**IEEE上关于dysarthria论文题目与摘要汇总**

**Quantitative Assessment of Syllabic Timing Deficits in Ataxic Dysarthria,**

Parametric analysis of Cerebellar Dysarthria (CD) may be valuable and more informative compared to its clinical assessment. A quantifiable estimation of the timing deficits in repeated syllabic utterance is described in the current study. Thirty-five individuals were diagnosed with cerebellar ataxia to varying degrees and twenty-six age-matched healthy controls were recruited. To automatically detect the local maxima of each syllable in the recorded speech files, a topographic prominence incorporated concept is designed. Subsequently, four acoustic features and eight corresponding parametric measurements are extracted to identify articulatory deficits in ataxic dysarthria. A comparative study on the behaviour of these measures for dysarthric and non-dysarthric subjects is presented in this paper. The results are further explored using a dimensionreduction tool (Principal Component Analysis) to emphasize variation and bring out the strongest discriminating patterns in our feature dataset.

**Non-invasive stroke diagnosis using speech data from dysarthria patients,**

Acute Ischemic Stroke (AIS) is a major cause of disability and can lead to death in severe cases. A common symptom of AIS, dysarthria, significantly impacts the quality of life of patients. In this study, we developed a deep learning model using dysarthria data for cost-effective and non-invasive brain stroke diagnosis. We utilized models such as ResNet50, InceptionV4, ResNeXt50, SEResNeXt18, and AttResNet50 to effectively extract and classify speech features indicative of stroke symptoms. These models demonstrated high performance, with Sensitivity, Specificity, Precision, Accuracy, and F1-score values reaching 96.77%, 96.08%, 92.82%, 95.52%, and 93.82%, respectively. Our approach offers a non-invasive, cost-effective alternative for early stroke detection, with potential for further accuracy improvements through additional research. This method promises rapid, economical early diagnosis, which could positively impact long-term treatment and healthcare options.

**Fine-Tuning Pre-Trained Audio Models for Dysarthria Severity Classification: A Second Place Solution in the Multimodal Dysarthria Severity Classification Challenge,**

Dysarthria, a neurological disorder affecting speech, poses significant challenges for automatic diagnosis and severity classification due to its complexity and the scarcity of relevant datasets. This study addresses these challenges by leveraging the dataset provided by the ISCSLP 2024 Multimodal Dysarthria Severity Classification Challenge. Our approach primarily involves two key steps: splitting the training set into training and validation subsets, and fine-tuning pre-trained au-dio models to adapt them to the task. To ensure robust evaluation and prevent label leakage, we employed the StratifiedGroupKFold function from sklearn for dataset splitting, ensuring that training and validation sets do not share participants. This method allowed us to maintain consistent statistical distributions across folds. For model finetuning, we selected two pre-trained audio models, w2v-bert-2.0 and whisper-Iarge-v3, and fine-tuned them on different folds of the dataset. Our results indicate that the w2v-bert-2.0 model fine-tuned on fold-2 achieved the highest performance, with a validation Fl-score of 0.699 and an online score of 7.7452. The experimental results, including confusion matrices and embedding distributions, highlight the model's ability to distinguish between different severity levels of dysarthria. Our approach achieved a commendable second place in the competition’ demonstrating the effectiveness of fine-tuning pre-trained audio models for this task. Future work will explore the integration of multimodal data and multi-task learning strategies to further enhance model performance.

**Automated Detection and Severity Assessment of Dysarthria using Raw Speech,**

Dysarthria is a medical condition that impairs an individual’s ability to speak clearly due to muscle weakness or paralysis. To diagnose and monitor dysarthria severity, this article proposes the use of a deep learning model that utilizes raw speech waveforms. This approach eliminates the need for feature engineering and enhances the model’s ability to handle noise and speech variability. The proposed system was compared to a standard convolutional neural network (CNN) model and was found to perform better in both dysarthria severity classification and dysarthria/healthy control classification tasks. The results indicate that the SincNet model achieved an accuracy of 95.7% and 99.6% in these tasks, respectively. The proposed system can aid clinicians in diagnosing and monitoring dysarthria severity and could have broader applications in speech-related disorders. The study emphasizes the importance of using raw waveform-based models for speech analysis and demonstrates the effectiveness of SincNet in automatic dysarthria/healthy control classification and dysarthria severity assessment.

