SSNCE Dysarthric Speech Corpus in Tamil for Rehabilitation Research

INTRODUCTION:

The SSNCE dysarthric speech corpus [1], [2] was developed by the Speech Lab, SSN College of Engineering, in collaboration with the National Institute of Empowerment of Persons with Multiple Disabilities (NIEPMD), Chennai, Tamil Nadu, India. The dysarthric speech data is collected in Tamil, a regional language of India. The corpus includes approximately 8 hours of speech data along with their time-aligned transcriptions collected from 20 dysarthric speakers (13 - male and 7 - female), and 10 non-dysarthric speakers (5-male and 5-female).

Dysarthria is a neurogenic motor speech disorder characterised by slow, weak, or uncoordinated movements of the muscles involved in speech leading to an unintelligible speech. Automatic speech recognition (ASR) systems can be considered as a suitable alternative for them to interact with computers and with their peers. The inadequacy of the modern ASR systems to recognize dysarthric speech is due to the fact that ASR systems for dysarthric speakers need to be tailor-made (speaker-dependent) specific to each dysarthric speaker's dysfunction which requires a sufficient amount of speech data to enable speaker-dependent statistical modelling. SSNCE dysarthric speech corpus is one of its kind, that has speech data with a minimum of 25 examples per phoneme that is sufficient to train a dysarthric speaker model. Moreover SSNCE dysarthric speech corpus is the first dysarthric speech corpus available for an Indian language.

The speech data was collected for the purpose of developing a speech-input speech-output communication aid (SISOCA) for dysarthric speakers as a part of a project funded by Department of Science Technology, Technology Interventions for Disabled and Elderly (DST-TIDE), Government of India (Ref. no. SEED/TIDE/027/2016/G).

DATA:

The speech data was collected between 2015 and 2017 in two sessions at NIEPMD. The dysarthric speakers of various age groups (between 12 and 37 years) of both male and female who reported a diagnosis of cerebral palsy were chosen. All the chosen dysarthric speakers were native speakers of Tamil. Most of the dysarthric speakers in our corpus are from rural regions of Kovalam, Tamil Nadu, India with a very minimal educational qualification. Furthermore, most of the dysarthric speakers in our corpus have poor knowledge in reading and writing. There are 7 mild, 10 moderate and 3 severe dysarthric speakers in the corpus.

The speech data includes 365 utterances in Tamil recorded from each dysarthric speaker. To enable statistical model training, the text data is formulated such that all the phonemes have more than 25 examples except for phonemes that occur very rarely (/h/ with 10 examples, /f/ and /au/ with null examples). The text includes 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.

Before recording from a dysarthric speaker, a discussion is made with the dysarthric speaker & their parents with the trainer (the person who record the speech data) to know information regarding their socio-demographic characteristics and risk factors. A consent form was also signed by every parent after the discussion.

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. During recording 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. For severe dysarthric speaker each sentence was recorded word by word.

The speech data is recorded in two sessions. In the first session 198 utterances (first 107 sentences and fist 91 isolated words) are 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.

The SSNCE dysarthric speech corpus, apart from dysarthric speech data, includes time-aligned phonetic transcription for the data collected from all the 20 dysarthric speakers. The phoneme boundaries, derived using forced-Viterbi alignment, are manually verified and adjusted. The phonemes used in the Tamil text are labelled using the TIMIT phoneset. However, certain speech sounds that are unique to Tamil language and not available in TIMIT are mapped using the common phoneset [3]. For a global use of our corpus beyond language boundaries, the phonemes are also mapped with the IPA labels.

For comparison, the SSNCE dysarthric speech corpus includes speech data collected from 10 normal speakers (5-female and 5-male) as well, who have uttered the same 365 utterances as the dysarthric speakers, in a laboratory environment at a sampling rate of 16000 Hz. The time-aligned phonetic and word transcriptions are available for the normal speech as well.

The data is organized speaker-wise based on the severity of the dysarthric speaker as mild, moderate and severe. Each speaker was assigned a code with three letters and given their own file directory. The code for female speakers begins with F, and the code for male speakers begins with M, the second and third letters are the first two letters from their name. A separate directory is maintained for non-dysarthric speakers with the code 'FC' for female and 'MC' for male, where C refers to control group.

