Emotional Machines
an overview

Dr. Felix Burkhardt

Emotional Machines
OUTLINE

- Why it is important?
- What is an emotion?
- How can it be recognized?
- How can it be synthesized?
- What are the main usecases?
- How does the market look?

Emotional Machines
TRENDS

- **Ubiquitous computing** accessible via
- Smart mobile devices: phones, glasses, watches, t-shirts, implants, etc.
- Home automation: central intelligence controlling media, communication, environment
- Aging society gets supported by technological interfaces

- **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

What is an emotion?

Emotional machines
EMOTIONS AND INTELLIGENCE

- “For decades, biologists spurned **emotion and feeling** as uninteresting. But Antonio Damasio demonstrated that they are **central to the life-regulating processes** of almost all living creatures.”
- Brain injuries specific to emotional processing robbed people of their **capacity to make decisions**, see the bigger picture, exercise common sense
- In opposition to Descartes, body and mind are not separated

Emotional machines
CATEGORIES

- …everyone except a psychologist knows what an emotion is (Young 1973)
- Charles Darwin: The Expression of the Emotions in Man and Animals
- The big four:
  - **Anger**
  - **Sadness**
  - **Joy**
  - **Fear**
- Needed to survive and „culturally universal“
- Many more catgorical models exist, e.g. Ekman‘s six ot Plutchik‘s emotion weel

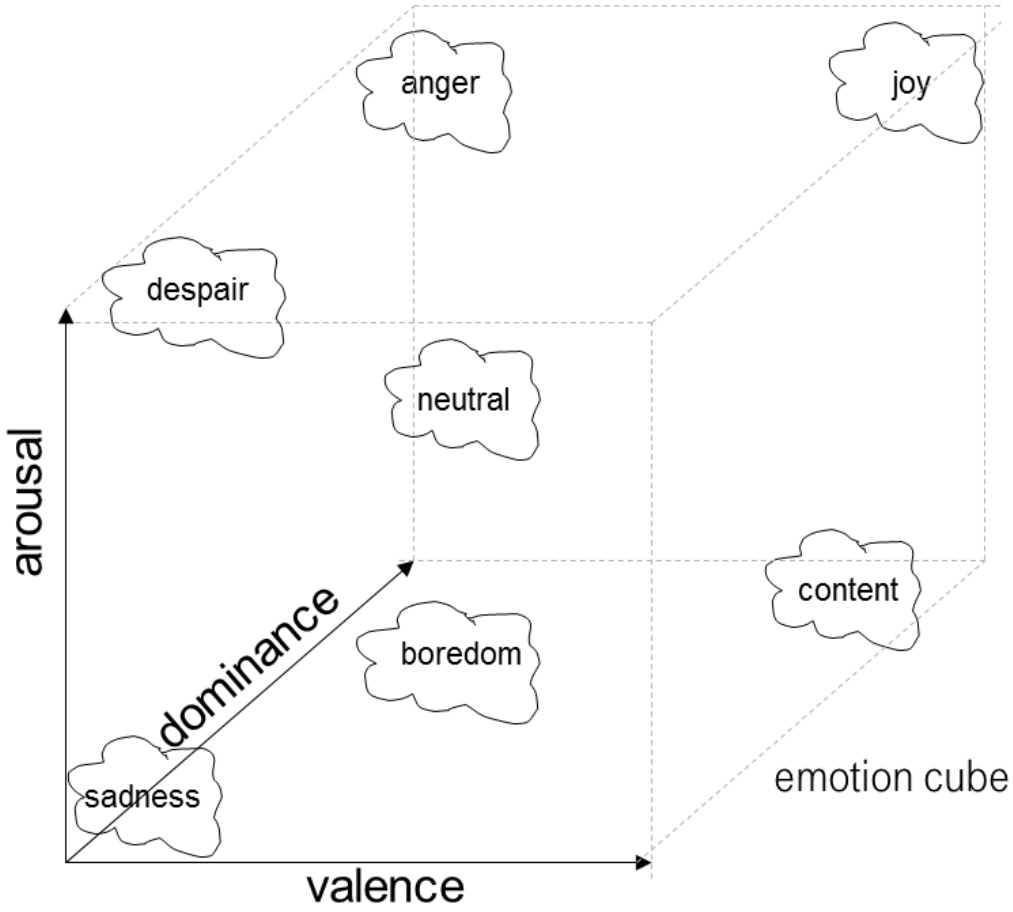
Emotions as characters in Pixar‘s „Inside Out“ (anger, fear, joy, envy, sadness)

Emotional machines
PANKSEPP‘S SEVEN PRIMAL EMOTIONS

- Jaak Panksepp was a neuro-scientist who suggested seven emotion categories in men an animals that can be localized in the brain.
- Search (anticipation, desire)
- Rage ((frustration, body surface irritation, restraint, indignation)
- Fear (pain, threat, foreboding)
- Panic/Loss ((separation distress, social loss, grief, loneliness)
- Play ((rough-and tumble carefree play, joy)
- Lust (copulation, mating)
- Care ((maternal nurturance)

