DAU-KDAH Dysarthic Multi-Lingual and Multimodal Speech Corpora for Indic Languages

Arth J. Shah *‡, Hiya Chaudhari *‡, Kavya Kumar *, Arushi Srivastava *, Priya J. Kaple *, RavindraKumar M. Purohit *, Dharmendra H. Vaghera *, Bhavna Singh †, Aparna Walanj †, Abhishek Srivastava †, and Hemant A. Patil *
* Speech Research Lab, Dhirubhai Ambani University (DAU)
E-mail: {202101154, 202101047, 202101017, 202215003, hemant_patil}@daiict.ac.in
† Kokilaben Dhirubhai Ambani Hospital-Medical Research Institute (KDAH-MRI), Mumbai, Maharashtra, India
‡ Equal Contribution

***Abstract*—Dysarthria is a disease that affects timing, rhythm, and pronunciation, which frequently makes traditional Automatic Speech Recognition (ASR) systems less effective. Most of the corpora for dysarthric research are in English language, which restricts multilingual users. In this paper, we aim to develop an multilingual speech corpora for ASR system (i.e., Assistive Speech Technology (AST)) that can adapt to various dysarthric severity-levels, improving accessibility and inclusivity in speech-based technologies. To that effect, speech samples in Hindi, Marathi, Gujarati, and Indian English are collected from both dysarthric *vs.* normal subjects in order to accomplish this goal. The initial goal of the research is to use spectral analysis to distinguish between dysarthric and typical speech, with the ultimate goal of creating a sophisticated ASR system for people with dysarthria. After obtaining clearance from ethics committee, data collection is conducted as a synergistic collaboration between DAU and KDAH-MRI. Finally, we present interim results in severity-level classification using spectral features and obtained accuracy of 60.46 %. We also employed transformers based features, in which we got highest accuracy of 58.13 %. For realistic approach of the proposed dataset, we performed analysis of latency period.**

***Index Terms*—Speech impairment, ethical guidelines, dysarthric severity, multilingual and multimodal corpora.**

I. Introduction

Motor speech disorders such as dysarthria impair the muscular control required for speech production, resulting in reduced clarity and intelligibility of spoken communication [1]. The development of robust Automatic Speech Recognition (ASR) systems relies heavily on the ability to understand and accurately transcribe dysarthric speech. However, existing datasets often fall short in representing the full range of grammatical diversity, acoustic characteristics, and the spontaneous nature of real-world dysarthric speech. Among the widely used corpora, UA-Speech includes only scripted speech from individuals with cerebral palsy [2], while TORGO offers both controlled and conversational samples from dysarthric and control speakers [3]. Despite their contributions, these resources have notable limitations, including constrained linguistic diversity and limited spontaneity in speech. One critical gap in the field is the lack of multilingual dysarthric speech corpora. Most existing datasets are restricted to English, ignoring the linguistic and phonetic diversity present in other languages. This limits the generalizability of current ASR models and hinders their application in linguistically diverse populations. To address these limitations, we introduce the DAU-KDAH dataset, a novel Indic-language-based dysarthric speech corpus. It comprises both normal and dysarthric speech samples collected from multilingual speakers, providing a more representative and inclusive resource. Notably, the dataset features spontaneous speech, capturing the real-world challenges faced by individuals with dysarthria more effectively than scripted datasets. This paper presents an interim analysis from our ongoing data collection drive (to be discussed Section II). Authors firmly believe that the DAU-KDAH dataset has the potential to enhance new research direction in speech technology by supporting the development of multilingual ASR systems for dysarthic patients.

To analyze the proposed dataset, and to propose a baseline system, we in this study also propose the experiments based on transformers-based pre-trained models, such as HuBERT, wav2vec, XLSR, and whisper, as well as commonly used acoustic features, such as Mel Frequency Cepstral Coefficients (MFCC), Linear Frequency Cepstral Coefficients (LFCC), and Gammatone Frequency Cepstral Coefficients (GFCC), for classification of severity-levels of dysarthric patients. Advance deep learning models, such as Bidirectional Long Short-Term Memory (BiLSTM), which records *temporal* relationships in speech signal, and Convolutional Neural Networks (CNNs), which extract *spatial* patterns, are used as classifiers tasks. By combining these approaches, proposed study aims to improve the categorization of dysarthric speech-severity for low resource languages.