**Investigation on articulatory and acoustic characteristics of dysarthria,**

Direct measurement of articulation contributes to figure out the mechanism of dysarthric speech production accurately, comparing to analyzing acoustic signal alone. This study makes a statistical comparison between dysarthric and normal speech, and investigates and analyzes dysarthric characteristics based on articulatory and acoustic data of continuous English utterance. The distribution of the articulatory point is calculated for each vowel, where the vowels show a relatively smaller region for dysarthria than for normal speech. This implies that the speakers with dysarthria may have some difficulties to fully use the articulatory space as the speakers without dysarthria. The rhythm and sound source are analyzed using autocorrelation function, power spectrum and linear prediction coding. The results reveal that the speakers without dysarthria can keep steady rhythm and energy in uttering consonant-vowel repetition sequences but it is hard for the speakers with dysarthria to maintain the stability. Meanwhile, the dysarthric sound source shows unstable period of the fundamental frequency, which reflects that the speakers with dysarthria could not control the glottis movement well.

**Evolving Diagnostic Techniques for Speech Disorders: Investigating Dysarthria Classification Through DenseNet201 CNN Framework,**

The primary subject of this research work pertains to dysarthria, a prevalent speech impairment observed in individuals diagnosed with cerebral palsy (CP) or amyotrophic lateral sclerosis (ALS). Prompt detection of dysarthria is essential for effective therapies and improved patient results, as it significantly impacts communication proficiency. The research employed a DenseNet201 Convolutional Neural Network (CNN) to categorize speech using the TORGO database, which has 2000 samples of individuals with and without dysarthria, representing various genders. The dataset comprises individuals of both genders, encompassing both dysarthric and non-dysarthric individuals. Each dataset consists of 500 samples, which were collected during distinct sessions. The DenseNet201 convolutional neural network (CNN) technique has a notable accuracy rate of 96% in distinguishing between cases of dysarthria and non-dysarthria. This outcome underscores the potential of deep learning technology in aiding the timely identification of dysarthria and facilitating fast interventions for affected individuals. This work provides valuable insights on the progress of diagnostic capabilities and offers a potential avenue for enhancing the well-being of those experiencing speech impairments.

**Analysis of Features for Dysarthria Severity Classification from Speech,**

Dysarthria is a disorder that affects the ability of an individual to speak clearly. Diagnosis of the severity of dysarthria can aid in providing appropriate therapy to the individual. Therefore, the current work focuses on identifying features that can be used to estimate the severity of dysarthria. In this regard, two feature sets are considered, namely DisVoice and OpenSMILE, and the genetic algorithm is used to identify the most suitable list of features from both feature sets using 3 speech corpora, namely SSN-TDSC, Torgo, and UA Speech corpora. In order to test the effectiveness of these features, classifiers are trained on each feature set as a whole and on the features identified by the genetic algorithm, and their performance compared. It is observed that articulatory and prosodic features are most important in the diagnosis of severity of dysarthria. A maximum accuracy of 85% is obtained with the shortlisted DisVoice features and 97% with the OpenSMILE features.

**Multi-Modal Dysarthria Severity Assessment Using Dual-Branch Feature Decoupling Network and Mixed Expert Framework,**

Dysarthria, a motor speech disorder resulting from neurological damage, significantly impairs an individual's ability to articulate words. Assessing the severity of dysarthria is crucial for effective therapeutic interventions and rehabilitation strategies. However, current methods are time-consuming and subjective, leading to potential inconsistencies. This study proposed a dual-branch feature decoupling network and mixed expert framework for the automatic diagnosis and assessment of dysarthria. The proposed hybrid network architecture integrates a ResNet-based branch for audio data with an attention module and a 3D ResNet branch for video data, and was evaluated on the Multimodal Dysarthria Severity Assessment Challenge (MDSA) dataset, achieving significant performance improvements and ranking first in the challenge with a best score of 8.7404. The study's findings highlight the efficacy of the proposed framework in capturing essential features from multimodal data, contributing to more accurate and reliable assessments of dysarthria.

