

Felix Burkhardt, Benjamin Weiss, Florian
Eyben, Jun Deng, Björn Schuller

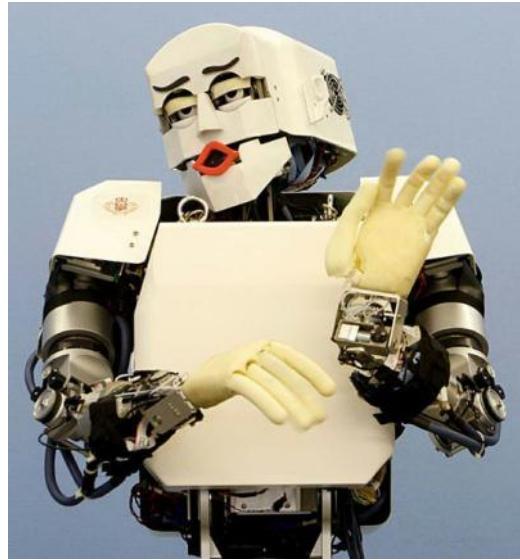
Detecting Vocal Irony



ERLEBEN, WAS VERBINDET.

Vocal irony data collection

- Motivation
- The app
- Acoustic irony classifier
- Textual sentiment classificationn
- Data collection
- Data labeling
- Experiments/results
- Wrap-up/Outlook



<http://www.takanishi.mech.waseda.ac.jp/top/research/kobian/KOBIAN-R/index.htm>

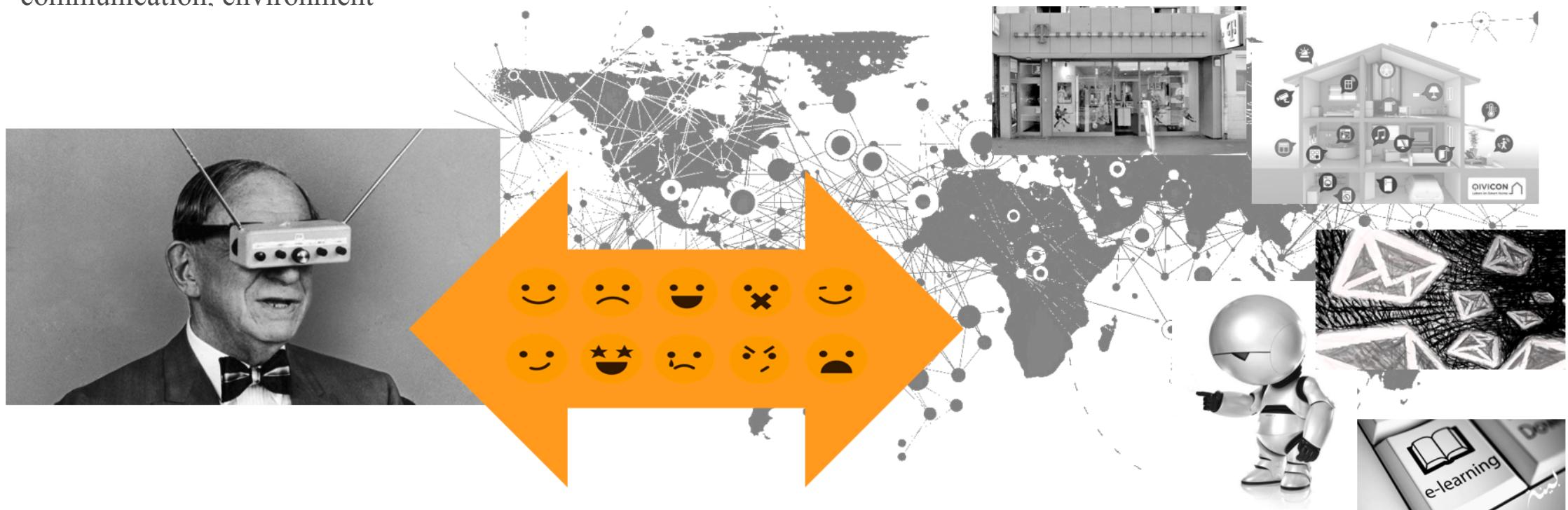
Affective Agents Trends

Ubiquitous computing accessible via

- Smart mobile devices: phones, glasses, watches, t-shirts, implants, etc
- Home automation: central intelligence controlling media, communication, environment