A. Related Works

English language speech datasets, such as UA-Speech and TORGO have been the main source of data for dysarthric speech recognition studies [4]. With the help of recordings from people with cerebral palsy, UA-Speech allows models to be developed for both ASR and the classification of dysarthric severity-level [4]. Similarly, speech samples from non-dysarthric and dysarthric speakers recorded in conversational and controlled environments are included in the TORGO

979-8-3315-7206-8/25/$31.00

dataset [2]. Although these datasets have made a substantial contribution to the development of ASR, their main drawback is that they are available only in English, which restricts their use for non-English dysarthric speech research [4] - [2]. As a result, resources for creating ASR models that take into account linguistic diversity—especially for Indic languages—are a few. The creation of ASR systems for underrepresented linguistic groups is made possible by this study's introduction of a novel dysarthric speech dataset in Indic languages, which addresses this constraint. The goal of this project is to make ASR more inclusive and accessible by developing newer datasets and carrying out baseline studies. This study offers the following novelty:

- Language diversity via multilingual data collection drive in four languages, namely, Hindi, Marathi, Gujarati, and Indian English.
- With respect to assistive speech technology of lip-to-wave conversion, simultaneous audio-video recordings (i.e., multimodal) are done.
- For *cross-lingual* experiments, recordings of 5 dysarthric speakers in Hindi and English languages is also done.

II. Proposed DAU-KDAH Dataset

A multi-class classification of dysarthric speech according to severity-levels—low, medium, high, and normal—is introduced by the suggested dataset. A more thorough examination of dysarthric speech traits is made possible by this classification, which guarantees an organized and thorough portrayal of speech impairments. Proposed dataset covers a wider range of dysarthric conditions than the existing datasets, which mainly concentrate on a particular dysarthria types (or etiology), such as *spastic* and *mixed*. Because of this diversity, more generalized ASR models that can handle different dysarthric speech patterns can be developed [5]. To guarantee a balanced representation of both dysarthric and normal subjects, the dataset collection process was carried out in five phases. In order to create a baseline for comparison, data from normal subjects was gathered in the first two phases. The following three phases recorded speech variations at various severity-levels with an emphasis on dysarthric speakers. The dataset is useful for research on ASR because of this phased approach, which guarantees high quality data while taking speaker variability into account.

A. Normal Speakers' Data

Four languages—Hindi, Indian English, Marathi, and Gujarati—were considered to collect data from normal subjects, guaranteeing linguistic diversity in the proposed dataset. ASR models can be applied more broadly across various Indic languages due to multilingual approach [7]. Six months were invested for the data collection process, which guaranteed a large enough sample size and speaker variability. The entire data collection procedure was carried out in a single room with a controlled setup in order to preserve consistency and reduce outside acoustic noise interference. Recordings from 129 normal subjects were collected at the Speech Research Lab@DAU. Strong ASR models that can distinguish between normal and impaired speech patterns are developed with the help of the gathered data, which provides an essential baseline for comparison with dysarthric speech [8]. Since a number of factors, including pitch (F_0), duration, and spectral features, can affect speech intelligibility, altering these aspects has been demonstrated to enhance perception in noisy environments [9].

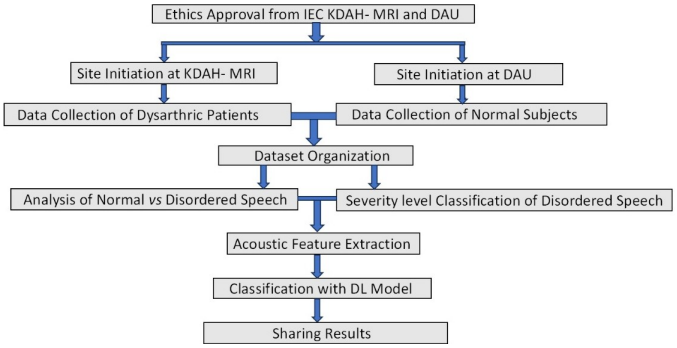

Fig. 1. Flow Adopted in the Proposed DAU-KDAH Data Collection Drive.

Fig. 2. (a) Recording setup, and (b) Zoom H4N microphone.

