A Dutch Dysarthric Speech Database for Individualized Speech Therapy Research

Emre Yılmaz[1], **Mario Ganzeboom**[1], **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,
l.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 CHASING 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 CHASING 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 (Menndez-Pidal et al., 1996) contains 74 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 Amyotropic Lateral Sclerosis (ALS). Motor functions of all dysarthric speakers were assessed using the perceptual measures of the Frenchay Dysarthria Assessment (Enderby, 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. (Choi et al., 2012) describes a Korean dysarthric speech database from at least 120 dysarthric speakers due to CP. The speech material contains up to 359 utterances per speaker. Utterances vary from individual speech sounds to full sentences.

French dysarthric speech data has been collected in the DesPho-APaDy project (Fougeron et al., 2010). This collection effort has resulted in two French databases that include recordings of 860 and 699 dysarthric speakers respectively. These databases contain dysarthric speech due to various etiologies, like PD, ALS, multiple sclerosis and stroke. Moreover, the COPAS database (Middag, 2012) of pathological speech in Flemish Dutch contains recordings from 75 dysarthric speakers with PD, CVA, TBI and MS. This database has been collected in the scope of SPACE project aiming to develop a reliable ASR-based speech assessment tool for pathological speech. Speech materials range from uttering individual speech sounds to short texts.

3. Data Collection

The data collection has been achieved both in face-to-face speech therapy sessions (Sanders et al., 2002) and via the web application developed for the EST research program (Beijer, 2012). In the face-to-face sessions, the patients are asked to read aloud the written material that is either presented on a paper or a computer screen. The written material includes Dutch numbers, 10 phonetically rich sentences, Plomp and Mimpen sentences (Plomp and Mimpen, 1979) and 50 most frequent utterances from the Dutch Polyphone database (Damhuis et al., 1994).

The web application served as another source of data collection and it has facilitated the interaction between the therapist and patient by helping the patient performing the speech training prescribed by the therapist in a computer-assisted manner. In the beginning of the training session, the patient accesses the prescribed speech tasks on the web service. Then, the patient is instructed about the target speech task and asked to perform it using a computer. After the patients response, the recording is sent to a central server and stored for medical examination. This procedure is detailed in (Beijer et al., 2010).

The data collection in the EST research program has been performed either using the web application or using a computer with a recording software. All recordings have been made in the presence of an expert to guide the patient. In this way, we aimed to maximize the recording quality by properly calibrating the recording setup and providing help to the patient in case of a technical problem. The responses of the patients have been recorded with a headset connected to a computer in a quiet environment to minimize the background noise.

The dysarthric speech collection for the EST program has been achieved in several experimental contexts. For this purpose, the speech tasks presented to the patients consist of numerous word and sentence lists with varying linguistic complexity. The speech tasks used in this phase of the data collection include 12 semantically unpredictable sentences with 6- and 13-word declarative sentences, 12 6-word interrogative sentences, 13 Plomp and Mimpen sentences, 5 short texts, 30 sentences with /t/, /p/ and /k/ in initial position and unstressed syllable, 15 sentences with /a/, /e/ and /o/ 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-

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

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

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.

Based on the meta-information, the age of the speakers is in the range of 34 to 75 with a median of 66.5. There are 10 PD patients, 4 CVA patients, 1 TBI patient and 1 patient with birth defect and the level of dysarthria varies from mild to moderate. The duration of recordings varies between 2 minutes to approximately an hour among different speakers. Type of dysarthria information such as non-progressive, (asymmetrical) hypokinetic, atactic, spastic and flaccid dysarthria is available for all speakers providing clues about the individual speech characteristics. All recordings have a sampling frequency of 16 kHz and were organized based on the type of the speech task content.

Using the transcriptions of the speech material, the word-level segmentation information was obtained by applying forced alignment through a conventional GMM-HMM-based recognizer. The segmentation and orthographic transcription information is stored in textgrid files which are obtained using the PRAAT software (Boersma and Weenink, 2015). Finally, these automatic annotations are manually verified by a group of linguistic students to maximize the quality of the transcriptions. The pronunciation errors are marked with an asterisk symbol to make the use of the proposed database viable for both automatic speech recognition and pronunciation error detection research.

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

Beijer, L. J. and Rietveld, A. C. M. (2011). Potentials of telehealth devices for speech therapy in parkinson's disease, diagnostics and rehabilitation of parkinson's disease. *InTech*, pages 379–402.

Beijer, L. J., Rietveld, A. C. M., van Beers, M. M. A., Slangen, R. M. L., van den Heuvel, H., de Swart, B. J. M., and Geurts, A. C. (2010). E-learning based Speech Therapy (EST): a web application for speech training. *Telemedicine Journal and eHealth*, 16(2):177–180, Mar.

