Introduction

- differentiation of the 2 speech registers neutral and intimate
- 3 different types of speakers:
  - mothers addressing their own children or an unknown adult
  - women without own children addressing an imaginary child or adult
  - children addressing a pet robot using both intimate and neutral speech

Data

1. Mothers addressing their own child or an unknown adult
   - study of the Department of Psychology, University of Stirling:
     - How do children follow instructions?
     - From which age on are they able to do so?
   - set of 6 instructions, partly ambiguous
   - Touch the horse with the spoon (non-ambiguous)
   - Touch the fish with the flower (non-ambiguous)
   - Touch the dog with the flower (ambiguous)
   - addresses:
     - child-directed: their own child (at the age of 2;0 - 3;8)
     - adult-directed: unknown adult
     - 24 mothers (23-46 years old, 35 years)
     - 192 recordings: 24 mothers - 4 instructions · 2 addressers
     - language: English

2. Non-mothers addressing an imaginary child or an imaginary adult
   - parallel communication task with 24 non-mothers (age: 21-42, 27)
   - addresses:
     - child-directed: imaginary child at the age of 2 - 3
     - adult-directed: imaginary adult (friend or acquaintance)

3. Children interacting with the Sony robot Aibo
   - subset of the AIBO Emotion Corpus (University of Erlangen-Nuremberg)
   - 21 children (10-13 years old)
   - language: German
   - level of analysis:
     - chunk level: 568 chunks motherese, 1998 chunks neutral
     - word level (only the word Aibo): 220 words

Features and Classification

Features

- typical acoustic low-level descriptors (LLD) on frame level
- applying functionals
- feature vectors on word, chunk, or file level

<table>
<thead>
<tr>
<th>low-level-descriptors (2 - 37)</th>
<th>functionals (19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(J) F0</td>
<td>mean, std. dev., centroid skewness, kurtosis zero-crossing-rate quartile 1-3 quartile 1 - minimum quartile 2 - quartile 1 quartile 3 - quartile 2 maximum - quartile 3 max., min. range position of rel. max./min. pos. 95 % roll-off-point</td>
</tr>
</tbody>
</table>

| 1,406 features: 2 · 37 LLD · 19 functionals |

Classification

- 2-class problem: neutral vs. intimate
- 2 classifiers:
  - SVM with linear kernel
  - random forests (RF)
- 3-fold cross-validation
- speaker independent
- normalization per speaker using the whole speaker context
- no explicit feature selection
- evaluation: F measure

Results

<table>
<thead>
<tr>
<th>Train</th>
<th>Test</th>
<th>F measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mothers</td>
<td>Mothers</td>
<td>76.6 %</td>
</tr>
<tr>
<td>Non-mothers</td>
<td>Non-mothers</td>
<td>70.3 %</td>
</tr>
<tr>
<td>Mothers</td>
<td>Mothers</td>
<td>72.4 %</td>
</tr>
<tr>
<td>M+N</td>
<td>M+N</td>
<td>68.8 %</td>
</tr>
<tr>
<td>Aibo (chunks)</td>
<td>Aibo (chunks)</td>
<td>72.8 %</td>
</tr>
<tr>
<td>Aibo (word)</td>
<td>Aibo (word)</td>
<td>71.4 %</td>
</tr>
</tbody>
</table>

- intimacy more pronounced for Mothers than for Non-mothers
- in accordance with subjective impression when listening to the data
- accuracy for children in the same range as for Non-mothers

Most Important Feature Types

Feature Types

- Duration (222)
- Energy (64)
- F0 (43)
- Formants (480)
- MFCC (512)
- Voice Quality (96)

Two Ways of Feature Selection

1. SVM-SFFS
   - 50 features per split (3-fold cross-validation)
   - reduction to < 11 %

2. approach using eigenvectors
   - 50 eigenvectors with the highest eigenvalues
   - reduction to 5 eigenvalues using SVM-SFFS
   - selection of 15 “original” features per eigenvector
   - reduction to < 16 %

Discussion

- no indication that age group or language are decisive factors
- segmental structure:
  - higher impact of duration if segmental structure is constant (M, N, W)
  - higher impact of energy and F0 features if segmental structure is variable (C)
  - F0 less important if focus only on one single short word (W)
- degree of intimacy: higher impact of formants (M) vs. higher impact of MFCC features (N)