

# Dysarthric speech corpus in Tamil for rehabilitation research

Mariya Celin T A\*, Nagarajan T\*\*, Vijayalakshmi P\*

\*Department of Electronics and Communication Engineering, \*\*Department of Information Technology  
SSN College Of Engineering  
(mariyacelinta,nagarajant,vijayalakshmiip)@ssn.edu.in

**Abstract**—This paper describes a speech data collected from 22 dysarthric speakers (7-female & 15-male) of various age groups in an Indian language, namely, Tamil. Dysarthric speakers who reported a diagnosis of cerebral palsy are chosen for speech data collection. The text for recording is chosen such that the influence of the articulatory errors can be observed in the initial, medial and end of the word. Each dysarthric speaker has uttered 262 sentences and 103 isolated words using a head mounted microphone. This corpus includes clinical data for all the 22 dysarthric speakers. The entire corpus is marked in sentence and word level. Time-aligned phonetic transcriptions are available for all the 22 speakers. The phonetic transcriptions are mapped with the IPA for a global use. The speech corpus is designed to act as a resource for the development of an automatic speech recognition system for speakers with neuromotor disability, research on articulatory dysfunctions, and rehabilitation research. The corpus includes unimpaired speech data, collected from 10 speakers (5-male & 5-female) for the same text used to collect dysarthric speech data along with the time-aligned phonetic transcription, for comparison. The corpus is available, on request, via secure FTP.

**Index Terms:** dysarthria, cerebral palsy, speech intelligibility assessment, assistive devices

## I. INTRODUCTION

Dysarthria is a motor speech disorder that results due to poor muscular control between the central and peripheral nervous system. This affects the basic process of speech production resulting in speech errors namely substitutions, deletions, and insertions. Based on the index of severity, dysarthria can be classified as mild, moderate, and severe. In India, there are approximately 33,000 people with cerebral palsy [1]. Worldwide, the incidence of cerebral palsy is 1 in 500 births. For most people with cerebral palsy, the cause is unknown and there is no known cure. Subjective measures namely speech therapy techniques can be a remedy for their speech errors, however it cannot be expected to recover an unimpaired speech quality and it is highly time dependent.

Some individuals are severely limited in their ability to communicate verbally, due to cognitive disabilities or physical impairments, for whom speech therapy may even fail. They can be recommended to an assistive device like speech input speech output communication aid (SISOCA), that recognizes their impaired speech using automatic speech recognition (ASR) system and synthesize the corrected text using speech synthesis system. The effectiveness of using an ASR for dysarthric speakers is supported from the documents of Kent [2] that states, most of the articulatory errors in dysarthric speech are primarily errors due to speech production of one distinctive feature. This provides a cue that the articulatory

errors are predictable and consistent. When these articulatory errors are addressed in a proper way, the development of an ASR system for dysarthric speakers is expected to be a successful tool. Hawley et al [3] worked in developing a voice-input voice-output communication aid (VIVOCA) for speech impaired people. This is the first model that combined both speech recognition and speech synthesis system and available as a prototype for severe speech impairments. The participants who used the prototype found it as a positive view towards independent living, provided they improve the accuracy of the ASR system. The ASR system accuracy highly depends on the training data available for each speaker.

For the development of a SISOCA, the focus relies on the requirement of a large vocabulary of dysarthric speech corpus. However, collecting a speech corpus from dysarthric speakers with a large vocabulary is highly tedious and tiring. Three English dysarthric speech corpora are publicly available. The first is the Nemours database of dysarthric speech [4] that includes speech data collected from 11 male dysarthric speakers along with the Frenchay dysarthric assessment. Second being the Universal Access speech corpus [5] collected from 13 male and 4 female dysarthric subjects. The TORGO database of acoustic and articulatory speech from speakers with dysarthria [6] is the first speech corpus having acoustic and articulatory features of 5 male and 3 female speakers with cerebral palsy. However, as of now there is no dysarthric speech corpus available for any of the Indian language. Thus, our work emphasizes on collecting a Tamil dysarthric speech corpus that is phonetically rich with sufficient examples to train an ASR system. Our corpus is not only designed to develop SISOCA, but also for researches in speech pathology to determine the articulatory dysfunction. The useful knowledge about the articulatory error can be then used for both clinical investigations and for the improvement in the performance of ASR system.

