

# Dysarthric Speech Database for Development of QoLT Software Technology

**Dae-Lim Choi<sup>1</sup>, Bong-Wan Kim<sup>1</sup>, Yeon-Whoa Kim<sup>2</sup>,  
Yong-Ju Lee<sup>2</sup>, Yongnam Um<sup>3</sup>, Minhwa Chung<sup>4</sup>**

<sup>1</sup>Speech Information Technology & Industry Promotion Center, Wonkwang Univ.

<sup>2</sup>Department of Computer Engineering, Wonkwang Univ.

<sup>3</sup>Department of English Language & Literature, The Catholic Univ. of Korea

<sup>4</sup>Department of Linguistics, Seoul National Univ.

E-mail: {dlchoi, tacanemo, kimyw, yjlee}@wku.ac.kr, umyongnam@empal.com, mchung@snu.ac.kr

## Abstract

This paper describes the creation of a dysarthric speech database which has been done as part of a national program to help the disabled lead a better life – a challenging endeavour that is targeting development of speech technologies for people with articulation disabilities. The additional aims of this database are to study the phonetic characteristics of the different types of the disabled persons, develop the automatic method to assess degrees of disability, and investigate the phonetic features of dysarthric speech. For these purposes, a large database of about 600 mildly or moderately severe dysarthric persons is planned for a total of 4 years (2010.06. 01 ~ 2014.05.31). At present a dysarthric speech database of 120 speakers has been collected and we are continuing to record new speakers with cerebral paralysis of mild and moderate severity. This paper also introduces the prompting items, the assessment of the speech disability severity of the speakers, and other considerations for the creation of a dysarthric speech.

**Keywords:** speech DB, Dysarthria, QoLT

## 1. Introduction

As recently the social interests in welfare and QoLT (Quality of Life Technology) are increasing and national supports are also following, various efforts are being made actively to develop technologies to help the disabled lead a better life. In Korea, Technology Innovation Program funded by the MKE/KEIT was organized last year to support QoLT industry and infrastructures. Its goal is to improve the environments for the disabled and develop technologies for the hardware and software for their better life.

Up to now the work for the speech recognizers has been focused on those for the people with paralyzed limbs or for the healthy people, and the study for speech recognition rate with respect to the speech of the disabled in articulation has rarely been done. However, speech recognition technology is indispensable to help the disabled communicate easily with others, and the distribution of a speech database which is created with the disabled in real environments is essential to develop this technology.

In this paper we will deal with designing and creating a dysarthric speech database which is necessary for developing QoLT software technology to aid the dysarthric speakers, who experience difficulty in articulation due to physical paralysis due to neurological injury, in using IT devices such as PCs, smart phones, etc.

## 2. Existing dysarthric speech DB

Dysarthria ('dys' meaning 'having a problem with'; 'arthr' meaning 'articulating') is a motor speech disorder resulting from the malfunction of the motor component of the motor-speech system due to neurological injury or disease of the central nervous system [1]. Those with dysarthria experience difficulty in articulating words.

Currently there are four databases of the speech of the disabled persons with dysarthria known in the field. One of them is the Whitaker Database of Dysarthric (Cerebral Palsy) Speech (1993), which comprises the speech of 6 speakers with cerebral palsy and one healthy speaker [2]. The prompting items comprise 46 words: alphabet letters, single digits, 10 control words, and 36 words from Grandfather Passage. The degrees of disability of the speakers are included in the database.

Nemours Database of Dysarthric Speech (1996) comprises the speech of 11 male speakers with cerebral palsy of various degrees of disability [3]. The sentences are nonsense ones which have the structure of "the N1 is Ving the N2" (N1 and N2: monosyllabic nouns; Ving: disyllabic verbs). Each speaker produced 74 nonsense sentences to obtain a total of 814 sentences. Before recording, a speech therapist performed FDA (Frenchay Dysarthria Assessment) to check the dysarthric disability of the speakers [4]. This database is quite reliable since a speech therapist assessed the dysarthric

disability of the speakers using an assessment tool, but disadvantageous in that much time and efforts should be spent in the assessment of disability.

