

# Analysis and Classification Dysarthric Speech



Siddhant Gupta and Hemant A. Patil

## 1 Introduction

A natural speech production mechanism works with a *synchronized harmony* of different human organs. These organs include lungs, larynx, vocal folds, jaw muscles, tongue, lips, teeth, soft palate, velum amongst others. The output received from such system is a distinct sound which is complex in nature and intelligible by the listeners. However, sometimes a disorder in one or more of the sub-systems results in disruption in the overall speech production mechanism rendering speech unintelligible and difficult to interpret. This gives rise to a completely different class of speech signals, which is impaired in general perception, and may not be analyzed considering normal healthy speech as a basis for comparison.

Dysarthria is one such speech impairment in which the muscles that help in speaking, such as vocal folds, jaw muscles, throat muscles, etc., becomes weak and coordination between them becomes difficult. Dysarthria has been rated amongst one of the most common types of speech impairments. Speech of dysarthric patients can be characterized as slow, slurry, monotonous, unnaturally whispered, etc. or a combination of such symptoms [1]. Analysis and classification of dysarthric speech is finding its applications in fields, such as biomedical speech signal processing [2], and voice-assisted electronic device manufacturing [3]. Dysarthria is directly associated with neurological diseases, such as Parkinson Disease, Cerebral Palsy, etc. Therefore, dysarthric speech analysis can help in the diagnosis and progression mapping of such diseases. However, the characteristics of dysarthric speech are different from that of normal speech. Therefore, it has been found that applications,

---

S. Gupta (✉) · H. A. Patil

Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar,  
Gujarat, India

e-mail: [Hemant\\_patil@daiict.ac.in](mailto:Hemant_patil@daiict.ac.in)

such as Automatic Speech Recognition (ASR) systems and Voice Privacy (VP) systems do not perform even considerably in the case of dysarthric speech [4, 5]. Analysis of dysarthric speech can help in development of more robust systems targeted at people suffering from dysarthria. This chapter focuses on understanding dysarthria as a signal processing problem.

Rest of the book chapter is organized as follows. Section 2 presents various types of dysarthria, whereas Sect. 3 presents time-domain and time-frequency domain analysis, such as Linear Prediction (LP) spectrum, Teager Energy Operator (TEO) profiles, spectrograms, waterfall plots, etc., of normal vs. dysarthric speech. Section 4 gives brief details of some standard and statistically meaningful dysarthric speech corpora. A discussion on the application of deep-learning methods in the classification of dysarthric speech from normal speech. Finally, the chapter concludes with potential future research directions.

## 2 Types of Dysarthria

Dysarthria shares many of its symptoms with the other neurological diseases, such as Aphasia, Dysphasia, and Apraxia [6]. However, it is distinct from these neurological diseases due to the organ of its origin [7]. While Aphasia and Dysphasia effect the ability of an individual to understand and produce speech, and Apraxia results from the damage to the parietal lobe of the brain that is responsible for planning of speech [8]; dysarthria resides in the muscles responsible for the production of speech. Patients with dysarthria do not show any deviations in perceptual processing and planning of speech, as compared to a healthy subject. However, the lack of synchronization amongst muscles causes the output speech to be damaged and unintelligible. This section describes the types of dysarthria that are widely recognized in the field of speech impairments:

### 2.1 *Spastic Dysarthria*

Spastic dysarthria is caused as a result of some damage to the Central Nervous System (CNS), which includes brain and spinal cord [9]. It is usually accompanied by weakening of muscles and abnormal reflexes in the other regions of the body as well. Hence, phonation is strained-strangled and articulation becomes weak. In addition, mouth opening seems to be restricted and speech is perceived to come from the back of the mouth. Furthermore, jaw jerk, gag reflex, and facial reflexes are also common with the patients of dysarthria.

## **2.2 *Flaccid Dysarthria***

Flaccid dysarthria is usually recognized by the difficulties faced by the patients in pronouncing consonants. It is caused by the damage to the Peripheral Nervous System (PNS), which connects brain and spinal cord to the rest of the body [9]. Flaccid dysarthria results in symptoms, such as hypernasality, breathiness in voice, and weak pressure consonants. Depending on which nerves are damaged, it affects phonation, respiration, resonance, and articulation.

## **2.3 *Ataxic Dysarthria***

Ataxic dysarthria is caused due to a damage to part of the brain called *Cerebellum* which is responsible for receiving sensory information and regulating movements [9]. It results in imprecise articulation with distorted vowels and inaccurate consonant production, disturbed speech prosody, and abnormal phoneme timing. There is inappropriate stress on syllables, loudness, and the pitch ( $F_0$ ) of the voice is deviant.