**Spectral Analysis of Vowels and Fricatives at Varied Levels of Dysarthria Severity for Amyotrophic Lateral Sclerosis,**

Dysarthria due to Amyotrophic Lateral Sclerosis (ALS) affects the acoustic characteristics of different speech sounds. The effects intensify with increasing severity leading to the collapse of the acoustic space of the affected individuals. With an aim to characterize such changes in the acoustic space, this paper studies the variations in band-specific and full-band spectral properties of 4 sustained vowels (/a/, /i/, /o/, /u/) and 3 sustained fricatives (/s/, /sh/, /f/) at different dysarthria severity levels. Effect of dysarthria on spectral features of these phonemes are not well explored. Statistical comparison of these features among different severities for the phonemes considered and among different vowels/fricatives for every severity level using speech data from 119 ALS and 40 healthy subjects indicate the followings. Though all band-specific and full-band features of the three fricatives and most of those features for the four vowels become statistically similar at high severity levels, certain features remain distinguishable. Spectral differences in 0-2 kHz band between /a/ and the other vowels and in the 2-6 kHz band between /a/ and /o/, /u/ persist through all severity levels. Moreover, properties of /f/ remain mostly unchanged with increasing dysarthria severity levels.

**Advancing Speech Disorder Diagnostics: A Comprehensive Study on Dysarthria Classification with CNN,**

This study article focuses on dysarthria, a speech problem that is often seen in patients with cerebral palsy (CP) or amyotrophic lateral sclerosis (ALS). Early identification of dysarthria is crucial for successful treatments and better patient outcomes, since it has a substantial influence on communication ability. Using the TORGO database, which consists of 2000 samples of persons with and without dysarthria, spanning different genders, the research use a Convolutional Neural Network (CNN) to classify speech. The dataset includes both dysarthric and non-dysarthric individuals of both genders, with 500 samples each, captured during separate sessions. The CNN-based technique demonstrates an impressive 96% accuracy in differentiating between dysarthric and non-dysarthric instances, highlighting the promise of deep learning technology in assisting with the early detection of dysarthria and enabling prompt therapies for afflicted people. This study makes significant contributions to the advancement of diagnostic capacities and presents a potential opportunity to improve the quality of life for those with speech difficulties.

**Progressing Speech Disorder Identification: A Thorough Investigation into Dysarthria Categorization Utilizing ResNet50 CNN,**

The speech issue of dysarthria, which is frequently observed in individuals with cerebral palsy (CP) or amyotrophic lateral sclerosis (ALS), is the subject of this research article. Timely detection of dysarthria is essential for effective interventions and improved patient results, as it significantly impacts communication proficiency. The study employs the TORGO database, which has 2000 samples of individuals with and without dysarthria, encompassing various genders. The researchers utilize a ResNet50 Convolutional Neural Network (CNN) to categorize speech. The dataset comprises individuals of both genders, encompassing both dysarthric and non-dysarthric individuals. Each group consists of 500 samples, which were recorded over distinct sessions. The CNN-based technique achieves a remarkable 96% accuracy in distinguishing between dysarthric and non-dysarthric instances. This underscores the potential of deep learning technology in aiding the early identification of dysarthria and facilitating timely treatments for affected individuals. This study provides valuable contributions to the enhancement of diagnostic capabilities and offers a potential avenue for enhancing the quality of life for individuals experiencing speech impairments.

**Enhancing the Diagnosis of Speech Disorders: An In-Depth Investigation into Dysarthria Classification Using the ResNet18 Model,**

This research article concentrates on dysarthria, a speech impediment commonly observed in individuals with conditions such as cerebral palsy (CP) or amyotrophic lateral sclerosis (ALS). Early detection of dysarthria is essential for effective interventions and improved patient outcomes, as it significantly impacts communication abilities. Employing the TORGO database, comprising 2000 samples from individuals with and without dysarthria, encompassing diverse genders, the study utilizes a Convolutional Neural Network (ResNet18 Model) for speech classification. The dataset encompasses both dysarthric and non-dysarthric subjects of both genders, with 500 samples each, recorded in separate sessions. The ResNet18 Model-based approach demonstrates an impressive 96% accuracy in distinguishing between dysarthric and non-dysarthric instances, underscoring the potential of deep learning technology in aiding early dysarthria detection and facilitating timely interventions. This investigation significantly contributes to advancing diagnostic capabilities and presents a promising avenue to enhance the quality of life for individuals grappling with speech difficulties.