Thus, to develop SISOCA, to study about articulatory errors and neuro-motor dysfunctions, the dysarthric speech corpus for Tamil language is collected. This would also encourage researches in rehabilitation that would help in the development and support for a better lifestyle of dysarthric speakers.

## II. RECORDING PROCEDURE

The speech corpus recording involved three segments namely identifying the dysarthric individuals, text formation, and recording as discussed below.

### A. Identifying the dysarthric individuals

Collaboration is made with National Institute of Empowerment of Persons with Multiple Disabilities (NIEPMD),

an institute for students with multiple disorders run by the Government of India. The dysarthric students from various age groups of both male and female who reported a diagnosis of cerebral palsy are identified. The dysarthric individuals from three different departments were involved in the speech corpus collection namely, the department of adult independent living (DAIL), the pre-vocational cerebral palsy (CP) and the secondary CP. Table I summarizes the details about the departments and students participated in the speech corpus collection.

TABLE I. IDENTIFYING DYSARTHRIC INDIVIDUALS AT NIEPMD FOR THE CORPUS

Department	Age group (years)	No. of students participated
DAIL	19-37	12 (10 male and 2 female)
Pre-vocational CP class	15-18	5 (2-male and 3-female)
Secondary CP class	12-14	5 (3-male and 2-female)

### B. Text formulation

Once the dysarthric individuals are identified, the next step towards recording is to formulate the text corpus. The Tamil text is formulated such that all the phonemes have examples more than 25 except for phonemes that occur very rarely (/h/ with 10 examples, /f/ and /au/ with null examples). The text include 262 sentences with the maximum of 6 words in a sentence and 103 isolated words. The words are chosen such that the effect of the phoneme articulatory errors can be observed in the initial, medial and end of a word. The sentences include a combination of common and uncommon Tamil phrases.

### C. Recording process

Before beginning with the recording from the dysarthric speakers, an initial assessment session is conducted. It includes two segments: (a) discussion on the dysarthric speaker with the speech pathologist to know about their oral health habits and (b) discussion with the dysarthric speaker & their parents with the trainer (the person who record the speech data - the first author) to know information regarding their socio-demographic characteristics and risk factors. A consent form is signed by every parent before the recording.

The speech corpus is recorded using h250-Logitech adjustable head-mounted microphone. The microphone is checked to maintain a distance range of 3-4 cm from the mouth. The recording is performed in a laboratory environment with only the trainer and the dysarthric speaker. The speech is recorded in Audacity at a sampling rate of 16000 Hz using a Windows 8 laptop. The speech data is recorded in two sessions. In the first session 198 text (107 sentences and 91 isolated words) is recorded and the rest were recorded in session 2 within a time period of 6 months from the first session. This is to observe the progressiveness of the disease.

Some of the dysarthric speakers in our corpus have poor or null knowledge in reading and writing. So, the trainer uttered the sentence or word three times and asked the dysarthric speaker to repeat them, making him/her familiarize with the text. The fourth utterance from the dysarthric speaker is recorded. This is not applicable for dysarthric speakers having severe speech impairments, in which the number of repetition of an utterance by the trainer is not predictable, also each sentence were recorded word by word.

After recording each sentence, an interval of 15 minutes is given for the dysarthric speaker. This pause in the recording helped in two ways: (a) it did not make the dysarthric speaker feel tired due to recordings and (b) it is observed that an immediate recording of the next utterance had a great influence from the previously recorded sentence. This is more prominent in severe and some of the moderate classes. It is also noted that though the dysarthric speaker is capable of pronouncing the phoneme of the new utterance, he either substitutes or inserts a phoneme due to the influence from the previous utterance. Having an interval between the recordings highly helped in getting a proper observation on the phoneme errors.

