Tesis doctoral

On Line Personalization and Adaptation to Disorders and Variations of Speech on Automatic Speech Recognition Systems

Oscar Saz

Director: Prof. Eduardo Lleida
1. Introduction
   - Introduction
   - Review of Previous Resources

2. Corpus and Baseline
   - “Alborada-I3A” Corpus
   - Experimental Framework and Baseline Results

3. Analysis of Disordered Speech
   - Acoustic Analysis of the Corpus
   - Lexical Analysis of the Corpus

4. Techniques for Personalization
   - Acoustic-Lexical Adaptation for ASR of Disordered Speech
   - Confidence Measuring for Detection of Mispronunciations
   - Unsupervised On-line Systems for ASR of the Speech Handicapped

5. Conclusions
   - Development of Speech Based Applications for the Handicapped
   - Discussion
   - Conclusion
Introduction
Motivation for the thesis arises from:

- **Communicative and digital gap**
  - Difficulties in communication limit strongly the possibilities of social development of handicapped individuals
  - Accessibility to new technologies can facilitate communication and knowledge

- **Interests from educational and assistive centers**
  - CPEE “Alborada”
  - CADIS-Huesca & ASPACE-Huesca

- **Multidisciplinary work**
  - Speech technologies
  - Linguistics, psycholinguistics
  - Education, pedagogy, psychology
Difficulties in speech-based systems for the handicapped

- Voice pathologies
  - Dysphonia

- Speech disorders
  - Functional disorders: Dyslalia
  - Organic disorders: Dysarthria, dysglossia
  - Audiogenous disorders: Hearing impairments

- Language impairments
  - Specific Language Impairment
Introduction

Objectives

- **Scientific objectives**
  - Acquisition of a fully functional corpus of disordered speech
  - Study of adaptation techniques for ASR
  - Techniques for the assessment of speech proficiency

- **Development objectives**
  - Devices for control and interaction based on speech
  - Computer-Aided Language Learning tools
Introduction

Methodology

Thesis methodology and organization

CORPUS + BASELINE

Confidence measuring

Evaluation

Acoustic analysis

Acoustic adaptation

Phonological analysis

Lexical adaptation

Unsupervised on-line adaptation

Acoustic-lexical adaptation
Review of Previous Resources
Corpora of Disordered Speech

Some corpora in English:

- **Whitaker database** [Deller et al., 1993]
  - 6 Dysarthric speakers and a control speaker
  - Isolated words (2430) and a 35-word passage (1215)

- **Nemours database** [Menéndez-Pidal et al., 1996]
  - 11 Dysarthric speakers
  - 74 Meaningless sentences and 2 paragraphs
  - Fully labeled and assessed

- **Universal Access database** [Kim et al., 2008]
  - 19 dysarthric speakers
  - 765 isolated words per speakers
  - Fully assessed and multimodal data

**Efforts in Spanish:** HACRO corpus [Navarro-Mesa et al., 2005]

- 43 impaired speakers + 19 unimpaired speakers
- 57 isolated words per speaker
- Fully labeled
Speech technology for the handicapped

- **Vocal Joystick** [Bilmes et al., 2006][Harada et al., 2008]
  - Vowel detection → Motion control

- **Stardust** [Green et al., 2003][Hawley et al., 2003][Hawley et al., 2005]
  - ASR → Home devices

- **VIVOCA** [Creer et al., 2009][Creer et al., 2010]
  - ASR → TTS → Feature extractor
CALL tools

Distinction of CALL tools

CALL

CASLT

CAPT

L2

Vocabulary

Syntax

Grammar
1. Introduction
2. Corpus
3. Analyses
4. Personalization
5. Conclusions

“Alborada-I3A” Corpus
Requirements

- Good signal quality (low noise)
- Naturalness in the speech
- Balance in gender
- Balance in ages
- Balance in degree of impairments
- Multiple sessions, recorded in different days
Recording environment

(a) Recording place

(b) Recording interface
## Speaker characterization

Speakers in the disordered speech corpus

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Age</th>
<th>Gender</th>
<th>Speaker</th>
<th>Age</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spk01</td>
<td>14 years</td>
<td>Female</td>
<td>Spk02</td>
<td>11 years</td>
<td>Male</td>
</tr>
<tr>
<td>Spk03</td>
<td>21 years</td>
<td>Male</td>
<td>Spk04</td>
<td>21 years</td>
<td>Female</td>
</tr>
<tr>
<td>Spk05</td>
<td>18 years</td>
<td>Male</td>
<td>Spk06</td>
<td>17 years</td>
<td>Male</td>
</tr>
<tr>
<td>Spk07</td>
<td>18 years</td>
<td>Male</td>
<td>Spk08</td>
<td>19 years</td>
<td>Male</td>
</tr>
<tr>
<td>Spk09</td>
<td>11 years</td>
<td>Female</td>
<td>Spk10</td>
<td>15 years</td>
<td>Female</td>
</tr>
<tr>
<td>Spk11</td>
<td>20 years</td>
<td>Female</td>
<td>Spk12</td>
<td>18 years</td>
<td>Male</td>
</tr>
<tr>
<td>Spk13</td>
<td>13 years</td>
<td>Female</td>
<td>Spk14</td>
<td>11 years</td>
<td>Female</td>
</tr>
</tbody>
</table>

- Physical and cognitive disorders
  - Down’s syndrome, attention deficit, cerebral palsy, . . .

