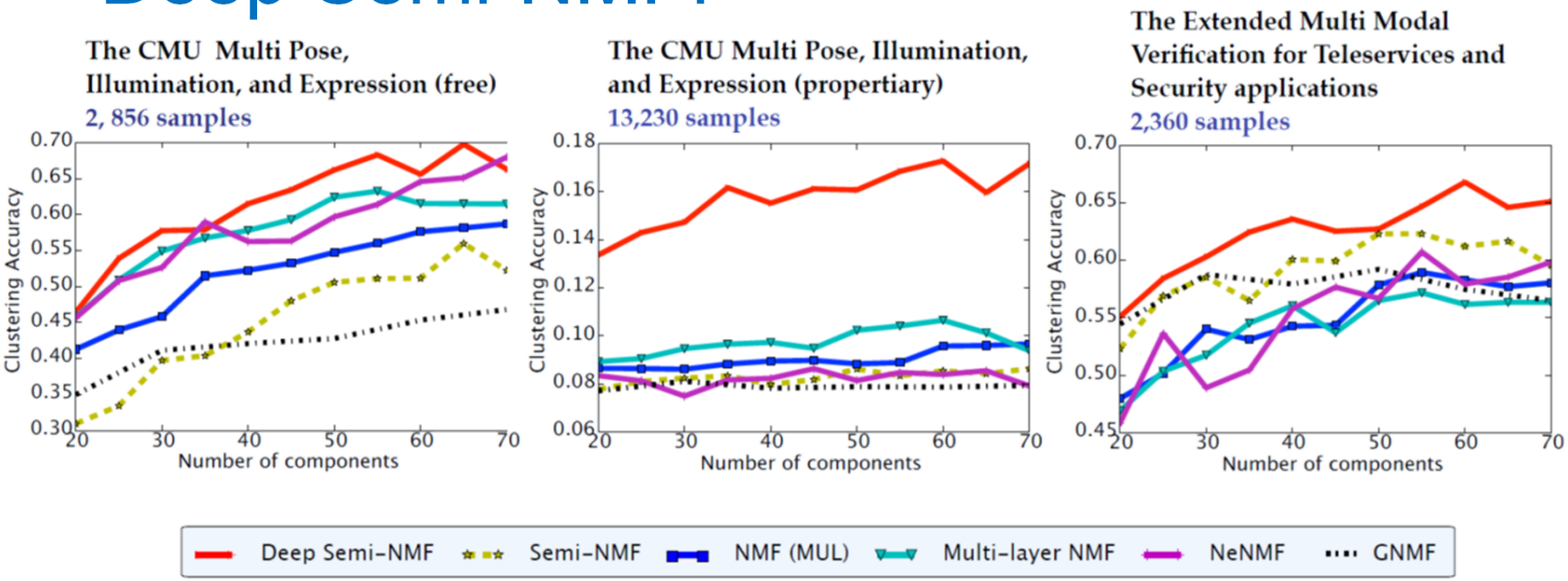
Better, if data is not distributed in spherical manner

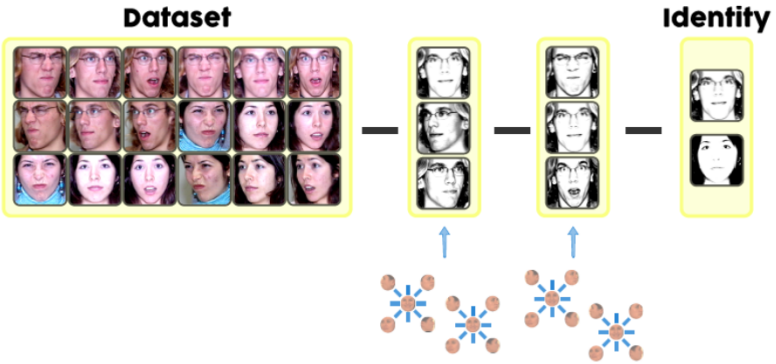
Deep Semi-NMF.



*"A Deep Semi-NMF Model for Learning Hidden Representations", **ICML**, 2014.*

Deep Semi-NMF.