The corpus also includes recordings from unimpaired people for comparison. Speech data from 10 unimpaired (orally healthy) people including 5-male and 5-female were recorded in the same environment for the same 365 text used for the dysarthric speakers.

TABLE II. SUMMARY OF DYSARTHRIC SPEAKER'S CLINICAL INFORMATION

Speaker id	Age	Speech Intelligibility score	Clinical diagnosis
MPK	27	2	Spastic quadriplegia
FAM	18	3	Alibetoid movement and presence of drooling
FSI	18	2	Spastic diplegia and quadriplegia, presence of drooling
FGA	15	3	Spastic quadriplegia and occasional drooling
MRI	13	6	Spastic quadriplegia and severe drooling
MPA	15	2	Mixed and frequent drooling
MAK	11	1	Spastic quadriplegia
FSP	13	3	Quadriplegia
FDH	16	3	Spastic diplegic
MRA	18	2	Quadriplegia
MSR	16	2	Spastic quadriplegia and presence of drooling
MMA	25	5	Infantile hemiplegic and frequent drooling
MGN	23	4	Quadriplegia and absence of lip closure
MER	26	6	Microcephaly
FVP	35	1	Spastic diplegia
MSU	35	3	Quadriplegia
MMU	35	3	Cerebral ataxia
MSA	21	5	Hypoxia and frequent drooling
MKA	19	3	Spastic quadriplegia, birth asphyxia and presence of drooling
FBF	21	4	Asphyxia and absence of lip closure
MVI	24	4	Spastic quadriplegia and frequent drooling
MPR	24	1	Cerebral ataxia and dysmetria

## III. DESCRIPTION OF THE CORPUS

### A. Clinical information of the speakers

Along with the speech data of the dysarthric speaker, the corpus also includes clinical information about them. Each speaker is identified by three letters. The first specifying the gender as M or F, the second and third letters are the first two letters from their name. The clinical information includes age, speech intelligibility score from speech therapists and clinical diagnosis (type of CP and presence of drooling). Table II summarizes the clinical information obtained for 22 dysarthric speakers.

1) *Speech intelligibility score*: The speech intelligibility for all the 22 dysarthric speakers were marked by learned speech therapists from NIEPMD. Two speech therapists were involved. One scored for the students from the pre-vocational & secondary CP class while the other scored for the DAIL. Each dysarthric speaker is assessed for about 3 minutes. They were asked to answer many uncommon questions, and were also asked to repeat few words uttered by the speech therapist.

TABLE III. MAPPING IPA LABELS WITH TIMIT PHONESET

IPA	/a:/	/aɪ/	/b/	/tʃ/	/d/	/ð/	/g/	/h/	/i:/	/dʒ/	/k/	/l/	/m/	/n/	/p/	/ŋ/	/r/	/s/	/ʃ/	/t/	/θ/	/w/	/j/
TIMIT phones	/aa/	/ay/	/b/	/ch/	/d/	/dh/	/g/	/h/	/iy/	/j/	/k/	/l/	/m/	/n/	/p/	/ng/	/r/	/s/	/sh/	/t/	/th/	/w/	/y/

TABLE IV. MAPPING IPA LABELS WITH COMMON PHONESET

IPA	/a/	/e:/	/i/	/o/	/u/	/w/	/e/	/o:/	/u/	/aʊ/	/ɪ/	/ɛ/	/ʊ/	/u/
Common phones	/a/	/ee/	/i/	/o/	/u/	/w/	/e/	/oo/	/u/	/au/	/ɪ/	/e/	/u/	/u/

Speech intelligibility score is marked according to the scaling procedure followed at NIEPMD [7], as mentioned below.