**Statistical Analysis of Speech Disorder Specific Features to Characterise Dysarthria Severity Level,**

Poor coordination of the speech production subsystems due to any neurological injury or a neuro-degenerative disease leads to dysarthria, a neuro-motor speech disorder. Dysarthric speech impairments can be mapped to the deficits caused in phonation, articulation, prosody, and glottal functioning. With the aim of reducing the subjectivity in clinical evaluations, many automated systems are proposed in the literature to assess the dysarthria severity level using these features. This work aims to analyse the suitability of these features in determining the severity level. A detailed investigation is done to rank these features for their efficacy in modelling the pathological aspects of dysarthric speech, using the technique of paraconsistent feature engineering. The study used two dysarthric speech databases, UA-Speech and TORGO. It puts light into the fact that both the prosody and articulation features are best useful for dysarthria severity estimation, which was supported by the classification accuracies obtained on using different machine learning classifiers.

**Identification of Cerebellar Dysarthria with SISO Characterisation,**

Quantitative identification of dysarthria plays a major role in the classification of its severity. This paper quantitatively analyses several components of cerebellar dysarthria. The methodology described in this study will be extended to other types of dysarthria via systematic analysis. The speech production model is characterized as a second-order singleinput and single-output (SISO), linear, time-invariant (LTI) system in our study. A comparative study on the behavior of the damping ratio and resonant frequency for dysarthric and non-dysarthric subjects is presented. The results are further analyzed using the Principal component analysis (PCA) technique to emphasize the variation and uncover strong patterns in the selected features. The effects of some other related factors like decay time and Q-factor are also highlighted.

**CNN-Driven Innovations in Dysarthria Classification and Speech Disorder Management,**

This research article addresses dysarthria, a speech disorder frequently observed in individuals with cerebral palsy (CP) or amyotrophic lateral sclerosis (ALS). Early detection of dysarthria is essential for effective treatment and improved patient outcomes, as it significantly impacts communication abilities. Utilizing the TORGO database, which comprises 2000 speech samples from both dysarthric and non-dysarthric individuals across various genders, this study employs a Convolutional Neural Network (CNN) for classification. The dataset includes 500 samples each from both groups, collected in different sessions. The CNN approach achieves an impressive 96% accuracy in distinguishing dysarthric from non-dysarthric speech, showcasing the potential of deep learning technology in facilitating early dysarthria diagnosis and enabling timely interventions. This study makes substantial contributions to enhancing diagnostic capabilities and offers a promising avenue for improving the quality of life for individuals with speech disorders.

**Constant Q Transform for Audio-Visual Dysarthria Severity Assessment,**

This paper describes an automatic dysarthria severity assessment system submitted for the MDSA2024 challenge. The system utilizes audio and visual features to classify the severity of dysarthria. We investigate the effectiveness of different acoustic feature combinations, such as Mel spectrogram, filterbank and constant Q transform. Different encoders for audio data processing are compared in the experiments. The results demonstrate that the Conformer encoder achieves superior performance compared to the Densenet encoder on pure audio features. We further explore model-level fusion by combining audio and visual embeddings, leading to improvements in precision’ recall, and accuracy metrics. The best-achieved score on the test set is 7.5579, with an individual F1-score of 0.7028. The findings suggest that the proposed system with Conformer encoder and model-level fusion holds promise for automatic dysarthria assessment.