UAAUC [%]	Pose	Emotion	ID
Semi-NMF	94.32	44.72	50.33
Deep Semi-NMF	99.78	73.33	74.79
A priori DSNMF	99.99	76.67	80.44

*"A Deep Semi-NMF Model for Learning Hidden Representations", **ICML**, 2014.*

Challenge Your Peers

The Challenge

- **Why at all?**

  Provide unified test-beds for evaluation

  Exchange of ideas across the community

  …on specific given tasks

  Help overcome often present lack of comparability

  …due to different data-sets, partitioning, evaluation measures, etc.

  May serve as long-standing reference

  Raises awareness for your field outside of it

  Raise acceptance for “low numbers” (the reality shock)

Risks

- **Focus on Optimisation**

  Novel paradigms often take development time
  Other aspects such as efficiency often neglected
  Awarding best paper can ease focus on optimisation

- **Over-Optimisation**

  Using new datasets for the challenge helps

Selecting a Corpus

Requirement	Example
Quantity	"There's no data like more data"
	High diversity with respect to diverse factors
	Reasonably balanced distribution
	Knowledge of natural distribution ('priors')
Quality	Adequate data
	Realistic data
	Ideal capture conditions
	Intended corruption
Modelling	Reasonable categorisation
	Well-defined mappings between models
Labelling	Unique and additional labelling (text+events, labeller tracks, context, ...)
	Provision of gold standard's reliability
Release	Documentation of side conditions
	Additional perception tests
	Freeing of the data
	Defined partitioning

The End User License Agreement

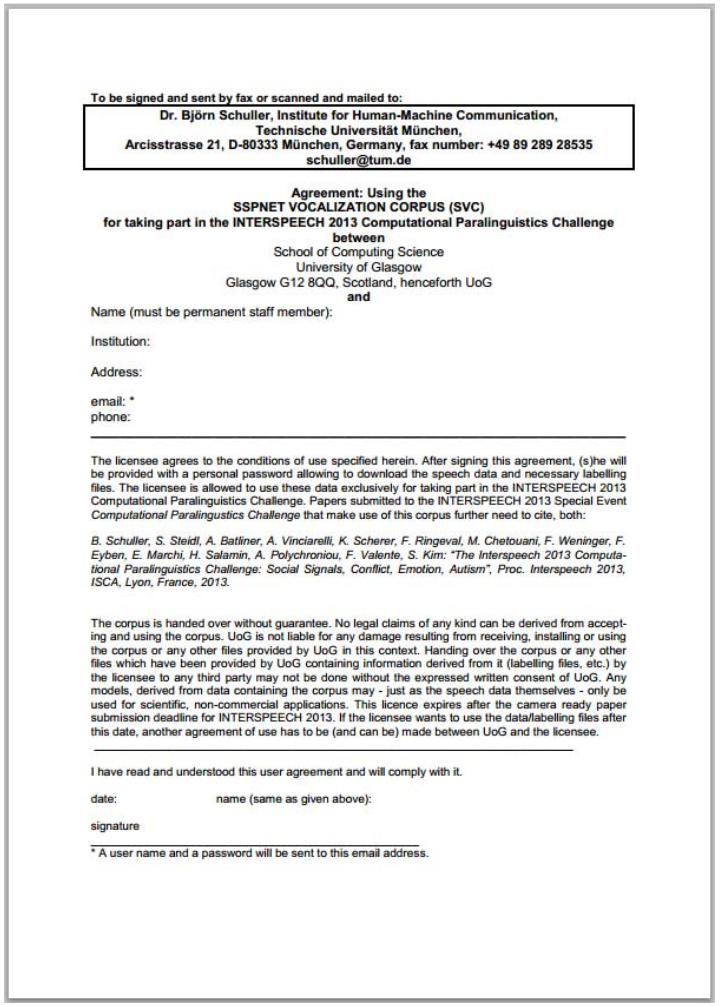
- **EULA should**

  Be the same for all corpora

  Fit one page (fax!)

