Predicting emotional speech with Nkululeko, a framework to automize machine learning experiments

Felix Burkhardt
audEERING GmbH, TU-Berlin

Outline

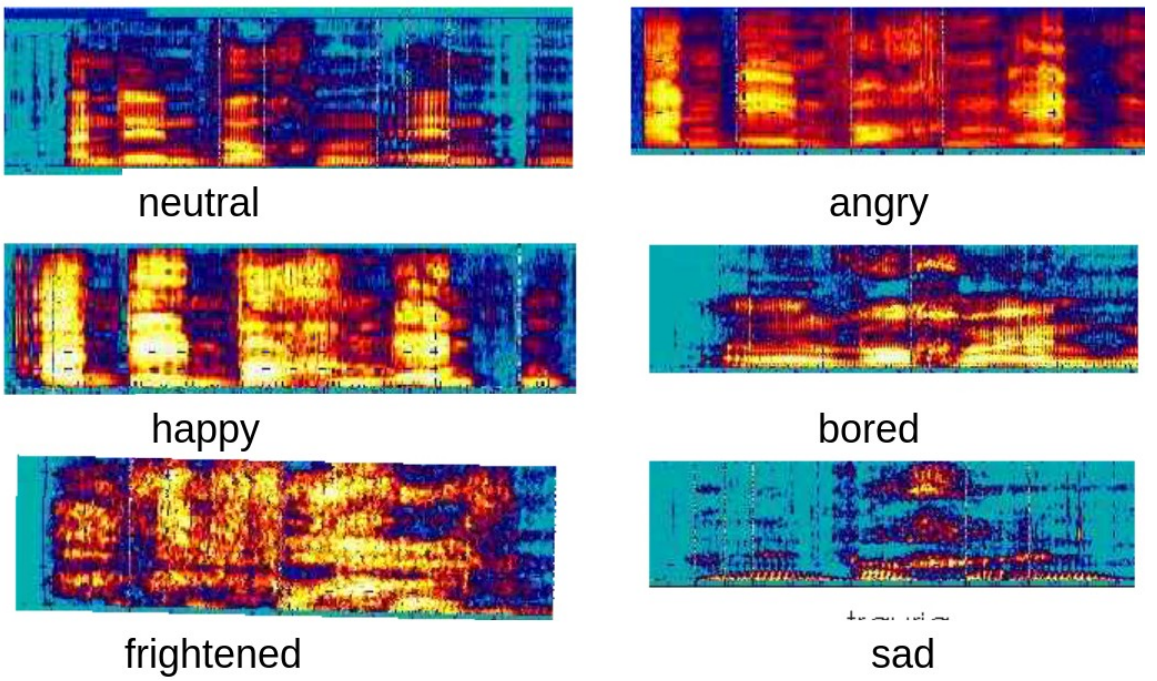
- How is emotion expressed in speech?
- what is Nkululeko
- how to use it
- example experiments

What the voice reveals

- sex
- age
- sociolect
- dialect
- health
- personality
- mood
- relationship
- weight, length
- heart rate
- …

How is emotion expressed in speech

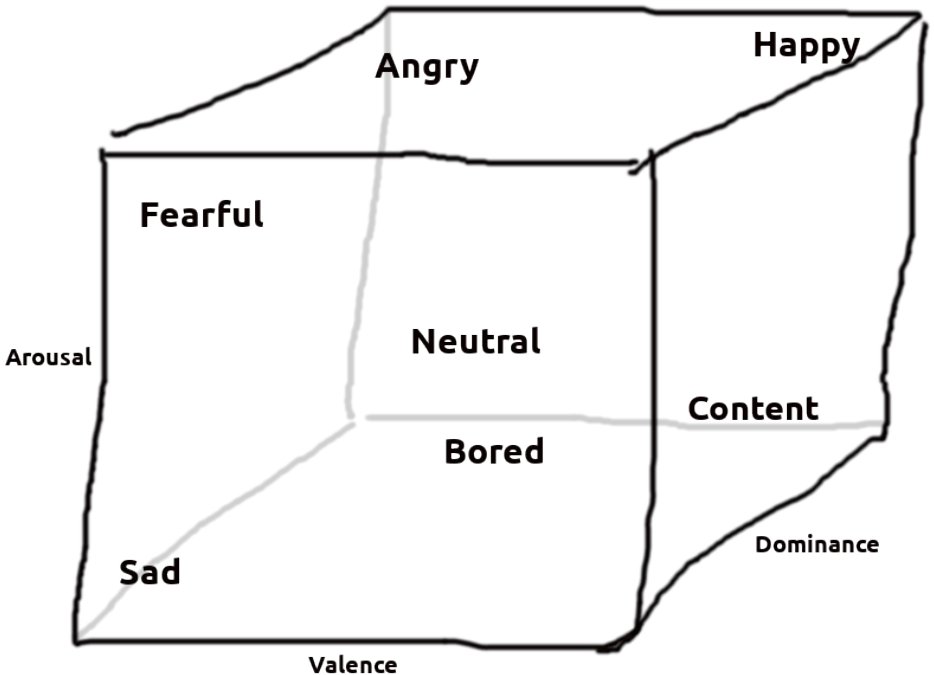
- Emotions influence body, body influences emotions...