Variations in recording conditions, equipment, and subject participation make obtaining accurate speech data from both normal and dysarthric subjects challenging. Controlled recording is crucial, as deep learning models and signal processing techniques rely on precise data. To minimize *claustrophobia* [6], sessions were conducted in a hospital room with the door open for patient comfort [10], though this introduced background noise from conversations, footsteps, and medical equipment. Noise reduction techniques were applied to enhance speech clarity while preserving essential features. Despite these challenges, recording in a real hospital setting provided valuable insights, improving the study's validity and applicability.

Three Zoom H4n Handy recorders known for their reliability and high audio quality were used, utilizing only the built-in tracks. To capture diverse acoustic effects, two recorders were paired with DSLR cameras at 90° (left) and

TABLE I
SELECTED DYSARTHRIC SPEECH CORPORA IN THE LITERATURE

Corpus Used	Method	Outcome	Limitations
TORGO [2]	Subjects read English text from a screen 60 cm in front.	Database of 7 dysarthric subjects (cerebral palsy, ALS).	Multiple microphones enable noise reduction unavailable in single-source speech.
Google Euphonia [6]	Participants recorded remotely, without specialized equipment.	Corpus $>$ 1 million utterances ($>$ 1300 hours).	No uniformity in data collection due to remote recording.
UA_Speech [4]	8-microphone array with 1.5 cm spacing.	Dysarthric speech from 19 individuals with cerebral palsy.	Limited to cerebral palsy patients, unbalanced gender distribution.
ALS-Dataset [5]	Study on 67 ALS patients with 36.7 hours of audio.	Good performance on dysarthric ALS and accented speech.	Robust ASR for strongly dysarthric speech.
Qolt [7]	Speaker adaptation scheme for dysarthric speech.	Minimizing training data critical for dysarthric speakers.	Dysarthric speakers tire quickly, requiring minimal data.
DAU-KDAH **(Proposed)**	Deep learning models for dysarthric speech classification.	Comparison of different feature extractors and classifiers.	Performance varies based on features and classifiers.

180° (front), with microphones placed 25 cm apart. An OPPO A-58 at 22.5° and an iPhone 14 Pro Max served as backup devices during quiet recording sessions. The zoom H4n recorders ensured clear audio with a 140 dB SPL and a -120 dB EIN noise floor. Generalized ASR models are made possible by the dataset's introduction of a multi-class classification of dysarthric severity-level (low, medium, high, and normal) [9]. Data collection followed five phases: the first two recorded normal speech for baseline comparison, while the final three captured varying dysarthric severity-levels. This structured approach ensured high quality audio and video data while considering speaker variability.

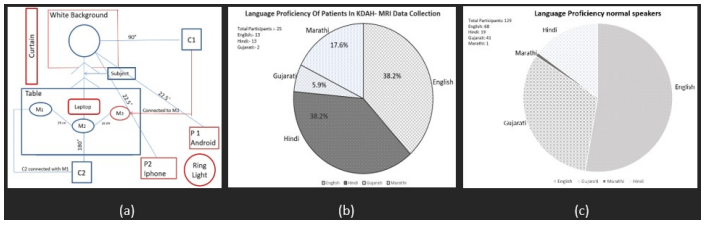

Fig. 3. (a) is the top view of a typical experimental setup developed for data collection, (b) is pie chart depicting statistic of corpora, and (c) is the statistics of normal subjects.

B. Dysarthric Speakers' Data

Neurological disorders that affect muscle control can result in dysarthria, a motor speech disorder that affects speech strength, speed, and clarity. Due to weak or uncoordinated articulation muscles, people with dysarthria may speak slowly, slurringly, or strainedly [12]. The recordings were made in a comparatively calm and controlled setting on the 6^{th} floor of the hospital, which is about 18 m above sea-level. Despite the equipment change, the microphone was positioned carefully in order to guarantee that speech characteristics were successfully recorded for additional analysis. The total intelligibility of a speaker is indicated by the overlap degree and vowel space area [13]. A structured document called the *metadata* form is used to gather vital information about people who have dysarthria. Important information is recorded, including the participant's personal details, the kind and cause of the dysarthria, the bodily parts affected, the degree of severity, and the intelligibility of their speech. It also contains details about the length of speech treatment, the quantity of recording sessions, and the language background. The clinical evaluation of speech, which assesses breathing, phonation, articulation, and intelligibility, is a standardized method used to measure the severity and impact of dysarthria. It facilitates the categorization of dysarthria categories and the tracking of alterations in speech function over time. Frequent evaluation is essential for monitoring the development of dysarthria, and developments in AI-powered ASR could improve assessment and treatment even more. Fig. 2 represents the expremental setup preview (a) and language distribution for normal and dysarthic speaker. The red instruments in Fig. 2 (a), were the objects which were unavailable during dysarthic speech collection.