  Be signed by permanent staff member

  Emailed or faxed…



To be signed and sent by fax or scanned and mailed to:

Dr. Björn Schuller, Institute for Human-Machine Communication,
Technische Universität München,
Arcisstrasse 21, D-80333 München, Germany, fax number: +49 89 289 28535
schuller@tum.de

Agreement: Using the
SSPNET VOCALIZATION CORPUS (SVC)
for taking part in the INTERSPEECH 2013 Computational Paralinguistics Challenge
between
School of Computing Science
University of Glasgow
Glasgow G12 8QQ, Scotland, henceforth UoG
and

Name (must be permanent staff member):

Institution:

Address:

email: *
phone:

The licensee agrees to the conditions of use specified herein. After signing this agreement, (s)he will be provided with a personal password allowing to download the speech data and necessary labelling files. The licensee is allowed to use these data exclusively for taking part in the INTERSPEECH 2013 Computational Paralinguistics Challenge. Papers submitted to the INTERSPEECH 2013 Special Event *Computational Paralingustics Challenge* that make use of this corpus further need to cite, both:

B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, H. Salamin, A. Polychroniou, F. Valente, S. Kim: "The Interspeech 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism", Proc. Interspeech 2013, ISCA, Lyon, France, 2013.

The corpus is handed over without guarantee. No legal claims of any kind can be derived from accepting and using the corpus. UoG is not liable for any damage resulting from receiving, installing or using the corpus or any other files provided by UoG in this context. Handing over the corpus or any other files which have been provided by UoG containing information derived from it (labelling files, etc.) by the licensee to any third party may not be done without the expressed written consent of UoG. Any models, derived from data containing the corpus may - just as the speech data themselves - only be used for scientific, non-commercial applications. This licence expires after the camera ready paper submission deadline for INTERSPEECH 2013. If the licensee wants to use the data/labelling files after this date, another agreement of use has to be (and can be) made between UoG and the licensee.

I have read and understood this user agreement and will comply with it.

date: name (same as given above):

signature

---

* A user name and a password will be sent to this email address.

The Data

- **Mandatory Things**

  Permissions to distribute the data

  An end-user license agreement

  There is a development and a test set

  The test set is not known to anybody previously (!)

  Train and test are sufficiently different…

  It contains the task and not rather noise

Finding a Venue

- **To satellite or not to satellite**

  ACM ICMI, ACM MM, HUMAINE / IEEE ACII, IEEE FG, Interspeech
  IEEE ICASSP (?), NIPS (?), ICDAR (?), ICPR (?), IJCAI (?), etc.

  Satellite often less effort and more attractive

  Potential to reach audience of main conference

  More restrictive, e.g., with proceedings

Rules Rule!

- **Rules…**

  Making rules clear from beginning
  Readable at any time to all (!) participants
  No change during the ongoing Challenge
  Sticking strictly (!) to them
  Clarifying immediately

  Organisers do not participate
  Clarify if (independent) paper acceptance is mandatory
  Limit number of trials
  Preserve right to re-evaluate winners‘ results