F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss: A Database of German Emotional Speech, Interspeech 2005

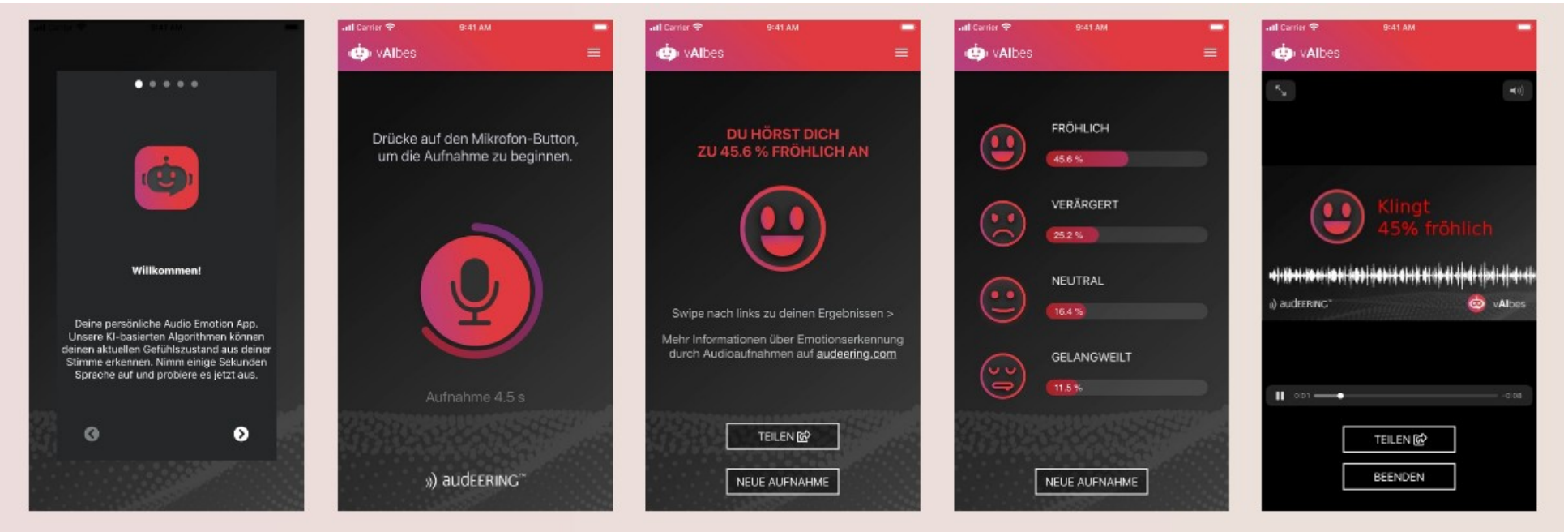
Emotion modeling

- Categories
  - Anger, sadness, joy, fear, ...
- Dimensions
  - Pleasure, arousal, dominance, ...
- Appraisals
  - novelty
  - pleasantness
  - goal
  - coping
  - norm compatibility (culturally dependent)