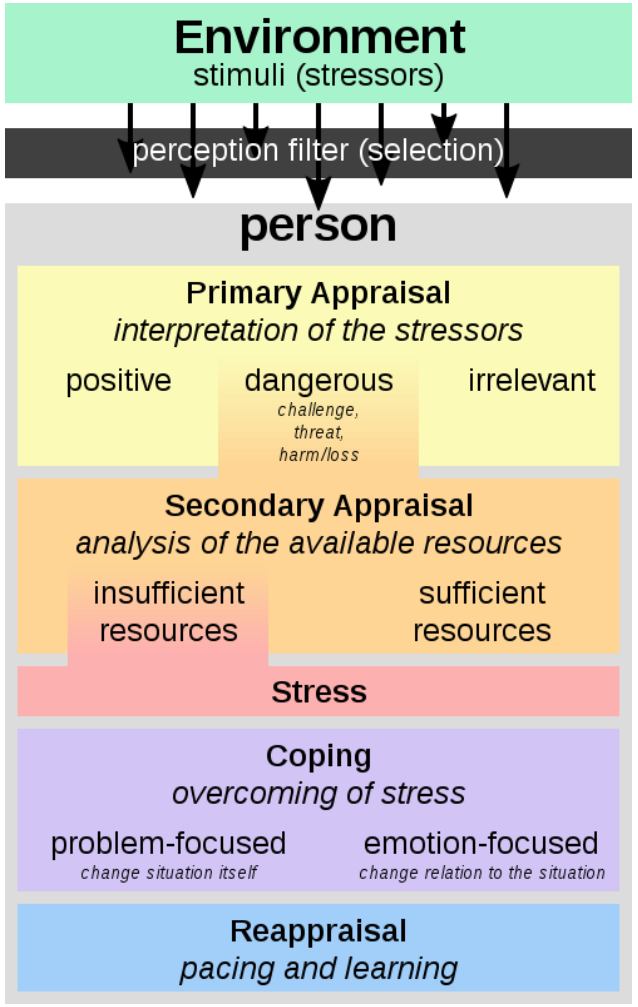
Emotional machines
DIMENSIONAL MODELS

- Dimensions consider an emotion as a point in an n-dimensional emotion space.
- One of the most well-known spaces is the PAD-space:
  - **Pleasure (valence)**
  - **Arousal (activation)**
  - **Dominance**
- Specific dimensions are better recognized by different modalities, e.g. activation in the speech but valence in the mimics



Emotional machines
APPRAISAL THEORY

- Appraisal theory means that emotions are extracted from our evaluations (appraisals or estimates) of events that cause specific reactions in different people.
- E.g. Scherer's multi-level sequential check model
- Three levels of processing are: innate (sensory-motor), learned (schema-based), and deliberate (conceptual)

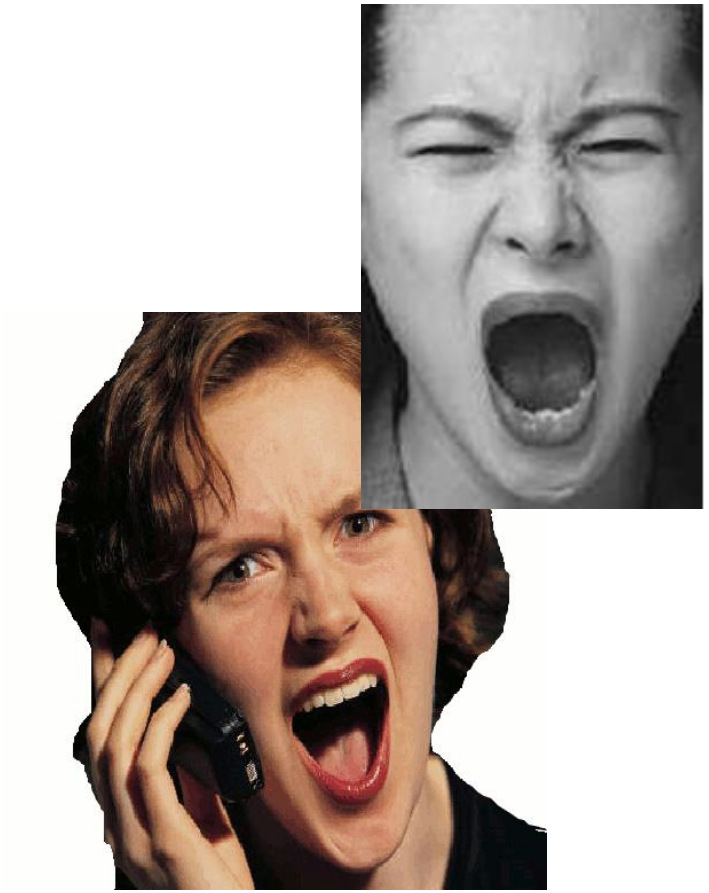


https://en.wikipedia.org/wiki/Appraisal_theory

http://emotionresearcher.com/the-component-process-model-of-emotion-and-the-power-of-coincidences/

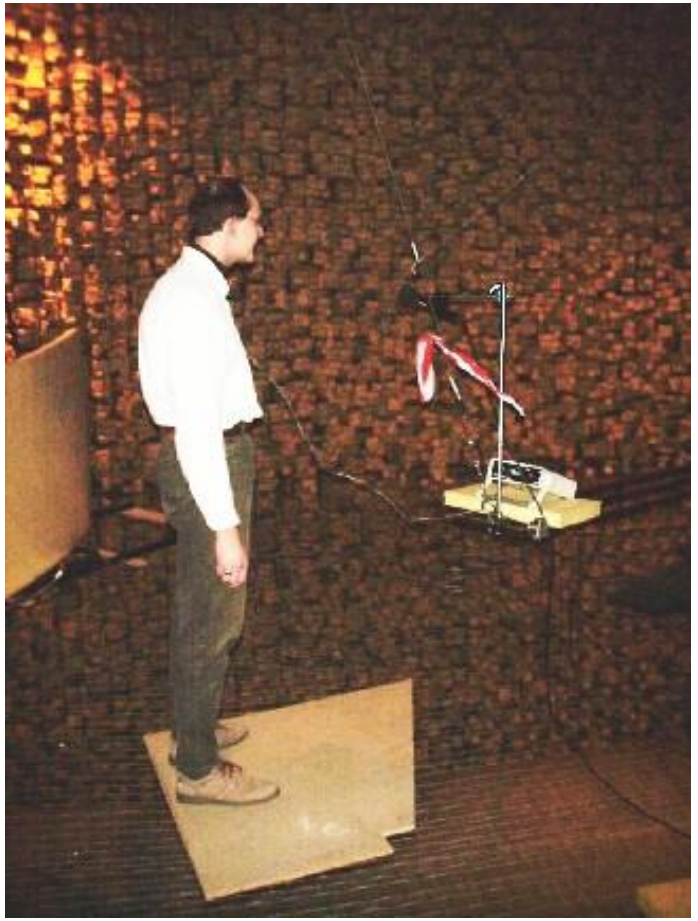
Emotional machines
MODALITIES

- User introspection: e.g. Emoticon, press button etc
- Text: sentiment analysis
- Audio: speech, extralinguistics
- Video: facial expression, gestures, posture
- Physiology: respiration rate, blood pressure, skin conductivity, neuronal activity, speech (held vowels)
- Behaviour, e.g. switched room often, typing speed
- Context: localization, weather, time of day, other people‘s moods etc.



Emotional Machines
WHERE DOES THE DATA COMES FROM?