## **2.4 *Hypokinetic Dysarthria***

Hypokinetic dysarthria is caused because of the malfunction in the extrapyramidal systems of brain, which consists of areas of the brain responsible for coordination of subconscious muscle movement [9]. It is characterized by reduced pitch ( $F_0$ ) variation, reduced loudness, variable speaking rate, imprecise consonants, breathy voice, and short rushed of speech. Patients with Hypokinetic dysarthria also have difficulties in swallowing and sometimes observe drooling.

## **2.5 *Hyperkinetic Dysarthria***

Hyperkinetic dysarthria is caused due to the damage to the part of the brain collectively known as Basal Ganglia, which is responsible for regulating involuntary muscle movements [9]. It is characterized by abnormal involuntary muscle movements that affect respiration, phonation, and articulatory structure impacting speech quality.

## 2.6 *Mixed Dysarthria*

Mixed dysarthria represents a heterogeneous group of speech disorders and neurological diseases [9]. Any combination of two or more types of dysarthria (discussed above) results in mixed dysarthria. This type of dysarthria is more common than one kind of dysarthria occurring in a patient. The symptoms of mixed dysarthria can be a mix of the symptoms discussed above.

Though different types of dysarthria are clearly defined, it is often a difficult task to distinguish amongst them because of the negligible perceptual differences in production of output speech. One has to be a trained expert to recognize one dysarthria type from the other and with a substantial probability of error. These experts are called Speech-Language Pathologists (SLPs), and they often use the pre-defined techniques and scales to distinguish between different kinds of dysarthria.

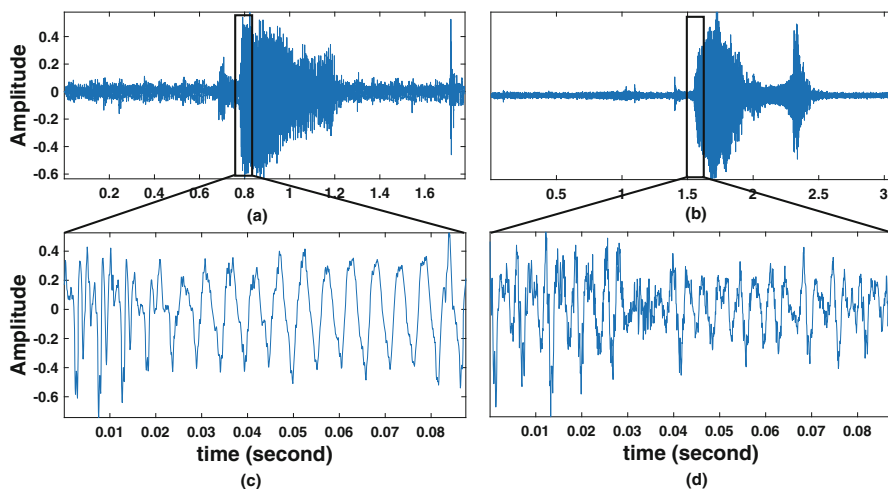
To better understand how dysarthria effects the production of speech and its intelligibility, acoustic analysis of dysarthric speech becomes necessary. Study of different acoustic features can help in distinguishing whether the speech is dysarthric or normal. This can help in early diagnosis of diseases associated with dysarthria.

## 3 Analysis of Dysarthric Speech

Since in patients with dysarthria, the speech is affected by the weakness in the muscles of the vocal tract system, a dysarthric speech output is very different from the speech output of a normal speaker. In addition, the acoustic features change with different severity of dysarthria from which a patient is suffering from. This section provides a comparative study of different acoustic features between dysarthric speech and normal speech and between different kinds of dysarthric speech.

### 3.1 *Time-Domain Analysis*

The time-domain waveform of dysarthric speech consists of useful information for the analysis of dysarthria. The pathological defects in the vocal tract system can be observed by looking at the time-domain waveform of dysarthric speech. Figure 1 shows the speech waveform of a normal person and a person suffering from dysarthria uttering the same word. It can be easily observed from the two waveforms that the dysarthric speech waveform is much longer as compared to the normal speech. Moreover, it can be that the variability in the acoustic pressure is more in dysarthric speech. Dysarthric speech also consists of regions of silence which are absent in case of normal speech, implying that these silent regions are not necessary for the speech wave but still exists. These silent regions represent the defects in motor control mechanism in the speech production system, where the vocal



**Fig. 1** Time-domain waveform. (a) normal, (b) dysarthric speech, and zoomed waveform for (c) normal, and (d) dysarthric speech

folds involuntary start/stop vibrating creating jitters and shimmers, representing variations in pitch period and volume airflow velocity, repeated across consecutive glottal cycles in the speech waveform.