The Task Design

- **Make the Task**

  As close to the real thing as possible

  Appealing, novel in some aspect

  Useful to the world

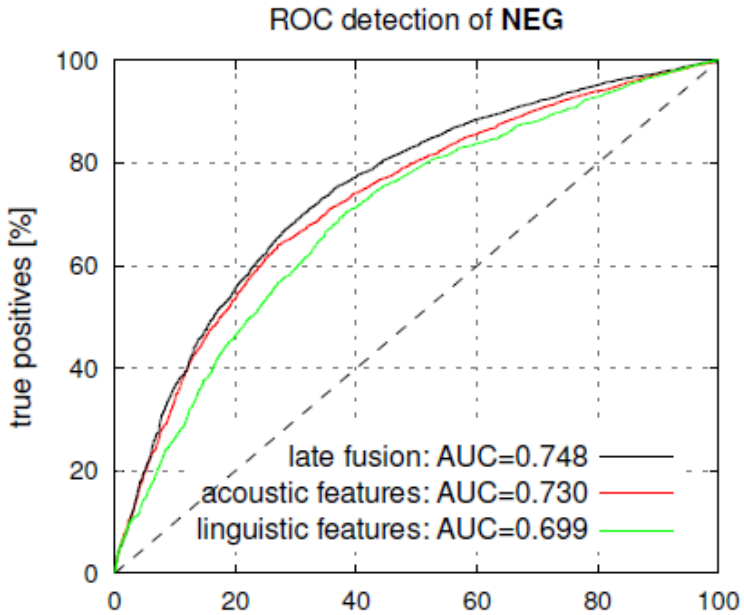
  Not too hard…

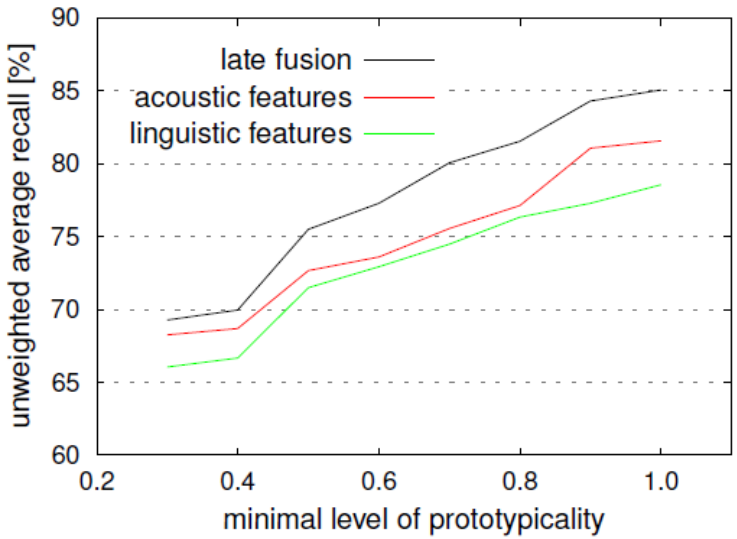
  Manageable in terms of data transfer

  Comprehensible

Testing

- **Measures**
  - Accuracy (Un-/Weighted Average Recall)
  - Precision, F-Measure, confusions
  - Area under ROC
  - Mean Linear Error, Cross Correlation
- **Partitioning**
  - J-fold Stratified Cross-Validation
  - Leave-One-Speaker (group)-Out
  - Train-Develop-Test
- **Discussion**
  - Significance
  - Ground truth
  - Prototypicality





Tools

- **Tasks:**

  Feature extraction
  Model training and classification
  Pre-trained models

- **Application Domains:**

  Batch evaluations / Off-line analysis
  On-line / realtime recognition

The Baseline

- **Has to be:**

  State-of-the-art
  Transparent („no“ optimisation)
  Reproducible
  Comprehensible

- **Should be:**

  Established by open-source tools
  Re-doable “at no effort” (e.g., batch scripts provided)
  Not overly optimised
  Beatable…

The Paper

- **The Essential**

  Describe it such that the participants know

- **The Secret Ingredient**

  Highlight one technical aspect to make it more interesting:
  IS 2009: Instance balancing / dynamic vs. static
  IS 2010: Multitask vs. single task / regression task
  IS 2011: Feature optimisation
  IS 2012: Effect of random seed (standard deviation)
  IS 2013: Detection

Evaluation

- **Saving work…**

  Have a web script for upload

  Not requiring codes and running them, but
  Keep right to ask for it

- **Ensure that**

  The servers is up

  You provide a “safe” deadline

The Workshop / Event / Session

- **Instruct the Participants**
  Write mail in advance
  Explain presentation format
  Potentially limit (e.g., no repetition of challenge conditions, data, etc.)

- **Choose format wisely**
  Poster session usually not suited – participants are curious to see others
  Short talks and intro / summary good choice

- **Keep the suspension**
  Consider announcing winners at end
  Lay out presentations in suited order

Awarding

- **Make a Certificate**
- **Make a Dummy Cheque**


INTERSPEECH 2012 Speaker Trait Challenge

AWARD CHEQUE

[EUR] €

Pay to

The order of (in words)

Drawer humaine

Authorized Signature

Monday, September 10, 2012

Portland, Oregon, U.S.A.