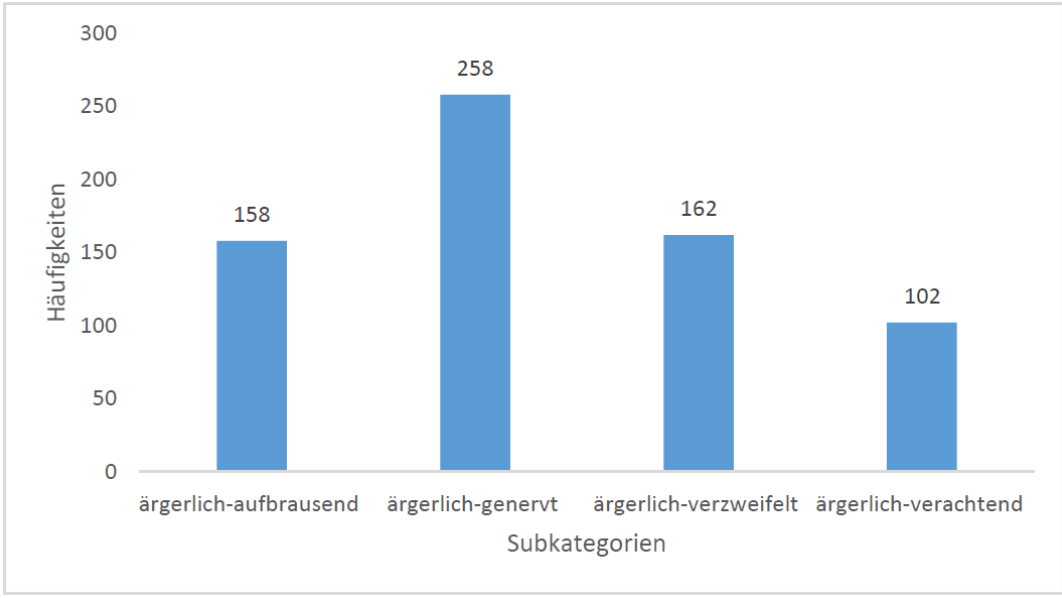
- Ideally from the application
- From an application similar to the target
- From Wizard of Oz scenario
- From field recordings („Vera am Mittag“)
- From induced emotions („Lost luggage“, „Aibo“)
- From actors



Felix Burkhardt, Astrid Paeschke, Miriam Rolfes, Walther F. Sendlmeier and Benjamin Weiss: A Database of German Emotional Speech, *Proc. Interspeech* 2005

Emotional Machines
WHICH EMOTION EXACTLY IS MEANT?

- 20 listeners judged the distribution of four anger sub-categories in a German customer voice portal data.
- The choice of subcategories was based on a listener experiment undertaken by Banse and Scherer (hot, cold anger, despair, contempt).
- All subcategories were attributed with a Kappa value of about 0.3 and they have distinct acoustic features.



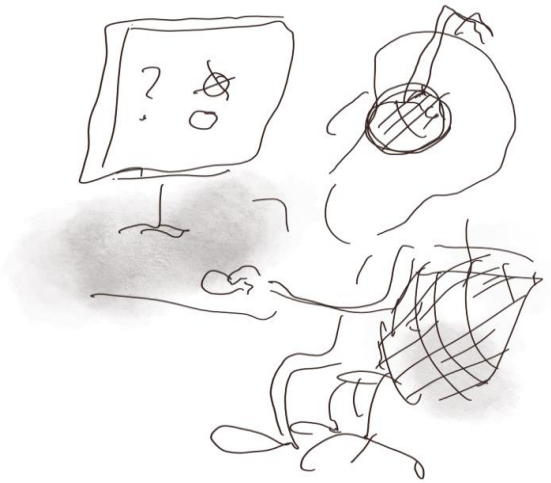
Maximilan Schmitt Masterthesis TU-Berlin 2018

Banse, Rainer; Scherer, Klaus (1996): Acoustic Profiles in Vocal Emotion Expression

Emotional machines
WHERE IS THE GROUND TRUTH?

- Five human labelers annotated the emotional content of textual data using four categories.
- A machine algorithm did the same classification.
- “majority” means the majority voting of the human labelers.
- The chart shows the Cohen’s kappa values for the so-called “inter rater agreement”, i.e. how much each rater agrees with all other raters.
- It is a statistical measure relating the agreement to the agreement with chance level, given the categories and their frequencies.

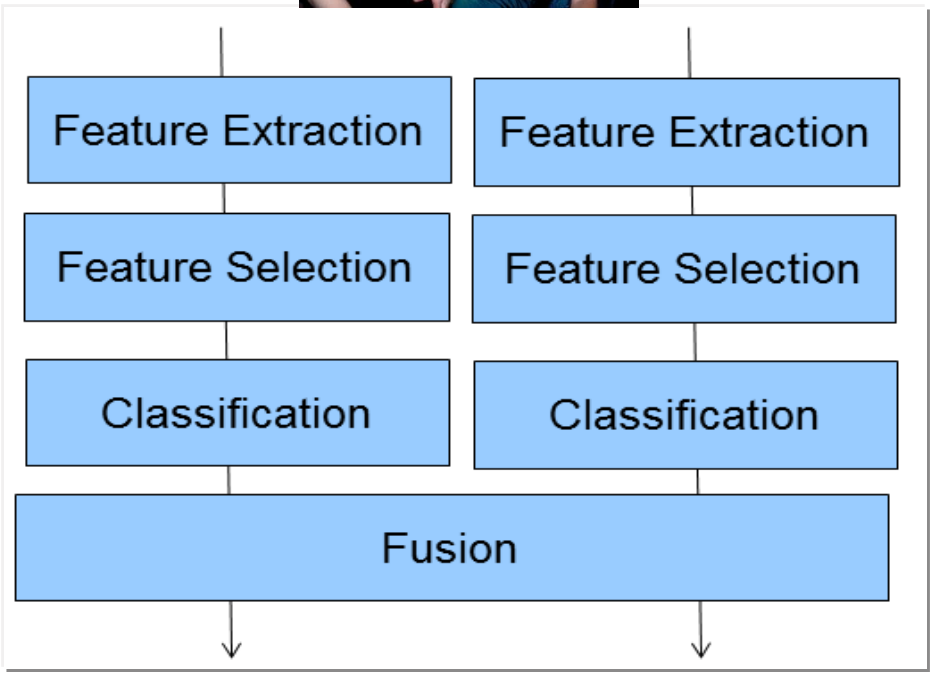
	labeler A	labeler B	labeler C	labeler D	labeler E	majority	machine
labeler A	1,00	0,20	0,19	0,10	0,24	0,27	0,15
labeler B		1,00	0,79	0,46	0,15	0,81	0,15
labeler C			1,00	0,47	0,19	0,83	0,14
labeler D				1,00	0,09	0,52	0,07
labeler E					1,00	0,29	0,10
majority						1,00	0,17
machine							1,00



How can it be recognized?

