The Alborada-I3A corpus of disordered speech

Oscar Saz, E. Lleida, C. Vaquero, W.-R. Rodríguez
Aragón Institute for Engineering Research (I3A)
University of Zaragoza, Spain
Index

- Introduction
- Impaired speakers corpus
- Extensions
- Experimentation with the corpus
- Conclusions
Introduction

- Interest in research in HLTs for the handicapped

- Collaboration in Zaragoza (Spain) between
  - Aragón Institute for Engineering Research (I3A)
  - Public School Special Education (CPEE) “Alborada”

- Aim
  - Development of assistance systems based on speech technology for the handicapped
  - Development of language learning tools for children with special linguistic needs
Introduction

- Lack of speech corpora, different requirements in different approaches
  - Whitaker database (Deller et al., 1993)
  - Nemours database (Menéndez-Pidal et al., 1996)
  - Universal Access database (Kim et al., 2008)
  - HACRO database (Navarro-Mesa et al., 2005)
  - Other languages…
Index

- Introduction
- Impaired speakers corpus
- Extensions
- Experimentation with the corpus
- Conclusions
Impaired speech corpus

Requirements for a corpus useful in speech recognition and assessment

- Variety of impairments and disorders
- Realistic speech
- Short and balanced vocabulary
- Several sessions per speaker
Impaired speakers corpus

Recording environment

- Facilities of the CPEE Alborada
- Each speaker supervised by member of I3A and Alborada
- Headset wireless microphone to reduce ambient noise, mounted in conventional laptop
- 16 kHz, 16 bit
Impaired speakers corpus

- Recording environment
  - Recording tool was Vocaliza (Vaquero et al., 2008)
  - Provides audio-visual prompting
Speaker selection

| Speaker | Gender | Age     | Speaker | Gender | Age     |
|---------|--------|---------|---------|--------|---------|
| Spk001  | Female | 14 years| Spk002  | Male   | 11 years|
| Spk003  | Male   | 21 years| Spk004  | Female | 21 years|
| Spk005  | Male   | 18 years| Spk006  | Male   | 17 years|
| Spk007  | Male   | 18 years| Spk008  | Male   | 19 years|
| Spk009  | Female | 11 years| Spk010  | Female | 15 years|
| Spk011  | Female | 20 years| Spk012  | Male   | 18 years|
| Spk013  | Female | 13 years| Spk014  | Female | 11 years|

Big impact of impairments and disorders:
- Down syndrome & other cognitive and physical impairments
- Dysarthria & other speech and language disorders
Impaired speakers’ corpus

- Session design
  - Isolated word sessions: 57 words per session, 4 sessions per speaker (3192 utterances – 2h 17m data)

| Prompt       | SAMPA   | Prompt       | SAMPA   | Prompt       | SAMPA   | Prompt       | SAMPA   |
|--------------|---------|--------------|---------|--------------|---------|--------------|---------|
| árbol        | [“ArBOI]| boca         | [“boka]  | bruja        | [“bruxa] | cabra        | [“kaBra]| |
| campana      | [kAm“pan~]| caramelo     | [“kara“melo] | casa        | [“kasa]  | clavo        | [“klaBo]| |
| cuchara      | [ku”tSara]| dedo         | [“deDo]  | ducha        | [“dutSa] | escoba       | [es“koBa]| |
| flan         | [“flAn]  | fresa        | [“fresa] | fuma         | [“fuma]  | gafas        | [“gafAs]| |
| globo        | [“gloBo] | gorro        | [“gorro] | grifo        | [“grifo] | indio        | [“indjo]| |
| jarra        | [“xarra] | jaula        | [“xawla] | lápiz        | [“lapIT] | lavadora     | [laBa“Dora]| |
| luna         | [“luna]  | llave        | [“LaBe]  | mariposa     | [mari“posa]| moto         | [“moto]| |
| niño         | [“ni”Jo]| ojo          | [“Oxo]   | pala         | [“pala]  | palmera      | [pAl“mera]| |
| pan          | [“pAn]   | peine        | [“pEjne] | periódico    | [pe“rjDiko]| pez          | [“peT]| |
| piano        | [“piano] | pie          | [“pje]   | piña         | [“piJa]  | pistola      | [p1s“tola]| |
| plátano      | [pla“tano~]| playa       | [“plajja]| preso        | [“preso] | pueblo       | [“pueB1o]| |
| puerta       | [“pwerta]| ratón        | [ra“ton] | semáforo     | [se“maforo]| silla        | [“siLa]| |
| sol          | [“sOL]   | tambor       | [tAm“BOr] | taza         | [“taTa]  | teléfono      | [te“lefono]| |
| toalla       | [to**aLa]| toro         | [“toro]  | tortuga      | [tOr“tuGa]| tren         | [“tren]| |
| zapato       | [Ta“pato] |              |         |              |         |              |         |

- RFI (Monfort & Juárez-Sánchez, 1989)
Impaired speakers’ corpus

Session design

- Meaningless sentence sessions: 4 speakers uttering 112 sentences (448 utterances – 25m of data)

  el/la [Word1] y el/la [Word2]

