A Dutch Dysarthric Speech Database for Individualized Speech Therapy Research

Emre Yilmaz\(^1\), Mario Ganzeboom\(^4\), Lilian Beijer\(^2,3\), Catia Cucchiarini\(^1\) and Helmer Strik\(^1,4\)

\(^1\)Centre for Language and Speech Technology (CLST), Radboud University, Nijmegen, Netherlands
\(^2\)Sint Maartenskliniek, Nijmegen, Netherlands
\(^3\)HAN University of Applied Sciences, Faculty of Health and Social Studies, Nijmegen, Netherlands
\(^4\)Centre for Language Studies (CLS), Radboud University, Nijmegen, Netherlands

\{e.yilmaz, m.ganzeboom, c.cucchiarini, h.strik\}@let.ru.nl,
1.beijer@maartenskliniek.nl

Abstract

We present a new Dutch dysarthric speech database containing utterances of neurological patients with Parkinsons disease, traumatic brain injury and cerebrovascular accident. The speech content is phonetically and linguistically diversified by using numerous structured sentence and word lists. Containing more than 6 hours of mildly to moderately dysarthric speech, this database can be used for research on dysarthria and for developing and testing speech-to-text systems designed for medical applications. Current activities aimed at extending this database are also discussed.

Keywords: Speech database, dysarthric speech, Parkinsons disease, automatic speech recognition, speech therapy

1. Introduction

Dysarthria is a speech disorder caused by difficulties in neuromuscular control (Duffy, 1995) that lead to decreased speech intelligibility and communication impairment (Kent and Kim, 2003). Consequently, the life quality of dysarthric patients is negatively affected (Walshe and Miller, 2011) and they run the risk of losing contact with friends and relatives and eventually becoming isolated from the society. Research has shown that intensive therapy can be effective in (speech) motor rehabilitation (Ramig et al., 2001; Bhogal et al., 2003; Kwakkel, 2006; Rijntjes et al., 2009), but various factors conspire to make intensive therapy expensive and difficult to obtain.

Recent developments show that therapy can be provided without resorting to frequent face-to-face sessions with therapists by employing computer-assisted speech training systems (Beijer and Rietveld, 2011). According to the outcomes of the efficacy tests presented in (Beijer et al., 2014), the user satisfaction appears to be quite high. However, most of these systems are not yet capable of automatically detecting problems at the level of individual speech sounds, which are known to have an impact on speech intelligibility (De Bodt et al., 2002; Yunusova et al., 2005; Van Nuffelen et al., 2009).

One of the problems in developing more advanced and effective automatic speech therapy systems is the lack of sufficient language resources to develop and test the algorithms. Motivated by this problem, this paper presents a new Dutch dysarthric speech database with recordings of word and sentence sets uttered by 16 speakers with mild to moderate dysarthria due to Parkinsons disease (PD), traumatic brain injuries (TBI) and cerebrovascular accident (CVA). These word and sentence sets are carefully designed to ensure phonetically rich and balanced utterances. The audio files are annotated with the orthographic transcriptions and detailed speaker information such as age, gender, speech intelligibility level and origin of dysarthria.

This database combines the output of two independent data collection efforts for Dutch as spoken in the Netherlands. The first dysarthric speech data collection of Dutch-speaking patients aimed to use the data for a pilot study to investigate the performance of speech-to-text systems on deviant speech (Sanders et al., 2002). A large-scale collection effort was made by Beijer between 2008 and 2011. This work was performed as a part of the e-health-based speech therapy (EST) research program (Beijer, 2012). The database presented in this paper includes all collected data which are annotated according to a common protocol to create a principled dysarthric speech corpus. To the best of our knowledge, this database is the first dysarthric speech database of Dutch as spoken in the Netherlands.

This database is useful for conducting fundamental research on dysarthric speech and for building speech-to-text systems which can be incorporated in various assistive applications for neurological patients in the Netherlands. One such application is being developed in the CHAS-ING project\(^1\), which is an extension of the EST re-search program, and the (ASR-based) e-Health and e-Learning projects previously carried out at the CLST group of RU Nijmegen\(^2\).

The goal of CHAS-ING is to develop automatic speech recognition (ASR)-based serious games for individualized speech training of neurological patients and to investigate their effectiveness. The dysarthric speech dataset will be used for the development and testing of the dedicated ASR system. Subsequently, through the ASR-based games, additional dysarthric speech will be collected which, in turn, will be incorporated in the database. The speech elicited through the game will constitute a new component in the database containing dysarthric speech in natural human-human interaction, a valuable addition to the already existing speech samples.