- Speech and language disorders
  - Dyslalia, dysarthria, SLI, . . .
Session characterization

1. Isolated word sessions
   - 4 sessions per speaker (intraspeaker variability)
   - 57 words per speaker and session: RFI [Monfort and Juárez-Sánchez, 1989]
   - 3,192 utterances - 2h17m

2. Simple meaningless sessions
   - 4 sessions of 28 sentences from speakers Spk01, Spk04, Spk06 and Spk11
   - el/la RFIWord1 y el/la RFIWord2
   - Each word appears twice as RFIWord1 and RFIWord2
   - 448 utterances - 25m30s

3. Complex meaningful sessions
   - 1 session of 10 sentences from speakers Spk01, Spk06 and Spk11
   - 3 RFI words per sentence, 9 words total as maximum
   - 40 utterances - 2m9s
Reference corpus

Unimpaired healthy children for child speech reference

- 232 speakers, 9576 utterances - 8h50m

<table>
<thead>
<tr>
<th>Age</th>
<th>Males</th>
<th>Females</th>
<th>Age</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 years old</td>
<td>15</td>
<td>16</td>
<td>12 years old</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>13 years old</td>
<td>15</td>
<td>15</td>
<td>14 years old</td>
<td>15</td>
<td>23</td>
</tr>
<tr>
<td>15 years old</td>
<td>11</td>
<td>21</td>
<td>16 years old</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>17 years old</td>
<td>15</td>
<td>9</td>
<td>18 years old</td>
<td>14</td>
<td>10</td>
</tr>
</tbody>
</table>

Courtesy of CEIP “Río Ebro”, IES “Tiempo Modernos” and IES “Félix de Azara”
Perceptual labeling of lexical mispronunciations

Description of the labeling process

árbol

Expert i

Expert n

Expert x

Poll

Untie

Expert 1

Expert N

2 0 2 2 0

2 0 2 2 1

2 0 2 2 X

1

2 1 2 2 2

2 0 2 2 1
Perceptual labeling of lexical mispronunciations

Labeling results

- Correct: 82.39%
- Substituted: 10.31%
- Deleted: 7.30%

...with 85.81% of interlabeler agreement
Experimental Framework and Baseline Results
HMM topology and feature extraction method

Different proposed HMM topologies

Feature extraction method
Baseline results

Measure of ASR performance: Word Error Rate
WER in the isolated-word case:

\[ WER = \frac{\text{Substitutions}}{\text{Words}} \] (1)
Experimental Framework and Baseline Results

Baseline results

Child unimpaired speech with models trained on
- Adult speech
- Child speech in the same task (cross validation)

<table>
<thead>
<tr>
<th>WER</th>
<th>Word units</th>
<th>Phone units</th>
<th>Sub-phone units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task Indep.</td>
<td>3.99%</td>
<td>9.91%</td>
<td>3.99%</td>
</tr>
<tr>
<td>Task Dep.</td>
<td>2.04%</td>
<td>2.77%</td>
<td>2.11%</td>
</tr>
</tbody>
</table>
Baseline results

Child impaired speech with models trained on
- Adult speech
- Child unimpaired speech in the same task
- Other impaired children in the same task

<table>
<thead>
<tr>
<th>WER</th>
<th>Word units</th>
<th>Phone units</th>
<th>Sub-phone units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task Indep.</td>
<td>36.69%</td>
<td>40.86%</td>
<td>36.69%</td>
</tr>
<tr>
<td>Task Dep.</td>
<td>25.85%</td>
<td>31.96%</td>
<td>28.20%</td>
</tr>
<tr>
<td>Domain Dep.</td>
<td>26.97%</td>
<td>31.86%</td>
<td>28.60%</td>
</tr>
</tbody>
</table>
Experimental Framework and Baseline Results

Baseline results

Confusion matrix

![Confusion Matrix Diagram]

Decoded word (1-arbol, 57-zapato)

Uttered word (1-arbol, 57-zapato)

dedo
taza
Speaker grouping

WER as a function of the mispronunciation rate

![Graph showing WER as a function of the mispronunciation rate for different groups of speakers. The graph includes points for Spk01 & 03 & 04 & 06 & 11, Spk02 & 10, Spk05, Spk12, and Spk13, grouped into three categories: Group A, Group B, and Group C. The regression line shows the trend.]
Speaker grouping

Three different groups of speakers according to their degree of speech impairments

<table>
<thead>
<tr>
<th></th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>5.26%</td>
<td>10.35%</td>
<td>6.32%</td>
</tr>
<tr>
<td>Group B</td>
<td>20.44%</td>
<td>28.25%</td>
<td>23.86%</td>
</tr>
<tr>
<td>Group C</td>
<td>57.90%</td>
<td>63.38%</td>
<td>59.43%</td>
</tr>
<tr>
<td>Total</td>
<td>25.85%</td>
<td>31.96%</td>
<td>28.20%</td>
</tr>
</tbody>
</table>
Connected speech

Simple meaningless sentences

- Proposed grammar

![Diagram of connected speech]

- Results obtained

<table>
<thead>
<tr>
<th>WER</th>
<th>Ph. TI</th>
<th>Ph. TD</th>
<th>Sub TI</th>
<th>Sub TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spk01</td>
<td>16.4%</td>
<td>15.0%</td>
<td>13.7%</td>
<td>14.2%</td>
</tr>
<tr>
<td>Spk04</td>
<td>26.6%</td>
<td>18.6%</td>
<td>24.8%</td>
<td>18.1%</td>
</tr>
<tr>
<td>Spk06</td>
<td>8.0%</td>
<td>5.3%</td>
<td>11.0%</td>
<td>6.2%</td>
</tr>
<tr>
<td>Spk11</td>
<td>25.2%</td>
<td>14.2%</td>
<td>24.3%</td>
<td>15.0%</td>
</tr>
</tbody>
</table>
Connected speech