Björn Schuller, Stefan Steidl, Anton Batliner, Elmar Nöth, Alessandro Vinciarelli, Felix Burkhardt, Rob van Son
(Organizers of the Speaker Trait Challenge)

Account Holder	Bank	IBAN	BIC	Signature

- **Award in the big Ceremony**
- **Put the Winners on the Web**


14th ACM International Conference on Multimodal Interaction (ICMI)
October 22-26, 2012

AVEC 2012
Multimodal Grand Challenge Award
Fully Continuous Sub-Challenge

"Robust continuous prediction of human emotions using multiscale dynamic cues"

Presented to

Jeremie Nicolle, Vincent Rapp, Kévin Bailly, Lionel Prevost, and Mohamed Chetouani

Social Signal Processing Network
PASCAL2
2012 · SANTA MONICA

AVEC 2012 Chairs

Louis-Philippe Morency, Dan Bohus and Hamid Aghajan
ICMI 2012 General Chairs


Special Event INTERSPEECH 2011 Speaker State Challenge

humaine
BAS

INTERSPEECH 2011
SPEAKER STATE
CHALLENGE
AWARD

The Intoxication Sub-Challenge Prize
is awarded to

Daniel Bone, Matthew P. Black, Ming Li, Angeliki Metallinou, Sungbok Lee, Shrikanth S. Narayanan

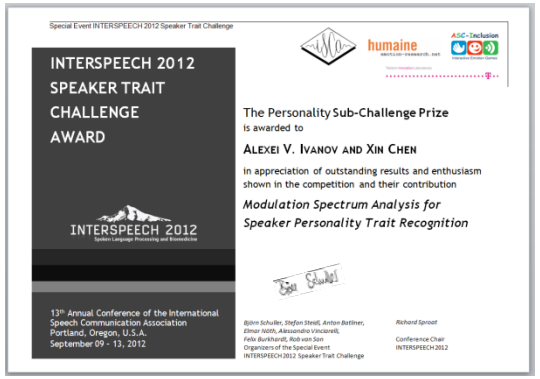
in appreciation of outstanding results and enthusiasm shown in the competition and their contribution

Intoxicated Speech Detection by Fusion of Speaker Normalized Hierarchical Features and GMM Supervectors

12th Annual Conference of the International Speech Communication Association
Florence, Italy
28 - 31 August 2011

Björn Schuller, Stefan Steidl, Anton Batliner, Florian Schiel, Janek Krajewski
Organisers of the Special Event
INTERSPEECH 2011 Speaker State Challenge

Piero Cosi
General Chair
INTERSPEECH 2011


Special Event INTERSPEECH 2012 Speaker Trait Challenge

humaine
ASC-Inclusion

INTERSPEECH 2012
SPEAKER TRAIT
CHALLENGE
AWARD

INTERSPEECH 2012

The Personality Sub-Challenge Prize
is awarded to

Alexei V. Ivanov and Xin Chen

in appreciation of outstanding results and enthusiasm shown in the competition and their contribution

Modulation Spectrum Analysis for Speaker Personality Trait Recognition

13th Annual Conference of the International Speech Communication Association
Portland, Oregon, U.S.A.
September 09 - 13, 2012

Björn Schuller, Stefan Steidl, Anton Batliner, Elmar Nöth, Alessandro Vinciarelli, Felix Burkhardt, Rob van Son
Organizers of the Special Event
INTERSPEECH 2012 Speaker Trait Challenge

