Pre-schoolers’ categorisation of speakers by phonological variables

Ella Jeffries, Paul Foulkes & Carmen Llamas
Department of Language and Linguistic Science
University of York
erwj500@york.ac.uk
How do pre-school children learn to group together speakers on the basis of regionally distributed features of pronunciation?

How is this ability impacted by the child’s Age and Sex and the linguistic Input they receive?
Background

• **Adults** can group speakers into broad perceptual regional accent categories
  → Williams et al. (1999), Clopper & Pisoni (2004, 2007)

• The age at which **children** can use **regional accent** features in order to **group** speakers and how this **develops** is not clearly understood
  - 7-year-olds (Floccia et al. 2009)
  - 5-year-olds (Beck 2014)

• Categorising speakers by regional accent is a **life-long** skill

• But is there any evidence of this ability **emerging** in **preschool** children?
Background

• Previous studies have uncovered the development of sociolinguistic skills in the pre-school years

• **Linguistic input** important
  - Children learn **community norms** of pronunciation
  - Children’s preference for **standard variables** is related to their exposure to standard forms
    → Smith et al. (2007), Barbu et al. (2013)
Background

• **Usage-based** theories of language acquisition best describe the importance of input

• Other theoretical models don’t show how **the indexical meaning of sociophonetic variability** is learned (cf. Foulkes and Docherty 2006)

  - **Storing** of specific **linguistic units** (cf. Tomasello 2003)
  - **Frequency of encounters** aids acquisition (cf. Chevrot et al. 2009)
  - Exemplars of **individual talker** differences → **broader groups** based on these differences (cf. Foulkes & Hay 2015)
  - More **transparent categories** easier to learn – direct exposure important (cf. Foulkes and Docherty 2006)
Research questions

(1) To what extent can 3-4 year-olds group speakers by phonetic variants indexing regional accents?

(2) To what extent does their ability in (1) vary with age, sex and input from different regional accents?

- Age: Improvement through pre-school years?
- Sex: Difference between boys and girls?
- Input: Those who have parents from outside the local area (and are therefore exposed to a wider variety of accents at home) better in this ability?
Methodology

Participants
- 20 pre-school children in York (+ 4 discarded)
- 12 girls, 8 boys
- Aged 3.1 years to 4.6 years

Experiment
- Sentences
- Two regional accents
- Single speaker
- Run on laptop in quiet corner of nursery or home
Methodology

- Regional accent differences

<table>
<thead>
<tr>
<th></th>
<th>NORTH</th>
<th>SOUTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>bath, grass</td>
<td>[a]</td>
<td>[ɑː:]</td>
</tr>
<tr>
<td>face, gate</td>
<td>[eː]</td>
<td>[eI]</td>
</tr>
</tbody>
</table>

(Hughes et al. 2012:71)
Difficulty level 1: Same word

“b[ɑː]sket”

“b[a]sket”
Difficulty level 2: Same phoneme

Daughters: grass [a]/[ɑː]

Mothers: path [a]/[ɑː]
Difficulty level 3: Different phoneme

Daughters: cake [e:]/[eɪ]

Mothers: after [a]/[ɑː]
Results for each DL

<table>
<thead>
<tr>
<th>Difficulty level</th>
<th>Mean % correct (SD)</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65 % (15)</td>
<td>4.44 (p&lt;0.001)</td>
</tr>
<tr>
<td>2</td>
<td>60 % (18)</td>
<td>2.01 (p=0.055)</td>
</tr>
<tr>
<td>3</td>
<td>63 % (27)</td>
<td>1.83 (p=0.089)</td>
</tr>
</tbody>
</table>
Age group and DL

Percentage Correct (95% CI)

3-year-olds

4-year-olds

Age group and Difficulty level

chance=50%
Sex and DL

Percentage Correct (95% CI)