0- Normal

1- Can understand with difficulty, however the clinician/listener feels the speech is not normal

2- Can understand with difficulty, occasionally needs repetitions

3- Can understand with concentration but sympathetic listener needs two or three repetitions

4- Can understand with difficulty

5- Can understand with effort if context is known

6- Cannot understand at all even if context is known

The scores evaluated for each dysarthric speaker is shown in Table II. From the scores it can be observed that scores 1 & 2 fall under the mild class; 3 & 4 under the moderate class and scores 5 & 6 under severe speech impaired class.

*B. Splicing and segmentation*

It is observed from the recording process, that the dysarthric speaker repeats his utterances after the trainer, hence a data preprocessing is required. This cleans the recorded speech by removing the trainer’s utterance from the recording. This is performed manually for all the 22 dysarthric speakers. Time-aligned phonetic transcriptions are obtained for all the 22 dysarthric speakers and 10 unimpaired speakers, from which word level boundaries are also obtained. The time aligned phonetic transcriptions are obtained using forced Viterbi alignment [8]. The phoneme boundaries derived need to be then manually verified and adjusted. Tamil language includes 40 phonemes, however our corpus is restricted with 38 phonemes, as explained in Section 2.2. These 38 phonemes are labeled using the TIMIT phoneset. However, for certain speech sounds that are unique to Tamil language and not available in TIMIT are mapped using the common phoneset [8]. For a global use of our corpus beyond language boundaries, the 38 phonemes are mapped with the IPA labels. Table III & IV summarizes mapping of the speech sounds to the IPA labels using TIMIT and common phoneset. The phoneme and word level segmentation is available for the unimpaired speech data collected.

IV. SPEECH INTELLIGIBILITY ASSESSMENT FROM NAIVE LISTENERS

Apart from the clinical assessments discussed in Section III-A a speech intelligibility assessment test is also performed using naive listeners. The procedure for the speech intelligibility assessment and findings from the assessment are discussed below.

*A. Procedure*

An open test is performed for judging each dysarthric speaker’s speech intelligibility. Each dysarthric speaker is assessed by 3 naive listeners. The assessment is performed using

a speech intelligibility assessment worksheet. The preparation of the assessment worksheet is discussed in Section IV-A1. Each naive listener is given a worksheet along with two folders through a web server. The two folders include the trial and test folder. The trial folder contains 5 speech wave utterances of the dysarthric speaker for which they will be assessing along with the text data. They were asked to listen to all the 5 speech wave utterances and then verify with the given text. The trial helps the listener to familiarize with the dysarthric speaker and with the assessment procedure. The test folder includes 80 speech wave utterances containing 45 sentences and 35 isolated words without the text. The sentences and words are chosen based on two criteria: (i) sentences that are uncommon, so that the listener may not predict next word in the sequence and (ii) such that the effect of phonemes can be studied in the start, medial and end of the word. They were asked to assess all the 80 speech wave utterances and score them using the worksheet.

*1) Speech intelligibility assessment worksheet preparation:*

The worksheet includes instructions to perform the assessment. The assessment sheet includes fields like listener name, age and sex. The age field can accept only values within 18 - 40 [5]. The worksheet includes few queries for the listener to answer namely (i) Native speaker of Tamil? (ii) Previous experience with person having speech disorder? (iii) Language knowledge (Tamil)? with preferences like (a) very good (b) good and (c) poor. Listeners who positively satisfy all the above criteria are chosen for assessing the dysarthric speakers. Once the pre-requisites are entered, they are instructed to fill the utterances and degree of speech intelligibility (DSI) for each test utterance. The following subsection describes the assessment procedure in detail.

*2) Speech intelligibility assessment test:*

The test is performed in a laboratory environment using a headphone. Listeners are informed that they would be listening to speech uttered by speech disordered people and they are instructed to write only his/her perception about the speech utterance without any discussions. When filling the utterance column, the listener is asked to enter the text what they perceive. They were requested to use phonetic transcription if possible. They were allowed to listen to the test utterance as many times as needed and use dashes for the words they weren’t sure. The listeners were also asked to write a number (0 to 2) for each test utterance, that indicates their view about the speech intelligibility of the dysarthric speaker. They indicated their choice as follows: 0 - not intelligible even after a serious listening of infinite times, 1 - somewhat intelligible after a serious and repetitive listening, 2 - intelligible after two or three attempts of listening.