Richard Sproat
Conference Chair
INTERSPEECH 2012

Evaluation

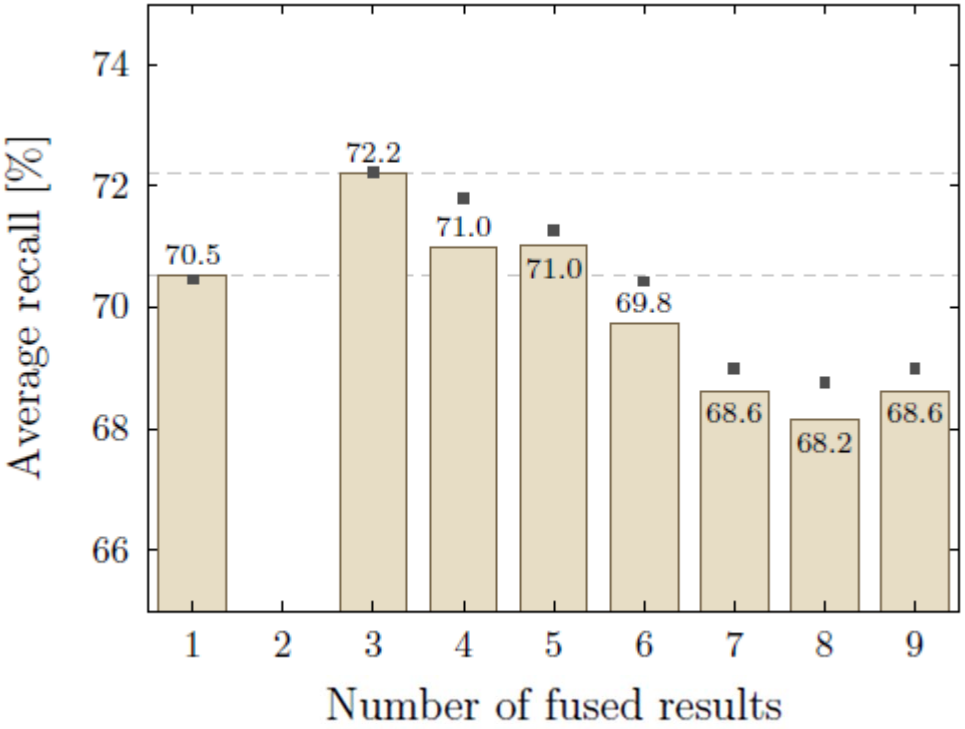
- **Fusion**

  Late fusion (vote)