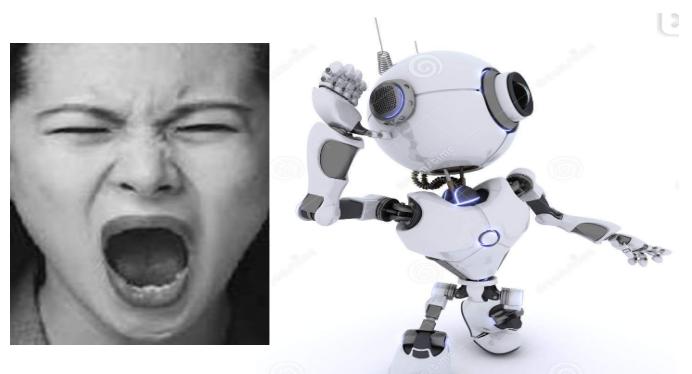
Uses natural interface: voice, gestures, wearables, ...

- Gets much nearer to user, unobtrusive
- Will be emotional because it's easier: emotion expression is a channel of communication



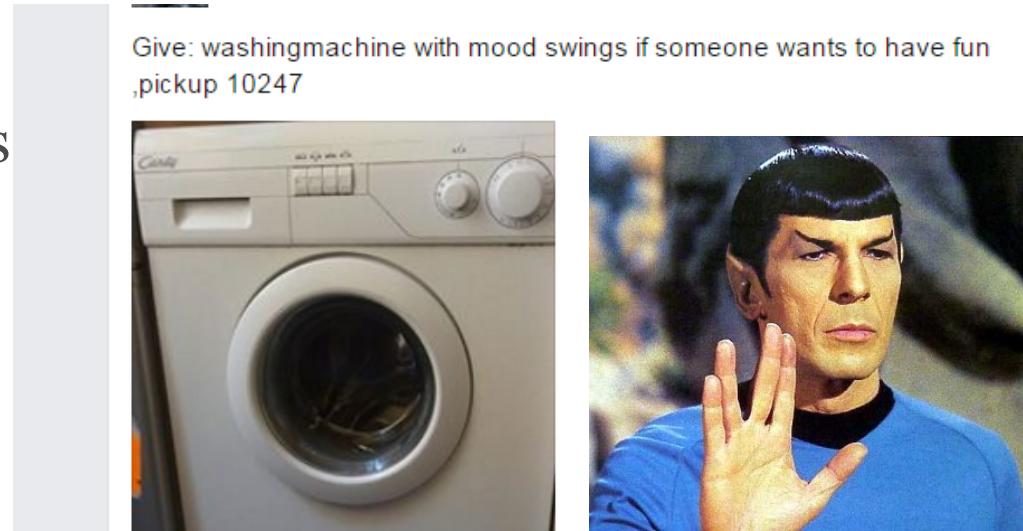
Motivation I

- Irony / Sarcasm detection: To analyze user opinion this is still a big problem
- Polarity detection: How do users feel about a new product?
- Emotional Monitoring: E.g. anger detection can be used to soothe disgruntled users or for automatic quality monitoring.
- Emotional Chat: Special channels can be provided to facilitate emotional computer mediated communication and analysis can be used to automate emotional labeling.