Complex meaningful sentences

- Proposed grammar

Results obtained

<table>
<thead>
<tr>
<th>WER</th>
<th>Ph. TI</th>
<th>Ph. TD</th>
<th>Sub. TI</th>
<th>Sub. TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spk04</td>
<td>63.3%</td>
<td>60.0%</td>
<td>53.3%</td>
<td>43.3%</td>
</tr>
<tr>
<td>Spk06</td>
<td>56.7%</td>
<td>36.7%</td>
<td>33.3%</td>
<td>36.7%</td>
</tr>
<tr>
<td>Spk11</td>
<td>46.7%</td>
<td>40.0%</td>
<td>43.3%</td>
<td>46.7%</td>
</tr>
</tbody>
</table>
Experimental Framework and Baseline Results

APD results

APD: ASR with phones as target units to decode

- Phone units
  - Rule-based grammar
  - Stochastic grammar [Koehn, 2005]
- Sub-phone units
  - Stochastic grammar [Koehn, 2005]

Performance is measured as PER:

\[
PER = \frac{Insertions + Deletions + Substitutions}{Phonemes}
\]
Experimental Framework and Baseline Results

APD results

Child unimpaired speech with models trained on
- Adult speech
- Child unimpaired speech in the same task (cross validation)

<table>
<thead>
<tr>
<th>PER</th>
<th>Phone models</th>
<th></th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rule-based</td>
<td>Stochastic</td>
<td>Stochastic</td>
</tr>
<tr>
<td>Task Indep.</td>
<td>33.40%</td>
<td>32.55%</td>
<td>26.42%</td>
</tr>
<tr>
<td>Task Dep.</td>
<td>16.24%</td>
<td>15.76%</td>
<td>8.48%</td>
</tr>
</tbody>
</table>
**APD results**

Child impaired speech with models trained on

- Adult speech
- Child unimpaired speech in the same task

<table>
<thead>
<tr>
<th>PER</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rule-based</td>
<td>Stochastic</td>
</tr>
<tr>
<td>Task Indep.</td>
<td>60.37%</td>
<td>57.73%</td>
</tr>
<tr>
<td>Task Dep.</td>
<td>49.66%</td>
<td>48.28%</td>
</tr>
</tbody>
</table>
APD results

APD as possible estimator of mispronunciations

Two measures of performance:

\[
FAR = \frac{\text{Accepted}}{\text{Mispronunciations}}
\]

\[
FRR = \frac{\text{Reject}}{\text{Correct}}
\]

FRR = PER of the correct phonemes

<table>
<thead>
<tr>
<th></th>
<th>Phoneme models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FAR</td>
<td>FRR</td>
</tr>
<tr>
<td>Task Ind.</td>
<td>11.84%</td>
<td>43.33%</td>
</tr>
<tr>
<td>Task Dep.</td>
<td>12.26%</td>
<td>31.27%</td>
</tr>
</tbody>
</table>
Influence of the disorders

WER in terms of phonemes and mispronounced phonemes in the word

9.1% WER
Influence of the disorders

Elements affecting WER

- **Children’ speech acoustic mismatch**: 36.7%
- **Lexical inaccuracies due to the speech disorders**: 28.2%
- **Acoustic distortion due to speech disorders**: 9.1%
- **Children speech, children models**: 2.1%
- **Disordered speech, children models**: 0.0%
Acoustic Analysis of the Corpus
Acoustic Analysis of the Corpus

Acoustic features

Vowels

- First and second formant frequencies
- Fundamental frequency (pitch)
- Intensity (Energy: SNR)
- Duration
Acoustic Analysis of the Corpus

Acoustic features

Feature extraction method

Input signal → Forced alignment → Vowel boundaries → Feature extraction → Output features

- Energy
- LPC analysis
- Formants
- Prediction error
- Pitch
Acoustic Analysis of the Corpus

Values in children’ speech

Unimpaired speech

(a) Formant map

(b) Pitch means

(a) Intensity histograms

(b) Duration histogram
Impaired speech: Correct vowels

(a) Formant map  
(b) Pitch means

(a) Intensity histograms  
(b) Duration histogram
Results in the impaired speakers

Impaired speech: substituted vowels

(a) Formant map

(b) Pitch means

(a) Intensity histograms

(b) Duration histogram
Acoustic Analysis of the Corpus

Measurement of the acoustic distortion

- **Kullback-Leibler Divergence**

\[
KL(A, B) = \sum_{i=0}^{n} \left[ \log\left( \frac{\sum A_i}{\sum B_i} \right) + \frac{(\mu_{A_i} - \mu_{B_i})^2}{\sum B_i} \right] + \frac{\sum A_i}{\sum B_i} \tag{3}
\]

\[
sKLD(A, B) = \frac{KL(A, B) + KL(B, A)}{2} \tag{4}
\]

- **Fisher’s Ratio**

\[
FR(A, B) = \sum_{i=0}^{n} \left( \frac{(\mu_{A_i} - \mu_{B_i})^2}{\sum A_i + \sum B_i} \right) \tag{5}
\]
Measurement of the acoustic distortion

- **Loss of formant separability**

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Unimp.</th>
<th>Imp. (cor.)</th>
<th>Imp. (subs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sKLD</td>
<td>FR</td>
<td>sKLD</td>
</tr>
<tr>
<td>/a/-/e/</td>
<td>18.59</td>
<td>6.74</td>
<td>17.43</td>
</tr>
<tr>
<td>/a/-/o/</td>
<td>12.04</td>
<td>4.42</td>
<td>6.49</td>
</tr>
<tr>
<td>/e/-/i/</td>
<td>5.91</td>
<td>2.37</td>
<td>11.36</td>
</tr>
<tr>
<td>/e/-/o/</td>
<td>22.89</td>
<td>7.13</td>
<td>16.34</td>
</tr>
<tr>
<td>/o/-/u/</td>
<td>4.99</td>
<td>2.27</td>
<td>7.42</td>
</tr>
</tbody>
</table>