vaibes

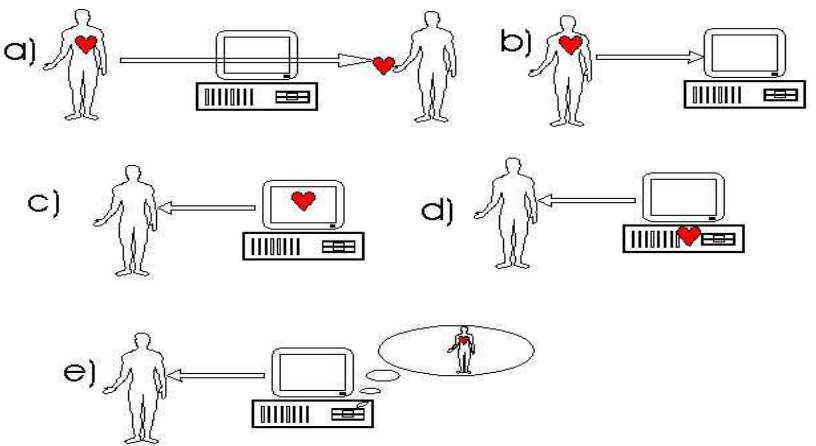
- App available in Google app store



Usecases

Emotions are a) social signal b) inner state

- Enhance dialog systems (Adaptation, Semantics)
- Monitor diseases (depression, schizophrenia, parkinson, MS)
- Make virtual beings more believable (gaming/training)
- Training/monitoring (panic in a crowd)



A. Batliner, F. Burkhardt, M. van Ballegooy, E. Nöth: A Taxonomy of Applications that Utilize Emotional Awareness, Proc. IS-LTC 2006

What is Nkululeko?

- A software written in Python hosted on github*
- A tool to do machine learning (ML) experiments on audio WITHOUT the need to program yourself
- Combinations of acoustic features and machine learners
- Data analysis

* https://github.com/felixbur/nkululeko

** blog.syntheticspeech.de/?s=nkululeko

Motivation

- With the success of Deep Learning, machine learning dominates science
- Empiricists sometimes struggle with programming
- Quick overview on databases
- Teaching students

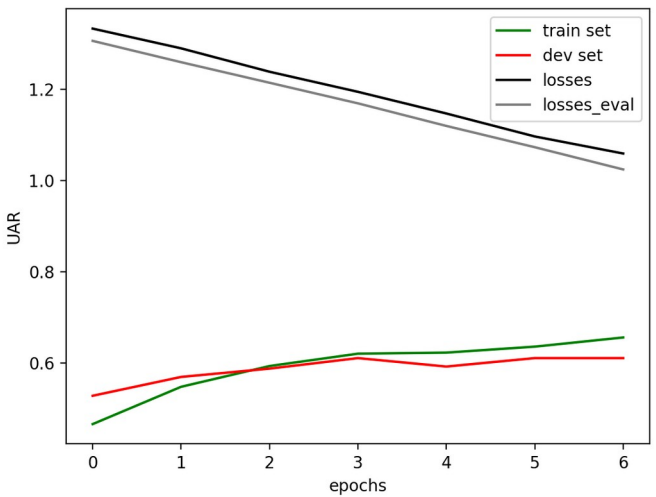
Input format

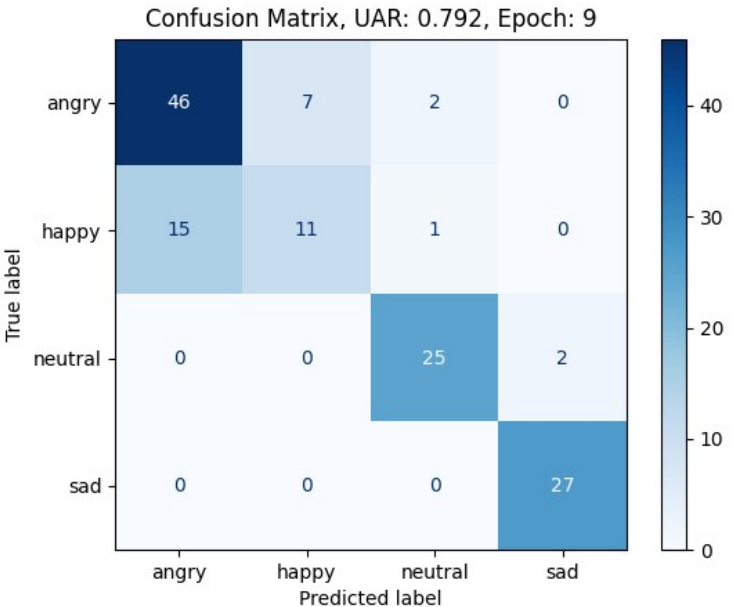
- Experiments are defined with configuration files

```
[EXP]
root = ./
name = exp_emodb
[DATA]
databases = ['emodb']
emodb = ./emodb/
emodb.split_strategy = speaker_split
target = emotion
labels = ['anger', 'boredom', 'disgust', 'fear']
[FEATS]
type = ['praat']
[MODEL]
type = svm
[EXPL]
model = tree
plot_tree = True
[PLOT]
combine_per_speaker = mode
```