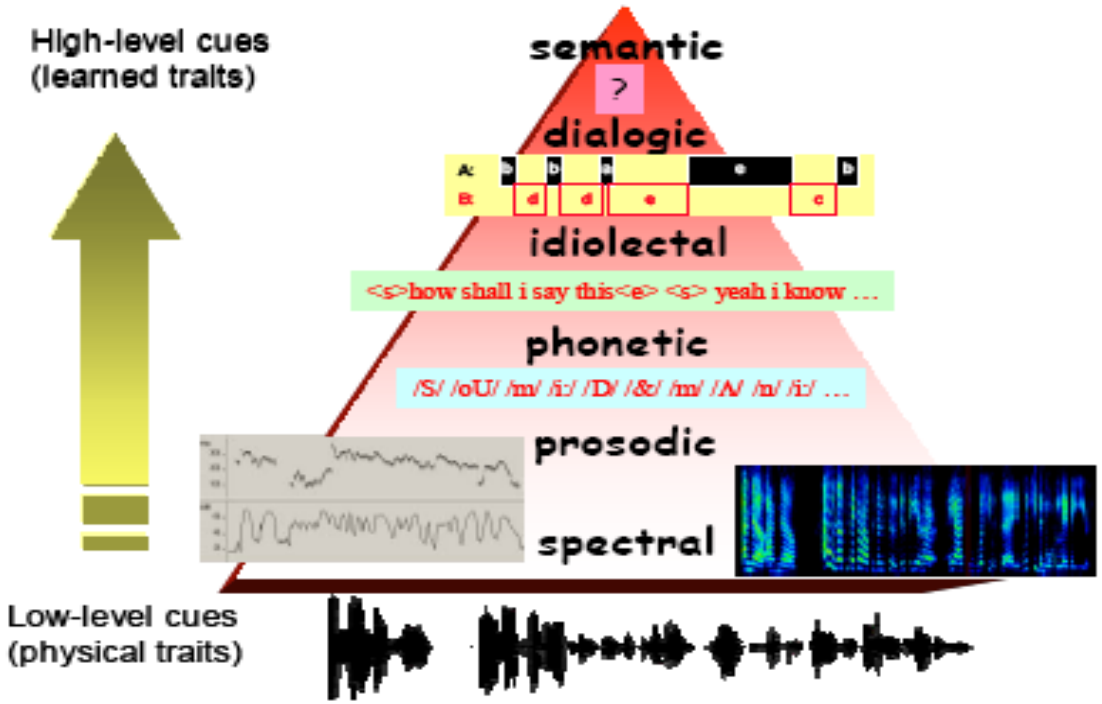
Emotional Machines
RECOGNITION BY STATISTICAL CLASSIFICATION

- Basic approach:
  - extract features,
  - select best ones,
  - classify features,
  - fuse classifier outputs
- Classifiers: **Gaussian Mixture Models**: model training data as Gaussian densities, **Artificial Neural Networks** (ANN), e.g. Multi Layer Perceptron, **Support Vector Machines** (SVM): use „kernel functions“ to separate non-linear decision boundaries, **Classification and Regression Trees** (CART), **Hidden Markov Models** (HMMs) used to model temporal structure
- Deep Neural Networks can operate directly on signal: end-to-end approach



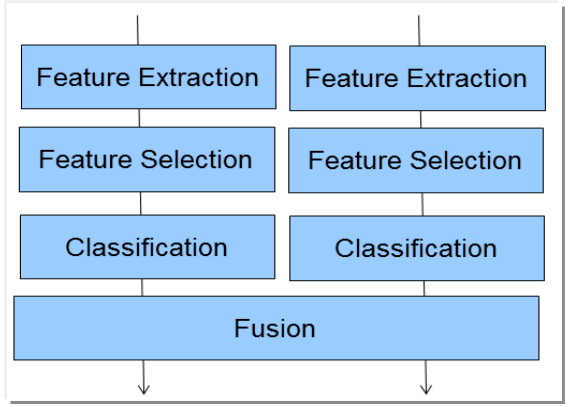
Emotional Machines
EXAMPLE: SPEECH FEATURES FOR CLASSIFICATION

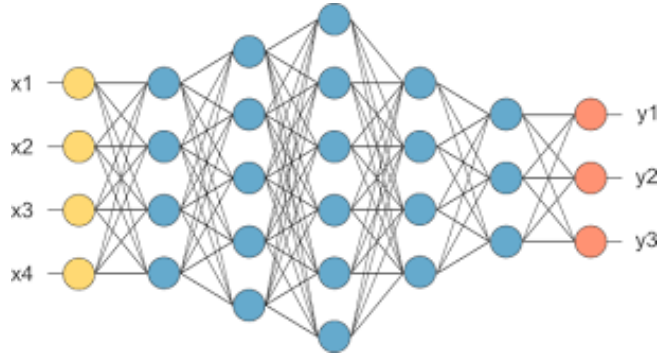
- Telephone speech: digital signal
- Base features, e.g. MFCCs
- Prosody
- Microprosody, e.g. jitter/shimmer
- Pitch durations and energy
- Functionals: Mean, max/min, deviation, regression, …,
- Text (as recognized by Automatic Speech Recognition)