Motivation II

- Emotion-aware Surrounding: Quite an old idea is a computer controlled environment that adapts automatically on the user's mood by e.g. playing 'just the right music' or adjusting automotive system reaction.
- Believable Agent: The naturalness of an artificial 'being' and the appearance of intelligence is highly altered by emotional expressions; especially gaming applications can benefit.
- Artificial intelligence models, use emotions for motivation modeling



Questions

Irony / Sarcasm

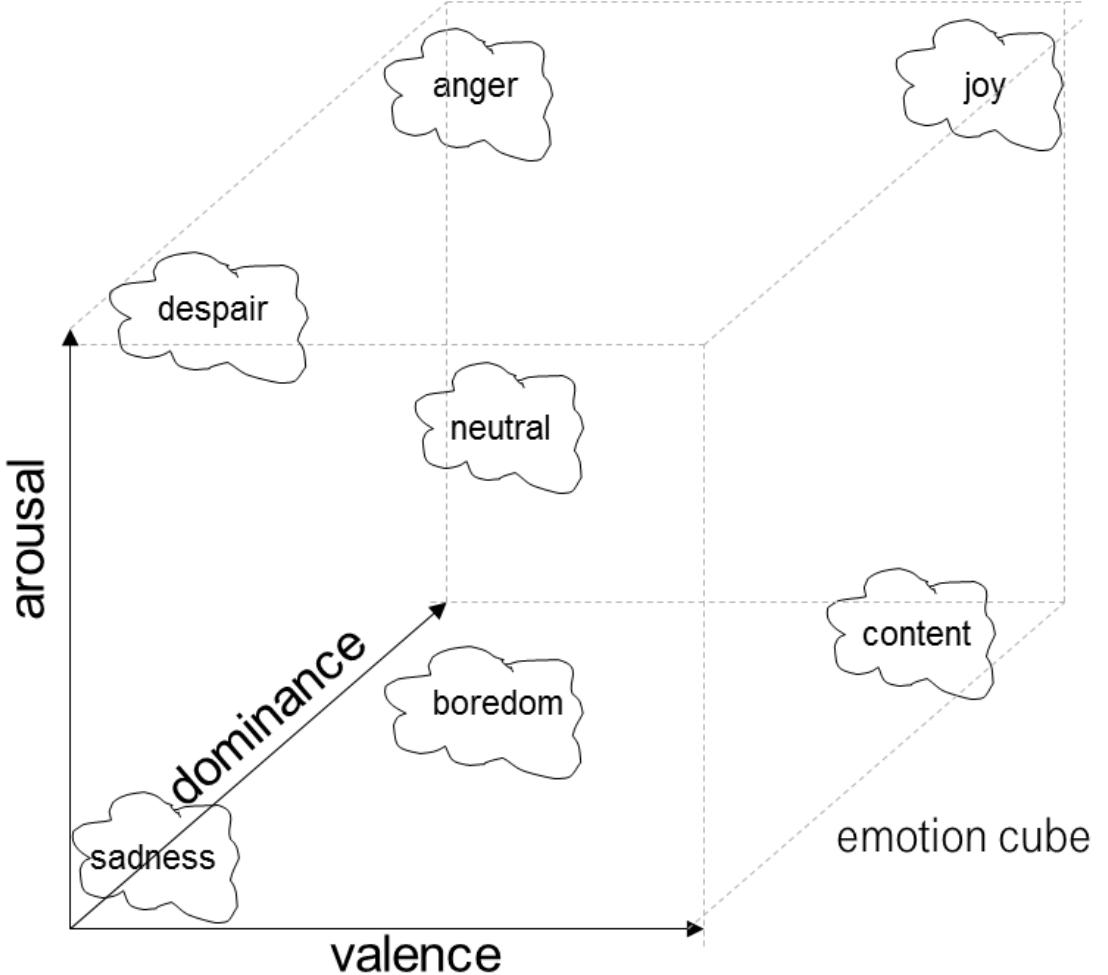
- Wording only
- Context
- Discrepancy between words and “tone of voice”?
- Acoustic correlates of ironic speech?
- Multimodal recognition?