C. Clinical Assessment of Dysarthria at KDAH-MRI

At Kokilaben Dhirubhai Ambani Hospital-Medical Research Institute (KDAH-MRI), a clinical evaluation of dysarthria was carried out to determine patients' speech difficulties. Using structured evaluation techniques, this assessment examined speech characteristics, severity-levels, and intelligibility. These results advance our knowledge of the effects of dysarthria, and how it varies among various patient populations.

D. Severity-Levels of Dysarthric Speech

Four severity categories are used to classify speech samples: normal, low, medium, and high. The most difficult speech for ASR models to process is high-severity speech, which consists of six 30-seconds audio files with severe distortion and poor articulation. These patients require models to handle severely impaired speech since they have trouble maintaining consistent pitch (i.e., fundamental frequency, (F_0)), loudness, and clarity [10]. With seventeen 30-seconds audio clips, medium-severity speech demonstrates inconsistent prosody, uneven articulation, and slower speaking rates. Although

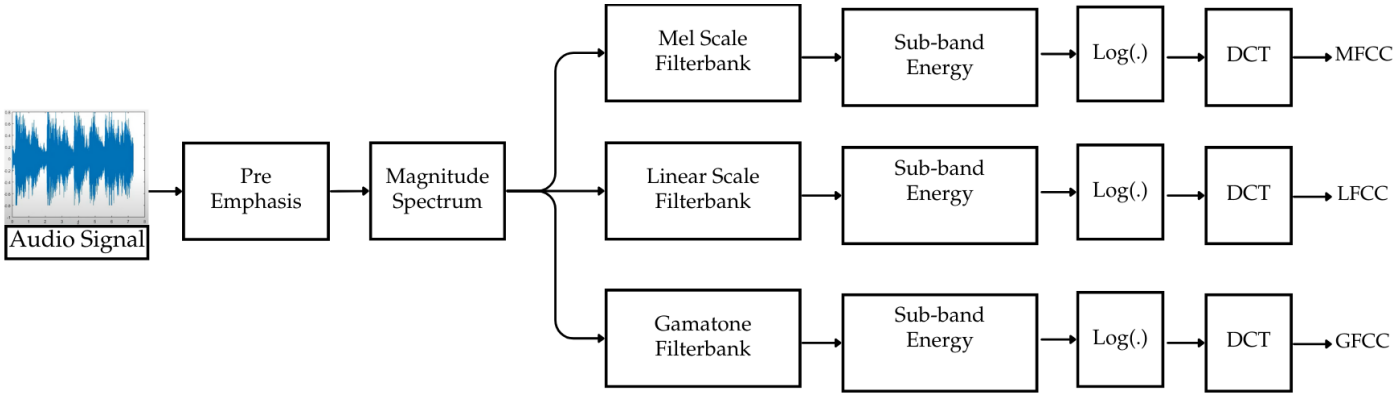

Fig. 4. Functional block diagram of MFCC, LFCC, and GFCC. After [10] –[11].

still largely understandable, there are noticeable differences in pronunciation, which presents a big challenge for ASR models. Ten 30-seconds audio clips of low-severity speech show largely comprehensible speech with a few slight rhythmic changes, slurred words, or articulation problems. ASR models should perform noticeably better in this category than in high and medium-severity situations, despite the presence of aberrations [11]. A vital reference point for comparison is the normal speech category, which consists of ten 30-seconds recordings from healthy people with normal articulation, pitch (i.e., fundamental frequency, (F_0)), tone, and loudness.