\(^1\)http://hstrik.ruhosting.nl/chasing/
\(^2\)http://hstrik.ruhosting.nl/projects/
The rest of the paper is organized as follows. Section 2 summarizes other dysarthric speech databases from the literature. Section 3 details the data collection effort that has been made over years. The database details such as the speaker statistics and the annotation procedure are described in Section 4. Section 5 concludes the paper.

2. Related Work

One of the first dysarthric speech databases described in literature is the Whitaker database (Deller et al., 1993). This database contains audio recordings of various isolated words spoken by six dysarthric speakers due to Cerebral Palsy (CP). The database was created for fundamental research into severely dysarthric speech and the development of recognition technology.

The Nemours database (Menendez-Pidal et al., 1996) contains 7 sentences and two connected-speech passages all spoken by 11 American English, young adult, male speakers with varying degrees of dysarthria due to CP or Traumatic Brain Injury (TBI). Both the sentences and paragraph speech data were labeled and segmented on the orthographic and phonetic level. Originally, Nemours was developed to test the effects of various signal processing methods on the intelligibility of dysarthric speech. Since then, it has been used in many other related fields of research (e.g. acoustic analysis, ASR, etc.).

The dysarthric speech database for Universal Access research (UA-speech) was primarily designed for research into ASR of pathological speech (Kim et al., 2008). It contains audiovisual recordings of 765 isolated words for 19 American English speakers having dysarthria due to CP. The speech tasks include uncommon words, digits, computer commands and the radio alphabet.

Acoustic and articulatory data of dysarthric speakers were combined in the TORGO database (Rudzicz et al. 2012). TORGO contains recordings of eight dysarthric speakers due to CP or Amyotrophic Lateral Sclerosis (ALS). Motor functions of all dysarthric speakers were assessed using the perceptual measures of the Frenchay Dysarthria Assessment (Endery, 1980) and the resulting data were added to the database. Speech of seven matched non-dysarthric speakers was also recorded for control purposes.

The speech material used for the recordings varied from words in different categories (e.g. non-words, digits, common words, etc.) to different kinds of sentences (e.g. restricted, unrestricted / spontaneous description of images).

All databases summarized above contain audio recordings of American-English of dysarthric speech primarily due to CP. In recent years several efforts have also been made to publish dysarthric speech databases in other languages. Notable are the multi-year efforts for collecting Korean and French dysarthric speech/data collection include 12 semantically unpredictable sentences with 6- and 13-word declarative sentences, 12 6-word interrogative sentences, 13 Plomp and Mimpan sentences, 5 short texts, 30 sentences with /t/, /p/ and /k/ in initial position and unstressed syllable, 15 sentences with /l/, /l/ and /l/ in unstressed syllables, production of 3 individual vowels /a/, /e/ and /o/, 15 bisyllabic words with /t/, /p/ and /k/ in initial position and unstressed syllable and 25 words with alternating vowel-consonant composition (CVC, CVCVCC, etc.).

A substantial amount of speech was elicited by reading aloud the presented target speech. This was the case for isolated vowels, for words of various consonant-vowel compo-
sitions, for semantically unpredictable sentences (e.g. Een baan ging door een vlucht. (A job went through a flight) or Een baas stormt naar een stem. (A boss storms to a voice)) and for texts containing connected speech. Another elicitation strategy was the imitation of auditorily presented speech, supported by an orthographic representation of the utterance. It was used to elicit sentences and bisyllabic words with /p/, /k/ and /t/ in initial word position and in unstressed syllables (e.g. Vroeger stonden hier grote huizen en mooie paleizen. (Here used to be big houses and beautiful palaces)).

A third elicitation strategy also consisted in imitating auditorily presented speech supported by orthography, but in this case the target sentence was a response to an aurally presented question. This strategy was used to elicit targeted sentence accents, thus resulting in sentences containing the vowels /e/, /a/ and /o/ in unstressed syllables (e.g. Question: Wat is de kleur van een tomaat? Response: Rood is de kleur van een tomaat. (Q: What is the color of a tomato? R: Red is the color of a tomato).

### 4. Data Details

The proposed database contains 376 minutes of dysarthric speech material from 16 speakers. The details of these speakers are presented in Table 1. S1 and S2 are the patients who attended a face-to-face recording session and read the written material described in the first paragraph of Section 3. The rest of the speakers have participated the EST research program and their speech was either recorded online via the web application (S3, S9, S10, S11, S12, S13, S14, S15, S16) or locally using the Audacity software (S4, S5, S6, S7, S8). The annotations are enriched with various meta-information which includes age, origin of dysarthria, gender, dialect, level of dysarthria, total duration and some additional remarks for each speaker.