Emotional Machines
EXAMPLE: IVR ANGER DATABASE

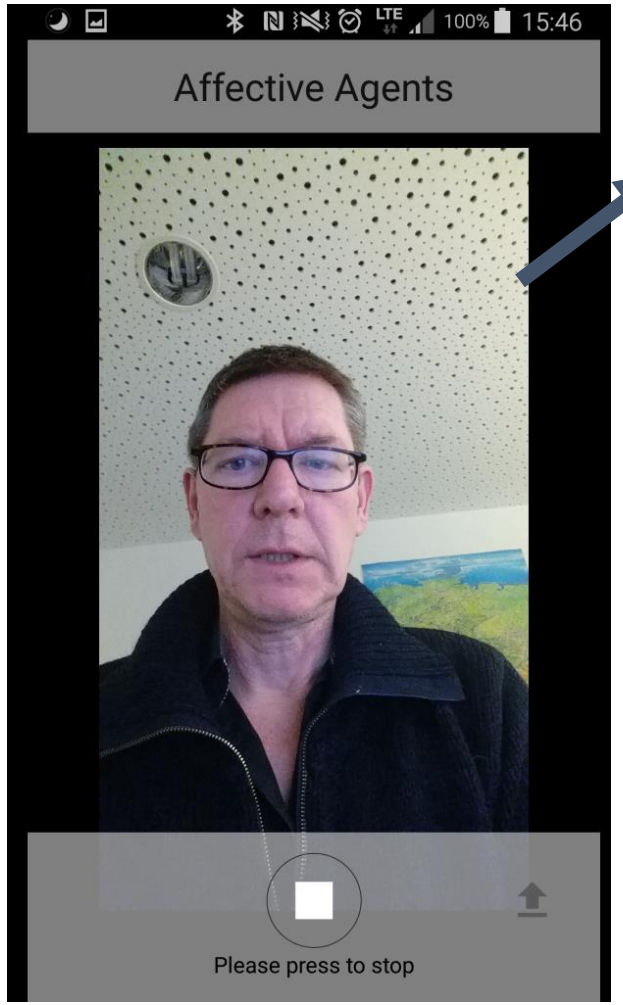
- A) Basic approach:
  - extract features, select best ones, classify features, fuse classifier outputs
  - Apply different classifiers like Gaussian Mixture Models (GMMs), Support Vector Machines (SVM)
- B) Deep Neural Networks
  - might operate directly on base features: end-to-end approach
- Results:
- Compared to 2009, the results improved 10% in average recall: 80.1 % with a bidirectional LSTM network (ComParE feature set) vs. 69.0 % with SVMs on a reduced feature set.

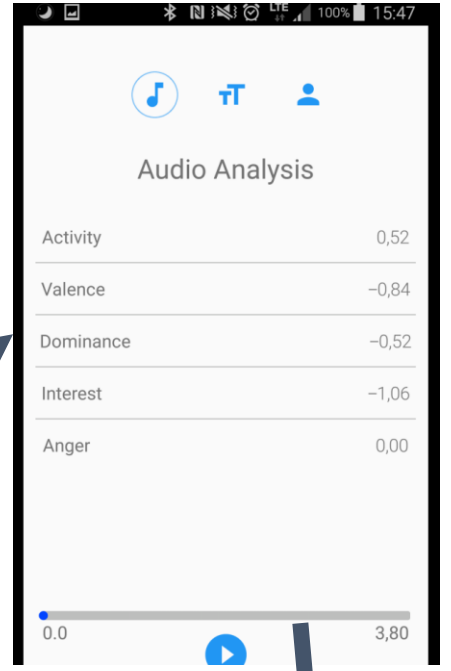


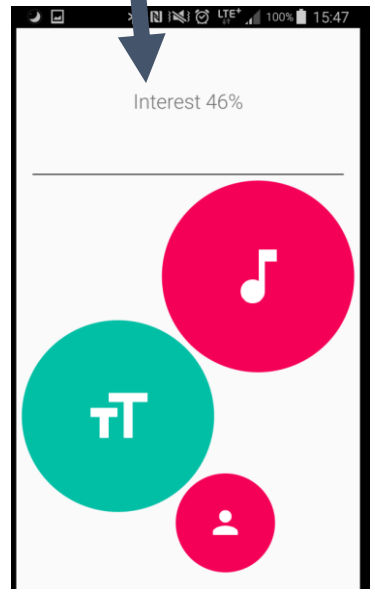


Emotional machines
DEMONSTRATOR

- Press the „record“ button at the bottom
- Look into the camera and say something of about three seconds length
- Try to smile while you speak
- Then press the stop bottom at the bottom



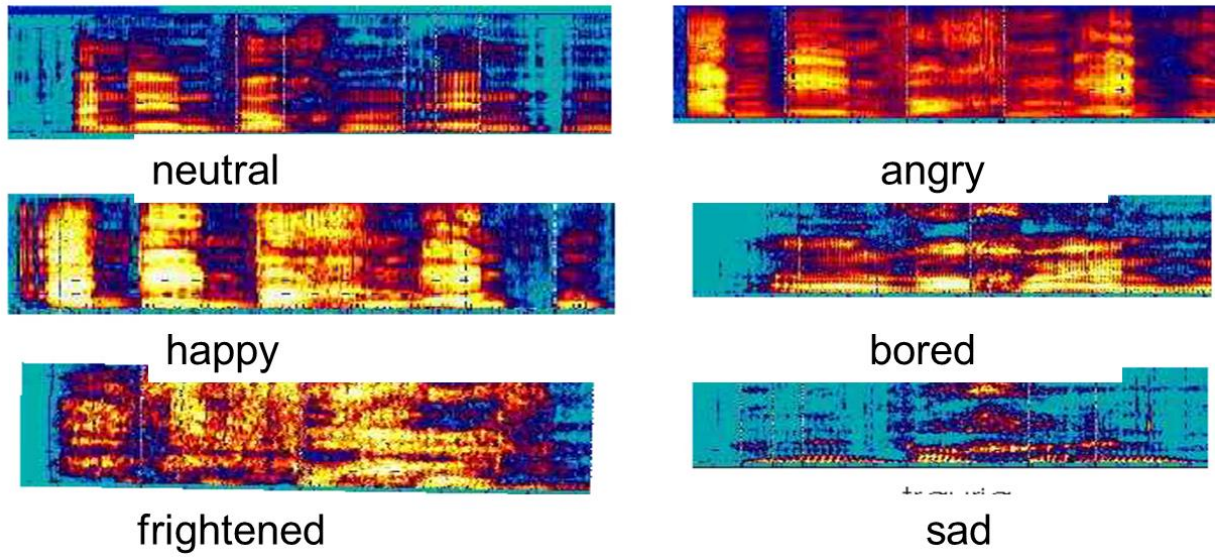




How can it be synthesized?