- **Maintaining of the pitch features**

<table>
<thead>
<tr>
<th></th>
<th>Unimp.</th>
<th>Imp. (cor.)</th>
<th>Imp. (subs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sKLD</td>
<td>FR</td>
<td>sKLD</td>
</tr>
<tr>
<td></td>
<td>0.85</td>
<td>0.38</td>
<td>1.91</td>
</tr>
</tbody>
</table>
Acoustic Analysis of the Corpus

Measurement of the acoustic distortion

- Loss of intensity feature
  
<table>
<thead>
<tr>
<th>Unimp.</th>
<th>Imp. (cor.)</th>
<th>Imp. (subs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sKLD</td>
<td>FR</td>
<td>sKLD</td>
</tr>
<tr>
<td>1.04</td>
<td>0.44</td>
<td>0.01</td>
</tr>
</tbody>
</table>

- Irregular production in vowel duration
  
  - 30% increase in mean duration (Longer vowels)
  - 100% increase in standard deviation (Less steady vowels)
Lexical Analysis of the Corpus
Lexical Analysis of the Corpus

Speakers’ consistency

Consistency of the impaired speakers’ pronunciations
- Probability of pronouncing the same phoneme within a same context in the same way.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Consistency</th>
<th>Speaker</th>
<th>Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spk01</td>
<td>98.00%</td>
<td>Spk02</td>
<td>87.10%</td>
</tr>
<tr>
<td>Spk03</td>
<td>91.04%</td>
<td>Spk04</td>
<td>95.77%</td>
</tr>
<tr>
<td>Spk05</td>
<td>72.26%</td>
<td>Spk06</td>
<td>98.74%</td>
</tr>
<tr>
<td>Spk07</td>
<td>88.47%</td>
<td>Spk08</td>
<td>81.62%</td>
</tr>
<tr>
<td>Spk09</td>
<td>94.29%</td>
<td>Spk10</td>
<td>83.85%</td>
</tr>
<tr>
<td>Spk11</td>
<td>94.29%</td>
<td>Spk12</td>
<td>77.96%</td>
</tr>
<tr>
<td>Spk13</td>
<td>73.06%</td>
<td>Spk14</td>
<td>94.24%</td>
</tr>
<tr>
<td>AVG</td>
<td>87.91%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Non-contextual studies

Mispronunciations rates

(a) per phoneme  (b) manner of articulation  (c) point of articulation
Influence of the syllable context

Mispromnunciations per syllable length

- Correct
- Substituted
- Deleted
Influence of the syllable context

Syllable construction in Spanish

- **V**: to-a-lía
- **CV**: bru-ja
- **VC**: ar-bol
- **CCV**: ca-bra
- **CVC**: cam-pa-na
- **CVV**: pue-blo
- **CVVC**: puer-ta
- **Others (VVC, CCVV, CVVV, CCVVC ...)**
Lexical Analysis of the Corpus

Influence of the syllable context

Vowels

- Nucleus vs glide

Consonant

- Onset vs Coda
- Onset vs Beginning of 2-consonant cluster
- Onset vs Ending of 2-consonant cluster
Influence of the syllable context

Vowels: Nucleus vs glide

(a) /i/: Glide

(b) /u/: Glide
Lexical Analysis of the Corpus

Influence of the syllable context

Consonants: Onset vs coda

(a) /m/: Coda

(b) /n/: Coda

(c) /T/: Coda

(d) /s/: Coda

(e) /r/: Coda

(f) /l/: Coda
Influence of the syllable context

Consonants: Onset vs beginning of consonant cluster

(a) /b/: Cluster
(b) /g/: Cluster
(c) /p/: Cluster
(d) /k/: Cluster
(e) /f/: Cluster
(f) /t/: Cluster
Influence of the syllable context

Consonants: Onset vs ending of consonant cluster

(a) /r/: Cluster

(b) /l/: Cluster
Lexical Analysis of the Corpus

Influence of the syllable context

Study of statistical significance

- Vowels
  - Nucleus vs glide

- Consonants
  - Onset vs Coda
  - Onset vs Beginning of 2-consonant cluster
  - Onset vs Ending of 2-consonant cluster

Tool:

- z-test

\[
z = \frac{p_{position} - p_{baseline}}{\sqrt{p_{phoneme}(1 - p_{phoneme})\left(\frac{1}{n_{position}} + \frac{1}{n_{baseline}}\right)}}
\]  

(3)
Influence of the syllable context

Study of statistical significance

- **Vowels**
  - Nucleus vs glide
    - Significant increase of substitutions and deletions for /i/ and /u/
    - Significant decrease of corrections for /i/ and /u/
  - Consonants
    - Onset vs Coda
      - Significant increase of deletions for /m/, /n/, /T/, /s/, /r/, /l/
      - Significant decrease of substitutions for /m/, /n/, /T/, /l/
      - Significant decrease of corrections for /m/, /n/, /s/, /r/, /l/
    - Onset vs Beginning of 2-consonant cluster
      - Significant increase of deletions for /b/, /g/
      - Significant decrease of corrections for /b/, /g/
    - Onset vs Ending of 2-consonant cluster
      - Significant increase of substitutions and deletions for /r/ and /l/
      - Significant decrease of corrections for /r/ and /l/
Language acquisition delays

Mispronunciations in 293 healthy children [Bosch-Galcerán, 2004]
Acoustic-Lexical Adaptation for ASR of Disordered Speech
Acoustic adaptation techniques

