Possibilities and limits of speaker-specific articulatory variability
"Idiosyncratic features of a person’s speech”
may
“be a part of an individual’s learned speech behaviour”
or due to
“anatomical and physiological considerations”

(Ladefoged & Broadbent 1957, p. 98)
Sociolinguistic studies focus on learned external non-biological influences on speech and the speaker’s (conscious or unconscious) choice to develop an individual speech style that mirrors “local history and personal desire”

(Foulkes & Docherty 2006, p. 85)
Idiosyncratic choices in speech production have limits.

Speech production process is affected by physiological and biomechanical parameters:
- e.g. larynx morphology, vocal tract geometry, palatal shape, and tongue muscles

(Stone 1991, Vorperian & Kent 2007)
Individual differences in these biological determinants confine **speaker-specific strategies** used in sound production

Why study individual differences?

• in general:
  - it can help us understand how the production and coarticulation of speech sounds works and by what it is influenced

• more specifically
  - it can facilitate our understanding of the possibilities and limits of inter-speaker variability
  - also with respect to the phoneme inventory of a language and its phonemic contrasts
Outline:

⇒ 3 studies investigating inter-speaker articulatory variability in German by means of Electromagnetic Articulography & Electropalatography

⇒ 1 and 2 deal with speech of monozygotic and dizygotic twin pairs

⇒ 3 highlights differences in male and female speech
Why twins?

⇒ standard method in psychology (since 1876, Sir F. Galton)
⇒ to investigate the impact of individual differences in
  • biological determinants
  • environmentally caused learned speech behaviour

Look at:
A) Speakers who are genetically identical but grew up in a different environment (‘twin adoption design’)
OR
B) Speaker who grew up together but differ in their genetic similarity (‘twin design’)

Twin Design

MZ (monozygotic) = 100\% gen. similarity, shared environment
DZ (dizygotic) = 50\% gen. similarity, shared environment

**shared physiology:**
- amount of genetic/physiological similarity differs between MZ and DZ
  (includes jaw, teeth, larynx, palate morphology, cf. Lundström 1948,
  Langer et al. 1999, Eguchi et al. 2004)

**shared environment:**
- amount of shared environmental factors is the same in MZ and DZ

=> IF DZ twins differ more than MZ = impact of **individual physiology**!
M. Weirich, FSU Jena / Speaker-specific articulatory variability
The twins: 6 pairs (4 MZ, 2 DZ)
Speech material:

1. realization of the phoneme contrast /s/ - /ʃ/

2. looping trajectories of the tongue dorsum during vowel-stop-vowel sequences

3. sizes and shapes of articulatory vowel spaces
Study 1
Realization of the phoneme contrast /s/ - /ʃ/ in twins and unrelated speakers

(Weirich & Fuchs, in press)
Different articulatory strategies for realizing /s/ - /ʃ/ contrast in French (Toda 2006)

Tongue retraction & lip rounding for [ʃ]

Tongue doming & less lip rounding for [ʃ]
Does **palatal morphology** influence individual differences in the articulatory realization of the /s/ - /ʃ/ contrast?
METHOD
<table>
<thead>
<tr>
<th>Exp.</th>
<th>Participants</th>
<th>Methodology</th>
<th>Speech material</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4 MZ twin pairs (2 female, 2 male) 2 DZ twin pairs (both female)</td>
<td>Electromagnetic Articulography (EMA)</td>
<td>/s/ in /kysə/ (1. p. sg. of ‘to kiss’) /ʃ/ in /vəʃə/ (1. p. sg. of ‘to wash’) embedded in carrier sentence ca. 32 reps for each speaker and phoneme</td>
</tr>
<tr>
<td>2</td>
<td>7 female 5 male unrelated speakers</td>
<td>Electro-palatography (EPG)</td>
<td>/s/ in nonsense word /zaʃə/ /ʃ/ in nonsense word /ʃəʃə/ ca. 30 times each embedded in carrier sentence</td>
</tr>
</tbody>
</table>
Electromagnetic Articulography (EMA)