Beijer, L. J., Rietveld, A. C. M., Ruiter, M. B., and Geurts, A. C. (2014). Preparing an E-learning-based Speech Therapy (EST) efficacy study: Identifying suitable outcome measures to detect within-subject changes of speech intelligibility in dysarthric speakers. *Clinical Linguistics and Phonetics*, 28(12):927–950.

Beijer, L. (2012). *E-learning based Speech Therapy (EST): Exploring the potentials of e-health in dysarthric speakers*. Ph.D. thesis, Radboud University, Nijmegen, Netherlands.

Bhogal, S. K., Teasell, R., and Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke*, 34(4):987–993.

Boersma, P. and Weenink, D., (2015). *Praat: doing phonetics by computer, Version 5.4.12*.

Choi, D.-L., Kim, B.-W., Kim, Y.-W., Lee, Y.-J., Um, Y., and Chung, M. (2012). Dysarthric speech database for development of qolt software technology. In *Proc. of the Eight International Conference on Language Resources and Evaluation (LREC)*, pages 3378–3381.

Damhuis, M., Boogaart, T., Veld, C. i. t., Versteijlen, M., Schelvis, W., Bos, L., and Boves, L. (1994). Creation and analysis of the dutch polyphone corpus. In *Proc. of International Conference on Spoken Language Processing (ICSLP)*, pages 1803–1806.

De Bodt, M. S., Hernandez-Diaz, H. M., and Van De Heyning, P. H. (2002). Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders*, 35(3):283–292.

Deller, J. R., Liu, M. S., Ferrier, L. J., and P., R. (1993). The Whitaker database of dysarthric (cerebral palsy) speech. *Journal of the Acoustical Society of America*, 93(6):3516–3518.

Duffy, J. R. (1995). *Motor speech disorders: substrates, differential diagnosis and management*. Mosby, St. Louis.

Fougeron, C., Crevier-Buchman, L., Fredouille, C., Ghio, A., Meunier, C., Chevrie-Muller, C., Bonastre, J.-F., Colazo-Simon, A., Delooze, C., Duez, D., Gendrot, C., Legou, T., Lvłque, N.and Pillot-Loiseau, C., Pinto, S., Pouchoulin, G., Robert, D., Vaissire, J., Viallet, F., and Vincent, C. (2010). The DesPho-APaDy Project: Developing an acoustic-phonetic characterization of dysarthric speech in French. In *Proc. of the International Conference on Language Resources and Evaluation (LREC)*, pages 2831–2838.

Kent, R. D. and Kim, Y. J. (2003). Toward an acoustic topology of motor speech disorders. *Clin Linguist Phon*, 17(6):427–445.

Kim, H., Hasegawa-Johnson, M., Perlman, A., Gunderson, J., Huang, T. S., Watkin, K., and Frame, S. (2008). Dysarthric speech database for universal access research. In *Proc. INTERSPEECH*, pages 1741–1744.

Kwakkel, G. (2006). Impact of intensity of practice after stroke: issues for consideration. *Disability and Rehabilitation*, 28((13-14)):823–830.

Menndez-Pidal, X., Polikoff, J. B., Peters, S. M., Leonzio, J. E., and Bunnell, H. T. (1996). The nemours database of dysarthric speech. In *Proc. International Conference on Spoken Language Processing (ICSLP)*, pages 1962–1966.

Middag, C. (2012). *Automatic analysis of pathological speech*. Ph.D. thesis, Ghent University, Belgium.

Plomp, R. and Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18:43–52.

Ramig, L. O., Sapir, S., Fox, C., and Countryman, S. (2001). Changes in vocal loudness following intensive voice treatment (LSVT) in individuals with Parkinsons disease: A comparison with untreated patients and normal age-matched controls. *Movement Disorders*, 16:79–83.

Rijntjes, M., Haevernick, K., Barzel, A., van den Bussche, H., Ketels, G., and Weiller, C. (2009). Repeat therapy for chronic motor stroke: a pilot study for feasibility and efficacy. *Neurorehabilitation and Neural Repair*, 23:275–280.

Sanders, E., Ruiter, M. B., Beijer, L. J., and Strik, H. (2002). Automatic recognition of dutch dysarthric speech: a pilot study. In *Proc. INTERSPEECH*, pages 661–664.

Van Nuffelen, G., Middag, C., De Bodt, M., and Martens, J.-P. (2009). Speech technology-based assessment of phoneme intelligibility in dysarthria. *International Journal of Language & Communication Disorders*, 44(5):716730.

Walshe, M. and Miller, N. (2011). Living with acquired dysarthria: the speaker's perspective. *Disability and rehabilitation*, 33(3):195–215.

Yunusova, Y., Weismer, G., Kent, R. D., and Rusche, N. M. (2005). Breath-group intelligibility in dysarthria: characteristics and underlying correlates. *J Speech Lang Hear Res.*, 48(6):1294–1310.