Each dysarthric/non-dysarthric speaker directory, include the following content:

Audio: These directories contain the speech wavefiles. Each file is a FLAC compressed WAVE audio file (linear PCM, 16 bit, mono, 16000 Hz)

Time_aligned_Transcriptions (label): These directories contain time-aligned phonetic and word transcriptions of audio data. Each file is plain text with a *.lab and *.wrdlab file extension and a filename referring to the utterance number. These files can be viewed in line with the wav file using the free Wavesurfer tool.

Prompts: These directories contain orthographic transcriptions of the audio data. Each file is a plain text in UTF-8 format with a *.txt file extension and a filename referring to the utterance number.

DOCUMENTS:

Along with the speech data the SSNCE dysarthric speech corpus also includes clinical data of each dysarthric speaker that includes age, clinical diagnosis and speech intelligibility score provided by speech therapist. The summary of the dysarthric speaker's clinical information is provided in Table 1.

The documentation directory in the corpus includes this clinical information, a README file and a file that provides mapping of the IPA labels with the TIMIT and common phoneset.

Table 1. Dysarthric speaker's clinical information

Sl. No.	Name	Intelligibility Score	Type of CP	Muscle tone	Other details
1.	MPK	2	Quadriplegia	Normal	Mild mental retardation with IQ level 60. Verbal performance was below average.
2.	FAM	3	CP with athetoid movement	Dystonic	Can understand but needs repetitions and concentration. Presence of drooling. Impression : dysarthria
3.	FSI	2	Spastic diplegia and quadriplegia	Hypertonic	Drooling present. Standing with support.
4.	FGA	3	Spastic quadraplegia	Hypertone	Poor sitting balance. Tongue restricted. Drools occasionally. Mild mental retardation. Has difficulty to speak; delayed speech and language; referred for speech therapy to improve speech clarity; oromotor exercise for drooling, lip, tongue and cheek muscles.

5.	MRI	6	Spastic quadriplegia ; Seizure disorder	Severe issue in all 4 limbs	Slurred speech. Lip closure feedback to be followed; Drooling present; inadequate speech and language development; he expresses in word level; delayed speech and language; was not cooperative during assessment.
6.	MPA	2	Athetoid, ataxia and quadriplegia	Dystone	Difficulty in speaking, walking and presence of drooling; Tongue –protrusion present. Can articulate consonants and vowels but the speech is stressful includes strained strangled voice; slow rate of speech; frequently drools; wet lips and chin.
7.	MAK	1	Spastic Quadriplegia, Hypoxic inflaly brain	Hypertone	Moderate mental retardation. The child can communicate verbally. Never drools.
8.	FSP	3	Quadriplegia	Normal	Birth time not active. Swelling in hands and legs. Child did not cry. Difficulty in vision, speaking and walking. can speak in words; missing teeth; Hard Palate: High arched; Soft palate: short; umla present; Expressive language disorder; She is been recommended to speech and language therapy for her speech is not clear.
9.	FDH	3	Spastic diplegic	Hypertone	Can express in words. Moderate mental retardation.
10.	MRA	2	Quadriplegia	Muscle tone all 4 - hypotone	Language disorder. expresses in simple words; can communicate verbally and non-verbally
11.	MMA	5	Infantile hemiplegic	Hypertone	Absence of Lip closure, Tongue movement and Jaw movements. Presence of drooling.