- **Maximum a Posteriori** [Gauvain and Lee, 1994]
  - Computationally simple
  - Convergent to ML with sufficient data
  - Understanding of the effect of lexical variants
  - Sensible to data sparsity

- **Maximum Likelihood Linear Regression** [Legetter and Woodland, 1995]
  - Requires less data for adaptation
  - Creates clusters of phonemes to “expand” data
Acoustic-speaker adaptation

Acoustic adaptation framework: Baseform transcription
## Acoustic speaker adaptation

### Acoustic adaptation framework: Baseform transcription (Results and Improvements)

<table>
<thead>
<tr>
<th></th>
<th>WER</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group A</strong></td>
<td>2.10%</td>
<td>2.45%</td>
<td>2.19%</td>
<td></td>
</tr>
<tr>
<td><strong>Group B</strong></td>
<td>7.54%</td>
<td>10.79%</td>
<td>9.65%</td>
<td></td>
</tr>
<tr>
<td><strong>Group C</strong></td>
<td>32.68%</td>
<td>40.79%</td>
<td>39.37%</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>12.78%</td>
<td>16.38%</td>
<td>15.49%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Improve</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group A</strong></td>
<td>60.08%</td>
<td>76.33%</td>
<td>65.35%</td>
<td></td>
</tr>
<tr>
<td><strong>Group B</strong></td>
<td>63.11%</td>
<td>61.81%</td>
<td>59.56%</td>
<td></td>
</tr>
<tr>
<td><strong>Group C</strong></td>
<td>43.56%</td>
<td>35.64%</td>
<td>33.75%</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>50.56%</td>
<td>48.75%</td>
<td>45.07%</td>
<td></td>
</tr>
</tbody>
</table>
Acoustic speaker adaptation

Acoustic adaptation framework: Human labels

TRAINING

Adaptation data → Human experts

transcriptions

MAP adaptation

acoustic signals

TESTING

Recognition data → acoustic models

signals

ASR → Results
Acoustic-Lexical Adaptation for ASR of Disordered Speech

Acoustic speaker adaptation

Acoustic adaptation framework: Human labels (Results and improvements)

<table>
<thead>
<tr>
<th>WER</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>-</td>
<td>2.89%</td>
<td>2.28%</td>
</tr>
<tr>
<td>Group B</td>
<td>-</td>
<td>13.07%</td>
<td>11.67%</td>
</tr>
<tr>
<td>Group C</td>
<td>-</td>
<td>45.29%</td>
<td>43.86%</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>18.64%</td>
<td>17.51%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Improve</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>-</td>
<td>72.08%</td>
<td>63.92%</td>
</tr>
<tr>
<td>Group B</td>
<td>-</td>
<td>53.73%</td>
<td>51.09%</td>
</tr>
<tr>
<td>Group C</td>
<td>-</td>
<td>28.54%</td>
<td>26.20%</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>41.68%</td>
<td>37.91%</td>
</tr>
</tbody>
</table>
Acoustic speaker adaptation

Acoustic adaptation framework: APD
Acoustic-Lexical Adaptation for ASR of Disordered Speech

## Acoustic speaker adaptation

Acoustic adaptation framework: APD (Results and improvements)

<table>
<thead>
<tr>
<th>WER</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>-</td>
<td>9.91%</td>
<td>7.46%</td>
</tr>
<tr>
<td>Group B</td>
<td>-</td>
<td>24.38%</td>
<td>23.07%</td>
</tr>
<tr>
<td>Group C</td>
<td>-</td>
<td>60.64%</td>
<td>61.40%</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>29.57%</td>
<td>28.45%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Improve</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>-</td>
<td>4.25%</td>
<td>18.04%</td>
</tr>
<tr>
<td>Group B</td>
<td>-</td>
<td>13.70%</td>
<td>3.31%</td>
</tr>
<tr>
<td>Group C</td>
<td>-</td>
<td>4.32%</td>
<td>3.31%</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>7.48%</td>
<td>0.89%</td>
</tr>
</tbody>
</table>
Acoustic-Lexical Adaptation for ASR of Disordered Speech

Acoustic speaker adaptation

Influence of the amount of adaptation data

- 8 sessions from Spk07 and Spk08
- Evaluation for all possibilities in cross adaptation
Proposals for lexical adaptation [Strik, 2001][Caballero et al., 2004]

- **Rule-based**
  - Based on a priori linguistic knowledge
  - Viable for dialectal lexical adaptation

- **Data-driven**
  - Based on observations
  - Required manual or automated transcription
  - Impossible generalization of disorders, need of personalization
Lexical adaptation framework: APD

TRAINING

Adaptation data

prompt

Adaptation data

SI-APD

TESTING

Recognition data

signals

Lexicon

ASR

Results
Lexical speaker adaptation

Lexical adaptation: APD (Results and improvements)

<table>
<thead>
<tr>
<th>WER</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>-</td>
<td>10.53%</td>
<td>6.75%</td>
</tr>
<tr>
<td>Group B</td>
<td>-</td>
<td>20.35%</td>
<td>20.18%</td>
</tr>
<tr>
<td>Group C</td>
<td>-</td>
<td>54.17%</td>
<td>51.54%</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>26.50%</td>
<td>24.34%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Improve</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>-</td>
<td>1.74%</td>
<td>6.80%</td>
</tr>
<tr>
<td>Group B</td>
<td>-</td>
<td>27.96%</td>
<td>15.42%</td>
</tr>
<tr>
<td>Group C</td>
<td>-</td>
<td>14.53%</td>
<td>13.25%</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>17.08%</td>
<td>13.69%</td>
</tr>
</tbody>
</table>
Lexical speaker adaptation

Influence of the amount of adaptation data

![Graph showing the influence of the amount of adaptation data](image)
Mixed speaker adaptation