Introduction – Twin studies – Gender study – Discussion

M. Weirich, FSU Jena   /   Speaker-specific articulatory variability
To analyse data:
Get positional data of a sensor (at a certain timepoint, e.g. at a target) in horizontal (x) and vertical (y) dimension

- Position of tongue back sensor (x,y) during /iː ɪ eː ɛ aː aː oː ɔ uː ʊ/
Electropalatography (EPG)

custom-made artificial palate

(from Fuchs et al. 2007)
To analyse data:
Calculate articulatory Center Of Gravity (COG):
Measure contact between tongue and hard palate
COG is a weighted index, adding more weight to rows which are more front than back

=> the higher the COG the more anterior the place of articulation
(Hardcastle, Gibbon & Nicolaidis 1991)
In both experiments measurements of:

a) morphological parameters:
   palate height, length, width
   shape (palatal steepness, alveo-palatal steepness)

b) articulatory realization of /s/ -/ʃ/ contrast
   in terms of horizontal/vertical distance
a) Morphology

- Palate height/width/length (EMA-twin study)
a) Morphology

- Palate height/width/length (EPG-study)
a) Morphology

- Palate shape (*EMA*-twin study)
a) Morphology

- Palate shape (EMA-twin study)
  - steepness of palatal rise: A) angle $\delta$
  - steepness of the alveo-palatal rise: B) angle $\gamma$
- Palate shape (EPG-study)
  - steepness of palatal rise: A) angle $\delta$
  - steepness of the alveo-palatal rise: B) angle $\gamma$
b) Realization of the /s/-/ʃ/ contrast

EMA-study:

- measure **horizontal** and **vertical** distance between targets
  (tongue tip sensor)

=> calculate the **average percent of horizontal distance** between phonemes (A: around 50 %, B: nearly 100%)
b) Realization of the /s/-/ʃ/ contrast

EPG-study:
- difference between articulatory Centre of Gravity (COG) between phonemes

\[ /\text{s}/ \quad \Rightarrow \text{distance in horizontal dimension} \]

\[ /\text{ʃ}/ \]
RESULTS
EMA Twin Study
a) Morphology

Twins:
- More variation in DZ than in MZ twins regarding
  - body measures
  - tongue measures and
  - palate measures (height, palatal and alveo-palatal steepness)
b) Articulatory realization of contrast
(in terms of percent horizontal distance)
Relationship between morphology & articulation
(no significant correlation between palate height and articulation)

BUT: ANGLEs (δ, γ)

Spearman correlations:
rho (δ) = - 0.53 (p < 0.05)
rho (γ) = -0.78 (p < 0.01)
EPG Unrelated Speakers Study
Relationship between morphology & articulation

- no sign. correlation between palate size measures (length, width, height) and articulation
- no correlation between general steepness $\delta$ and articulation
Summary and main findings:

- Speaker specific variability in realizing phoneme contrasts

  - in particular: steepness of alveolo-palatal ridge ($\gamma$)
 Speakers with flat inclination angle produce contrast by retracting tongue horizontally (*tongue position strategy*).

 Speakers with steep angle not only retract, but also elevate tongue (vertical tongue displacement, *tongue shape adjustment*) to maintain contact with palate and build an air channel that directs air flow towards incisors.
Look at **contrasts** not only targets!

⇒ not exact place of articulation of a sound might be most meaningful parameter but relation between produced sounds within a particular phoneme inventory (= realization of phoneme contrasts)
Study 2
Inter-speaker articulatory variability during vowel-consonant-vowel sequences in twins and unrelated speakers

(Weirich, Lancia & Brunner, in press)
Why vowel-consonant-vowel sequences?