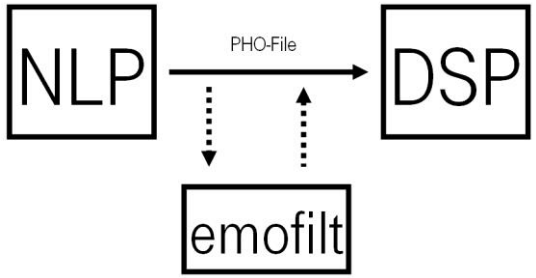
Emotional machines
EXAMPLE: SIMULATION OF AFFECTIVE BEHAVIOUR BY SPEECH SYNTHESIS

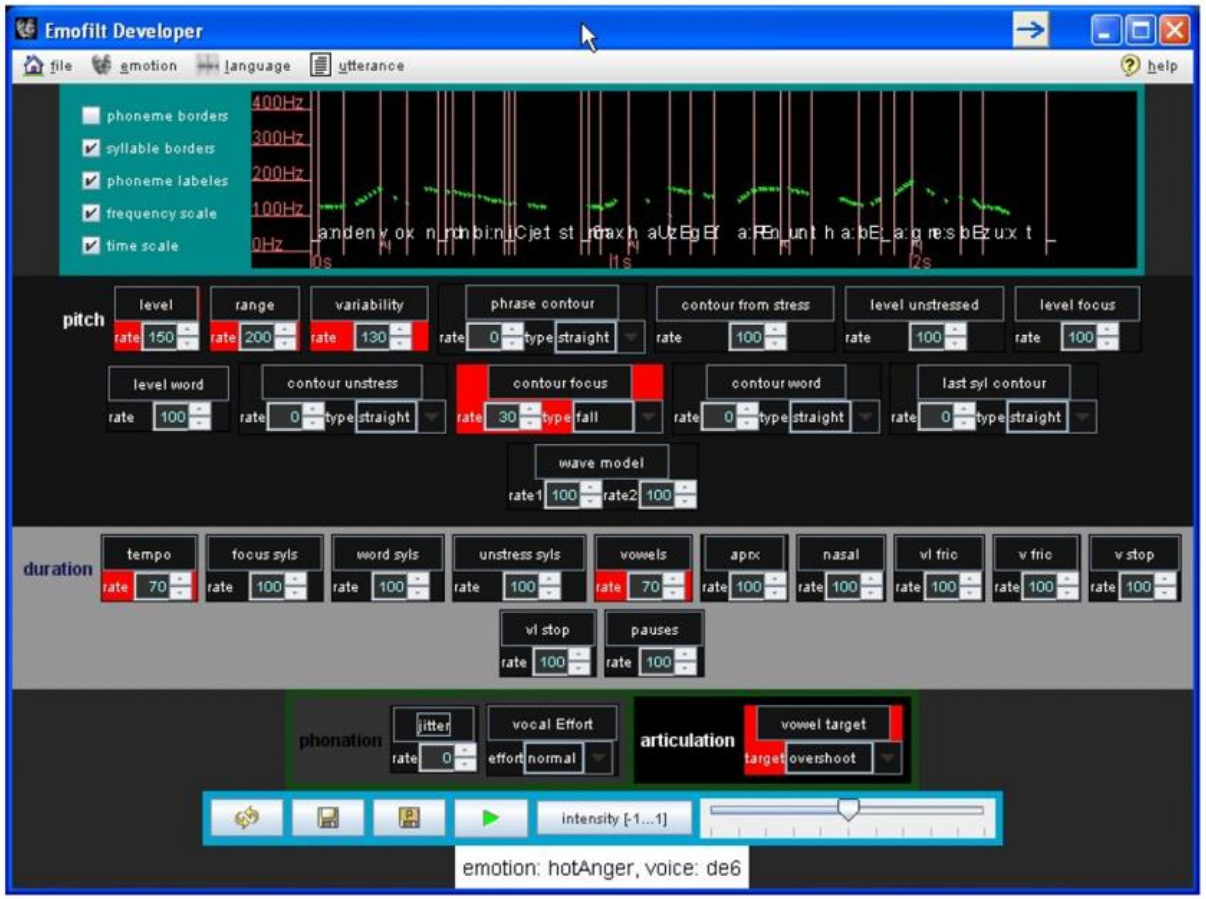
- DNN Synthesis: TTS with neural nets has been done since many decades. They replace the HMM approach to predict the best acoustic parameters for a given sequence of symbols representing text.
- HMM Synthesis: Synthesis based on Hidden Markov Models, a statistical approach to model the transition probabilities of the acoustic parameters based on the speech to be generated.
- Non-uniform unit-selection: Best chunks of speech get concatenated, minimizng a double cost-function: best fit to next unit and best fit to target prosody.
- Diphone-synthesis: Speech concatenated from diphone-units (two-phone combinations), prosody-fitting done by signal-manipulation
- Formant-synthesis: Speech synthesized by physical models (formants are resonance frequencies in vocal-tract).
- Articulatory synthesis models the human voice tract by mathematical models. Modeling dynamics is difficult.



Emotional machines
EXAMPLE EMOFILT

- Emofilt is a tool to transform the prosody of a given utterance in order to simulate emotional expression
- It is based on Mbrola for speech generation and an arbitrary phonemization generator like MARY or Txt2Pho
- Mbrola is a diphone synthesizer from the University of Mons with databases for 34 languages





What are the main usecases?