Acoustic-lexical adaptation framework: Baseform transcription
## Mixed speaker adaptation

### Acoustic-lexical adaptation framework: Baseform transcription

(Results and improvements)

<table>
<thead>
<tr>
<th>WER</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>-</td>
<td>3.42%</td>
<td>2.45%</td>
</tr>
<tr>
<td>Group B</td>
<td>-</td>
<td>11.49%</td>
<td>10.44%</td>
</tr>
<tr>
<td>Group C</td>
<td>-</td>
<td>41.56%</td>
<td>41.12%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>-</td>
<td>17.20%</td>
<td>16.35%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Improve</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>-</td>
<td>66.96%</td>
<td>61.23%</td>
</tr>
<tr>
<td>Group B</td>
<td>-</td>
<td>59.33%</td>
<td>56.24%</td>
</tr>
<tr>
<td>Group C</td>
<td>-</td>
<td>43.77%</td>
<td>30.81%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>-</td>
<td>46.18%</td>
<td>42.02%</td>
</tr>
</tbody>
</table>
Mixed speaker adaptation

Acoustic-lexical adaptation framework: Human labels
## Mixed speaker adaptation

Acoustic-lexical adaptation framework: Human labels (Results and improvements)

<table>
<thead>
<tr>
<th></th>
<th>WER</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>3.42%</td>
<td>-</td>
<td>12.11%</td>
<td>42.76%</td>
</tr>
<tr>
<td>Group B</td>
<td>12.11%</td>
<td>-</td>
<td>11.75%</td>
<td>42.65%</td>
</tr>
<tr>
<td>Group C</td>
<td>42.76%</td>
<td>-</td>
<td>42.65%</td>
<td>42.65%</td>
</tr>
<tr>
<td>Total</td>
<td>17.76%</td>
<td>-</td>
<td>17.29%</td>
<td>17.29%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Improve</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>66.96%</td>
<td>-</td>
<td>59.81%</td>
<td>59.81%</td>
</tr>
<tr>
<td>Group B</td>
<td>57.13%</td>
<td>-</td>
<td>50.75%</td>
<td>50.75%</td>
</tr>
<tr>
<td>Group C</td>
<td>32.53%</td>
<td>-</td>
<td>28.23%</td>
<td>28.23%</td>
</tr>
<tr>
<td>Total</td>
<td>44.43%</td>
<td>-</td>
<td>38.69%</td>
<td>38.69%</td>
</tr>
</tbody>
</table>
Mixed speaker adaptation

Acoustic-lexical adaptation framework: APD
Mixed speaker adaptation

Acoustic-lexical adaptation framework: APD (Results and improvements)

<table>
<thead>
<tr>
<th></th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td></td>
<td>7.19%</td>
<td>7.81%</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td>16.90%</td>
<td>17.15%</td>
</tr>
<tr>
<td>Group C</td>
<td></td>
<td>51.53%</td>
<td>48.36%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>23.33%</td>
<td>22.71%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td></td>
<td>30.53%</td>
<td>23.58%</td>
</tr>
<tr>
<td>Group B</td>
<td></td>
<td>40.18%</td>
<td>28.12%</td>
</tr>
<tr>
<td>Group C</td>
<td></td>
<td>18.70%</td>
<td>18.63%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>27.23%</td>
<td>19.47%</td>
</tr>
</tbody>
</table>
Considerations on the mixed approach

<table>
<thead>
<tr>
<th>Improve</th>
<th>Acoustic adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No adaptation</td>
</tr>
<tr>
<td>No adapt.</td>
<td>-</td>
</tr>
<tr>
<td>APD</td>
<td>17.08%</td>
</tr>
</tbody>
</table>
Confidence Measuring for Detection of Mispronunciations
The PV task

Speaker verification vs pronunciation verification

Speaker verification

Target utterance

Target model

Speaker i

Competing models

UBM
Speaker 1
Speaker M
Speaker j

Pronunciation verification

Target utterance

Target model

Phoneme n

Competing models

UBM
Phoneme 1
Phoneme m
Phoneme N
The PV task

Sources of variability
Baseline scoring

Baseline scoring system

Score histograms
Confidence Measuring for Detection of Mispronunciations

Baseline scoring

Measure of performance

- Detection curve: FAR and FRR for all possible decision thresholds
- EER: Working point in the detection curve where FAR=FRR

EER of the baseline log-likelihood score: 44%
Scoring system with score normalization
Score normalization

- **T-norm** [Auckenthaler et al., 2000]

\[
LL_{T-norm}(p) = \frac{LL(p) - \mu_{utterance}}{\sigma_{utterance}}
\]  

- **Z-norm**

\[
LL_{Z-norm}(p) = \frac{LL(p) - \mu_{speaker}}{\sigma_{speaker}}
\]  

- **GOP-like normalization** [Witt and Young, 2000]

\[
LL_{GOP_{norm}}(p) = LL(p) - LL\left(\frac{1}{N} \sum_{n=1}^{N} (P(s|\lambda_n))\right)
\]  

- **1-best normalization**

\[
LL_{1-best}(p) = LL(p) - LL(l)
\]
Score normalization

- T-norm
- Z-norm

- GOP-like normalization
- 1-best normalization
Score normalization

- T-norm
- Z-norm
- GOP-like normalization
- 1-best normalization

(a) Detection curve  (b) EER zone
Speaker adaptation

Acoustic adaptation framework: Baseform transcription
Acoustic adaptation framework: Baseform transcription

(a) Detection curve  (b) EER zone
Acoustic adaptation framework: Human labels

![Acoustic adaptation framework diagram]

- **TRAINING**
  - **Adaptation data**
  - **Human experts**
  - **transcriptions**
  - **MAP adaptation**