**Dysarthria Detection with Deep Representation Learning for Patients with Parkinson’s Disease,**

Dysarthria is a very common motor speech symptom in Parkinson’s disease impairing normal communications of patients. Detection of dysarthria could assist clinical diagnosis and intervention of Parkinson’s disease, provide monitoring approach for treatment-related side effects, and lead to effective speech therapy to prevent further communication and social deficits. Applying machine learning techniques to speech analysis for patients offers more resource-efficient, accessible and objective tools for screening and assessment of dysarthria. In this study, we constructed a multi-task speech dataset recorded from 600 participants with high data diversity to facilitate the development of detection models. We established a remote data acquisition and end-to-end prediction pipeline with deep representation learning, compared the performance with different feature-based learning and classification methods, and achieved a superior accuracy of over 90%. Different affecting factors were analyzed for model performance. Our proposed framework demonstrates the potential of developing and deploying an automated self-monitoring approach of dysarthria for patients with Parkinson’s disease, which could benefit a large-scale population and their disease managements.

**Automatic assessment of dysarthria severity level using audio descriptors,**

Dysarthria is a motor speech impairment, often characterized by speech that is generally indiscernible by human listeners. Assessment of the severity level of dysarthria provides an understanding of the patient's progression in the underlying cause and is essential for planning therapy, as well as improving automatic dysarthric speech recognition. In this paper, we propose a non-linguistic manner of automatic assessment of severity levels using audio descriptors or a set of features traditionally used to define timbre of musical instruments and have been modified to suit this purpose. Multitapered spectral estimation based features were computed and used for classification, in addition to the audio descriptors for timbre. An Artificial Neural Network (ANN) was trained to classify speech into various severity levels within Universal Access dysarthric speech corpus and the TORGO database. An average classification accuracy of 96.44% and 98.7% was obtained for UA speech corpus and TORGO database respectively.

**Automated Dysarthria Severity Classification Using Deep Learning Frameworks,**

Dysarthria is a neuro-motor speech disorder that renders speech unintelligible, in proportional to its severity. Assessing the severity level of dysarthria, apart from being a diagnostic step to evaluate the patient's improvement, is also capable of aiding automatic dysarthric speech recognition systems. In this paper, a detailed study on dysarthia severity classification using various deep learning architectural choices, namely deep neural network (DNN), convolutional neural network (CNN) and long short-term memory network (LSTM) is carried out. Mel frequency cepstral coefficients (MFCCs) and its derivatives are used as features. Performance of these models are compared with a baseline support vector machine (SVM) classifier using the UA-Speech corpus and the TORGO database. The highest classification accuracy of 96.18% and 93.24% are reported for TORGO and UA-Speech respectively. Detailed analysis on performance of these models shows that a proper choice of a deep learning architecture can ensure better performance than the conventionally used SVM classifier.

**Deep Learning Based Speech Recognition for Hyperkinetic Dysarthria Disorder,**

Speech recognition is a technology that aims to transform human speech into text and has applications in a variety of fields, including information technology, healthcare, automobiles, and others. This research explores the advantages of employing deep learning methods to provide individualized assistance technology solutions for people with dysarthria. We proposed a Convolutional Temporal Bidirectional Network (CTBNet) model for translating Russian Hyperkinetic Dysarthria Disorder (RHDD) speech into text. CTBNet encodes input audio features using 1D convolution layers, and BidirectionalGRU (Bi-GRU) layers. The encoder effectively captures both short-term and long-term spatial temporal information. More importantly, CTBNet includes with an attention-CTC decoder, that is responsible for producing output texts. We utilized a dataset of 2797 recordings of RHDD speech, with a cumulative recorded duration of 2 hours and 33 minutes, for training purposes. The CTBNet model has achieved a 15% Character Error Rate (CER) and a 59% Word Error Rate (WER) on the dataset proposed. Our experimental findings show superior performance compared to state-of-the-art models, namely, 3xCNN, and 3xLSTM, when evaluated on the same dataset.

**Developing a Hyperkinetic Dysarthria Speech Classification System using Residual Learning,**

Hyperkinetic dysarthria abnormalities are exceedingly widespread across the worldwide populace. One of the most effective methodologies for disordered audio speech identification is classical deep learning technique. However, the efficiency of classification is impacted by the limitations imposed by the quality of the training dataset, such as expense and limited resources, data imbalance, and data annotation difficulties. Accordingly, we suggest implementing a residual learning framework (ResN et architecture) with various conditions and hyperparameters (batch size and learning rate) that improve the training process of deeper networks compared to earlier approaches. Pre-emphasis, windowing, and lifting techniques are applied to improve the quality of Mel-Frequency Cepstral Coefficients. In addition, we have introduced the hyperkinetic dysarthria speech dataset in the Russian language, consisting of 4 hours and 4 minutes of recorded speech. The empirical findings demonstrate that ResNet40 has an overall validation accuracy of 86.7%, surpassing other research models. It is anticipated that it will be recognized as an alternative diagnostic instrument for physicians in the future as research continues to progress.