- Meaningful sentence sessions: 3 speakers uttering 10 full sentences with 3 RFI words
Index

- Introduction
- Impaired speakers corpus
- Extensions
- Experimentation with the corpus
- Conclusions
Extensions: Further data

- Speakers *Spk007* and *Spk008* were recorded again 2 years after the initial recordings
  - Stored as speakers *Spk107* and *Spk108*
  - Repetition of the 4 RFI isolated word sessions
  - Possibility for longitudinal studies
  - More data for adaptation
Extensions: Reference corpus

- Recordings of age-matched unimpaired peers

| Age     | Males | Females | Age     | Males | Females |
|---------|-------|---------|---------|-------|---------|
| 10 years| 15    | 16      | 11 years| 15    | 16      |
| 12 years| 15    | 15      | 13 years| 15    | 23      |
| 14 years| 11    | 21      | 15 years| 11    | 11      |
| 16 years| 15    | 9       | 17 years| 14    | 10      |
| All     | 111   | 121     |         |       |         |

- One RFI isolated word session per speaker (13224 utterances – 8h50m data)
  - CEIP Río Ebro, IES Tiempos Modernos, IES Félix de Azara
Extensions: Human labeling

A set of 12 experts were requested to perform perceptual labeling of lexical mispronunciations.
Extensions: Human labeling

- Final results marked more than 17% of phonemes as substituted (10%) or deleted (7%)

| Speaker | Correct | Substituted | Deleted | Speaker | Correct | Substituted | Deleted |
|---------|---------|-------------|---------|---------|---------|-------------|---------|
| Spk001  | 98.88%  | 0.94%       | 0.17%   | Spk002  | 78.42%  | 12.41%      | 9.16%   |
| Spk003  | 94.78%  | 4.54%       | 0.68%   | Spk004  | 96.83%  | 2.05%       | 1.11%   |
| Spk005  | 56.51%  | 26.11%      | 17.38%  | Spk006  | 99.32%  | 0.51%       | 0.17%   |
| Spk007  | 87.07%  | 7.36%       | 5.57%   | Spk008  | 69.18%  | 17.72%      | 13.10%  |
| Spk107  | 82.11%  | 9.59%       | 8.30%   | Spk108  | 69.43%  | 18.15%      | 12.41%  |
| Spk009  | 91.78%  | 5.31%       | 2.91%   | Spk010  | 78.51%  | 13.10%      | 8.39%   |
| Spk011  | 93.24%  | 5.15%       | 2.05%   | Spk012  | 74.32%  | 13.96%      | 11.73%  |
| Spk013  | 43.58%  | 30.48%      | 25.94%  | Spk014  | 91.01%  | 5.14%       | 3.85%   |

- Interlabeler agreement: 85%
Index

- Introduction
- Impaired speakers corpus
- Extensions
- Experimentation with the corpus
- Conclusions
Experimentation with the corpus

- Analysis of speech disorders
  - Degradation of the acoustic quality in the impaired speakers compared to the unimpaired peers

- Patterns of lexical mispronunciation: Reduction of diphthongs, codas and consonant clusters
Experimentation with the corpus

- Speech recognition and speaker adaptation
  - Results with different algorithms for adaptation

|        | Baseline | MAP  | MLLR  | MLLR+MAP |
|--------|----------|------|-------|----------|
| WER    | 28.20%   | 15.48%| 14.69%| 12.53%   |

- Also results in lexical adaptation to the speaker (up to 20% relative improvement)

- Pronunciation verification and assessment
  - Precision curves around 15% Equal Error Rate
Index

- Introduction
- Impaired speakers corpus
- Extensions
- Experimentation with the corpus
- Conclusions
Conclusions

- Interest in sharing speech data in this area
  - Available, contact authors (oskarsaz@unizar.es, http://oscar.vivolab.es)
  - Restrictions due to conditions of the speakers

- Our corpus includes
  - Sufficient data
  - Wide range of disorders and linguistic affections
  - Extra data for work (labeling…)

- Inclusion in the LREC2010 Map
Conclusions

Further reading:

- O. Saz, J. Simón, W.-R. Rodríguez, E. Lleida, & C. Vaquero, 2009. Analysis of acoustic features in speakers with cognitive disorders and speech impairments. EURASIP Jounal of Advances in Signal Processing.

- O. Saz, E. Lleida, & A. Miguel, 2009. Combination of acoustic and lexical speaker adaptation for disordered speech recognition. In Interspeech, Brighton, UK.

- O. Saz, S.-C. Yin, E. Lleida, R. Rose, W.-R. Rodríguez, and C. Vaquero. 2009. Tools and technologies for computer-aided speech and language therapy. Speech Communication, 51(10):948–967

- S.-C. Yin, R. Rose, O. Saz, & E. Lleida,2009. A study of pronunciation verification in a speech therapy application. In ICASSP, Taipei, Taiwan.
The Alborada-I3A corpus of disordered speech

Oscar Saz, E. Lleida, C. Vaquero, W.-R. Rodríguez
Aragón Institute for Engineering Research (I3A)
University of Zaragoza, Spain