- **TESTING**
  - **Recognition data**
  - **ASR**
  - **Results**
  - **acoustic models**
Speaker adaptation

Acoustic adaptation framework: Human labels

(a) Detection curve

(b) EER zone
Speaker adaptation

Acoustic adaptation framework: APD
Confidence Measuring for Detection of Mispronunciations

Speaker adaptation

Acoustic adaptation framework: APD

(a) Detection curve

(b) EER zone
## Comparison of techniques

### EER values for all cases

<table>
<thead>
<tr>
<th>Norm.</th>
<th>No adapt.</th>
<th>Baseform</th>
<th>Human labels</th>
<th>APD</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-norm</td>
<td>22.80%</td>
<td>21.29%</td>
<td>18.16%</td>
<td>20.46%</td>
</tr>
<tr>
<td>GOP</td>
<td>21.44%</td>
<td>18.87%</td>
<td>15.71%</td>
<td>19.48%</td>
</tr>
<tr>
<td>1-best</td>
<td>22.01%</td>
<td>19.17%</td>
<td>15.63%</td>
<td>19.55%</td>
</tr>
</tbody>
</table>
2-pass system

Confidence Measuring for Detection of Mispronunciations

- 2-pass system

- Adaptation Data and Baseform Transcription

- Speaker adaptation

- Test data

- Evaluation decision

- Pronunciation verification

- Pronunciation verification
2-pass system

First pass of PV working on: Three working points
Unsupervised On-line Systems for ASR of the Speech Handicapped
Baseline unsupervised adaptation

Baseline unsupervised scheme

Unsupervised On-line Systems for ASR of the Speech Handicapped
Unsupervised On-line Systems for ASR of the Speech Handicapped

Baseline unsupervised adaptation

<table>
<thead>
<tr>
<th>Group</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>3.77%</td>
<td>3.95%</td>
<td>3.86%</td>
</tr>
<tr>
<td>Group B</td>
<td>17.11%</td>
<td>17.98%</td>
<td>17.90%</td>
</tr>
<tr>
<td>Group C</td>
<td>59.43%</td>
<td>57.46%</td>
<td>57.78%</td>
</tr>
<tr>
<td>Total</td>
<td>24.44%</td>
<td>24.25%</td>
<td>24.28%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>28.33%</td>
<td>61.84%</td>
<td>38.92%</td>
</tr>
<tr>
<td>Group B</td>
<td>16.29%</td>
<td>36.35%</td>
<td>24.98%</td>
</tr>
<tr>
<td>Group C</td>
<td>2.64%</td>
<td>9.34%</td>
<td>2.78%</td>
</tr>
<tr>
<td>Total</td>
<td>5.45%</td>
<td>24.12%</td>
<td>15.10%</td>
</tr>
</tbody>
</table>
Prior WER affecting Posterior WER

Ideally: Flattening of the curve

![Graph showing the relationship between WER before and after adaptation](image-url)
Real situation: Supervised adaptation
Prior WER affecting Posterior WER

Real situation: Unsupervised adaptation
CM-based unsupervised adaptation

Confidence measuring for unsupervised adaptation
CM-based unsupervised adaptation

Confidence measuring for unsupervised adaptation (results and improvements)

<table>
<thead>
<tr>
<th></th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group A</strong></td>
<td>-</td>
<td>5.00%</td>
<td>3.68%</td>
</tr>
<tr>
<td><strong>Group B</strong></td>
<td>-</td>
<td>20.61%</td>
<td>19.91%</td>
</tr>
<tr>
<td><strong>Group C</strong></td>
<td>-</td>
<td>59.65%</td>
<td>60.75%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>-</td>
<td>26.19%</td>
<td>25.78%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Word models</th>
<th>Phone models</th>
<th>Sub-phone models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group A</strong></td>
<td>-</td>
<td>51.69%</td>
<td>41.77%</td>
</tr>
<tr>
<td><strong>Group B</strong></td>
<td>-</td>
<td>27.04%</td>
<td>16.55%</td>
</tr>
<tr>
<td><strong>Group C</strong></td>
<td>-</td>
<td>5.89%</td>
<td>2.22%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>-</td>
<td>18.05%</td>
<td>8.58%</td>
</tr>
</tbody>
</table>
On-line personalization
On-line personalization (results and improvements)

<table>
<thead>
<tr>
<th>Spk</th>
<th>lt 1</th>
<th>lt 2</th>
<th>lt 3</th>
<th>lt 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>6.32%</td>
<td>4.59%</td>
<td>4.05%</td>
<td>3.51%</td>
</tr>
<tr>
<td>Group B</td>
<td>23.86%</td>
<td>21.46%</td>
<td>19.97%</td>
<td>18.54%</td>
</tr>
<tr>
<td>Group C</td>
<td>60.97%</td>
<td>60.77%</td>
<td>60.20%</td>
<td>59.72%</td>
</tr>
<tr>
<td>AVG</td>
<td>28.20%</td>
<td>26.67%</td>
<td>25.78%</td>
<td>24.94%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spk</th>
<th>lt 1</th>
<th>lt 2</th>
<th>lt 3</th>
<th>lt 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>-</td>
<td>27.37%</td>
<td>35.92%</td>
<td>44.46%</td>
</tr>
<tr>
<td>Group B</td>
<td>-</td>
<td>10.06%</td>
<td>16.30%</td>
<td>22.30%</td>
</tr>
<tr>
<td>Group C</td>
<td>-</td>
<td>0.34%</td>
<td>1.26%</td>
<td>2.05%</td>
</tr>
<tr>
<td>AVG</td>
<td>-</td>
<td>5.43%</td>
<td>8.58%</td>
<td>11.56%</td>
</tr>
</tbody>
</table>
Development of Speech Based Applications for the Handicapped
Speech interfaces for the handicapped