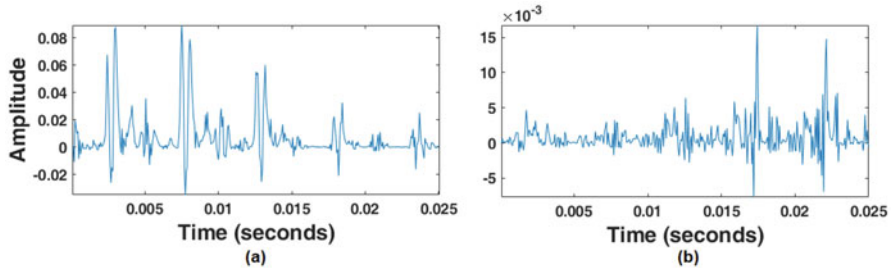
### 3.1.1 Fundamental Frequency ( $F_0$ )

The Fundamental Frequency ( $F_0$ ) of a speech signal is the average number of oscillations per second, in Hertz, of the voiced region of the speech. It arises due to the vibrations of the vocal folds which in turn oscillates the air flowing through the vocal tract system. Since the oscillations arise in an organic structure, it consists of some fluctuations, rather than being perfectly periodic, in particular, jitters and shimmers as discussed above. Due to the weakening of vocal fold muscles, a patient with dysarthria has less control over his vocal fold vibrations and hence, these jitters and shimmers are much more significant in dysarthric speech as compared to normal speech and can change the overall nature of the fundamental frequency.

### 3.1.2 Teager Energy Operator (TEO)

Teager Energy Operator (TEO) is a non-linear operator which helps in the analysis of speech waveform from an energy point of view. For a speech signal  $s(n)$ , TEO profile is given by [10]:

$$\text{TEO}\{s(n)\} = (s(n))^2 - s(n-1) \cdot s(n+1). \quad (1)$$



**Fig. 2** TEO Profile. (a) normal and (b) dysarthric speech

From TEO, we can observe that three consecutive speech samples are required to find the running estimate of signal energy and thus, it is known to have excellent time-resolution. TEO is very efficient in capturing the non-linearity in the speech, which is captured by the airflow in the vocal tract which will change for dysarthric vs. normal speech. Figure 2 shows the corresponding TEO profile for the normal vs. dysarthric speech. We can observe from Fig. 2 that, as in LP residual, TEO is also highly irregular for dysarthric speech, as compared to normal speech, indicating abnormal changes in pitch period, i.e.,  $T_0$ , and, hence, pitch frequency. In particular, TEO gives high energy pulses corresponding to GCIs due to its capability to capture characteristics of impulse-like excitation which are known to have higher signal-to-noise (SNR) ratios.

### 3.2 Linear Prediction (LP) Residual

Linear Prediction (LP) residual can be a very good method for the analysis of the characteristics of the speech excitation source. LP analysis deconvolves the speech signal into its source excitation and speech system components. For a speech signal  $s(n)$ , LP residual  $[r(n)]$  is given by [11]

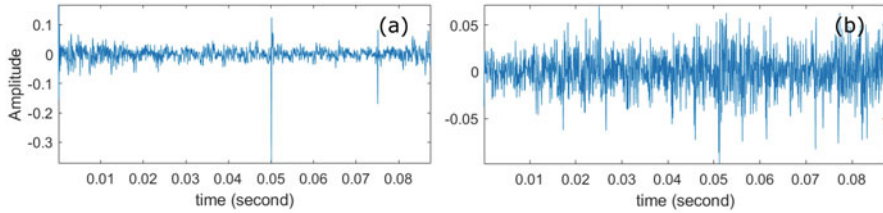
$$r(n) = s(n) - \hat{s}(n), \quad (2)$$

where

$$\hat{s}(n) = \sum_{k=1}^p a_k s(n-k), \quad (3)$$

and  $a_k$  corresponds to  $k$ th Linear Prediction Coefficient (LPC).