Emotion models

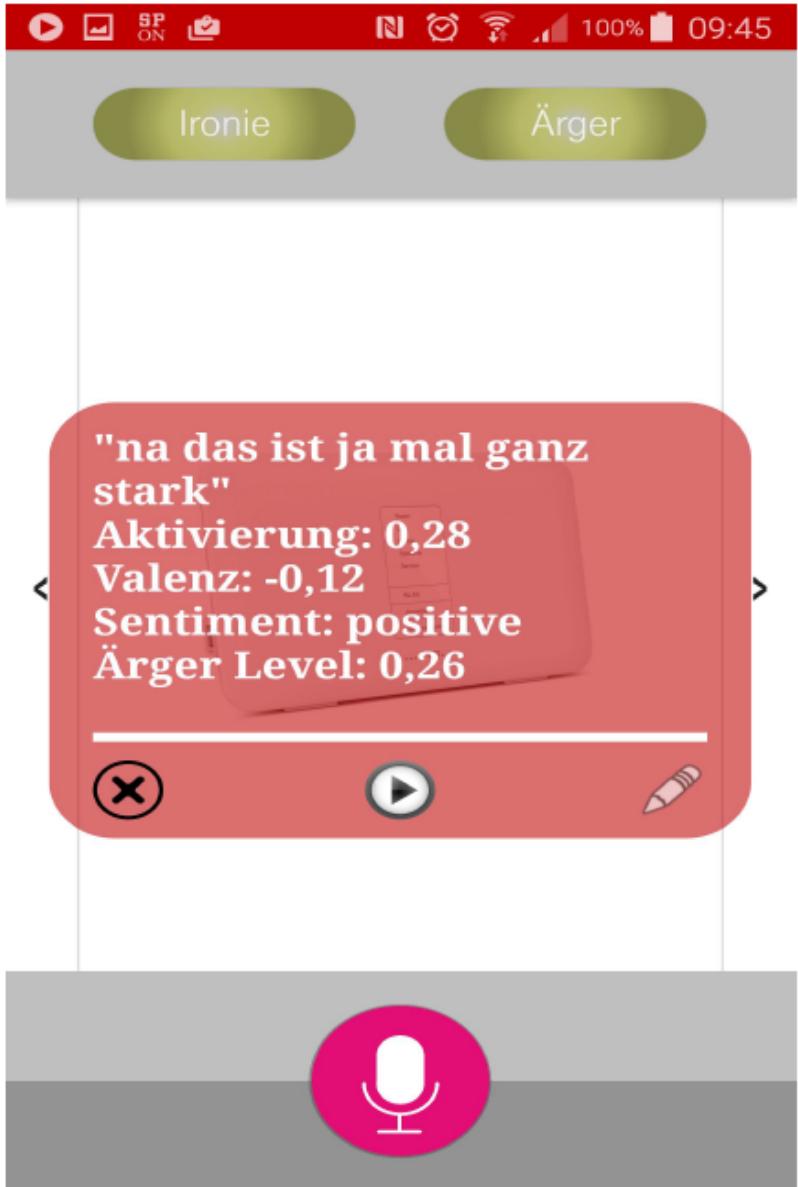
Dimensional models

- Dimensions consider an emotion as a point in a N-dimensional emotion space.
- How many dimensions are needed to model human emotional behaviour is open to dispute , but the most well-known spaces are
 - The two dimensional activation-pleasure space
 - The PAD-spache
 - Pleasure (valence)
 - Arousal (activation)
 - Dominance,



The App

- We developed an Android app to demonstrate the technology and collect user data.
- The app is meant to analyze short utterances of about 1 to 5 seconds.
- The result is a red field with values for
 - The Google ASR result
 - A value for the activity in the voice [-1- 1]
 - A value for the valence of the voice [-1- 1]
 - The sentiment (polarity of the text) (neut / pos / neg)
- An anger level: Compared with a database of call center customers [0 - 1].
- If sentiment and valence diverge, the irony lamp starts blinking.



