

Towards Speaker Characterization: Identifying and Predicting Dimensions of Person Attribution

Laura Fernández Gallardo, Benjamin Weiss
Quality and Usability Lab, Technische Universität Berlin, Germany
{laura.fernandezgallardo, benjamin.weiss}@tu-berlin.de

Introduction

- The Big-Five personality traits is a prevalent and widely accepted model for person characterization
- Our work aims at establishing a space of speaker attribution
 - dimensions detected from voice utterances
 - in zero-acquaintance scenarios (based on first impression)
 - what are the perceptual factors of (unfamiliar) speaker attributions?
 - how are they related to attributed Big-Five traits?
 - can they be predicted from speech features?

Semantic Differential

- Spontaneous dialog (mean = 19.6 s, sd = 5.2 s) from 15 male German speakers from the Nautilus Speaker Characterization (NSC) Corpus

<http://www.qu.tu-berlin.de/?id=nsc-corpus>



- Questionnaire involving a 28-item semantic differential rating scale
- 33 German listeners (17 m, 16 f), mean age = 26,9, sd = 3.7, unfamiliar with the speakers
- ICC(2,k): mean = .83, sd = .07
- Factor analysis performed with 5 factors. Items were retained if loadings > .5 and no cross-loading < .3 (14 out of 28 remained). An additional factor analysis of the remaining items explains 61% of variance

Perceptual factors

- attractiveness
- confidence
- apathy
- serenity
- incompetence

(attractiveness correlates negatively with incompetence: $r = -.88$)

| Retained items (German/English) | | Factor loadings |
|---------------------------------|-------------|-----------------|
| attraktiv | attractive | 0.77 |
| angenehm | pleasant | 0.75 |
| unsympathisch | dislikable | -0.70 |
| hässlich | ugly | -0.60 |
| sicher | confident | 0.82 |
| unentschieden | indecisive | -0.71 |
| dominant | dominant | 0.67 |
| gelangweilt | bored | 0.76 |
| emotional | emotional | -0.74 |
| gleichgültig | indifferent | 0.49 |
| ruhig | calm | 1.00 |
| entspannt | relaxed | 0.53 |
| inkompetent | incompetent | 0.74 |
| intelligent | intelligent | -0.69 |

Speaker Attributions and Personality

- Pair-wise Pearson correlations among attributed Big-Five personality traits and perceptual factors of speaker attribution