*B. Findings*

For each listener’s transcription from the utterance column the word intelligibility in % and utterance intelligibility in % [9] are calculated using equations (1) and (2) respectively. A

word or utterance is considered intelligible only if the entire word/utterance can be understood.

$$\text{word intelligibility in \%} = \frac{\text{Intelligible words}}{\text{Total no. of words}} * 100\% \quad (1)$$

$$\text{utterance intelligibility in \%} = \frac{\text{Intelligible utterances}}{\text{Total no. of utterances}} * 100\% \quad (2)$$

TABLE V. SUMMARY OF SPEECH INTELLIGIBILITY ASSESSMENT TEST FOR 10 DYSPARTHIC SPEAKERS

Speaker ID	Degree of speech intelligibility	% intelligibility		Speech intelligibility score from speech therapist	Index of severity
		words	utterance		
MPK	1.62	79	62	2	mild
FAM	1.41	71	55	3	
MSU	1.39	67	48	3	moderate
FGA	1.2	58	45	3	
MMU	1.1	42	36	3	
MVI	0.96	38	29	4	
MGN	0.91	31	28	4	
FBL	0.6	29	26	4	
MER	0.2	6	3	6	severe
MRI	0.12	4	3	6	

The word and utterance intelligibility in percentage for a dysarthric speaker is calculated by the average across the percentage received from 3 naive listeners who assessed that dysarthric speaker. The DSI is also calculated in the same manner. The assessment test is performed for all the 22 dysarthric speakers, among that, 10 are tabulated. The Table 5 is sorted based on the index of severity known from the assessment test. A dysarthric speaker with (i) (0-0.5) of DSI score, (0-25)% and (0-15)% of word and utterance intelligibility form the severe class, (ii) (0.5-1.5) of DSI, (25-75)% and (15-60)% of word and utterance intelligibility belong to moderate class and (iii) (1.5-2) DSI score, (>75)% and (>60)% of word and utterance intelligibility form the mild class. The classification based on the average percentage (words and utterances) and DSI score is observed to match with the classification based on the speech intelligibility scores obtained from speech therapists.

For visual observations, regarding the speech intelligibility levels of dysarthria, examples of waveforms and spectrograms of 4 dysarthric speakers with their decreasing order of speech intelligibility for the word 'balam' (meaning strength) are shown in Figure 1. The spectrogram of dysarthric speaker MPR shows clear distinctions of the sound units (/b/, /a/, /l/, /a/ and /m/) in the word, whereas for speaker FGA, the distinctions become weak. For speaker MGN and MRI the sound unit /b/ is deleted. Speaker MGN, thus begins the word with /a/ leading to a prolonged /a/ sound, and in severe dysarthric speaker MRI who is comfortable with only vowel sounds, replaces the consonants in the word with mixed combinations of vowels. The sound unit distinctions in the spectrogram of MGN and MRI becomes even more weaker as the speech intelligibility level decreases. More detailed phonological analysis to evaluate articulatory errors of each dysarthric speaker can be performed using ASR with different recognition architectures.

## V. CONCLUSIONS

The present paper has described a dysarthric speech corpus collected from 22 cerebral palsy individuals in Tamil. The corpus includes 365 speech utterances from 22 dysarthric speakers along with their clinical data. Time-aligned phonetic and word transcriptions are available for all the 22 dysarthric speakers. To use the corpus across the board, the Tamil phonemes are mapped with the IPA label. For comparison, the corpus also