Acoustic Irony classification

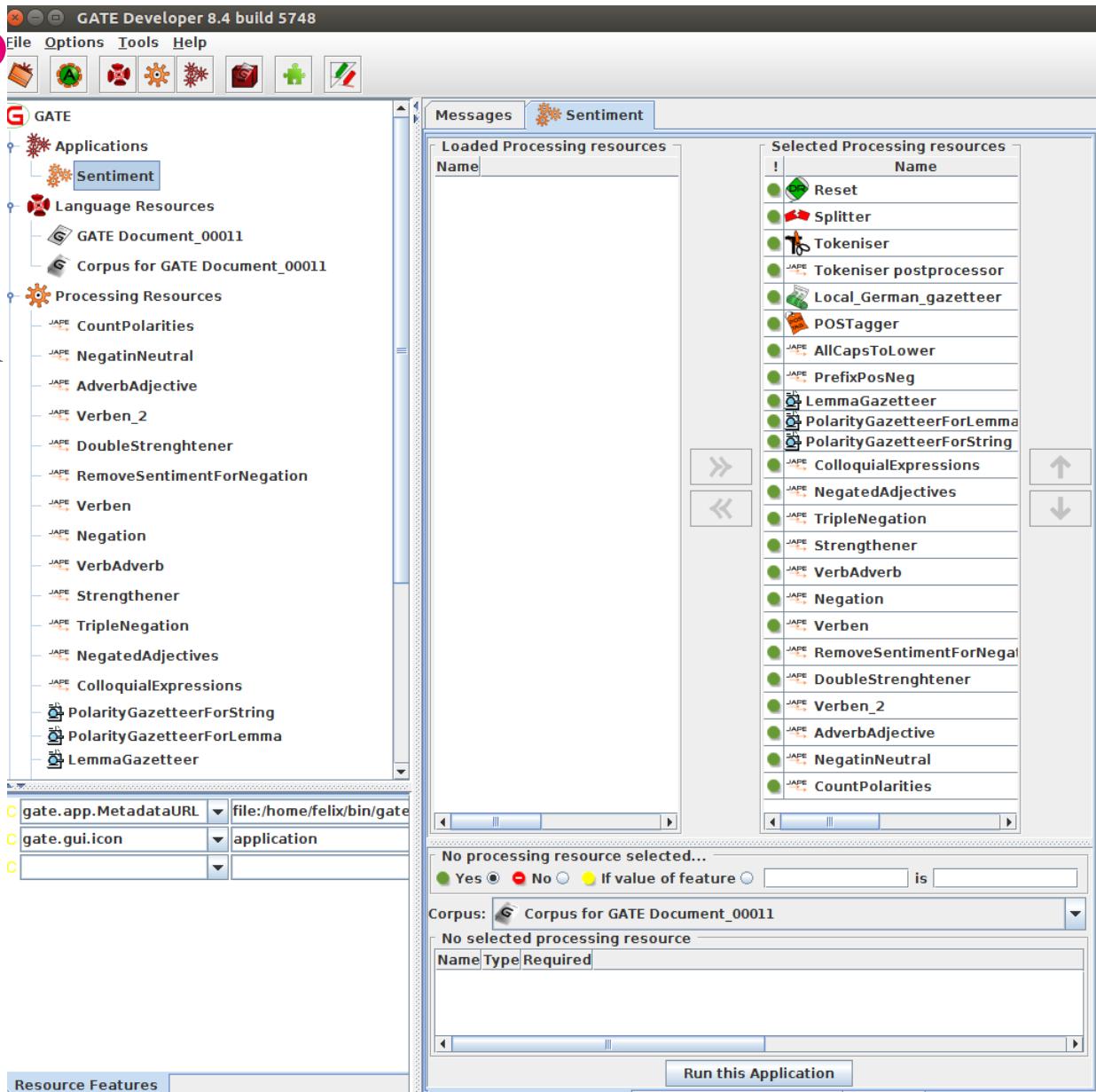
- Acoustic feature extraction tool openSMILE using two features sets:
 - extended Geneva Minimalistic Acoustic Parameter Set (eGEMAPS) - 88 features
 - Interspeech 2013 Computational paralinguistic Challenge feature set (ComParE) – 6373 features
 - scaled all features to zero mean and unit standard deviation derived from the training set.
- Support Vector Machines, from Scikit-learn (Pedregosa et al., 2011)