Nolan et al. (2006): speech signal contains

a) **linguistically determined targets**, which are constrained by the shared language system, and

b) **organically determined and speaker-specific transitions**, which link the adjacent linguistic targets

⇒ for an investigation of influence of physiology, the *dynamic* trajectory (rather than the *static* parameters) important

⇒ a specifically interesting kind of trajectory is the looping movement during velar stops
**LOOPS** = elliptical movements of tongue back

Curved paths (in speech and other human movements) potentially explained by **anatomical factors and muscle mechanics**

(cf. Flanagan et al. 1993; Gribble and Ostry 1996; Gribble et al. 1998 for arm movement, Perrier et al. 2003; Perrier and Fuchs 2008 for orofacial movements)

⇒ if loops are the result of biomechanical arrangement of the muscles (e.g. Perrier et al. 2003)
⇒ they should be very similar in MZ but differ in DZ and unrelated speakers
METHOD
Speech Material and Participants:

“Ich grüße/wasche Haka/Haga/Haku/Hagu im Garten”
(I greet/wash Haka/Haga/Haku/Hagu in the garden).

- 10 speakers: 3 MZ (2f, 1m), 2 DZ (f)
- on average 9.4 (SD=1.1) repetitions for /aga/, 9 (SD=1.4) for agu, 9.6 (SD=.97) for /aku/ and 9.8 (SD=.63) for /aka/
- EMA, tongue back coil
Procedure:

1) looping trajectories of each speaker examined **visually**

2) **curvature** for each articulatory trajectory (Tasko & Westbury 2004)
   - rate of change of a trajectory’s direction
   - increases as trajectory becomes more sharply curved
   - minimizes interfering effect of sensor positioning
   ⇒ well suited to investigate specific *shape* of loop

3) curvature trajectories were **time aligned** through registration method proposed in framework of Functional Data Analysis (FDA)
   (see Lucero et al. 1997, Lancia & Tiede 2012)
Introduction – Twin studies – Gender study – Discussion

CURVATURE DATA

POSITIONAL DATA

MZf1
MZm1
DZf2
DZf2

Twin A
Twin B

M. Weirich, FSU Jena / Speaker-specific articulatory variability
4) **mean absolute differences** measured between aligned curvature trajectories of each pair-wise speaker comparison and each /aCV/ sequence separately
=> 3 groups of speaker pairs: MZ, DZ, unrelated speakers

5) linear mixed model (LMM) to investigate potential effect of **physiology** on pattern of looping trajectory
RESULTS
Mean looping trajectories of MZ twin pairs  

**STERNE**

- **MZf1-aka**
- **MZf2-aka**
- **MZm1-aka**
- **MZf1-aku**
- **MZf2-aku**
- **MZm1-aku**
Mean looping trajectories of **DZ twin pairs**

![Graphs of DZf1-aka, DZf2-aka, DZf1-aku, DZf2-aku](image)
After time aligning and measuring the distances between pairs

**Statistical analysis:**

**Linear mixed model**
- dependent variable:
  - log transformed mean absolute distance between the aligned curvature data (for each sequence and pair)
- fixed factors
  - GROUP (levels MZ, DZ, UN)
  - VOWEL (levels /a/ and /u/) and
  - VOICE (levels voiced and voiceless)
- a pair specific random intercept for VOWEL and VOICE
distance in curvature between speaker pairs formed out of MZ or DZ or UN

GROUP:
- MZ vs. DZ: pMCMC < .001
- DZ vs. UN: n.s.

VOWEL:
- u vs. a: pMCMC < .001

VOICE:
Tendency for /a/ for a greater distance between MZ and DZ in voiced condition
Summary and main findings:

1) Inter-speaker articulatory variability in looping trajectories
   ⇒ Differences in number and magnitude of curvature peaks, in amount of horizontal sliding at palate, overall size and shape of loop

2) Impact of physiology:
   Significantly more inter-speaker variation (in terms of curvature, i.e. shape of loop) within unrelated speaker pairs and DZ twin pairs than within MZ twin pairs
   ⇒ Results in line with assumption of influence of biomechanics and physiology on looping pattern in VCV sequences (cf. Perrier et al. 2003, Birkholz et al. 2011)
Look at **dynamic** aspects when investigating influence of biomechanics and physiology on speech!
Study 3
Sex-specific differences in articulatory vowel spaces