12.	MGN	4	Quadriplegia	Normal	Moderate mental retardation. Speech not clear. Absence of lip closure.
13.	MER	6	Micro-cephaly	Normal	Can speak in words though not clear
14.	FVP	1	Spastic Diplegia	Hypertone	Can communicate verbally in meaningful sentences
15.	MSU	3	Quadriplegia	Normal	Difficulty in speaking; Can do meaningful counting; Lack of verbal skills
16.	MMU	3	Cerebral ataxia	Hypotone	Physical deformity, Moderate mental retardation, Ataxic gait; can express himself in sentences but speech is not intelligible to outsiders.
17.	MKA	3	Spastic quadriplegia and birth asphyxia	Hypertone	Inability to walk since birth, Mild Mental retardation, mild impairment in intellectual functioning. Vocal: no clear speech, speaks in slurred voice, low pitch, Drooling present, recommended for speech therapy. Poor oromotor mechanisms.
18.	FBL	4	asphyxia	Hypertone	speech word level; speech is unclear
19.	MVI	4	Spastic quadriplegia	Hypertone	Poor in speech, Accident with electric shock Oromotor functions : Drooling present ; Improper function of Lip closure, jaw movements and tongue.
20.	MPR	1	Cerebral ataxia & dysmetria	Hypotone	Speech meaningfully in sentence , Slurred speech (mild).

Table2: Clinical diagnosis of the dysarthric speakers

Sl. No.	Name	Age	Intelligibility Score	Clinical diagnosis
1.	MPK	27	2	Spastic Quadriplegia
2.	FAM	18	3	Athetoid movement and presence of drooling
3.	FSI	18	2	Spastic diplegia and quadriplegia, presence of drooling
4.	FGA	15	3	Spastic quadriplegia and occasional drooling
5.	MRI	13	6	Spastic quadriplegia ; Severe drooling
6.	MPA	15	2	Mixed and frequent drooling
7.	MAK	11	1	Spastic Quadriplegia,
8.	FSP	13	3	Quadriplegia
9.	FDH	16	3	Spastic diplegic
10.	MRA	18	2	Quadriplegia
11.	MMA	25	5	Infantile hemiplegic and frequent drooling
12.	MGN	23	4	Quadriplegia and absence of lip closure
13.	MER	26	6	microcephaly
14.	FVP	35	1	Spastic Diplegia
15.	MSU	35	3	Quadriplegia
16.	MMU	35	3	Cerebral ataxia
17.	MKA	19	3	Spastic quadriplegia, birth asphyxia and presence of drooling
18.	FBL	21	4	Asphyxia and absence of lip closure

19.	MVI	24	4	Spastic quadriplegia and frequent drooling
20.	MPR	24	1	Cerebral ataxia & dysmetria

USEFULNESS OF THE CORPUS:

We believe that this corpus will help in several areas of rehabilitation research such as developing a Speech-input Speech-output communication aid (SISOCA), determination of articulatory dysfunctions for clinical assessments, and to develop any speech-enabled assistive devices as communication aids that would uplift the lives of dysarthric speakers.

References:

[1] T.A. Mariya Celin, T. Nagarajan, and P. Vijayalakshmi, "Dysarthric speech corpus in Tamil for rehabilitation research", In 2016 IEEE Region 10 Conference TENCON, pp. 2610 - 2613, 2016.

[2] T.A. Mariya Celin, G. Anushiya Rachel, T. Nagarajan, and P. Vijayalakshmi, "A Weighted Speaker-Specific Confusion Transducer-Based Augmentative and Alternative Speech Communication Aid for Dysarthric Speakers", IEEE Transactions on Neural Systems and Rehabilitation Engineering Vol. 27, no. 2 pp. 187-197, Feb. 2019.

[3] G.Anushiya Rachel, V. Sherlin Solomi, K. Naveenkumar, P. Vijayalakshmi, and T. Nagarajan, "A small-footprint context-independent HMM-based synthesizer for tamil", International Journal of Speech Technology, vol. 18, no. 3, pp. 405–418, 2015.

Citing the speech corpus:

While using the corpus the users should cite the following papers;

[1] T.A. Mariya Celin, T. Nagarajan, and P. Vijayalakshmi, "Dysarthric speech corpus in Tamil for rehabilitation research", In 2016 IEEE Region 10 Conference TENCON, pp. 2610 - 2613, 2016.

[2] T.A. Mariya Celin, G. Anushiya Rachel, T. Nagarajan, and P. Vijayalakshmi, "A Weighted Speaker-Specific Confusion Transducer-Based Augmentative and Alternative Speech Communication Aid for Dysarthric Speakers", IEEE Transactions on Neural Systems and Rehabilitation Engineering Vol. 27, no. 2 pp. 187-197, Feb. 2019.