Nkululeko modules: nkululeko

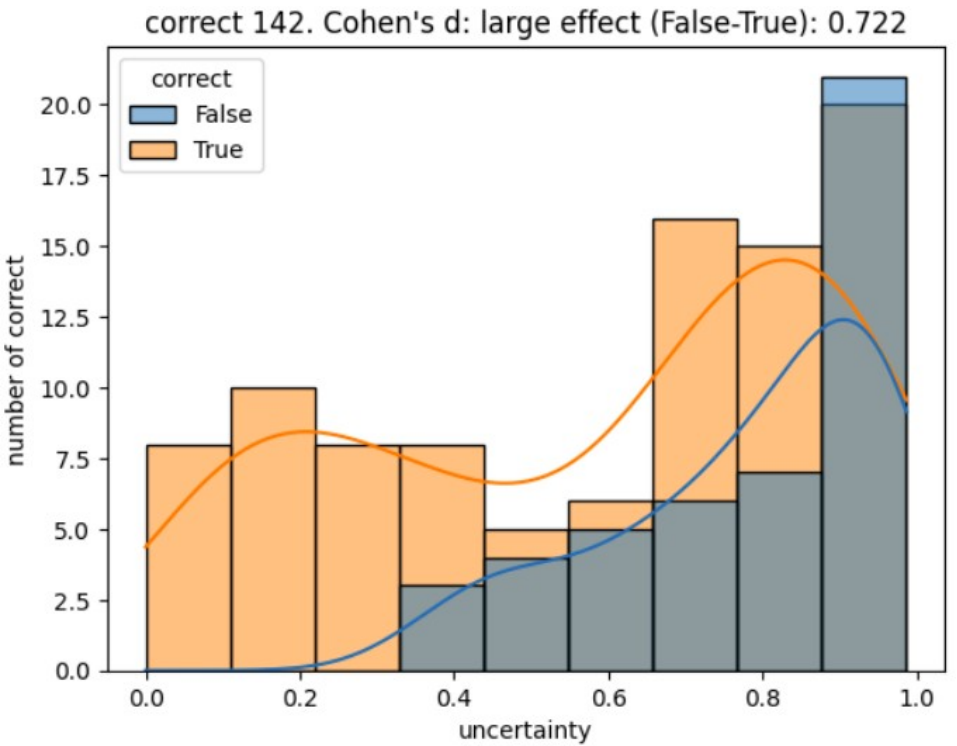
Do machine learning experiments, combining features and learners