The speech production system has its own inertia. The Glottal Closure Instants (GCIs), are the instances when the glottis closes to provide a sudden burst of air pressure through the vocal folds, act as an excitation signal in the form of input



**Fig. 3** LP Residual Plot. (a) normal, and (b) dysarthric speech

impulse to the system. Sudden bumps can be observed in LP residual at *periodic* locations. These locations are called GCIs. GCIs denote the time instants, where there is sudden closing of vocal folds, which acts as impulse-like excitation, during speech production. The GCIs have been estimated using Hilbert transform in [12, 13]. Figure 3 shows the LP residual plot of normal *vs.* dysarthric speech. It can be clearly observed from these plots that the LP residual for dysarthric speech is highly irregular as compared to the LP residual plot for the normal speech signal. This shows the abnormality in the dysarthric speech signal in terms of pitch period ( $T_0$ ) and therefore, pitch frequency ( $F_0$ ).

### 3.3 Time-Frequency Analysis

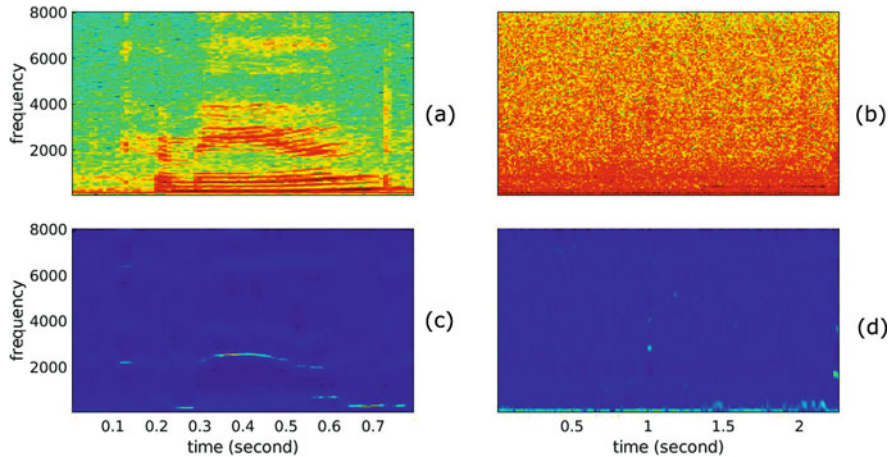
Due to non-stationary nature of speech signals, spectrograms can be used for the time-frequency analysis of a speech signal. A spectrogram is a visual representation of how spectral energy density varies with respect to different frequencies and time instances. The energy in a spectrogram is computed using Short-Time Fourier Transform (STFT) on windowed speech signal. Let  $x(n)$  be the input signal. STFT is calculated as :

$$X(\omega, \tau) = \sum_{n=-\infty}^{\infty} x(n) \cdot w(n, \tau) \cdot e^{-j\omega n}, \quad (4)$$

$$X(\omega, \tau) = \sum_{n=-\infty}^{\infty} x(n, \tau) \cdot e^{-j\omega n}, \quad (5)$$

where  $x(n, \tau) = x(n) \cdot w(n, \tau)$  is the windowed speech segment. Now spectrogram (spectral energy densities) is obtained by calculating the magnitude square of  $X(\omega, \tau)$ , i.e.,

$$S(\omega, \tau) = |X(\omega, \tau)|^2. \quad (6)$$



**Fig. 4** Comparison of spectrogram for (a) normal and (b) dysarthric speech vs. LP Spectrum for (c) normal and (d) dysarthric speech

It can be seen from Fig. 4 that the energy is concentrated in the lower spectrum of the frequency in contrast with the normal speech, in which the energy is observed to be spread in the high as well as low frequency spectrum. This is due to the fact that abnormal functioning of the speech production results in a great amount of energy loss for higher frequencies [14, 15]. This can be the reason why speaking intelligibly is a challenging task for a person suffering from dysarthria, which usually comes very naturally to the normal speakers. Figure 4 show the plot of Short-Time Fourier Transform (STFT) vs. LP spectrum for the normal vs. dysarthria speech case. Waterfall plot is also shown in Fig. 5 to emphasize the corresponding joint time-frequency characteristics during the production of dysarthric speech. From the waterfall plots, We can observe that the formant structure is severely damaged for dysarthric speech as compared to its normal counterpart, where formant peaks and their evolving structures are clearly visible. Thus, the analysis presented in this section indicates that  $F_0$ , its harmonics, formants, and their structures are severely affected due to dysarthria.

