



Update to the DELAD Project

Henk van den Heuvel¹, Martin J. Ball², Nicola Bessel³, Alice Lee³, Nicole Müller³, Satu Saalasti⁴
¹Radboud University, Nijmegen, The Netherlands; ²Bangor University, Wales; ³University College Cork, Ireland;
⁴University of Helsinki, Finland

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 project has been named the DELAD project (from 'shared' in Swedish)
- The first workshop to set up the project was held in mid-October 2015, the second held in June 2016, both in Sweden, supported by the Riksbankens Jubileumsfond (project number 8726201004, awarded to M J Ball).
- From 2017, the project has come under the auspices of CLARIN ERIC (Common Language Resources and Technology Infrastructure - European Research Infrastructure for Language Resources and Technology).
- CLARIN funded a third workshop in Cork in 2017, and will be funding a fourth in Utrecht in 2019.
- The second CLARIN-DELAD workshop is in preparation, and will deal with IPR (intellectual property rights), Ethics, GDPR, and infrastructure

Aims

- 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.

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.
- 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.
- We imagine that material can be archived at local CLARIN centres and made findable through a central portal via their (harvested) metadata. CLARIN can offer the standards, best practices and services

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.
- 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.

Links

- For more information about the workshop (incl. video lectures) and about the project, visit:
- <https://www.clarin.eu/blog/report-clarin-delad-workshop>
- <https://delad.ruhosting.nl/wordpress/>