Another is the Dysarthric Speech Database for Universal Access Research (2008, UA-Speech) created by the University of Illinois, which was designed for ASR application for the persons with neuromotor disability [5]. UA-Speech DB is a collection of recordings from a total of 19 speakers with cerebral palsy through a 8-channel microphone array with a video camera. Each speaker produced 765 words which are composed of 155 uncommon words including 10 digits (x 3, or three times), 26 radio alphabet letters, 19 computer commands, 100 common words, words from Grandfather Passage and PBS (TIMIT sentences) words. Recordings from a total of 19 speakers were checked by 5 listeners, who gave the points about the speech intelligibility and thereby classified the speakers' degree of disability. In this case, recording process may not be complicated or delayed since recordings are checked after recording, but much human resources and time will be required for this method if the number of speakers increases because human listeners check the intelligibility.

The last one is the TORGO database of dysarthric speech (2011) created by the University of Toronto, which consisted of aligned acoustic and articulatory recordings [6]. TORGO database includes data from seven subjects (4 male, 3 female) with speech impediments caused by cerebral palsy or amyotrophic lateral sclerosis and age- and gender-matched control subjects. The stimuli are obtained from a variety of sources including the TIMIT database, lists of identified phonetic contrasts, and assessments of speech intelligibility. The motor functions of each experimental subject were assessed according to the standardized FDA by a speech-language pathologist.

### 3. Collection of dysarthric speech DB

A large database of the speech of the disabled persons with dysarthria is needed to develop a speech recognition system for them. With the corpus the error patterns depending on the degrees of disability can be modeled and various methods can be tried to determine the most suitable speech recognition method for a specific group. For this purpose the information about the degree of disability of each speaker should be included in the database.

#### 3.1 Planning

A database of about 600 mildly or moderately severe dysarthric persons is planned for a total of 4 years (2010.06.01 ~ 2014.05.31). The work scheduled for the first year (2010.06.01 ~ 2011.05.31) is the speech database of 120 dysarthric persons of low or mild degree of disability and 40 healthy speakers. The work scheduled for the second year (2011.06.01 ~ 2012.05.31) is about a total of 106 persons,

for which we will readjust degrees and areas of disability and ranges of age of the speakers after assessing the work of the first year. In the third and fourth years, the speech of a total of 200 persons will be added to be used for research.

#### 3.2 Prototype DB

A small-sized prototype dysarthric speech database has been created to be used as a pilot for discussing the details about the creation of a dysarthric speech database. Prototype DB was shared by the researchers on dysarthric disability in the field of QoLT, who have studied the environment, process and financial resources for collecting the speech and finally determined the requirements for the speech database.

The prototype database is the recording of 97 words produced two times by each of 13 adult speakers with brain disability (12 males and 1 female). The prompting items comprise 61 words from APAC (Assessment of Phonology and Articulation for Children) and U-TAP (Test of Korean Articulation/ Phoneme) and 36 Korean phonetic alphabets, or code words, which are used for identifying the Korean alphabet letters in voice communication. They were recorded in the mono format in the environment of a quiet office with a notebook computer and a Plantronics Audio 750 DSP Stereo Headset.

The findings from the prototype database are as follows. Though speaking disability degree such as 'speaking disability degree 1 with brain lesion' is marked specifically for some of the speakers with dysarthric disability, there is no information about previous diagnosis for most of them who were diagnosed in the past. Thus it is not possible to differentiate the speakers' disability in articulation by looking at their given type and degree of disability, and there is a difference in disability between the speakers with the same disability degree 1 of brain lesion. There are two ways to assess the speakers' type and degree of disability. A specialist such as a speech therapist can assess the speakers' disability during recording, or a specialist or others can check later after recording. The latter way has been determined to be more efficient in that large-sized speech database is created and used for engineering. Then the prompting items should be designed considering later assessment of disability type and degree.

#### 3.3 Prompting items

The following items are taken into consideration in designing the prompting items required for the research into speech disability in the fields of speech engineering, phonetics and speech pathology.

- (1) APAC 37 words for Assessment of Degree of Disability with Dysarthria

Assessment of degree of disability with dysarthria is indispensable in the speech analysis and development of the

speech recognition system for the disabled persons. The word list used in APAC is used for this purpose.