Emotional machines
FIVE TYPES OF APPLICATIONS

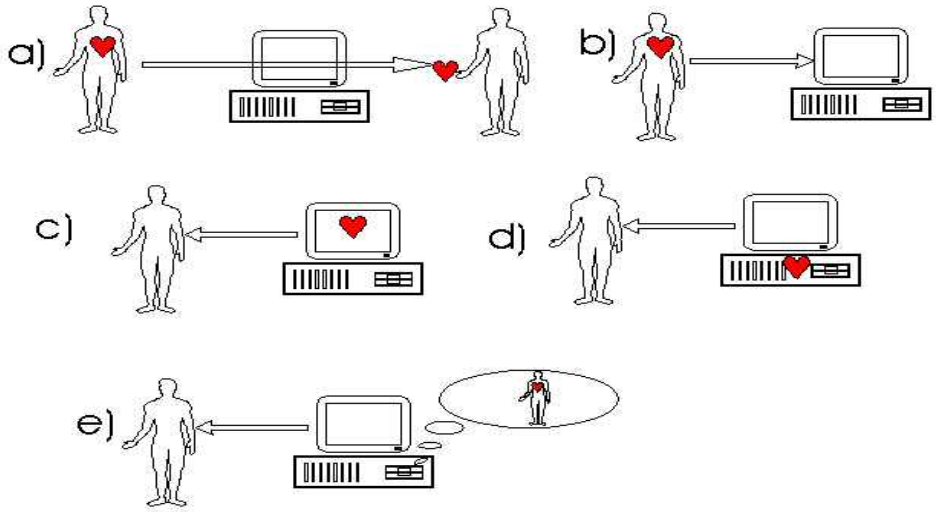
a)**Mediated emotion**

b)Affect **recognition**

c)Affect **simulation**

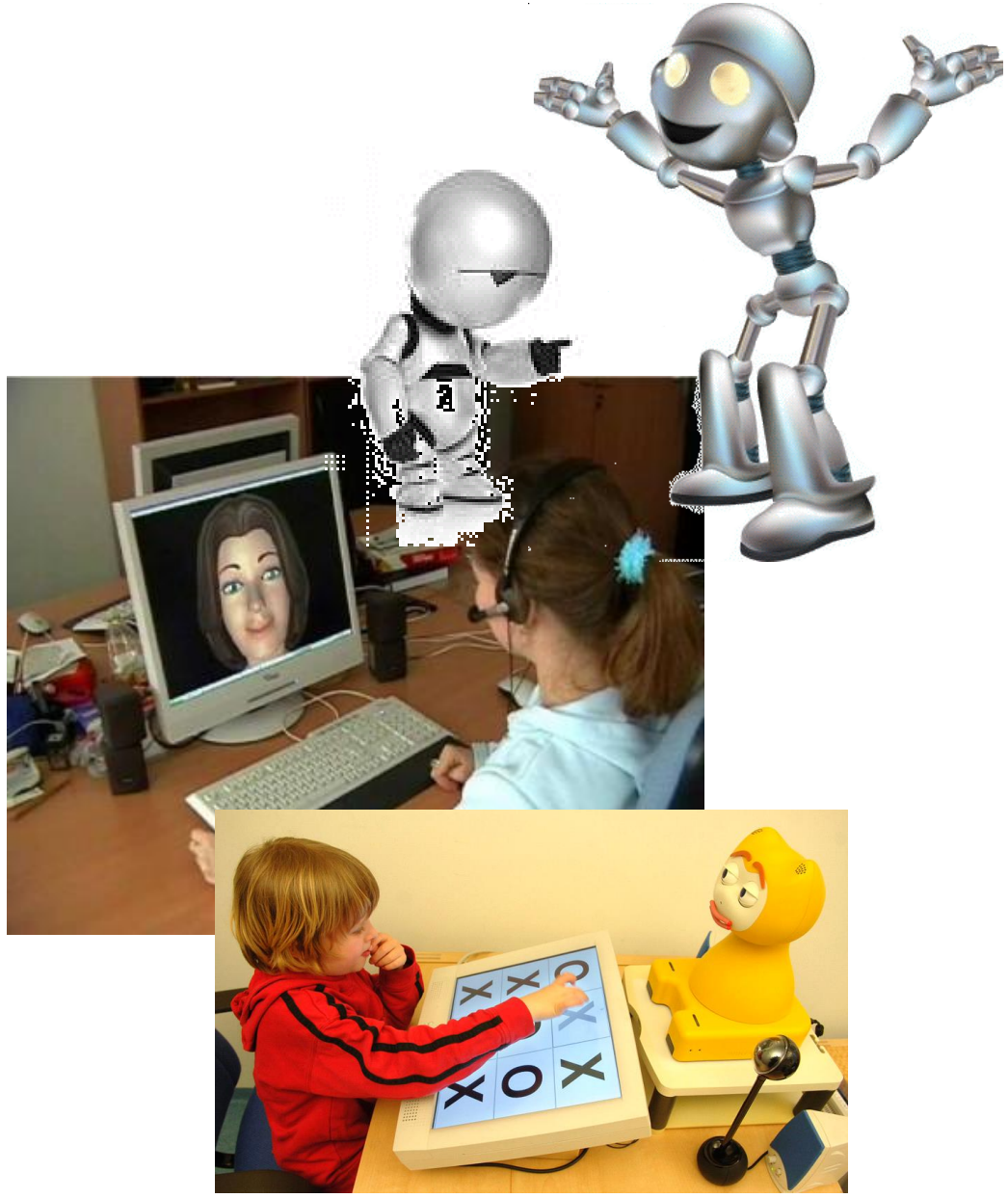
d)Modeling emotional **intelligence**

e)Modeling human emotional **behavior**



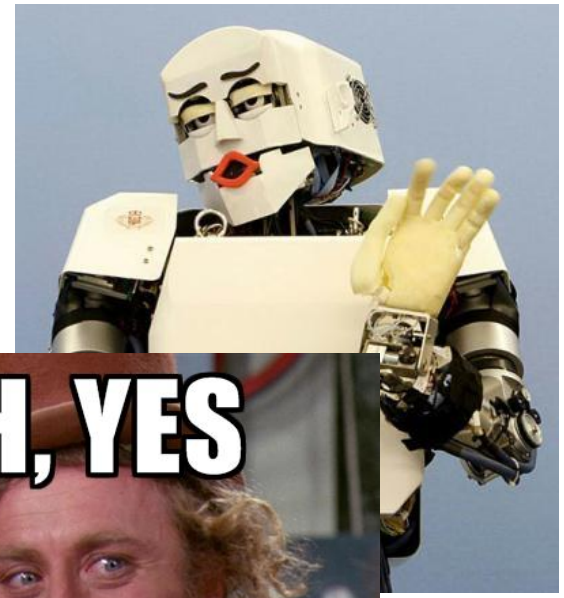
Emotional machines
SOME APPLICATIONS IN GENERAL

- Market research: analyze users response on new products
- Emotional Diary
- Comedy club charges per laugh, price advertising depending on how people respond to it
- Detect stress/interest in tutoring application
- Make artificial beings believable



Emotional machines
APPLICATIONS IN THE TELEKOM DOMAIN

- Irony / Sarcasm detection: To analyze user opinion this is still a big probl
- Call center support: distribute aggressive callers, support training.
- Automated dialog support: Anger detection can be used for churn prevention or for automatic quality monitoring.
- Emotional Chat: Facilitate emotional computer mediated communication.
- Emotion-aware Surrounding: Computer controlled environment that adapts automatically on the user's mood.
- Search–by-emotion, e.g. Entertain product.
- Believable Agent: The naturalness of an artificial 'being' and the appearance of intelligence is highly altered by emotional expressions.
- Artificial intelligence models, use emotions for motivation modeling



How does the market look?