## 4 Datasets on Dysarthric Speech

### 4.1 TORGO Database

TORGO Database [16] was developed through a collaboration between the departments of Computer Science and Speech-Language Pathology, University of Toronto; Holland-Bloorview Kids Rehab Hospital, Toronto; and The Ontario Federation for Cerebral Palsy with an aim to develop Automatic Speech Recognition



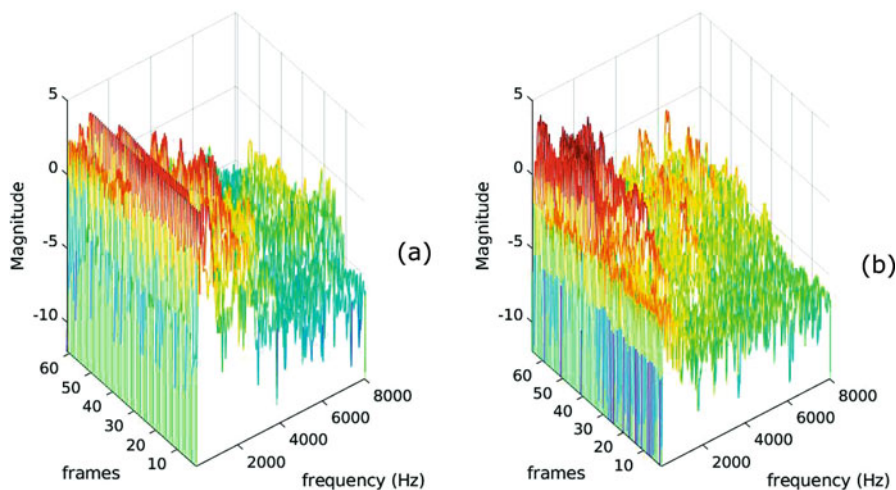


Fig. 5 Waterfall Plot. (a) normal and (b) dysarthric speech

(ASR) systems for people suffering from dysarthria as the unintelligibility in the dysarthric speech results in a Word Error Rate (WER) of as high as 97.5% as compared to a WER of 15.5% for normal speech when both the speech are tested on modern ASR systems trained on normal speech [17].

The database consists of speech samples from 7 dysarthric subjects (4 males and 3 females), between the age 16 and 50, who are chosen by a speech-language pathologist at Bloorview Research Institute, Toronto. Dysarthria of 6 subjects resulted from Cerebral Palsy while 1 subject developed dysarthria due to Amyotrophic Lateral Sclerosis (ALS), which are two of the most common cause of speech impairment [18]. The database also includes speech samples from non-dysarthric subject which were age and gender matched with the dysarthric subjects.

All the subjects were assessed using the standard Frenchay Dysarthria Assessment (FDA) [19] by a speech-language pathologist. The FDA measured 28 perceptual dimension of speech production and articulation which are rated on a 9-point scale. The database is divided into the speech samples of the following categories:

- **Non-Words** consists of 5–10 repetitions of /iy-p-ah, ah-p-iy and p-ah-t-ah-k-ah/, respectively. In addition, utterances with high and low pitch vowels are also recorded, e.g., pronouncing “eee” for 5 s (also used in [20]).
- **Short Words** consists of repetition of English digits 1–10 along with some other words like yes, no, left, right, etc. In addition, 50 words from word intelligibility section of FDA [19] and 360 words from Yorkston-Beukelman Assessment of Intelligibility of Dysarthric Speech (YBAIDS) [21] are chosen. Ten most common words from the British National Corpus were also recorded by the subjects.

- **Restricted sentences** consists of preselected phoneme rich sentences, The Grandfather Passage from Nemours Database [22], 162 sentences from sentence intelligibility section of YBAIDS [21], and 460 sentences from MOCHA database.
- **Unrestricted Sentence** consists of unscripted sentences by the subjects recorded while describing 30 images of interesting situation chosen randomly from Webber Photo Cards: Story Starters Collection.

## 4.2 *Universal Access (UA) Corpus*

The UA database [23] consists of speech samples from 19 dysarthric subjects (15 males and 5 females) [24]. Each subject is scored on a scale of 0–100% on the basis of their speech intelligibility rated by the human listeners. The recording is done using an eight-channel microphone arrays.

The database was recorded into three blocks of words. Each block consists of a total of 255 words, out of which 155 words are repeated across the blocks while the remaining 100 uncommon words are unique for every block. The repeated 155 words include 10 digits (one, two, etc.), 26 radio alphabets (Alpha, Bravo, etc.), 19 computer commands (enter, delete, etc.) and 100 most common words chosen from the Brown corpus of written English. The 100 common words (naturalization, exploit, etc.) were chosen from children’s novels. In this way, each subject was recorded for a total of 765 speech samples out of which 300 samples are distinct uncommon words and remaining 465 are 3 repetitions of 155 distinct words across the blocks.