  Ask for confidences

  Determining optimal N

  Usually long-standing
  Upper benchmark



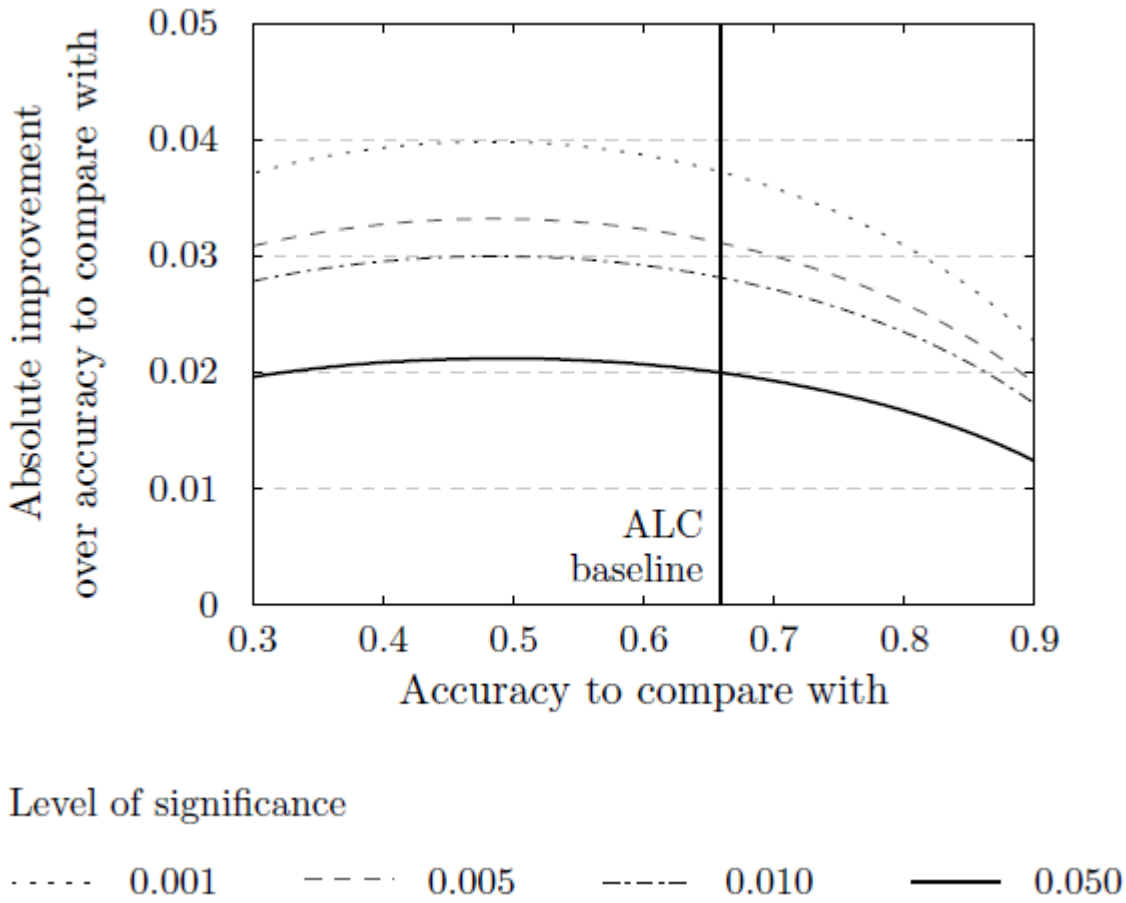
Message: *"Together we're best"*

Evaluation

- **Significance**

  Heavily depends on:

  # test instances
  % baseline



"Medium Term Speaker Traits – A Review on Intoxication, Sleepiness, and the First Challenge", ***Computer Speech & Language***, 2013.

Evaluation

- **Q-Statistics**

Pairwise measuring whether participants commit same errors on test

	Avg	#2	#3	#4	#5	#6	#7	#8	#9 (Nogueiras)
#1 (Bone et al.)	.427	.598	.425	.461	.460	.428	.433	.395	.217
#2 (Bocklet et al.)	.657		.710	.752	.757	.690	.740	.651	.356
#3 (Hönig et al.)	.754			.822	.941	.906	.946	.864	.418
#4 (Gajšek et al.)	.716				.836	.845	.880	.775	.358
#5 (Huang et al.)	.776					.917	.956	.885	.452
#6 (Ultes et al.)	.749						.967	.826	.417
#7 (Baseline)	.782							.903	.427
#8 (Montacié and Caraty)	.706								.346
#9 (Nogueiras)	.374								

"Medium Term Speaker Traits – A Review on Intoxication, Sleepiness, and the First Challenge", ***Computer Speech & Language***, 2013.

Continuity

- **To turn it into a continued success…**

  Make it consistent

  Use same EULA, format, conditions, etc.

  Alter … depending on last experiences

The Repository

- **Easy access to data**

  Make a page with info on past challenges

  Provide info for data access (contact persons, EULAs, etc.)

  Provide baseline feature sets (the ML community will appreciate it!)

  Make it a value way beyond the singular event

  Save yourself 1000 emails…

  And … really do it! (coming up some day soon for ComParE / AVEC…)

Wrap-Up

- **Organise a Challenge**

  Find great data
  Set clear rules
  Provide transparent redoable baselines
  Provide good infrastructure for data access, result upload, etc.

- **In the long run**

  Make a Special Issue
  Ensure consistency
  Make a repository
  Participate in other challenges

Experience ComParE / AVEC

- **# Participating Teams**

  Interspeech ComParE: 32 – 65 (2009 – 2015)

  AVEC:12 – 31 (2011 – 2015)