**Multilingual Speaker-Invariant Dysarthria Severity Assessment Using Adversarial Domain Adaptation and Self-Supervised Learning,**

Traditional assessments for dysarthria are subjective and time-consuming, highlighting the need for automated, objective approaches that can be scaled for remote and resource-constrained environments. This paper introduces an adversarial domain adaptation framework tailored for dysarthria severity assessment, addressing the challenges of high intra-class variability and limited availability of dysarthric speech data. By framing speaker variability as a domain adaptation problem, we utilise an adversarially trained feature extractor to derive speaker-invariant yet discriminatively powerful representations utilising speech features learned through self-supervised learning. Experiments on previously unseen and diverse speakers reveal that the proposed approach yields a 7.86% average improvement across multilingual datasets compared to traditional severity-discriminative training, outperforming competitive baselines by 12.33% on average. Additionally, the method inherently supports privacy-preserving applications by minimising reliance on speaker-specific information. The results demonstrate strong alignment with clinical assessments, reinforcing our model’s clinical relevance and effectiveness.

**AI-Enabled Speech Monitoring for Dysarthria Detection in Consumer Electronics,**

Speech is essential in human communication, and recent advancements in artificial intelligence (AI) have enabled new applications for voice data, particularly in health monitoring. Building on this progress, a home-based health detection system that utilizes voice signals provides an accessible method for users to independently identify and respond to conditions such as dysarthria, without the need for hospital or clinic visits. In this study, we developed and evaluated convolutional neural network (CNN) models to classify audio data from both healthy individuals and individuals with dysarthria. The dataset comprised command-based speech samples, which were segmented into sentence-specific input. Preprocessing steps included segmentation, downsampling, and feature extraction tailored for CNN input. The models exhibited high potential in distinguishing between normal and dysarthric states, with the most effective model achieving an f1-score of 96.18±4.7% and an accuracy of 96.01±5.02%. Furthermore, sentence structure and composition significantly influenced model performance, with specific sentences demonstrating higher classification accuracy. These findings suggest that AI-driven speech analysis may support detecting and managing specific health conditions in non-clinical environments. However, further work is needed to address data imbalance and sentence variability limitations. Future research could expand on real-world applications, particularly in settings with limited traditional health monitoring.

**Dysarthria Severity Classification Using Phase Based Features of LP Residual,**

Classifying the severity of speech impairment due to dysarthria is crucial for optimizing care and enhancing communication abilities for affected individuals. This study explores the use of the Modified Group Delay Function (MGDF) of LP residual signal in classifying dysarthria severity-levels. Evaluations were conducted using standard UA-Speech and TORGO datasets. A stratified Convolutional Neural Network (CNN) with 5-fold cross-validation validated the results. Baseline features included Linear Frequency Cepstral Coefficients (LFCC), Mel Frequency Cepstral Coefficients (MFCC), and Whisper module. Key performance evaluation metrics were accuracy, precision, recall, and F1-score. Finally, the latency period was analyzed for practical deployment of the system, system’s ability to accurately recognize and process speech from any speaker, without needing to be specifically trained or adapted to individual voice characteristics.

**Automatic Early Detection of Dysarthria using Deep Neural Network,**

Dysarthria is dis-coordination among articulatory muscles responsible for articulation in the production of speech. A person with dysarthric disorder cannot control their tongue, hence, are prone to slurred speech or swallowing words which lead to improper pronunciation and less intelligibility of speech. This work focuses on the automatic early detection of Dysarthria. To elucidate the main idea, it must be understood that the whole detection procedure revolves around the speech intelligibility of a person. It becomes an arduous job to distinguish between a healthy person's speech and a person who is slowly evolving with the disorder. The prime motive of the work is to provide a proper classification of the disorder in its earliest stages. The common words and phrases recorded in the Universal Access database are used for this work. Mel Frequency Cepstral Coefficients (MFCC) are extracted from both controlled and high intelligible speech data, and fed to a Deep Neural Network (DNN) classifier. MFCC fed to the DNN classifier has provided the best classification accuracy for the above experimental setup.