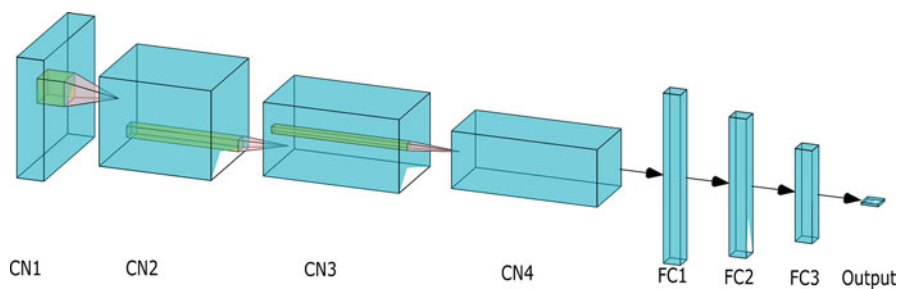
## 4.3 *HomeService Corpus*

The homeService corpus [25] is created as a part of the bigger homeService project, whose objective is to provide the people with speech and motor disabilities with the ability to operate home appliances with voice commands [26]. The project is motivated by the fact that there is a shortage of dysarthric speech data which is recorded in a real life environment within the research community. The project enables its user to operate their home appliances, such as TV, lamps, etc. using voice commands, which are recorded and transferred using a cloud-based environment to a data collection center. The dataset consists of speech data of 5 dysarthric patients (3 males and 2 females). The speech samples were recorded with an 8-channel microphone array at a sampling rate of 16 kHz and consists of two types of speech data. In particular,

- **Enrollment Data**—This data is recorded in a controlled research environment and is used to train the ASR system which the user can use to operate their

**Table 1** Comparison of various corporas for dysarthric speech [17], [24], [26]

Dataset	# Speakers	Male/Female	Text material	Dysarthria inducing disease	Application domain
TORGO corpus	7	4/3	Words & Sentences	Cerebral Palsy, Amyotrophic Lateral Sclerosis	ASR
UA corpus	19	14/5	Words	Cerebral Palsy	ASR
HomeService corpus	5	3/2	Voice commands	Cerebral Palsy, Motor-Neuron Disease	Voice assistants



**Fig. 6** CNN architecture for classification task. After [27]

appliances. The user was asked to read from a list. Therefore, the speech recorded is less natural. Annotation is done using the transcriptions in the reading list.

- **Interaction Data**—This data is recorded at the house of the users while they control their devices. Identity of each word in this data is not known and is therefore, annotated by the human listeners. The speech in the data is more natural.

## 5 Classification of Dysarthric and Normal Speech

Recently, there has been significant increase in the popularity of deep-learning based approach to solve complex task by the computers. For that reason, computers are now able to efficiently perform tasks, such as image classification [28–30], image recognition [31, 32], computer vision, etc. [33, 34]. Focus of deep-learning based algorithms are also increasing in Speech recognition and classification. Convolutional Neural Network (CNN) is one such algorithm which can efficiently detect complex pattern from a set of matrices, such as images. It can be used to classify normal vs. dysarthric speech by learning the patterns in the spectrogram of the speech samples (Fig. 6).

### 5.1 Experimental Setup

For the experiment, spectrograms were obtained for the speech utterances of both normal and dysarthric speakers, keeping a window size of 25 ms and an overlap of 10 ms. These spectrograms were stored in the form of images which is to be fed to network.

Our CNN [35] comprises 4 convolutional layers followed by 3 fully connected layers. Each convolution layer performs a convolution operation with a kernel size of 5x5 keeping step size of 1 and no padding. This convolution operation is followed

by a maxpooling operation with a kernel size of  $4 \times 4$ . The number of output channels obtained by the 4 convolutions channels are 8, 16, 32, and 64, respectively. The output of the final convolutional layer is given to the fully connected layer. Sigmoid activation is used on the final output from the final fully connected layer which provide us with probabilistic value of the input. The threshold value is kept at 0.5, i.e., if output value is greater than 0.5 then the input will be classified as dysarthric speech, and as normal speech if output is less than 0.5. ReLU activation function is used to activate the hidden layers in the network. In addition, Stochastic Gradient Descent (SGD) is used as optimization algorithm and binary cross entropy is taken as the loss function.