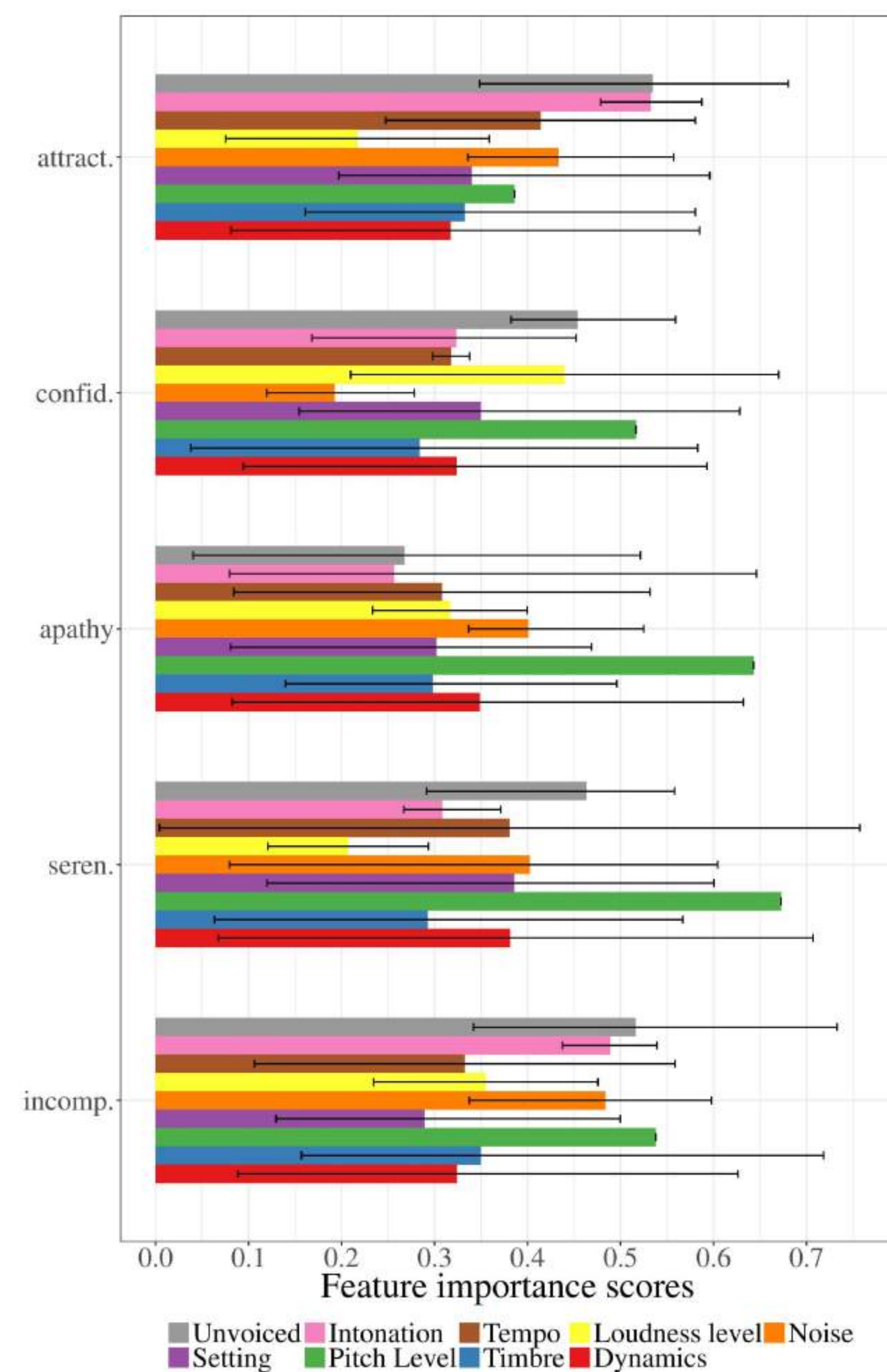
| | attract. | confid. | apathy | seren. | incomp. |
|-------------------|--------------|---------------|-----------------|---------------|----------------|
| Extraversion | -0.02 | 0.37 | -0.42 | -0.21 | 0.05 |
| Agreeableness | 0.53* | -0.55* | -0.29 | -0.26 | -0.35 |
| Conscientiousness | 0.52* | 0.18 | -0.87*** | -0.29 | -0.74** |
| Neuroticism | -0.32 | -0.70* | -0.03 | -0.62* | 0.31 |
| Openness | 0.53 | -0.13 | -0.94*** | -0.53* | -0.60* |

* $p < .05$; ** $p < .01$; *** $p < .001$

- attractiveness is positively correlated with agreeable, conscientious, and open-to-experience personalities. The correlation attractiveness-extraversion is not significant as in other studies

Prediction of Speaker Attributions

- 88 speech features: eGeMAPS extracted using OpenSMILE, assigned to 9 perceptual categories
- Average feature importance measure for regression
 - Loess smoother fit between the outcome and the predictor. The *importance* measure is an R^2 statistic calculated for this model against the intercept only null model



- SVM-based regression with polynomial kernel using a leave-one-speaker-out cross-validation (LOSO-CV) scheme
- 88 eGeMAPS features
- selected best-performing parameters C , $scale$, $degree$

| Factor | C | $scale$ | $degree$ | R^2 | RMSE |
|----------------|-----------|-----------|----------|-------|------|
| attractiveness | 10^{-4} | 10^{-4} | 1 | 0.68 | 0.42 |
| confidence | 10^{-4} | 10^{-4} | 1 | 0.65 | 0.58 |
| apathy | 10^{-4} | 10^{-4} | 1 | 0.58 | 0.53 |
| serenity | 0.1 | 10^{-4} | 6 | 0.69 | 0.42 |
| incompetence | 0.1 | 10^{-3} | 2 | 0.63 | 0.49 |

Conclusions

- Five perceptual dimensions have been identified that describe speaker attributions
- The relations to Big-Five personality traits have been analyzed
- Relevance of intonation features for *attractiveness* and *competence*; of mean F_0 for *confidence*, *apathy*, and *serenity*; of features over unvoiced segments for all dimensions, except for *apathy*
- SVM regression performances varying from $R^2 = .58$ to $R^2 = .69$
- Ongoing work: data collected to validate our findings and to obtain more meaningful models for the prediction of perceptual dimensions