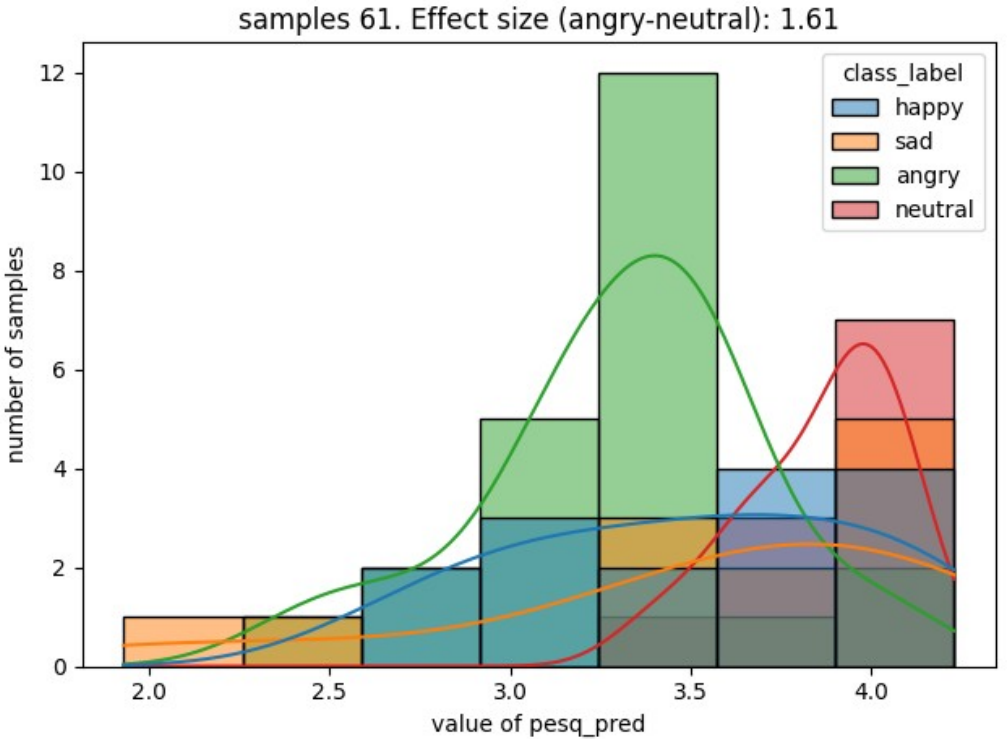
Nkululeko modules: nkululeko

Estimate uncertainty with entropy over class outputs per sample



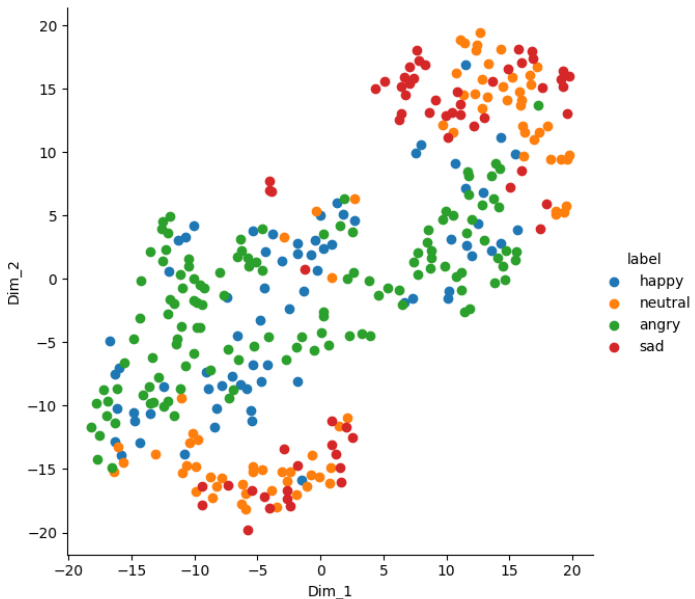
Nkululeko modules: predict

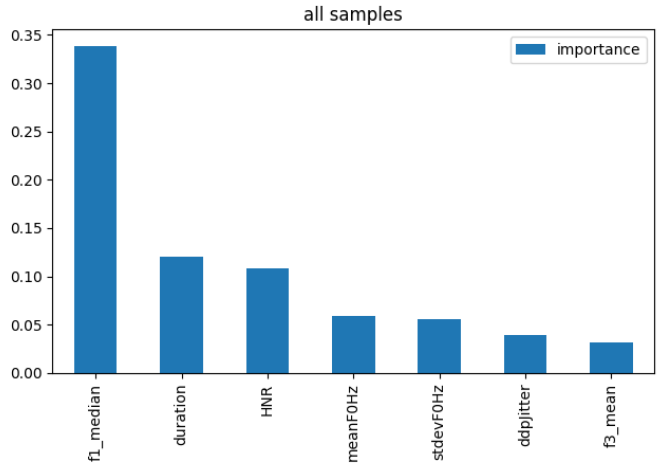
Predict features like SNR, MOS, arousal/valence, age/gender, with DNN models