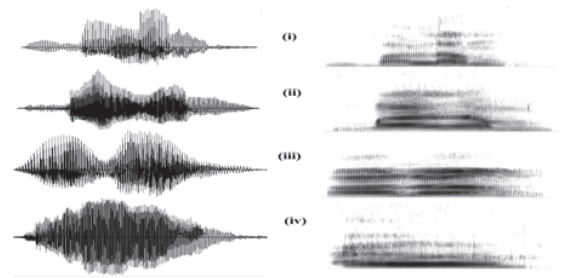


Fig. 1. Waveforms and spectrograms uttered by 4 dysarthric speakers (i) MPR (ii) FGA (iii) MGN (iv) MRI with varying index of speech intelligibility

includes speech data collected from unimpaired speakers (5-female and 5-male) for the same 365 text used for dysarthric speakers. We also performed a speech intelligibility assessment test using naive listeners. The consolidated results from the two intelligibility tests (from speech therapist and naive listeners) helped in classifying the dysarthric speakers as mild, moderate and severe. We believe that this corpus will help in several areas of rehabilitation research like SISOCA, determination of articulatory dysfunctions for clinical assessments and also for other speech communication aids that would uplift the lives of such beautiful children.

## VI. ACKNOWLEDGEMENTS

The authors would like to thank the National Institute of Empowerment of Persons with Multiple Disabilities (NIEPMD) for their constant support in collecting dysarthric speech data from the DAIL, pre-vocational and secondary CP class. Also, we are extremely grateful to all the dysarthric speakers and their parents for the participation during this venture. We also thank the tutors, speech therapists and pathologists of NIEPMD for their assistance and patience during the collection of clinical data.

## REFERENCES

- [1] The Indian institute of cerebral palsy (IICP), ESTD. 1974 Online: [www.iicpindia.org/PDF/What is Cerebral Palsy.pdf](http://www.iicpindia.org/PDF/What%20is%20Cerebral%20Palsy.pdf), accessed on 1.6.2016.
- [2] R. D. Kent, G. Weismer, J. F. Kent, J. K. Vorperian and J. R. Duffy, "Acoustic studies of dysarthric speech: Methods, progress and potential", *Journal of Communication Disorders*, vol. 32, pp. 141-186, 1999.
- [3] M. Hawley, S. Cunningham, P. Green, P. Enderby, R. Palmer, S. Sehgal and P. O'Neill, "A voice-input voice-output communication aid for people with severe speech impairment", *IEEE Trans. on Neural systems and Rehabilitation Engineering*, vol. 21, pp. 23-31, 2013.
- [4] X. Menendez Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio, and H. T. Bunnell, "The Nemours database of dysarthric speech", in *Proceedings of Fourth Int. Conf. on Spoken Language (ICSLP)*, vol. 3, pp. 1962-1965, 1996.
- [5] H. Kim, M. Hasegawa-Johnson, A. Perlman, J. Gunderson, T. Huang, K. Watkin, and S. Frame, "Dysarthric speech database for universal access research", in *Proceedings of INTERSPEECH* pp. 1741-1744, 2008.
- [6] Frank Rudzicz, Aravind Kumar Namasivayam and Talya Wolff, "The TORGO database of acoustic and articulatory speech from speakers with dysarthria", *Language Resources and Evaluation*, vol. 46.4 pp. 523-541, 2012.
- [7] Scaling procedure followed at NIEPMD for speech intelligibility test by speech therapist is referred by a national body Ali Yavar Jung National Institute for the Hearing Handicapped (AYJNIIH), Online: [ayjnihh.nic.in/index.asp](http://ayjnihh.nic.in/index.asp), accessed on 1.6.2016.
- [8] Ramani B, S. Lilly Christina, G. Anushiya Rachel, Sherlin Solomi V, Mahesh Kumar Nandwana, Anusha Prakash, Aswin Shanmugam, Raghava Krishnan, S. Kishore Prahalad, K. Samudravijaya, P. Vijayalakshmi, T. Nagarajan and Hema Murthy, "A Common Attribute based Unified HTS framework for Speech Synthesis in Indian Languages", in *Proc. of 8th ISCA Speech Synthesis Workshop*, pp. 291 - 296, 2013.
- [9] Kenneth G. Shipley, Julie G. McAfee, "Assessment in speech-language pathology", A Resource manual, 4th edition, Delmar Cengage learning, 2009.