(2) 100 Machine Control Commands and 36 Korean Phonetic Alphabets

After investigating the requirements for speech recognition for dysarthric persons while considering the environment in which speech recognition technology is applied, the recognition of the basic machine-control commanding words and Korean phonetic codes are considered to be required. Machine-control commanding words comprise the commands which are commonly used for PC, cell phone, TV, radio, other electronic appliances and subsidiary equipments. And Korean phonetic and numeric alphabets or codes are characteristic and distinct codes for each Korean alphabet letter and number reflecting the speaking features of the disabled persons who experience difficulty in articulation. The latter are included because they are thought to be useful for recognition.

(3) 452 Phonetically Balanced Words (PBW)

Phonetically balanced words are required for designing acoustic models of the phonemes in phonetic environments in constructing a speech recognition system. Phonetically balanced words which are incorporated in the database are used for research into phonemic characteristics of the speakers and adaptation to speakers as well as for constructing acoustic models.

(4) 100 Words and 5 Sentences for Investigating Korean Consonants and Vowels Reflecting Phonetic Environment

To find phonetic characteristics in the disabled speech is necessary for constructing speech recognition models depending on the types and degrees of disability of the disabled persons. Thus vowels and consonants reflecting phonetic environment are included in the utterance items for the database.

### 3.4 Speakers and recording

Participating dysarthric speakers are limited to those with cerebral palsy due to brain damage before or during birth who have weak limbs and muscles and experience difficulty in articulation due to neurological injury or disease. Those who have severe disability are excluded though those who are not so severe are included. Those who experience difficulty in speaking due to stroke are not included. The age range of the speakers is from 30 to 40, and the rate of male and female speakers is 2:1. Healthy speakers are also recorded additionally so that their speech can be compared with disabled speech. The dysarthric speakers were recruited from Seoul National Cerebral Palsy Public Welfare. Recordings were made in a quiet office with a Shure SM12A microphone. The composition of prompting items and speakers for the databases by use is shown in Table 1.

DB for dysarthric speech recognition	
Dysarthric: 100 speakers - 37 APAC words - 100 machine commands and 36 Korean phonetic alphabets (two times each of both sets) - A ninth (1/9) of the PBW set	359 utterances per speaker
Healthy: 30 speakers - 37 APAC words - 100 machine commands and 36 Korean phonetic alphabets (three times each of both sets) - A third (1/3) of the PBW set	595 utterances per speaker
DB for studying phonetic features of dysarthric speech	
Dysarthric: 20 speakers Healthy: 10 speakers - 100 words for investigating Korean consonants and vowels and 5 sentences (two times each of both sets)	210 utterances per speaker

Table 1: Composition of Prompting Items and Speakers for the Database by Use

### 3.5 Assessment of severity of difficulty in articulation

It will be the most valid to collect and assess the spontaneous speech for assessment of severity of disability in articulation [7]. However, all the phonetic environments may not be included in spontaneous speech, and many hours will be spent in collecting and transcribing spontaneous speech. Thus, a standardized assessment is used which includes words and sentences designed so that they can include all the environments needed for assessment.

There are two standardized assessments of dysarthria (U-TAP and APAC) for Korean. In this paper APAC is used to assess severity of dysarthria of the speakers. APAC, which is made up of 37 words, is for checking 19 Korean consonants with 70 speech sounds – word-initial, word-final, word-medial onset and word-medial coda consonants – to assess clarity, degree of response to stimulus, error pattern, and comprehensibility. A speech therapist who has a first or top level license for speech therapy and has worked in the field for 5 years assessed PCC (Percentage of Consonant Correct) while listening to the speech of the 100 dysarthric speakers. Based on this assessment, 4 groups - mild (PCC: 85~100%), mild to moderate (PCC: 65~84.9%), moderate to severe; (PCC: 50~64.9%), and severe (PCC: less than 50%) – were distinguished according to diagnostic classification of phonological disorders suggested by Shirber & Kwiatkowski (1982) [8]. The numbers of speakers by severity of dysarthria are given in Table 2 below.