(Weirich & Simpson, in progress)
Male – female differences:

- fundamental frequency (Stevens 1998, Mennen et al. 2012)
- voice quality (Henton & Bladon 1985)


  o Who speaks faster? Males or females?
Male – female differences:

- particular sounds:
  - sibilants (Schwarz 1968, Strand & Johnson 1996, Fuchs & Toda 2010)

Acoustic vowel spaces of 20 male and 20 female speakers /i: ɛ aʊ/ contained in "Wie lang hat es denn gedauert?" (=> How long did it take?) (from Weirich & Simpson 2013)

=> biological or behavioural reasons?
- **articulatory analysis** (26 f, 22 m, Wisconsin X-ray Microbeam Database) (Westbury 1994)
  - examines male and female differences in tongue movements in the diphthong in ‘light’ and the vowel sequence in ‘they all’
  ⇒ smaller acoustic space but greater articulatory distances in males

Simpson (1998)
- sex-specific differences in terms of **(non-)** existing correlations between formant values and duration (male speakers showed expected correlation in read speech between duration & F1 of /aː oː ɔː/ but females not)
Are there sex-specific differences in terms of undershoot?

Do females reach their articulatory targets more often than males?
METHOD
Articulatory recordings (EMA, NDI-Wave) at University of Potsdam, Department Linguistik
⇒ vertical and horizontal movements of tongue, lip and jaw
⇒ concentrate on tongue back coil

- 9 speakers
- 5 males and 4 females
- coming from Eastern Central German dialect area
- showing very little dialectal influence
- 23 – 43 years old
- no known speech or hearing problems
Speech material:

a) IAA, AUU, BII
- embedded in carrier sentence:
  “Sie fuhren letzte Woche zur IAA ganz schnell”
- 10 repetitions of /i:/, /a:/ and /u: /

⇒ to get speaker-specific extreme articulatory vowel spaces
(/i: a: u:/, temporally privileged)
b) /gV/- sequences

- embedded in carrier sentence: “Ich sah GVbi an.”
- 10 repetitions with /ɪ ɛ a ɔ ʊ/ and /iː eː aː oː uː/

3 stress conditions:

Ich sah Gabi an. \hspace{1cm} (n)
Sahst du Gabi oder Gabbi an? – Ich sah GABI an. \hspace{1cm} (s)
Siehst du oder sahst du Gabi an? – Ich SAH Gabi an. \hspace{1cm} (u)
Articulatory analysis:  
*(mview by M. Tiede)*

**a)** Parameterization of articulatory vowel space for each speaker:  
- horizontal and vertical tongue back position during acoustic midpoint of double vowel sequences in IAA, AUU, BII
b) Analysis of vertical and horizontal position of the tongue back during opening gesture /gV/ => labeling function finds 7 landmarks oriented on tangential velocity:

- start of gesture:
  /g/-target (velocity min, assumed to be max constriction)
- end of gesture:
  15 % of the range between PVEL2 and following minimum
RESULTS
**a)** Articulatory vowel spaces /i: a: u:/ of three speakers (y and x position of TBACK sensor)
a) Translated articulatory vowel spaces (5 m and 4 f)
Articulatory dimensions for mean IAU vowel space:

