COMPUTATIONAL PARALINGUISTICS

AND

WHAT WE MIGHT GET FROM PHONETICS / SPEECH SCIENCE

Anton Batliner
May 7th, 2015
NSF, Arlington
Paralinguistics: not what but how ⇒ the person(s) behind

The Interspeech Computational Paralinguistic Challenges

- 2009: emotion (children's speech)
- 2010: age & gender, affect (level of interest)
- 2011: intoxication (+/- alcoholised), sleepiness
- 2012: personality (big 5), likability, pathology
- 2013: social signals, conflict, emotion, autism
- 2014: physical load, cognitive load
- 2015: degree of nativeness, Parkinson's condition, eating condition

The Book

Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing
Phonetics (Speech Science)  

*phonetics/knowledge-based interpretation:* we don't really know what's happening because: only what we are looking for is what we get.

- small, laboratory, controlled
- manual (labels, segmentation)
- few (low resolution, high generalisation)
- basic, (M)anova, mixed models
- inferential, ⇒ significance
- description, explanation, models

Speech Processing  

*brute force:* we don't know what's happening but we know how good we can be (roughly).

- large, real-life
- automatic
- many, brute forcing, MFCC (high resolution)
- ML / Pattern Recognition
- (fusion of) classifiers / regression ⇒ effect size
- performance, applications

*both:* what can we model, convey, teach?
What to do: **CP Challenges ⇒ challenges**

- ML procedures, multi-modality, acoustic normalisation
- cross-corpus/language/culture databases
- speaker normalisation/adaptation

- confusions: hits vs. ± severe wrong assignments

- 'most important' features (from phonetics)
  - hybrid approach: same constellation, a few features based on tradition / phonetic evidence vs. brute force feature sets with/without feature reduction/selection

- interests: performance, interpretation, usability in applications
  - **loudness** in Parkinson's Condition – primary feature, to teach
  - **speech tempo** in non-nativeness – secondary feature, not to teach
  - **speaker overlap** in conflict – primary but: different cultures! – to teach
  - **variability** in depression or autism – cover feature, maybe to teach
Features: Hybrid approach

brute force

huge feature vector

? processing x

phonetics

hand-picked, few features

processing y

hybrid

huge feature vector

phonetic knowledge

processing x

processing y = x

performance

interpretation

performance

interpretation

performance

interpretation

usability in applications
A Bandanna Approach

Thank you for your attention