Emotional machines
W3C STANDARD FOR EMOTION ANNOTATION

- As the Web is becoming ubiquitous, interactive, and multimodal, technology needs to deal increasingly with human factors, including emotions.
- The specification of Emotion Markup Language 1.0 aims to strike a balance between practical applicability and scientific well-foundedness.
- The language is conceived as a "plug-in" language suitable for use in three different areas:
- manual annotation of data
- automatic recognition of emotion-related states from user behavior
- generation of emotion-related system behavior

Emotion Markup Language (EmotionML) 1.0

W3C Recommendation 22 May 2014

Authors:

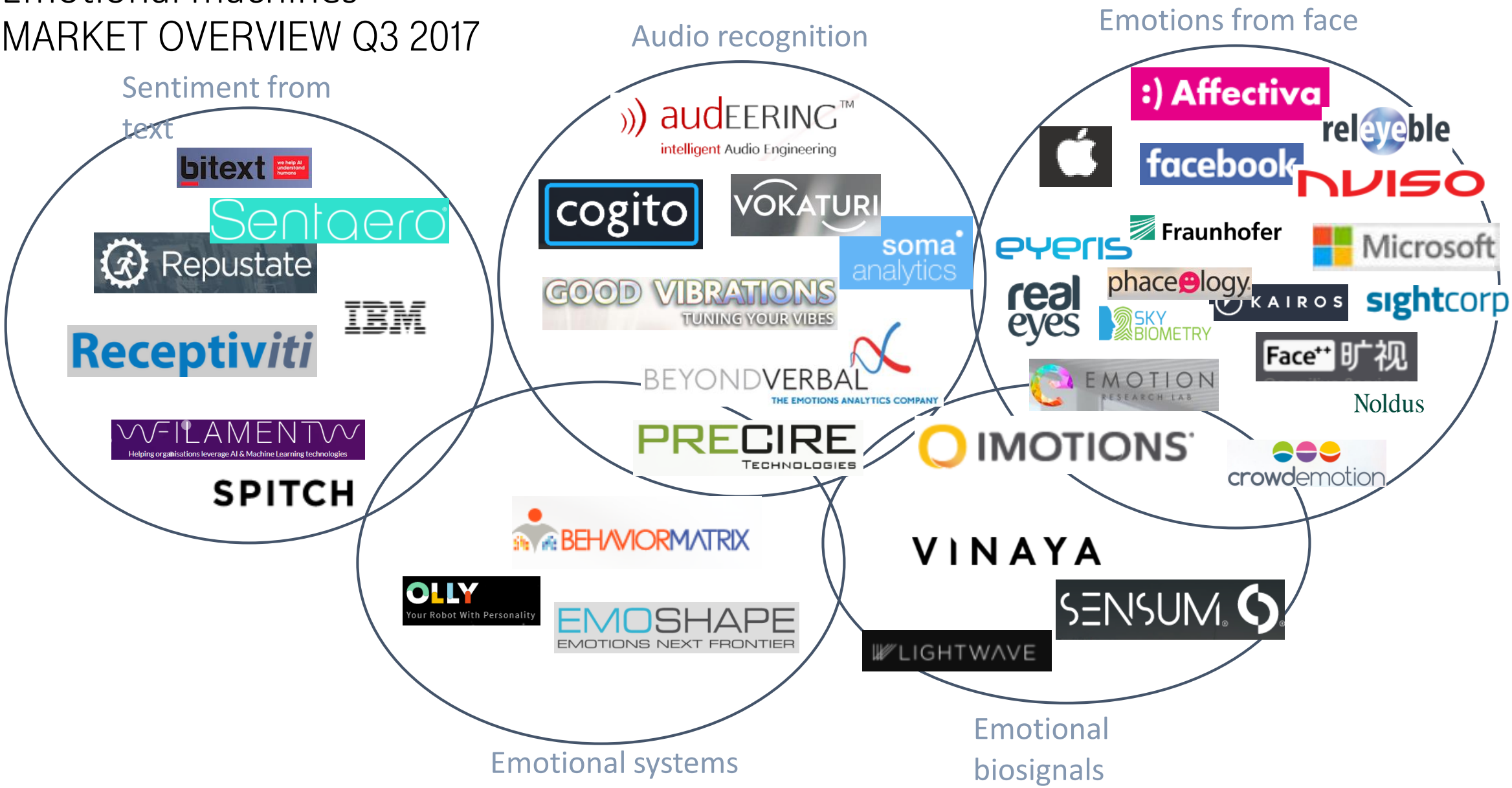
Felix Burkhardt (Deutsche Telekom AG)
Marc Schröder (until July 2012, while with DFKI GmbH)

(in alphabetic order)
Paolo Baggia (while at Loquendo, currently Nuance Communications)
Catherine Pelachaud (Telecom ParisTech)
Christian Peter (Fraunhofer Gesellschaft)
Enrico Zovato (while at Loquendo, currently Nuance Communications)


```
<emotion action-tendency-set="http://www.example.com/custom/action/robot.xml#voc">
    <action-tendency name="charge-battery" value="0.9"/> <!-- need to charge battery soon -->
    <action-tendency name="pickup-boxes" value="0.3"/>   <!-- feeling tired, avoid work -->
</emotion>
```

https://www.w3.org/TR/emotionml/

Emotional machines
MARKET OVERVIEW Q3 2017



Emotional machines
WRAP UP

- Emotional processing comes with **pervasive computing**
- It can be used with intuitive interfaces, **more natural mediated communication and sophisticated AI models**
- Emotional categories contrast with more complex models
- The nature of **emotion is dictated by application**
- The market is growing, all the **big players are already there**
- **Facial detection comes first** because emotional expression is most easily detected in the mimics