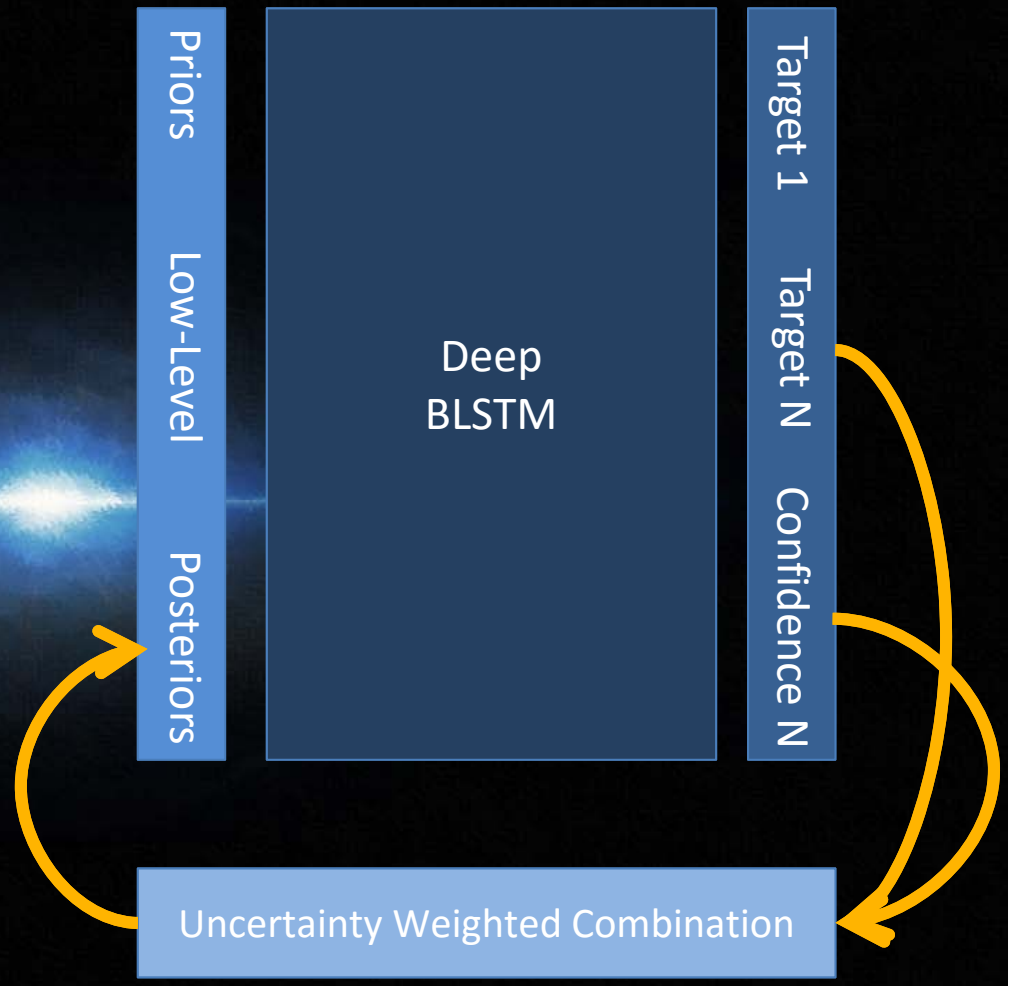
- **Some Observations**

  Fusion of best participant results exceeds individual winners throughout
  These “multiple-site” results are so far not reached by individual attempts
  Supra-segmental modelling of information prevails by far
  Only sparsely dynamic algorithms such as HMMs or DBNs
  Cross disciplines seems to be successful
  Great submissions outside competition, e.g., perception studies

Outlook

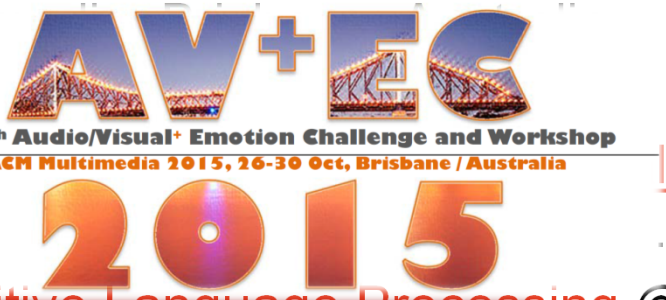
And Next…

- **Big Data Multi-task DL**
- **24/7 Coop Learning**
- **“Green” Learning**
- **Reinforced Learning**
- **Evolving Machines**
- **New Challenges**
- **Products!**



Better Ideas?

- ACII 2015,
- Special Iss ses for Natural Language ms, Elsevier
- "Emotion Representati n Techno M4 ICMI 2015, Seattle, .11
- AV+EC CM 26.-30.1
- 1st Inter orks lysis in the Wild (WASA 2015) @ A .2015.
- Special Ses sing @ 7th Internati stems (IWSDS), Riekonlinna



CALL FOR PAPERS

1st Int. Workshop on Automatic Sentiment Analysis in the Wild @ ACII 2015

21 September 2015, Xi'an / P.R. China