Perceived Interpersonal Speaker Attributes and their Acoustic Features

Laura Fernández Gallardo and Benjamin Weiss

Quality and Usability Lab, Technische Universität Berlin, Germany
Outline

unknown speaker

pitch
F1,F2,F3 freq
speech tempo
jitter
mfccs
...

features

Relationship?

likable
attractive
competent
childish
...

subjective
perceptions
Unknown speaker

Pitch
F1, F2, F3 freq
Speech tempo
Jitter
Mfccs
...

Features

Relationship?

Likable
Attractive
Competent
Childish
...

Subjective perceptions
Nautilus Speaker Characterization (NSC) Corpus

300 speakers (126 m, 174 f) Native German

acoustically-isolated room

Scripted, semi-spontaneous and spontaneous conversational speech
Fs = 48 kHz

Released for non-commercial research and teaching purposes only

Outline

unknown speaker

features

pitch
F1,F2,F3 freq
speech tempo
jitter
mfccs
...

Relationship?

likable
attractive
competent
childish
...

subjective
perceptions
NSC labels

Interpersonal Speaker Characteristics
SC-Questionnaire

• 34-item semantic differential rating scale, completed ~15 times for each speaker by different listeners
• Stimuli: Dialog 6 (semi-spontaneous, pizza order)
• 114 labellers (70 m, 44 f)
Factor analysis for male and for female speakers separately

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>herzlich</td>
<td>0.85</td>
<td>0.84</td>
<td>0.59</td>
<td>-0.58</td>
<td>-0.52</td>
</tr>
<tr>
<td>mitfuehlend</td>
<td>-0.76</td>
<td>-0.78</td>
<td>0.56</td>
<td>-0.49</td>
<td>-0.45</td>
</tr>
<tr>
<td>distanziert</td>
<td>0.59</td>
<td>0.56</td>
<td>0.48</td>
<td>0.83</td>
<td>-0.83</td>
</tr>
<tr>
<td>freundlich</td>
<td>0.84</td>
<td>0.56</td>
<td>0.48</td>
<td>0.83</td>
<td>-0.83</td>
</tr>
<tr>
<td>verstaendnislos</td>
<td>-0.58</td>
<td>-0.49</td>
<td>0.49</td>
<td>-0.45</td>
<td>-0.45</td>
</tr>
<tr>
<td>unsympatisch</td>
<td>-0.52</td>
<td>-0.45</td>
<td>0.49</td>
<td>-0.45</td>
<td>-0.45</td>
</tr>
<tr>
<td>nicht.genervt</td>
<td>0.51</td>
<td>0.51</td>
<td>0.48</td>
<td>0.83</td>
<td>-0.83</td>
</tr>
<tr>
<td>attraktiv</td>
<td>0.59</td>
<td>0.59</td>
<td>0.48</td>
<td>0.83</td>
<td>-0.83</td>
</tr>
<tr>
<td>haesslich</td>
<td>-0.79</td>
<td>-0.81</td>
<td>0.59</td>
<td>-0.45</td>
<td>-0.45</td>
</tr>
<tr>
<td>angenehm</td>
<td>0.58</td>
<td>0.56</td>
<td>0.48</td>
<td>0.83</td>
<td>-0.83</td>
</tr>
<tr>
<td>interessant</td>
<td>0.48</td>
<td>0.48</td>
<td>0.48</td>
<td>0.83</td>
<td>-0.83</td>
</tr>
<tr>
<td>sicher</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>unentschieden</td>
<td>-0.6</td>
<td>-0.6</td>
<td>-0.6</td>
<td>-0.6</td>
<td>-0.6</td>
</tr>
<tr>
<td>gehorsam</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>zynisch</td>
<td>-0.71</td>
<td>-0.71</td>
<td>-0.71</td>
<td>-0.71</td>
<td>-0.71</td>
</tr>
<tr>
<td>alt</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>kindlich</td>
<td>-0.73</td>
<td>-0.73</td>
<td>-0.73</td>
<td>-0.73</td>
<td>-0.73</td>
</tr>
</tbody>
</table>

13. Tagung Phonetik und Phonologie
im deutschprachigen Raum.
September 28-29, Berlin, Germany
NSC labels

Factor analysis for male and for female speakers separately

1: warmth
2: attractiveness
3: confidence
4: compliance
5: maturity

1: warmth
2: attractiveness
3: compliance
4: confidence
5: maturity
Speech features

For each speaker…

linear correlation
“Feature importance”

x 2 genders
x 5 dimensions

Box Cox transformed
centered
scaled
collinear features removed

88 eGeMAPS features

openSMILE:
by audEERING™
unknown speaker

pitch
F1,F2,F3 freq
speech tempo
jitter
mfccs
...

features

Relationship?

likable
attractive
competent
childish
...

subjective
perceptions
Acoustic correlates of speaker characteristics
Acoustic correlates of speaker characteristics

• Higher warmth:
  1. Higher F0 range
  2. Higher spectral slope 0-500Hz
  3. Lower std of F1 and F3 frequencies

• Higher attractiveness:
  1. Higher F0 range
  2. Higher std of Hammarberg Index
  3. Lower length of unvoiced segments
  4. Higher std of F0
  5. Lower median F0

• Higher confidence:
  1. Lower median F0
  2. Higher F0 range

• Higher compliance:
  1. Lower std length voiced segments
  2. Lower std of F1 frequency
  3. Lower std of falling slope for loudness
  4. Higher F1 and F2 bandwidth

• Higher maturity:
  1. Lower median F0
  2. Higher std of F3 frequency
  3. Higher std of F3 bandwidth

Most warm+attractive

Least warm+attractive
Acoustic correlates of speaker characteristics

- **Higher warmth:**
  1. Higher F1 frequency
  2. Higher F0 range
  3. Higher std of F2 bandwidth
  4. Higher std of spectral flux

- **Higher attractiveness:**
  1. Higher F1 frequency
  2. Higher std of F2 bandwidth
  3. Lower std of F1 frequency
  4. Higher F0 range
  5. Lower spectral slope 0-500Hz

- **Higher compliance:**
  1. Lower std of F1 frequency
  2. Higher F1 frequency
  3. Lower loudness range

- **Higher confidence:**
  1. Higher std of falling slope for loudness
  2. Higher F0 range

- **Higher maturity:**
  1. Lower median F0
  2. Higher mean mfcc4
  3. Lower F1 frequency
  4. Higher mean mfcc2
Summary

- Nautilus Speaker Characterization (NSC) Corpus
  - 300 German speakers
  - Speaker and interlocutor speech, interactions
  - Labels: speaker characteristics

- Perceived speaker characteristics
  - 5 traits: warmth, attractiveness, confidence, compliance, maturity

- Most contributing speech features
  - Pitch and spectral features (formants, mfcc, …)
  - More investigation needed

- Future work
  - Automatic recognition of subjective speaker traits
Thank you for your attention!

Laura Fernández Gallardo, Ph.D.
laura.fernandezgallardo@tu-berlin.de
Questions?

Laura Fernández Gallardo, Ph.D.
laura.fernandezgallardo@tu-berlin.de