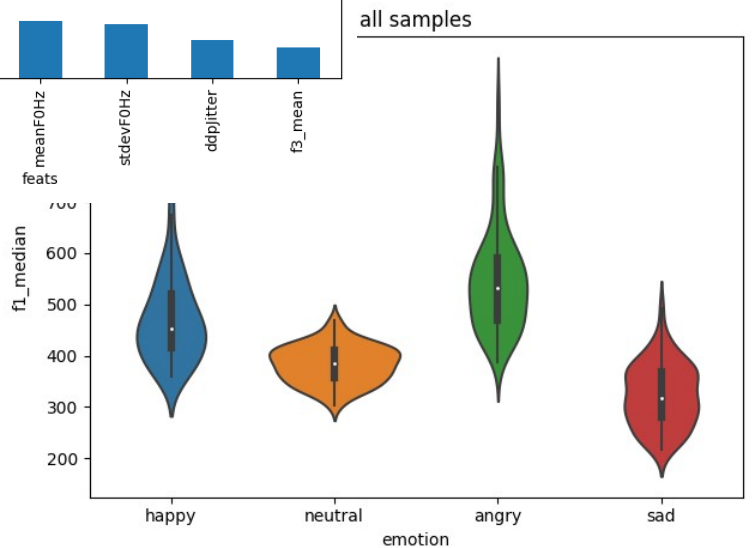


Nkululeko modules: explore

Perform data exploration





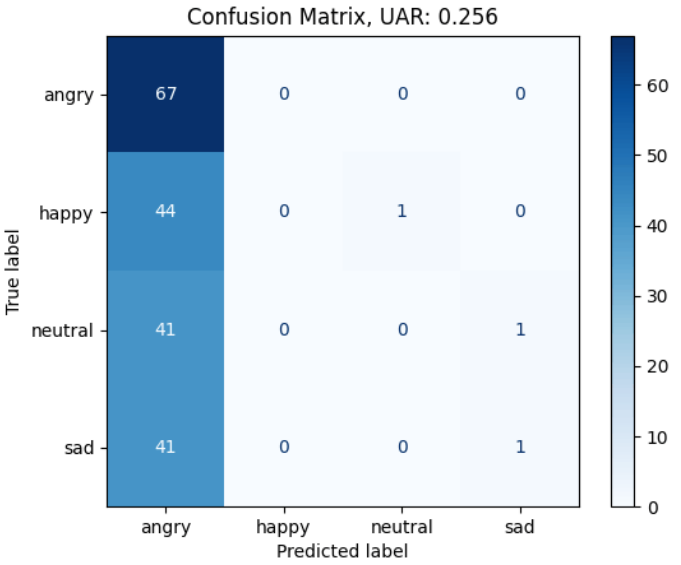


Nkululeko modules: segment/augment/resample

- nkululeko.augment: augment the current training data
- nkululeko.segment: segment a database based on VAD (voice activity detection)
- nkululeko.resample: filter and check on all sampling rates and change to 16kHz