4 energy related LLD	Group
RMS energy, zero-crossing rate	Prosodic
Sum of auditory spectrum (loudness)	Prosodic
Sum of RASTA-filtered auditory spectrum	Prosodic
55 spectral LLD	Group
MFCC 1–14	Cepstral
Psychoacoustic sharpness, harmonicity	Spectral
RASTA-filt. aud. spect. bds. 1–26 (0–8 kHz)	Spectral
Spectral energy 250–650 Hz, 1 k–4 kHz	spectral
Spectral flux, centroid, entropy, slope	Spectral
Spectral Roll-Off Pt. 0.25, 0.5, 0.75, 0.9	Spectral
Spectral variance, skewness, kurtosis	spectral
6 voicing related LLD	Group
F_0 (SHS and Viterbi smoothing)	Prosodic
Prob. of voicing	Voice qual.
log. HNR, jitter (local and δ), shimmer (local)	Voice qual.

ComParE acoustic feature set: 65 low-level descriptors (LLD)

Textual Sentiment Classification

- Based on Gate (Univ. of Sheffield)
- The first steps involve splitting and tokenization
- We added the Stanford Part-of-Speech Tagger
- Lemmatizer based on lexicon derived from the Morphy project
- A German polarity lexicon (Waltinger, 2010)
- Further gazetteers annotate negation tokens (like “not great”) and strengthener words (like “super bad”).
- Many JAPE rules....



Data collection

- We conducted a workshop with lay people
- During the workshop, 12 test users got introduced to the app and tried it out.
- There were nine male and three female participants, aged 25 to 48
- After the try-out period of 1 week a set of 937 labeled samples had been collected,