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Sex</th>
<th>Polygon (mm²)</th>
<th>ED_ia (mm)</th>
<th>ED_iu (mm)</th>
<th>relation ED_ia/ED_iu</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN</td>
<td>f</td>
<td>53.96</td>
<td>12.59</td>
<td>11.64</td>
<td>1.08</td>
</tr>
<tr>
<td>JB</td>
<td>f</td>
<td>56.26</td>
<td>13.37</td>
<td>13.48</td>
<td>0.99</td>
</tr>
<tr>
<td>BR</td>
<td>f</td>
<td>82.14</td>
<td>12.29</td>
<td>19.81</td>
<td>0.62</td>
</tr>
<tr>
<td>AS</td>
<td>f</td>
<td>75.05</td>
<td>15.32</td>
<td>13.33</td>
<td>1.15</td>
</tr>
<tr>
<td><strong>mean</strong></td>
<td>f</td>
<td><strong>66.85</strong></td>
<td><strong>13.39</strong></td>
<td><strong>14.56</strong></td>
<td><strong>0.96</strong></td>
</tr>
<tr>
<td>TL</td>
<td>m</td>
<td>162.8</td>
<td>19.87</td>
<td>11.93</td>
<td>1.67</td>
</tr>
<tr>
<td>HN</td>
<td>m</td>
<td>71.19</td>
<td>17.51</td>
<td>11.96</td>
<td>1.46</td>
</tr>
<tr>
<td>SG</td>
<td>m</td>
<td>60.94</td>
<td>12.37</td>
<td>13.73</td>
<td>0.90</td>
</tr>
<tr>
<td>AM</td>
<td>m</td>
<td>87.68</td>
<td>18.33</td>
<td>15.04</td>
<td>1.22</td>
</tr>
<tr>
<td>FH</td>
<td>m</td>
<td>81.31</td>
<td>18.25</td>
<td>13.75</td>
<td>1.33</td>
</tr>
<tr>
<td><strong>mean</strong></td>
<td>m</td>
<td><strong>92.78</strong></td>
<td><strong>17.26</strong></td>
<td><strong>13.28</strong></td>
<td><strong>1.32</strong></td>
</tr>
</tbody>
</table>

Welch two sample t-tests:

\[ t = -2.7, \text{df} = 5.9, \quad p < .05 \]
Euclidean distance from midpoint (MP) to /i: a: u:/

<table>
<thead>
<tr>
<th></th>
<th>F Mean (SD)</th>
<th>M Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>6.50 (1.01)</td>
<td>8.74 (1.62)</td>
</tr>
<tr>
<td>i</td>
<td>8.65 (1.08)</td>
<td>10.00 (1.59)</td>
</tr>
<tr>
<td>u</td>
<td>7.28 (2.2)</td>
<td>7.22 (2.27)</td>
</tr>
</tbody>
</table>

Welch two sample t-tests

$t = -2.5, \, df = 6.7, \, p < .05$

n.s.
Significant differences between male and female speakers
- in vertical TB position for /a:/ and /i:/ (pMCMC < .001)
- in horizontal TB position for /a:/ (pMCMC < .05)

⇒ Female vowel space is smaller in terms of a higher and more fronted position for /a:/ and a lower position for /i:/
b) Opening gesture of /gV/ for all vowels and conditions for three speakers (1 male and 2 female)

- start of opening gesture (maximal closure, /g/-target)
- offset of opening gesture

● mean articulatory vowel space of IAU
Overall vowel space size of 
/gV/ sequence  
- Polygon area -
# Articulatory dimensions for mean gi-ga-gu vowel space:

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Sex</th>
<th>Polygon (mm²)</th>
<th></th>
<th></th>
<th>Polygon (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n</td>
<td>s</td>
<td>u</td>
<td>n</td>
<td>s</td>
<td>u</td>
</tr>
<tr>
<td>AS</td>
<td>f</td>
<td>68.05</td>
<td>53.82</td>
<td>43.25</td>
<td>90.67</td>
<td>71.71</td>
<td>57.63</td>
</tr>
<tr>
<td>BR</td>
<td>f</td>
<td>70.38</td>
<td>51.12</td>
<td>58.62</td>
<td>85.69</td>
<td>62.22</td>
<td>71.37</td>
</tr>
<tr>
<td>JB</td>
<td>f</td>
<td>50.35</td>
<td>55.79</td>
<td>30.80</td>
<td>89.49</td>
<td>99.16</td>
<td>54.75</td>
</tr>
<tr>
<td>SN</td>
<td>f</td>
<td>58.38</td>
<td>47.19</td>
<td>44.29</td>
<td>108.19</td>
<td>87.45</td>
<td>82.09</td>
</tr>
<tr>
<td>mean</td>
<td>f</td>
<td>61.79</td>
<td>51.98</td>
<td>44.24</td>
<td>93.51</td>
<td>80.14</td>
<td>66.46</td>
</tr>
<tr>
<td>AM</td>
<td>m</td>
<td>99.54</td>
<td>115.16</td>
<td>73.27</td>
<td>113.53</td>
<td>131.34</td>
<td>83.56</td>
</tr>
<tr>
<td>FH</td>
<td>m</td>
<td>70.18</td>
<td>57.90</td>
<td>39.71</td>
<td>86.31</td>
<td>71.22</td>
<td>48.84</td>
</tr>
<tr>
<td>HN</td>
<td>m</td>
<td>31.79</td>
<td>44.18</td>
<td>29.33</td>
<td>44.67</td>
<td>62.06</td>
<td>41.20</td>
</tr>
<tr>
<td>SG</td>
<td>m</td>
<td>51.71</td>
<td>59.67</td>
<td>31.26</td>
<td>84.85</td>
<td>97.92</td>
<td>51.29</td>
</tr>
<tr>
<td>TL</td>
<td>m</td>
<td>88.43</td>
<td>93.04</td>
<td>63.11</td>
<td>54.32</td>
<td>57.15</td>
<td>38.77</td>
</tr>
<tr>
<td>mean</td>
<td>m</td>
<td>68.33</td>
<td>73.99</td>
<td>47.33</td>
<td>76.73</td>
<td>83.93</td>
<td>52.73</td>
</tr>
</tbody>
</table>
Linear mixed model with stress condition (n, s, u) and sex (f, m) as fixed factors (and speaker as random factor)
⇒ Significant interaction of stress condition and sex
⇒ males do differ between s and u, females not
⇒ more undershoot in males
On the edge vs. centralisation

- ED between midpoint of polygon & corner vowels /iː aː uː/ of /gV/ sequence -
Differences between m and f get smaller (only tendency for s, p= .08)

Significant interaction between stress and sex (males do significantly less in u than in s)
Normalised values
(expressed as percent of ED between MP and /i: a: u:/ of IAU-polygon

⇒ Differences between m and f either disappear or are even reverse (f vs. m for u p = .09)

⇒ Again significant interaction between stress and sex (males differ between n vs. u and s vs. u)
Summary and main findings:

- Articulatory vowel space differs between females and males especially in vertical dimension (ED between /i/ and /a/).

- Differences between females and males vary between absolute values and percentages of maximally used space males do not use the space they could use.
Summary and main findings:

- Interaction of sex and stress condition
  \(\Rightarrow\) **males** show undershoot in unstressed condition, **females** not
To analyse differences in articulatory spaces of male and female speakers:

- include interactions with *stress* condition
- look at realisation of gestures relative/with respect to the speaker-specific *extreme* articulatory space
Discussion
Investigating **idiosyncratic variation** can help us understand what is possible and what are the limits of normal variation in speech.

Speaker specific variability is not just random noise:
- it is determined by various reasons (which we need to understand)
- besides behavioural aspects don’t forget physiological/morphological/biomechanical sources of variation
o Look at speaker-specific variation with respect to the specific nature of speech:

Investigate...

• ...not only targets but **phoneme contrasts**
• ...not only static but **dynamic patterns**
• ...not only absolute parameters but **relational aspects**
Thanks to the organizers of this summer school!

Thanks to Department Linguistik, Universität Potsdam!

Thanks to my incredible co-workers Leonardo Lancia, Jana Brunner, Susanne Fuchs and Adrian Simpson!

Thank you!
REFERENCES:


Weirich, M. & Fuchs, S. (acc.) Palatal morphology can influence speaker-specific realizations of phonemic contrasts. *Journal of Speech Language and Hearing Research*