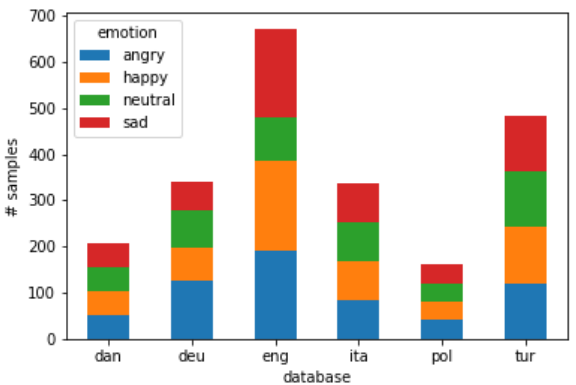
Visualizing results

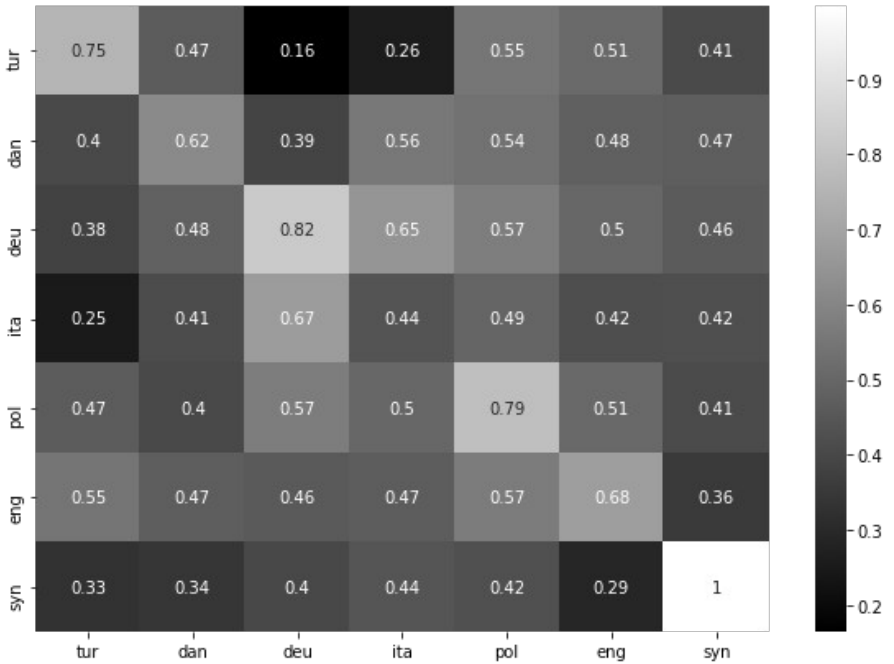
- Confusion matrices



Experiments

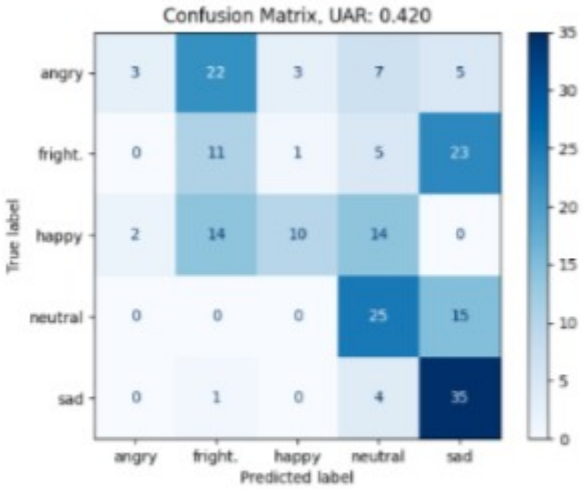
- 6 European acted emotional databases + synthesized emotion



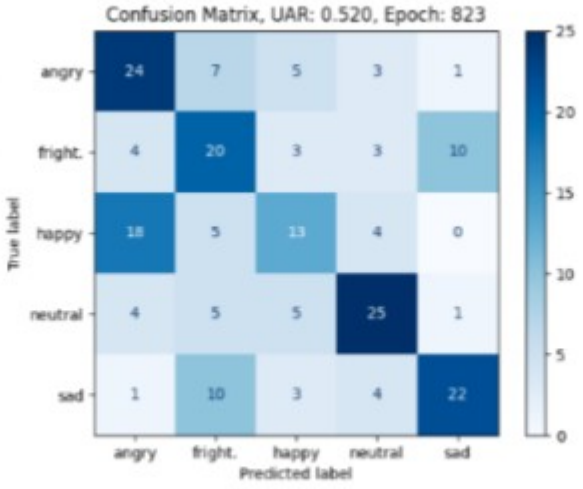


Burkhardt, F., Hacker, A., Reichel, U., Wierstorf, H., Eyben, F., and Schuller, B. (2022).
A comparative cross language view on acted databases portraying basic emotions utilising machine learning.
In Proceedings of LREC 2022

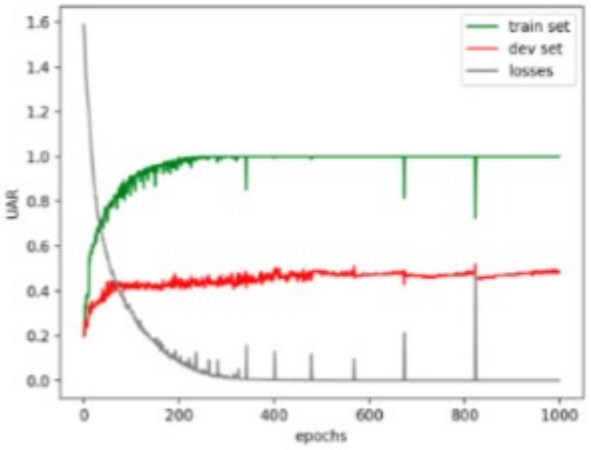
Experiments cont.



XGB with eGeMAPS



MLP with Wav2vec 2.0



Progression of epochs

- comparing expert with learned features on cross databases acted emotion learning
- Berlin EmoDB vs. Polish data

Wrap up

- Introduced Nkululeko
- A software to do machine learning experiments on spoken data without programming
- Combines features and learners
- Available open source with MIT license