- Speech control of automatized wheelchair [Alcubierre et al., 2005]
- Voice triggered mouse events [Saz et al., 2009]
- Speech control of virtual mouse [González, 2009]
CASLT tools: “Comunica”

- “PreLingua” [Rodríguez et al., 2007]
- “Vocaliza” [Vaquero, 2006]
- “Cuéntame” [Escartín, 2008]
Key points in “Comunica”

- Robust use of speech technologies: ASR, TTS, PV, speech analysis
- Intensive use of AAC elements: Pictographs, text and audio
- Ease of use for therapists
- Creation of user profiles
L2 tools: “VocalizaL2”

“VocalizaL2”: Modification of “Vocaliza”
- Phone level evaluation instead of word level
- Pronunciation verification instead of ASR
L2 tools: “VocalizaL2”

Real-world experience with CALL tools [Rodríguez, 2008]
- Performed at the Vienna International School
- 12 students in 6th grade (11 years old)
- 5 sessions of 45 minutes each (10 words per session)
- First year learning Spanish
- Variety of mother tongues, also proficient in English and German
L2 tools: “VocalizaL2”

Qualitative results

- **From the students**
  - Enthusiastic with the application
  - Found evaluation coherent with their effort
  - Enjoyed competition in session 5
  - Found evaluation “weird” sometimes
  - TTS voice was little natural

- **From the teacher**
  - Interest in the use of user profiles
  - Animations matched the children age
  - Children were less shy than in class
  - Compatibility problems
  - Minor bugs
L2 tools: “VocalizaL2”

Quantitative results

(a) Trials

(b) Sessions

(a) Words

(b) Phonemes
Discussion
Origins of the speech disorders

Elements affecting WER

- Children’s speech acoustic mismatch: 36.7%
- Disordered speech, adult models

- Lexical inaccuracies due to the speech disorders: 28.2%
- Disordered speech, children models

- Acoustic distortion due to speech disorders: 9.1%
- Disordered speech, children models, lex. correct words

- WER: 2.1%
- Children speech, children models

- WER: 0.0%
- Children speech, children models
Origins of the speech disorders

- **Acoustic degradation**
  - Loss of acoustic properties is reasonably low
  - Except for substitutions (which relate more to lexical variants)

- **Lexical variants**
  - High level of lexical variability
  - Study carried out points out the effect of delayed language acquisition
Issues on Personalization

ASR case: Acoustic vs acoustic lexical adaptation
Issues on Personalization

PV case: Influence of mispronunciations in training

Acoustic Adaptation

Train: [arbol]
Test: [arbol], rejection of [r] and [l]

Acoustic and Lexical Adaptation

Train: [abo]
Test: [arbol], rejection of [r] and [l]
From research to development

Cloud-based flow of work
From research to development

Examples: ARASAAC

- Provides free pictographs to “Comunica” and other free-distribution technical aids
- Created by the CATEDU, CPEE “Alborada” and Sergio Palao
- Creative Commons (CC) alternative to licensed pictographs
From research to development

Our effort in dissemination

- Collaborations
  - CPEE “Alborada”
  - Autismo Galicia & others
  - CADIS-Huesca & ASPACE-Huesca
  - Research groups at UZ

- Distribution
  - http://www.vocaliza.es
  - +5000 registered users
  - +500 exchanged mails
  - 400 visitors averaged daily, up to 1400 max.

- Participation
  - Invited talks: Xornadas Autismo Galicia
  - Publications in specific forums: ESAAC, Aitadis
Conclusion
Summary

Thesis summary

CORPUS + BASELINE

- Confidence measuring
- Evaluation
- Acoustic analysis
- Acoustic adaptation
- Phonological analysis
- Lexical adaptation
- Acoustic-lexical adaptation
- Unsupervised on-line adaptation
Conclusion

Fulfillment of objectives

- **Research objectives**
  - Corpus has shown to be useful and some interest has been shown by the community
  - Personalization techniques have been evaluated and the performance of them has been framed
  - Confidence measuring has achieved significant improvements and solid results

- **Development objectives**
  - Speech interfaces still to be worked: Now we have more knowledge
  - “Comunica” has been a great success as CALL tools
Fulfillment of objectives

Publication of results

- Journals of impact

- Peer-reviewed conferences
  - Saz, Lleida et al. (2009). *In Interspeech 2009*
  - Saz, Miguel et al. (2009). *In Interspeech 2006*
  - Vaquero, Saz et al. (2009). *In ICASSP 2008*
  - Saz, Lleida et al. (2009). *In AVFA 2009*
  - Saz, Lleida et al. (2009). *In WOCCI 2009*
  - Saz, Rodríguez et al. *In WOCCI 2008*
  - Saz, Rodríguez et al. *In SLaTE 2009*
Fulfillment of objectives

Scientific collaborations

- Internship June-September 2006: Department of Computer Science, The University of Sheffield
- Internship June-September 2007: Department of Electrical and Computer Engineering, McGill University
  - Yin, Rose et al. (2008). *In Interspeech 2008*
  - Yin, Rose et al. (2009). *In ICASSP 2009*
Future work

- Improvements in unsupervised adaptation techniques
  - MLLR-based techniques
- Improvements in confidence measuring
  - Mutual Information, Large Margin, etc
- Evaluations in different corpora
- Speech controlled “MICE”
- Dissemination of “Comunica”
Tesis doctoral

On Line Personalization and Adaptation to Disorders and Variations of Speech on Automatic Speech Recognition Systems

Oscar Saz

Director: Prof. Eduardo Lleida