Affective Agents

Hinweise für die Testphase

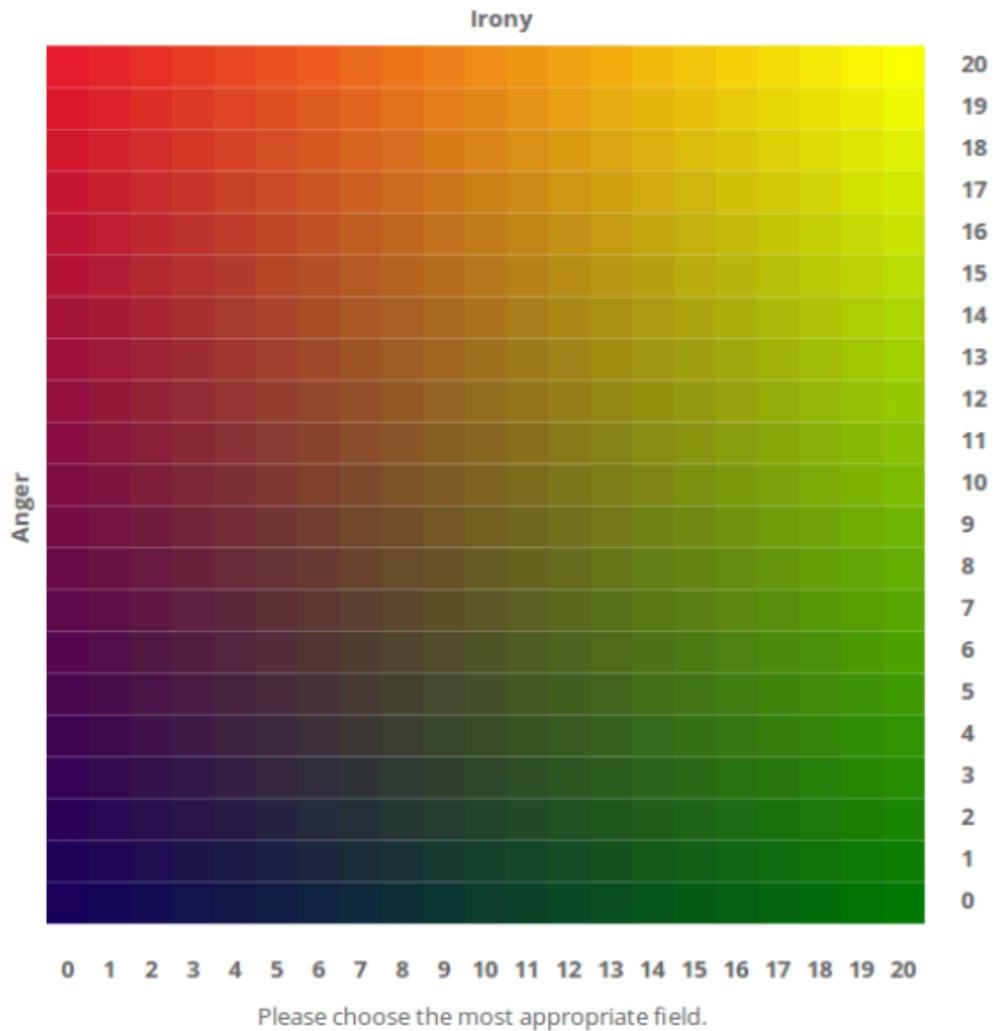
1. Nehmt die App mit nach Hause und behaltet sie bis zum 23.02. auf Eurem Smartphone. Danach wird der Server-Link abgeschaltet und Ihr könnt die App wieder löschen.
2. Versucht bitte, die App einmal täglich zur Hand zu nehmen und zu den angebotenen Bildern Sprachproben abzugeben.
3. Probiert dabei aus, ärgerlich und ironisch zu sprechen. Lasst Euch dabei von den Bildern inspirieren! Ihr könnt aber jederzeit auch frei von den vorgegebenen Bildern Äußerungen abgeben.
4. Ihr könnt so viele Äußerungen machen, wie Ihr wollt!
5. Ihr braucht beim Testen Internetzugang! Am besten funktioniert die App im W-LAN, aber auch im mobilen Internet.
6. Treten Probleme bei der Verwendung auf, könnt Ihr Euch direkt an das Innovationsforums-Team (030/835358580) innovationsforum@labs.telekom.de wenden.

Wichtig:

- **Eure Sprachproben werden aufgezeichnet, Ihr könnt aber entscheiden, ob Ihr einzelne Äußerungen freigeben wollt oder nicht!**
- **Wir werten Eure Sprachproben ohne Rückschluss auf Eure Person aus.**
- **Fertigt keine Kopien der App an und gebt den Link zum Download nicht an Dritte weiter.**
- **Für die legale Verwendung der App seid Ihr selbst verantwortlich!**

Data Labeling

- The judgements were given web-based as a point in a two dimensional matrix with the dimensions “irony” and “anger”, with twenty points on each axis for quasi-continuous rating
- In order to obtain, from the six individual ratings, a single gold standard for automatic classification, the Evaluator Weighted Estimator (EWE) is used
- The EWE is a weighted average of the individual ratings, where the weights are based on the average inter-rater reliability of each rater.



Data Labeling

rater #		1	2	3	4	5	6	AVG
EWE		0.89	0.68	0.70	0.82	0.87	0.73	0.81
1			0.52	0.59	0.66	0.77	0.67	0.73
2				0.44	0.48	0.45	0.42	0.57
3					0.54	0.59	0.56	0.63
4						0.68	0.58	0.68
5							0.65	0.72
6								0.66
Mean ρ	0.57							
Min ρ	0.42							
Max ρ	0.77							

Table 1: Pairwise cross correlation (Pearson correlation coefficient ρ) for the dimension irony. Mean, Min., Max. values exclude the Evaluator Weighted Estimator (mean) (EWE) values (first row).