Female
Male

Sex and Difficulty level
Input (Yorkshire parents) and DL

![Graph showing percentage correct (95% CI) for different difficulty levels (DL1, DL2, DL3) with and without Yorkshire parents. The graph compares the performance of participants with no Yorkshire parents to those with 1+ Yorkshire parents. The x-axis represents the Yorkshire parents and Difficulty level, while the y-axis shows the percentage correct with a chance level of 50%. The bars represent the 95% confidence intervals for each category.]
Results: statistical analysis

• Stepwise backward regression method in binary, mixed effects logistic models, run in R

• Three binary independent variables
  – **Age**: 3 or 4
  – **Sex**: F or M
  – **Input**: 1+ Yorkshire parent or no Yorkshire parent
    – default: **3-year-old girl with no Yorkshire parent(s)**

• Two-way interactions: **Age*Sex, Age*Input**

• Random effect: individual child

• Separate models for **DL1, DL2 & DL3**
Logistic mixed effects model: Same word (DL1) results

- Two significant **main effects** – Age, Sex
- No significant **interactions**

| Factor                        | Estimate | Std. Error | z    | Pr(>|z|) | Sig |
|-------------------------------|----------|------------|------|----------|-----|
| (Intercept)                   | 0.78     | 0.25       | 3.16 | 0.002    | **  |
| Four-Year-Old                 | 0.52     | 0.27       | 1.94 | 0.05     | *   |
| Male                          | -0.54    | 0.27       | -2.05| 0.04     | *   |
| With Yorkshire Parent(s)      | -0.43    | 0.25       | -1.68| 0.09     |      |
Same word (DLI) results: Age

Raw data
(Total correct answers)

Model prediction
(Predicted probability of correct answer)
Same word (DL1) results: Sex

Raw data
(Total correct answers)

Model prediction
(Predicted probability of correct answer)
Logistic mixed effects model: Same phoneme (DL2) results

- One significant **main effect** - Age
- No significant **interactions**

| Factor                        | Estimate | Std. Error | z     | Pr(>|z|) | Sig |
|-------------------------------|----------|------------|-------|----------|-----|
| (Intercept)                   | 0.26     | 0.53       | 0.50  | 0.62     |     |
| Four-Year-Old                 | 1.31     | 0.55       | 2.36  | 0.02     | *   |
| Male                          | -0.53    | 0.50       | -1.06 | 0.29     |     |
| With Yorkshire Parent(s)      | -0.44    | 0.46       | -0.96 | 0.34     |     |
Same phoneme (DL2) results

**Raw data**
(Total correct answers)

**Model prediction**
(Predicted probability of correct answer)
Logistic mixed effects model: Different phoneme (DL3) results

- One significant **main effect** – Yorkshire parents
- No significant **interactions**

| Factor                        | Estimate | Std. Error | z    | Pr(>|z|) | Sig |
|-------------------------------|----------|------------|------|----------|-----|
| (Intercept)                   | 1.61     | 0.45       | 3.58 | 0.0004   | *** |
| With Yorkshire Parent(s)      | -1.6     | 0.54       | -3.0 | 0.003    | **  |
Different phoneme (DL3) results

**Raw data**
(Total correct answers)

**Model prediction**
(Predicted probability of correct answer)
Summary

• **Pre-school** children score above chance level in the ability to group together speakers based on **regionally distributed phonetic variants**

  - Same word
  - Same phoneme
  - Different phoneme

  \[ \text{Easier} \rightarrow \text{Harder} \]

• But with significant effects of **Age, Sex and Input**
Discussion: Age

• **Age improvement between 3 and 4 years**
  - Younger age group than previously investigated
  - In line with other sociolinguistic developments and indexical learning
  - Accent aids 2-4-year-olds in the recognition of familiar speakers, ability improved with age (Jeffries, in press)

• Most robust for **DL 2 (same phoneme condition)**
  - Shows development in the understanding of a phoneme category and its variable realisations
Discussion: Sex

• **Sex: girls outperform boys**
  - Girls better at tasks requiring phonological and semantic information in long-term memory and perceptual speed ([Sternberg 2004](#), [Halpern 1997](#))

