

# The DisorderedSpeechBank: A multilingual digital archive of disordered speech

Martin J. Ball<sup>1</sup>, Lorraine Baqué<sup>2</sup>, Janet Beck<sup>3</sup>, Lilian Beijer<sup>4</sup>, May Bernhardt<sup>5</sup>, Tim Bressmann<sup>6</sup>, Ingo Hertrich<sup>7</sup>, Sara Howard<sup>8</sup>, Anu Klippi<sup>9</sup>, Kristian Emil Kristoffersen<sup>10</sup>, Alice Lee<sup>11</sup>, Vesna Mildner<sup>12</sup>, Henk van den Heuvel<sup>13</sup>, Nicole Müller<sup>1</sup>, Inger Lundeborg Hammarström<sup>1</sup>, Christina Samuelsson<sup>1</sup>  
 1 Linköping University, Sweden; 2 Autonomous University of Barcelona, Spain; 3 Queen Margaret University, Edinburgh, Scotland; 4 Institute for Health Studies, The Netherlands; 5 University of British Columbia, Canada; 6 University of Toronto, Canada; 7 University of Tübingen, Germany; 8 University of Sheffield, England; 9 University of Helsinki, Finland; 10 University of Oslo, Norway; 11 University College Cork, Ireland; 12 University of Zagreb, Croatia; 13 Radboud University, Nijmegen, The Netherlands

## Introduction

- The DisorderedSpeechBank will be a digital archive of disordered speech and we have established an international team of researchers in clinical linguistics to plan and collect material for the archive.
- The first workshop to set up the project was held in mid-October 2015, with the second held in early June 2016.
- The authors listed on this poster all attended the first workshop
- At the second workshop, Lorraine Baqué, Janet Beck, Lilian Beijer, Sara Howard, Kristian Kristoffersen and Vesna Mildner were unable to attend.
- Their places were taken by Pernille Hansen (Oslo), Marko Liker (Zagreb), Joaquim Llisterra (Barcelona), Joseph Stemberger (UBC) and Traci Walker (Sheffield).

## Content

- As noted, DisorderedSpeechBank will consist of a digital archive of sound files and video files representing samples of disordered speech, in a variety of languages.
- These sound and video files will be accompanied by high-fidelity transcripts, and acoustic analysis files together with imaging files (such as ultrasound imaging), as appropriate.
- Disorders of speech is interpreted widely, thus including:
  - Protracted phonological development
  - Articulatory and phonological disorders in children
  - Childhood apraxia of speech
  - Motor speech disorders
  - Speech disorders from other neurogenic conditions
  - Speech disorders resultant from physiological impairments
  - Speech disorders resultant from various genetic conditions
- As wide a variety as possible of languages will be included, and currently the team consists of researchers working with Catalan, Croatian, Dutch, English, Finnish, French, German, Irish, Norwegian, Spanish, Swedish, and Welsh.
- As well as the team named so far, help has been offered by Brian MacWhinney and Yvan Rose (of Talkbank and Phonbank), and Sebastian Drude of CLARIN ERIC (Common Language Resources and Technology Infrastructure; European Research Infrastructure Consortium).
- The archive will have a common digital format, but the overall format of entries will vary.
- This is because some may be audio only, others audio and video, yet others may simply be phonetic transcriptions of speech data.
- While the current team are working to produce a pilot archive for testing purposes, the eventual archive will be open to other researchers to upload material in a similar way to other archives such as Aphasiabank.

## Ethical Considerations

- The workshops laid considerable emphasis on the obtaining ethical approval for collecting and submitting disordered speech data.
- For this reason, very little already existing data can be grandfathered in.
- A detailed set of ethical guidelines and suggestions was drawn up by the first Workshop and is available on the project's Google Groups site (please apply to become a member): <https://groups.google.com/forum/#!forum/delad-disordered-speech-bank>

## Availability

- The archive will be made available to researchers, as well as speech-language pathology educators and students.
- For researchers, the attraction of such an archive is the ability to refine analysis methods and formulate and test hypotheses about disordered speech without having to 'reinvent the wheel' in terms of primary data collection.
- For educators and students, an archive of high quality speech data allows them to learn and practise analysis of disordered speech and the application of diagnostic tools on a variety of cases.
- The ultimate goal of this research is the improvement of evidence-based therapy for developmental and acquired speech disorders, as well as the improvement of research opportunities.

## Conclusion

- The benefits of an internationally accessible archive, which can serve as a tool both for researchers, and for the training of future speech-language pathologists, have been amply demonstrated by existing archives:
  - in the area of child language acquisition (MacWhinney, 2000, see: <http://childes.psy.cmu.edu/>),
  - and adult-acquired disorders of language (MacWhinney, Fromm, Forbes & Holland, 2011, see <http://talkbank.org/APhasiaBank/>).
- Our planned digital archive of disordered speech will complement these other initiatives, and make a major contribution to the systematic, data-driven analysis of speech and language around the world.

## Acknowledgement

- This project was supported by Riksbankens Jubileumsfond project number 8726201004, awarded to the first author.

## References

- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B., Fromm, D., Forbes, M. & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, 25, 1286-1307.