<table>
<thead>
<tr>
<th>Speaker id</th>
<th>Age</th>
<th>Dysart. Origin</th>
<th>Gender</th>
<th>Dialect</th>
<th>Dysart. Level</th>
<th>Duration (m)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>57</td>
<td>CVA</td>
<td>Male</td>
<td>Southern</td>
<td>Mild</td>
<td>8.4</td>
<td>Non-progressive dysarthria</td>
</tr>
<tr>
<td>S2</td>
<td>34</td>
<td>Birth defect</td>
<td>Male</td>
<td>Southern</td>
<td>Mild</td>
<td>12.7</td>
<td>Non-progressive dysarthria</td>
</tr>
<tr>
<td>S3</td>
<td>69</td>
<td>PD</td>
<td>Male</td>
<td>Southern</td>
<td>Moderate</td>
<td>22.9</td>
<td>Asymmetrical hypokinetic synd.</td>
</tr>
<tr>
<td>S4</td>
<td>70</td>
<td>PD</td>
<td>Male</td>
<td></td>
<td>Mild</td>
<td>2.2</td>
<td>Hypokinetic dysarthria</td>
</tr>
<tr>
<td>S5</td>
<td>53</td>
<td>PD</td>
<td>Male</td>
<td></td>
<td>Mild</td>
<td>2.3</td>
<td>Hypokinetic dysarthria</td>
</tr>
<tr>
<td>S6</td>
<td>69</td>
<td>PD</td>
<td>Male</td>
<td></td>
<td>Mild</td>
<td>2.5</td>
<td>Hypokinetic dysarthria</td>
</tr>
<tr>
<td>S7</td>
<td>70</td>
<td>PD</td>
<td>Male</td>
<td></td>
<td>Mild</td>
<td>2.3</td>
<td>Hypokinetic dysarthria</td>
</tr>
<tr>
<td>S8</td>
<td>71</td>
<td>PD</td>
<td>Male</td>
<td></td>
<td>Mild</td>
<td>2.2</td>
<td>Hypokinetic dysarthria</td>
</tr>
<tr>
<td>S9</td>
<td>61</td>
<td>CVA</td>
<td>Male</td>
<td></td>
<td>Moderate</td>
<td>57.5</td>
<td>Spastic dysarthria</td>
</tr>
<tr>
<td>S10</td>
<td>75</td>
<td>PD</td>
<td>Male</td>
<td></td>
<td>Moderate</td>
<td>35.3</td>
<td>Hypokinetic dysarthria</td>
</tr>
<tr>
<td>S11</td>
<td>34</td>
<td>TBI</td>
<td>Male</td>
<td></td>
<td>Mod.-Severe</td>
<td>40.1</td>
<td>Atactic dysarthria</td>
</tr>
<tr>
<td>S12</td>
<td>64</td>
<td>PD</td>
<td>Male</td>
<td>Southern</td>
<td>Moderate</td>
<td>29.8</td>
<td>Hypokinetic dysarthria</td>
</tr>
<tr>
<td>S13</td>
<td>73</td>
<td>PD</td>
<td>Male</td>
<td>Southern</td>
<td>Moderate</td>
<td>41.7</td>
<td>Hypokinetic dysarthria</td>
</tr>
<tr>
<td>S14</td>
<td>71</td>
<td>PD</td>
<td>Male</td>
<td>Southern</td>
<td>Moderate</td>
<td>27.0</td>
<td>Hypokinetic dysarthria</td>
</tr>
<tr>
<td>S15</td>
<td>63</td>
<td>CVA</td>
<td>Male</td>
<td></td>
<td>Moderate</td>
<td>44.1</td>
<td>Spastic dysarthria</td>
</tr>
<tr>
<td>S16</td>
<td>57</td>
<td>CVA</td>
<td>Male</td>
<td></td>
<td>Moderate</td>
<td>45.3</td>
<td>Flaccid dysarthria</td>
</tr>
</tbody>
</table>

Table 1: Details of the speakers contributed to the proposed database

5. Conclusions and Future Perspectives

The new database of Netherlandic Dutch dysarthric speech presented in this paper contains more than 6 hours of speech data uttered by patients suffering from Parkinsons disease, traumatic brain injuries and cerebrovascular accident. Various pre-designed word and sentence sets have been used for a diverse and balanced phonetic and lexical content. This database will serve as a useful data source for pathological speech research and is expected to facilitate the development of various assistive systems incorporating an ASR component. In particular, its use in the CHASING project will stimulate additional dysarthric speech data collection through an ASR-based serious game that has been developed to conduct research experiments and to offer alternative, sustainable speech therapy beyond the project lifecycle.
6. Acknowledgments

This research is funded by the NWO research grant with Ref. no. 314-99-101 (CHASING).

7. Bibliographical References