Severity of dysarthria	PCC	# Speakers
Mild	85 ~ 100	65
Mild to moderate	65 ~ 84.9	23
Moderate to severe	50 ~ 64.9	8
Severe	< 50	4

Table 2: The Numbers of Speakers by Severity of Dysarthria

The speech data of 20 speakers (20 % of the total data) were rechecked to find the intra-rater and inter-rater reliability. The intra-rater and inter-rater reliability was calculated using Pearson's product moment correlation. It was found that the intra-rater reliability was .957 and the inter-rater reliability was .901.

### 3.6 DB application

The database is for the development of an embedded keyword spotting speech recognition system individually customized for disabled persons with dysarthria, and it will be used for a speech recognition system for those who have weak limbs and muscles due to neurological injury or disease like cerebral palsy and experience difficulty in articulation. The research goals which are being pursued are as follows:

- Analyzing requirements for dysarthric speech recognition technology
- Analyzing speech characteristics of different disability types
- Studying automatic method to differentiate degrees of disability
- Studying the phonetic features of dysarthric speech and developing the method to treat noise
- Designing and implementing the software prototype for speech recognition system customized individually for the persons who experience difficulty in articulation.

### 4. Current recording for DB

Increasing the number of repetition of the utterances in order to check the consistency of utterances of the speakers of different degrees of disability is required for creation of the database for the second year. According to diagnostic classification of phonological disorders, 36 speakers are to be sampled from the speakers who participated in recording for the first year. They will record a set of 37 APAC words, 100 machine commands and 36 Korean phonetic alphabet a total of ten times on different days. In addition, 111 words and 5 sentences for investigating Korean consonants and vowels reflecting phonetic environment (the modified items of the first year) will be collected two times from new 70 dysarthric speakers.

### 5. Conclusion

In this paper we have described the work for a dysarthric speech database created by SiTEC to develop speech

technology for disabled persons, which has been done as part of a national program to help better life for Korean people. At present a dysarthric speech database of a total of 120 speakers has been completed, and we are continuing to record new speakers with cerebral paralysis of mild and moderate severity this year. The database has been provided to the team who develops the customized speech recognition system for the persons who experience difficulty in articulation, and their findings are being shared by researchers.

### 6. Acknowledgment

This work has been supported by the Technology Innovation Program (100036459, Development of center to support QoLT industry and infrastructures) funded by the MKE/KEIT, Korea.

### 7. References

- [1] Darley, F. C., Aronson, A. E., & Brown, J. R., "Differential diagnostic patterns of dysarthria," *Journal of Speech and Hearing Research*, pp. 246-269, 12, 1969.
- [2] Deller, J. R., Liu, M. S., Ferrier, L. J., and Robichaud, P., "The Whitaker database of dysarthric (cerebral palsy) speech", *Journal of the Acoustical Society of America*, 93(6):3516-3518, 1993.
- [3] X. Menéndez-Pidal, J.-B. Polikoff, S.-M. Peters, J. Lorenzo, and H.-T. Bunnell, "The Nemours database of dysarthric speech," *In Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP-Interspeech)*, pp. 1962-1965, Philadelphia (PA), USA, October, 1996.
- [4] Enderby, Pamela M. "Frenchay Dysarthria Assessment," College Hill Press, 1983.
- [5] Heejin Kim, Mark Hasegawa-Johnson, Adrienne Perlman, Jon Gunderson, Thomas Huang, Kenneth Watkin, and Simone Frame, "Dysarthric Speech Database for Universal Access Research," Interspeech 2008.
- [6] Rudzicz, F., Namasivayam, A.K., Wolff, T. "The TORGO database of acoustic and articulatory speech from speakers with dysarthria", *Language Resources and Evaluation*, pp. 1-19, 2011.
- [7] Stoel-Gammon, C., & Dunn, C. "Normal and abnormal phonology in children," Austin Texas: Pro-Ed. Inc, 1985.
- [8] Shirberg, L. D. & Kwiatkowski, J., "Phonological disorders III: A procedure for assessing severity of involvement", *Journal of Speech and Hearing Disorders*, Vol. 47, pp. 256-270, 1982.