E. Efforts and Experiences During Data Collection

During the data collection process, we faced multiple challenges that made it difficult to gather a comprehensive and diverse dataset. In particular,

- One of the primary obstacles was that many subjects were unwilling to provide video recordings due to *privacy* concerns and personal discomfort. Since video recordings could offer valuable insights into facial expressions and lip movements for possible lip-to-wave conversion (another assistive technology) [3], their absence limited the scope of our study [14].
- Additionally, some participants, especially patients suffering from dysarthria, experienced emotional distress during the recording sessions. Many of them struggled to articulate words or complete the tasks, which led to frustration and sadness. This emotional involvement made it challenging to ensure consistency in the recordings. In this context, emotion recognition of disabled people is technologically challenging and less researched area, where the first study of its kind is published [15].
- Another major difficulty was the limited availability of dysarthria speech corpora for research purposes. This scarcity of resources highlighted the need for the creation of new, multilingual dysarthria datasets to facilitate further research and improve ASR models for affected individuals. Overcoming these challenges required patience, ethical considerations, and adaptability in our data collection approach, ensuring that participants felt comfortable, while still obtaining useful data for analysis.

TABLE II
COMPARISION OF DATASETS ACROSS VARIOUS DATASETS

Parameter	DAU-KDAH	TORGO	UA-Speech
Total Files	307	1982	3573
Total Languages	4	1	1
Total Duration(mins)	103.50	133.55	182.68
Avg. Duration per Lang.(mins)	25	5	19
Total Dysarthric speakers	25	5	19
Total Size (in MB)	970	244	334
Sampling Frequency (Hz)	16,000	16,000	44,100
Bit Resolution (bits)	16	16	16

III. EXPERIMENTAL SETUP

A. Classifier Used

1) Bidirectional Long Short-Term Memory (BiLSTM): The BiLSTM network is employed to model the temporal dynamics of speech. It extends the conventional LSTM by processing the input sequence in both forward and backward directions, enabling the model to capture past and future context simultaneously. This is particularly beneficial for speech signals where contextual dependencies across time are important for understanding impairments such as those caused by dysarthria. The architecture and parameters of the classifier were same as employed in [16].

B. Convolutional Neural Network (CNN)

CNNs are widely used for speech classification tasks due to their ability to learn hierarchical feature representations. In our setup, CNNs are used to capture local time-frequency patterns from spectrogram-like inputs. They are efficient and effective for detecting spatial patterns that may be indicative of articulatory deficits present in dysarthric speech.

1) Features Used: We experimented with both traditional signal processing-based features and modern deep learning-based embeddings to represent the audio data.

2) MFCC: Mel-Frequency Cepstral Coefficients (MFCC) are among the most widely used features in speech processing. They model the short-term power spectrum of speech based on the mel scale, which closely approximates human auditory perception.

3) LFCC: Linear-Frequency Cepstral Coefficients (LFCC) differ from MFCCs by using a linear rather than mel filterbank, preserving more high-frequency information. This can be useful for detecting subtle spectral deviations in dysarthric speech.

4) GFCC: Gammatone-Frequency Cepstral Coefficients (GFCC) use a gammatone filterbank that mimics the human cochlear filtering mechanism more closely. They are particularly robust in noisy environments and are effective for capturing perceptual features relevant to speech disorders.

5) Whisper: Whisper is a robust ASR and speech representation model trained on a large multilingual and multitask dataset. We extract intermediate representations from Whisper as deep audio features. These embeddings capture both phonetic and linguistic information, making them valuable for dysarthric speech analysis.

6) HuBERT: HuBERT (Hidden-Unit BERT) is a self-supervised speech model trained to predict masked units. It provides high-level, context-aware audio embeddings that are effective for downstream speech tasks, particularly in low-resource or impaired speech scenarios [17].

7) XLSR: XLSR (Cross-Lingual Speech Representations) is based on the Wav2Vec 2.0 architecture but trained on multilingual datasets [18]. Its ability to generalize across languages makes it well-suited for our multilingual dysarthric corpus.

8) Wav2Vec 2.0: Wav2Vec 2.0 is a self-supervised model that learns speech representations directly from raw audio. Its powerful contextual embeddings capture acoustic and phonetic information, making it effective for impaired speech detection.

9) Performance Metrics: We evaluated performance of the system using accuracy as performance metrics.

IV. Experimental Results

While the data collection drive is ongoing, here we present interim results and severity-level classification. Fig. 5 represents the results obtained on the proposed dataset . Acoustic features, such as MFCC [19], LFCC [19], and GFCC [20] perform better with a BiLSTM classifier, whereas pre-trained deep learning model features perform poorly on BiLSTM but excel on CNN due to its strength in image-based tasks. On TORGO and UA-Corpus, similar features yielded 96.48 % and 93.78 % accuracy, indicating easier severity detection on existing datasets. A key factor in result variation is the recording environment—while most datasets were collected in acoustic rooms, ours was recorded in-the-wild hospital environment at 75 feet above sea-level. Interestingly, results varied for HuBERT, presenting an open research problem. According to observations, individuals' pronunciation varied; some spoke clearly, while others had mild-to-severe difficulties. Responses were influenced by emotional support and differing levels of engagement. The majority could read Hindi and English, and some were even familiar with Marathi, but one had trouble pronouncing the letters. These results demonstrate how participation, emotional support, and recording conditions affect the processing of dysarthric speech.