Experiments Acoustic classification

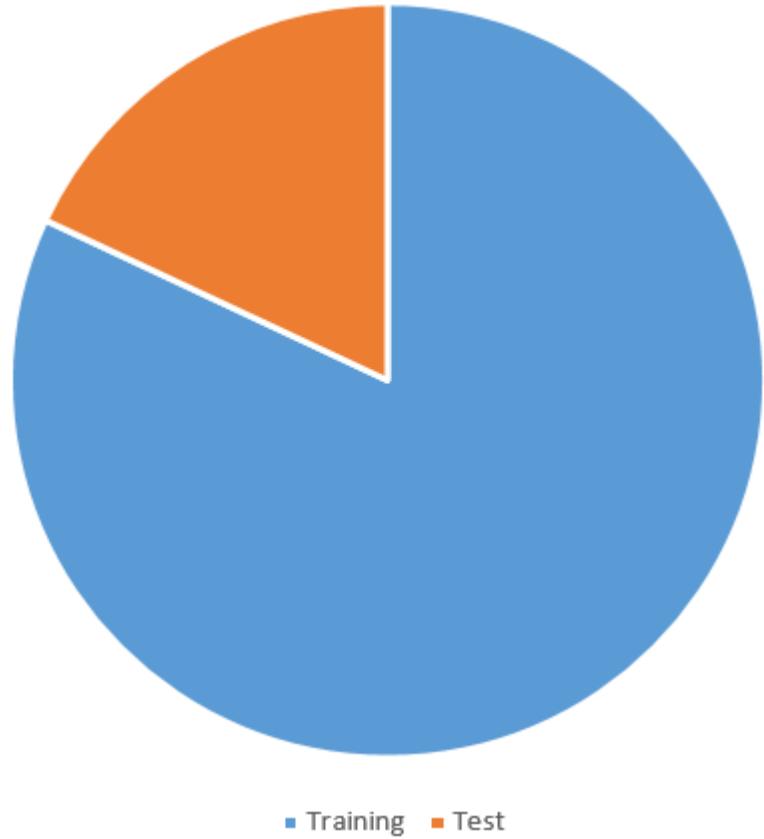
- Tests:
 - 5-fold cross validation (5-fold CV)
 - Vs. leave-one-speaker-out cross-validation (LOSOCV).
- Two feature sets
- Different complexity of the linear SVM

Feature	C (SVM)	UAR	
		LOSO-CV	5-fold CV
ComParE	1e-4	61.4	69.3
ComParE	1e-3	61.3	69.6
ComParE	1e-2	60.6	67.9
ComParE	1e-1	60.3	67.9
ComParE	1.0	60.4	67.9
eGEMAPS	1e-4	48.9	53.4
eGEMAPS	1e-3	53.9	65.5
eGEMAPS	1e-2	54.6	65.6
eGEMAPS	1e-1	54.0	64.1
eGEMAPS	1.0	52.6	64.8

Unweighted Average Recall (UAR, %) for the binary irony classification

Experiments Sentiment classification

- Two people labeled the text data and discussed divergent cases
- Split 750 training/165 test
- Different complexity of the linear SVM
- Out of the box 33.93 % error rate
- After tuning 27.27 % error rate.



Wrap-up/Outlook

- We presented an app and a dataset about ironic voices
- Enhance the textual sentiment classifier by machine learning.
- Investigate the influence of ASR errors on the sentiment classification
- Differentiate between categories of vocal and non-vocal expressions of irony
- Add more modalities, for example facial recognition.
- Investigate in how far prosodic expression of irony is culture dependent
- Validate hypotheses by re-synthesis experiments