## ***5.2 Dataset Used for This Study***

UA corpus [24] is used as the dataset for the experiment. Data from one dysarthric speaker (M07) and one normal speaker (CM01) is used. The data for each speaker was divided into 3 blocks out of which data from block 1 is chosen for training and utterances were taken from mic 3 of the 7-channel microphone array. For testing the accuracy of the model, testing was done with the data from block 3. Specifically, 100 distinct Uncommon Words (UW) were chosen for testing.

## ***5.3 Results and Analysis***

It is observed that the model was able to provide an accuracy of 65.68% on the testing data. The performance of the network is effective given the fact that the experiments were performed on a small training set. The model was able to recognize the variability in the spectrogram that differentiates dysarthric speech from normal speech. In addition, it can be said that the model was also able to learn that the low energy that is associated with the spectrogram of the dysarthric speech.

## **6 Conclusion**

In this chapter, we have discussed dysarthria as a speech technology problem. A number of analysis have been done on normal *vs.* dysarthric speech, such as  $F_0$ , TEO profile, LP residuals, spectrograms, and waterfall plot to provide the reader with an insight of the difference between normal *vs.* dysarthric speech. In addition, some widely used datasets are also discussed along with their key features. Furthermore, an experiment has also been presented for the classification of normal and dysarthric speech using a deep neural networks approach based on CNN.

The effectiveness of the classification task is dependent on the training data on which it is trained on. A model trained on the speech sample of a speaker having high severity-level of dysarthria may not be effective for a speaker with low dysarthria severity-level and vice versa. Therefore, this becomes a prominent limitation of the current methods of classification. Furthermore, research on severity-based classification of dysarthric speaker are very limited. In the future, more sophisticated deep neural networks can be used for classification of normal vs. dysarthric speech and classification based on the severity-level of dysarthria.

**Acknowledgments** The authors would like to thank the authorities at DA-IICT Gandhinagar, India for providing resources and kind support towards the completion of this book chapter. The authors would also like to thank Ms. Priyanka Gupta for providing useful suggestions for further improvement of our this book chapter.

## References

1. BENT, T., BAESE-BERK, M., BORRIE, S.A. and MCKEE, M. (2016) Individual differences in the perception of regional, nonnative, and disordered speech varieties. *The Journal of the Acoustical Society of America (JASA)* **140**(5): 3775–3786.
2. CASTILLO-GUERRA, E. (2009) Acoustic study of dysarthria. *International Journal of Biomedical Engineering and Technology* **2**(4): 352–369.
3. BALLATI, F., CORNO, F. and DE RUSSIS, L. (2018) “hey siri, do you understand me?”: Virtual assistants and dysarthria. In *7th International Workshop on the Reliability of Intelligent Environment (WoRIE), Rome, Italy*.
4. DE RUSSIS, L. and CORNO, F. (2019) On the impact of dysarthric speech on contemporary ASR cloud platforms. *Journal of Reliable Intelligent Environments* **5**(3): 163–172.
5. HWANG, Y., SHIN, D., YANG, C.Y., LEE, S.Y., KIM, J., KONG, B., CHUNG, J. et al. (2012) Developing a voice user interface with improved usability for people with dysarthria. In *International Conference on Computers for Handicapped Persons* (Springer): 117–124.
6. FLOWERS, H.L., SILVER, F.L., FANG, J., ROCHON, E. and MARTINO, R. (2013) The incidence, co-occurrence, and predictors of dysphagia, dysarthria, and aphasia after first-ever acute ischemic stroke. *Journal of Communication Disorders* **46**(3): 238–248.
7. MCNEIL, M.R., ROBIN, D.A. and SCHMIDT, R.A. (1997) Apraxia of speech: Definition, differentiation, and treatment. *Clinical Management of Sensorimotor Speech Disorders* : 311–344.
8. GOLDENBERG, G. (2009) Apraxia and the parietal lobes. *Neuropsychologia* **47**(6): 1449–1459.
9. DUFFY, J.R. (2013) *Motor Speech disorders-E-Book: Substrates, differential diagnosis, and management* (Elsevier Health Sciences).
10. KAISER, J.F. (1990) On a simple algorithm to calculate the ‘energy’ of a signal. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Albuquerque, USA* (IEEE): 381–384.
11. ATAL, B.S. and HANAUER, S.L. (1971) Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acoustical Society of America (JASA)* **50**(2B): 637–655.
12. ANANTHAPADMANABHA, T. and YEGNANARAYANA, B. (1975) Epoch extraction of voiced speech. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **23**(6): 562–570.
13. ANANTHAPADMANABHA, T. and YEGNANARAYANA, B. (1979) Epoch extraction from linear prediction residual for identification of closed glottis interval. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **27**(4): 309–319.