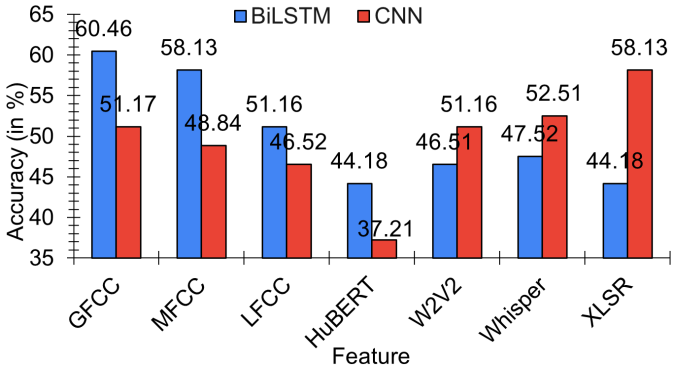

Fig. 5. Obtained results on proposed DAU-KDAH dataset.

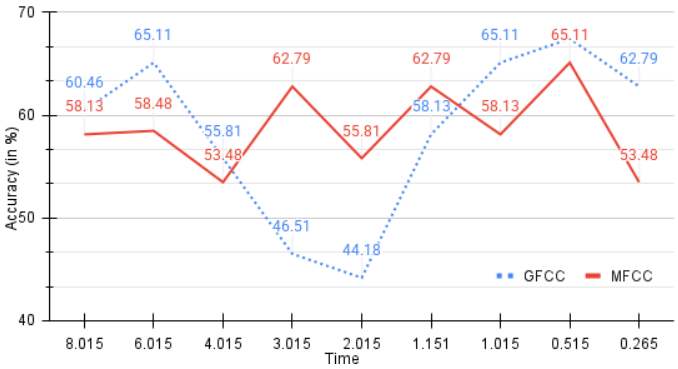

Fig. 6. Latency analysis on MFCC and GFCC features on proposed DAU-KDAH dataset.

A. Analysis of Latency Period

In order to analyze the minimum duration of speech wave required to obtain reliable results, we performed latency analysis on the proposed dataset with 4 class classification, i.e., high, medium, low, and normal. Both classes, i.e., dysarthic speech and normal speech contains all 4 languages. Such type of experiments helps to give reliability on real-life scenarios. In Fig. 6, it can be observed that even if we choose minimum speech duration (i.e., 0.265 seconds of the audio), the model is able to give similar performance as of 8 seconds. This observation helps us to conclude that the features being captured are sole dysarthic speech-based features, which tends to hold its properties even with lowest speech duration.

V. Summary and Conclusions

This study investigated development of dysarthric speech corpus with an emphasis on Indic languages in the multilingual settings and simultaneous audio-video multimodal recordings. Our dataset contains spontaneous, genuine speech changes, which makes it more applicable to real-world circumstances than the existing datasets, such as UA-Speech and TORGO, which are restricted to English and structured speech. The results show that creating inclusive future ASR systems is feasible even in the face of obstacles, such as resource limitations, data imbalance, and language imbalance. To further increase recognition accuracy and accessibility, future research will

concentrate on growing the dataset, improving models across a range of dysarthric severity-levels, and implementing cutting-edge DL approaches to develop assistive speech technologies. Correlating the quality of dysarthric speech using subjective measures (such as, listening tests and clinical evaluation of speech) *vs.* objective measures remains open research problem in the literature.

ACKNOWLEDGEMENTS

The authors sincerely thank the MeitY, for funding this study under project 'BHASHINI', (Grant ID: 11(1)2022-HCC(TDIL)).

REFERENCES

[1] P. Enderby, "Disorders of communication: Dysarthria," *Handbook of Clinical Neurology*, vol. 110, pp. 273–281, 2013.