• **Only significant for DL1 (same word condition)**
  - Boys needed longer to understand the task?
  - But also a much larger range of results for the boys in DL3
    - individual variation
Discussion: Input

• **Input:** children with parents from outside of Yorkshire perform better in DL3 (different phoneme condition)

• Exposure to speakers with **different accents** at home helps in the forming of **categories** based on **regionally distributed phonetic variants**

• Predicted by **Usage-based models**
  - Exposure to multiple accents generates more **robust categories**
    (cf. Logan et al. 1991: multiple speakers leads to more robust categories in L2 learning)
Conclusion

• **Development** between the ages of 3 and 4 in children’s ability to group speakers according to **regionally** distributed features of **pronunciation**

• **Varied input** helps in the creation of more **robust** categories

• Supports a **usage-based** model of language acquisition in which **speaker categories** are based on **experienced exemplars**
Thanks for listening!

Ella Jeffries, Paul Foulkes, Carmen Llamas
Department of Language and Linguistic Science
University of York
erwj500@york.ac.uk
References


References


References


References

Appendices

Methodology

• Positioning of mummy bears/mothers and baby bears/daughters randomised

• 8 sets of stimuli featuring [a]/[ɑː] and [eː]/ [eɪ]
  → 2 for DL1, 4 for DL2, 2 for DL3

• 20 children completed DL1
• 15 children completed DL2 and DL3
# Children’s details

<table>
<thead>
<tr>
<th>Sex</th>
<th>Age</th>
<th>Yorkshire parents</th>
<th>DLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>4.38</td>
<td>0</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>F</td>
<td>4.41</td>
<td>0</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>F</td>
<td>3.7</td>
<td>1</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>M</td>
<td>3.11</td>
<td>1</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>F</td>
<td>3.07</td>
<td>2</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>F</td>
<td>4.52</td>
<td>0</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>M</td>
<td>4.27</td>
<td>1</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>F</td>
<td>3.61</td>
<td>1</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>M</td>
<td>3.54</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>3.38</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>3.37</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>3.59</td>
<td>2</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>F</td>
<td>3.56</td>
<td>0</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>M</td>
<td>3.2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>M</td>
<td>4.5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>4.44</td>
<td></td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>M</td>
<td>4.41</td>
<td>2</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>F</td>
<td>4.64</td>
<td>2</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>M</td>
<td>4.2</td>
<td>0</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>M</td>
<td>4.59</td>
<td>1</td>
<td>1, 2, 3</td>
</tr>
</tbody>
</table>

**Totals** 12 F, 7 M 10 4yo, 10 3yo 10 with YP, 9 without 15 all DLS, 5 just DL1
Results

- Age divide between 3-year-olds and 4-year-olds for results from all difficulty levels
Logistic mixed effects model: All results

- One significant main effect
- One significant interaction

| Factor                        | Estimate | Std. Error | z value | Pr(>|z|)  | Sig |
|-------------------------------|----------|------------|---------|-----------|-----|
| (Intercept)                   | 0.45     | 0.18       | 2.48    | 0.01      | *   |
| Four-Year-Old                 | 1.27     | 0.25       | 5.19    | 2.14e-07  | *** |
| Male                          | 0.19     | -0.26      | 0.71    | 0.48      |     |
| With Yorkshire Parent(s)      | -0.34    | 0.17       | -2.0    | 0.05      | *   |
| Four-Year-Old:Male            | -1.14    | 0.37       | -3.09   | 0.002     | **  |
All results: Age & Sex

Raw data
(Total correct answers)

Model prediction
(Predicted probability of correct answer)

Age*Sex interaction
All results: Yorkshire parents

Raw data
(Total correct answers)

Model prediction
(Predicted probability of correct answer)
DL3: Southern parents

Key
- at least one Yorkshire parent
- no Yorkshire parents
- no Yorkshire parents, one Southern parent

Correct answers for DL3 results(%) by Child