14. VÁSQUEZ-CORREA, J.C., OROZCO-ARROYAVE, J.R. and NÖTH, E. (2017) Convolutional neural network to model articulation impairments in patients with Parkinson's disease. In *INTERSPEECH, Stockholm, Sweden*: 314–318.
15. GUPTA, S., PATIL, A.T., PUROHIT, M., PARMAR, M., PATEL, M., PATIL, H.A. and GUIDO, R.C. (2021) Residual neural network precisely quantifies dysarthria severity-level based on short-duration speech segments. *Neural Networks* .
16. Torgo Database. Available at. <http://www.cs.toronto.edu/~complingweb/data/TORGO/torgo.html>. {Last Accessed 23/03/2021 10:05PM}.
17. RUDZICZ, F., NAMASIVAYAM, A.K. and WOLFF, T. (2012) The torgo database of acoustic and articulatory speech from speakers with dysarthria. *Language Resources and Evaluation* **46**(4): 523–541.
18. KENT, R.D. (2000) Research on speech motor control and its disorders: A review and prospective. *Journal of Communication Disorders* **33**(5): 391–428.
19. ENDERBY, P. (1980) Frenchay dysarthria assessment. *British Journal of Disorders of Communication* **15**(3): 165–173.
20. BENNETT, J.W., VAN LIESHOUT, P.H. and STEELE, C.M. (2007) Tongue control for speech and swallowing in healthy younger and older adults .
21. YORKSTON, K.M. and BEUKELMAN, D.R. (1978) A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders* **11**(6): 499–512.
22. MENENDEZ-PIDAL, X., POLIKOFF, J.B., PETERS, S.M., LEONZIO, J.E. and BUNNELL, H.T. (1996) The Nemours database of dysarthric speech. In *Proceeding of Fourth International Conference on Spoken Language Processing (ICSLP), Philadelphia, USA (IEEE)*, **3**: 1962–1965.
23. Universal Access (UA) Corpus. Can be requested at. <http://www.isle.illinois.edu/sst/data/UASpeech/>. {Last Accessed 23/03/2021 10:30PM}.
24. 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 *Ninth Annual Conference of the International Speech Communication Association, Brisbane, Australia*.
25. HomeService Corpus. Can be requested at. <http://www.isle.illinois.edu/sst/data/UASpeech/>. {Last Accessed 23/03/2021 10:38PM}.
26. NICOLAO, M., CHRISTENSEN, H., CUNNINGHAM, S., GREEN, P. and HAIN, T. (2016) A framework for collecting realistic recordings of dysarthric speech-the homeservice corpus. In *Proceedings of LREC 2016 (European Language Resources Association)*.
27. KRIZHEVSKY, A., SUTSKEVER, I. and HINTON, G.E. (2012) Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **25**: 1097–1105.
28. LIU, S., TIAN, G. and XU, Y. (2019) A novel scene classification model combining ResNet based transfer learning and data augmentation with a filter. *Neurocomputing* **338**: 191–206.
29. WANG, F., JIANG, M., QIAN, C., YANG, S., LI, C., ZHANG, H., WANG, X. *et al.* (2017) Residual attention network for image classification. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), Hawaii, USA*: 3156–3164.
30. ZHU, C., SONG, F., WANG, Y., DONG, H., GUO, Y. and LIU, J. (2019) Breast cancer histopathology image classification through assembling multiple compact CNNs. *BMC Medical Informatics and Decision Making* **19**(1): 1–17.
31. HE, K., ZHANG, X., REN, S. and SUN, J. (2016) Deep residual learning for image recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA*: 770–778.
32. LU, Z., JIANG, X. and KOT, A. (2018) Deep coupled resnet for low-resolution face recognition. *IEEE Signal Processing Letters* **25**(4): 526–530.
33. JUNG, H., CHOI, M.K., JUNG, J., LEE, J.H., KWON, S. and YOUNG JUNG, W. (2017) Resnet-based vehicle classification and localization in traffic surveillance systems. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Hawaii, USA*: 61–67.

34. LIU, X., ZHOU, Y., ZHAO, J., YAO, R., LIU, B., MA, D. and ZHENG, Y. (2020) Multiobjective ResNet pruning by means of EMOAs for remote sensing scene classification. *Neurocomputing* **381**: 298–305.
35. FUKUSHIMA, K. and MIYAKE, S. (1982) Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and Cooperation in Neural Nets* (Springer), 267–285.