[2] F. Rudzicz, A. K. Namasivayam, and T. Wolff, "The torgo database of acoustic and articulatory speech from speakers with dysarthria," *Language Resources and Evaluation (LRE)*, vol. 46, pp. 523–541, 2012.

[3] S. B. Hegde, K. Prajwal, R. Mukhopadhyay, V. P. Namboodiri, and C. Jawahar, "Lip-to-speech synthesis for arbitrary speakers in the wild," in 30^{th} *ACM International Conference on Multimedia*, Lisbon, Portugal, 2022, pp. 6250–6258.

[4] H. Kim, M. Hasegawa-Johnson, A. Perlman, *et al.*, "Dysarthric speech database for universal access research," in *INTERSPEECH*, vol. 2008, Brisbane, Australia, 2008, pp. 1741–1744.

[5] R. Dubbioso, M. Spisto, L. Verde, *et al.*, "Voice signals database of als patients with different dysarthria severity and healthy controls," *Scientific Data*, vol. 11, no. 1, p. 800, 2024.

[6] R. Booth and S. Rachman, "The reduction of claustrophobia—i," *Behaviour Research and Therapy*, vol. 30, no. 3, pp. 207–221, 1992.

[7] S. Kim, E. J. Yeo, and M. Chung, "Design and creation of dysarthric speech database for development of qolt software technology," in *Proceedings of the 2012 International Conference on Language Resources and Evaluation (LREC)*, 2012, pp. 1234–1238.

[8] S. Rathod, M. Charola, and H. A. Patil, "Noise robust whisper features for dysarthric severity-level classification," in 10^{th} *International Conference of Pattern Recognition and Machine Intelligence, PReMI 2023, December 12–15*, vol. LMCS 14222, Kolkata, India: Springer, 2023, pp. 708–715.

[9] J. R. Green, R. L. MacDonald, P.-P. Jiang, *et al.*, "Automatic speech recognition of disordered speech: Personalized models outperforming human listeners on short phrases," in *INTERSPEECH*, vol. 2021, Brno, Czech Republic, pp. 4778–4782.

[10] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *The Journal of the Acoustical Society of America (JASA)*, vol. 24, no. 2, pp. 175–184, 1952.

[11] X. Zhao, Y. Shao, and D. Wang, "Robust speaker identification using a casa front-end," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, pp. 5468–5471.

[12] J. Shor, D. Emanuel, O. Lang, *et al.*, "Personalizing asr for dysarthric and accented speech with limited data," *arXiv preprint arXiv:1907.13511*, 2019, {Last accessed: 22^{nd} June, 2025.

[13] J. N. Saba and J. H. L. Hansen, "The effects of lombard perturbation on speech intelligibility in noise for normal hearing and cochlear implant listeners," *Journal of the Acoustical Society of America (JASA)*, vol. 151, no. 2, p. 1007, 2022.

[14] C. Ittichaichareon, S. Suksri, and T. Yingthawornsuk, "Speech recognition using mfcc," in *International Conference on Computer Graphics, Simulation and Modeling*, vol. 9, Bangkok, Thailand, 2012, p. 2012.

[15] S. Hantke, H. Sagha, N. Cummins, and B. Schuller, "Emotional speech of mentally and physically disabled individuals: Introducing the emotass database and first findings," in *INTERSPEECH*, Sweden, Stockholm, 2017, pp. 2943–2947.

[16] A. J. Shah, N. V. Mandaviya, and H. A. Patil, "Voice liveness detection using linear frequency residual cepstral coefficients," in *Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2024, Macau, China, pp. 1–6.

[17] W.-N. Hsu, B. Bolte, Y.-H. H. Tsai, K. Lakhotia, R. Salakhutdinov, and A. Mohamed, "Hubert: Self-supervised speech representation learning by masked prediction of hidden units," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 29, pp. 3451–3460, 2021.

[18] A. Babu, C. Wang, A. Tjandra, *et al.*, "XLS-R: self-supervised cross-lingual speech representation learning at scale," H. Ko and J. H. L. Hansen, Eds., INTERSPEECH, 2022, Incheon, Korea, pp. 2278–2282.

[19] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.

[20] H. Wang and C. Zhang, "The application of gammatone frequency cepstral coefficients for forensic voice comparison under noisy conditions," *Australian Journal of Forensic Sciences*, vol. 52, no. 5, pp. 553–568, 2020.