**Learning to Detect Dysarthria from Raw Speech,**

Speech classifiers of paralinguistic traits traditionally learn from diverse hand-crafted low-level features, by selecting the relevant information for the task at hand. We explore an alternative to this selection, by learning jointly the classifier, and the feature extraction. Recent work on speech recognition has shown improved performance over speech features by learning from the waveform. We extend this approach to paralinguistic classification and propose a neural network that can learn a filterbank, a normalization factor and a compression power from the raw speech, jointly with the rest of the architecture. We apply this model to dysarthria detection from sentence-level audio recordings. Starting from a strong attention-based baseline on which mel-filterbanks outperform standard low-level descriptors, we show that learning the filters or the normalization and compression improves over fixed features by 10% absolute accuracy. We also observe a gain over OpenSmile features by learning jointly the feature extraction, the normalization, and the compression factor with the architecture. This constitutes a first attempt at learning jointly all these operations from raw audio for a speech classification task.

**Modeling fundamental frequency dynamics in hypokinetic dysarthria,**

Hypokinetic dysarthria (Hd), which often accompanies Parkinson's Disease (PD), is characterized by hypernasality and by compromised phonation, prosody, and articulation. This paper proposes automated methods for detection of Hd. Whereas most such studies focus on measures of phonation, this paper focuses on prosody, specifically on fundamental frequency (F0) dynamics. Prosody in Hd is clinically described as involving monopitch, which has been confirmed in numerous studies reporting reduced within-utterance pitch variability. We show that a new measure of F0 dynamics, based on a superpositional pitch model that decomposes the F0 contour into a declining phrase curve and (generally, single-peaked) accent curves, performs more accurate Hd vs. Control classification than simpler versions of the model or than conventional variability statistics.

**Automatic detection of speech disorder in dysarthria using extended speech feature extraction and neural networks classification,**

This paper presents an automatic detection of Dysarthria, a motor speech disorder, using extended speech features called Centroid Formants. Centroid Formants are the weighted averages of the formants extracted from a speech signal. This involves extraction of the first four formants of a speech signal and averaging their weighted values. The weights are determined by the peak energies of the bands of frequency resonance, formants. The resulting weighted averages are called the Centroid Formants. In our proposed methodology, these centroid formants are used to automatically detect Dysarthric speech using neural network classification technique. The experimental results recorded after testing this algorithm are presented. The experimental data consists of 200 speech samples from 10 Dysarthric speakers and 200 speech samples from 10 age-matched healthy speakers. The experimental results show a high performance using neural networks classification. A possible future research related to this work is the use of these extended features in speaker identification and recognition of disordered speech.

**Analysis of Time Domain Features of Dysarthria Speech,**

In general, Speech can be described as the ability to express thoughts and feelings by articulating sounds in order to communicate with the person who understand the speech and accordingly interpret their action. The types of communication disorders that affects a person’s ability to produce speech sounds due to the lack of control over motor functions and articulators are speech disorders. Many people of varying age groups suffer from speech disorders and require early treatment in order to correct them. This paper analyses the time domain features such as jitter and shimmer that are extracted from the speech samples of dysarthria and healthy people. This is done by looking for dissimilarities between the different speech features preset in the normal healthy person and the disordered one by employing steps such as pre-processing followed by feature extraction.

**Emotional Communication Assist Interface App for People with Dysarthria,**

Text to speech (TTS) helps people who have speech disorders communicate with other people. The current TTS can express emotion, but this requires extra time and effort. We developed a TTS application by which users can render messages with emotion with ease by changing the pitch angle of a smartphone. We designed two types of TTS: free input and phrase selection. Also, we calculated the TTS editing parameters for each degree of emotion on the basis of the Online Game Voice Chat Corpus (OGVC).

**Fuzzy Expectation Maximization Phoneme Prediction in Diffusion Model-based Dysarthria Voice Conversion,**

In this paper proposed an effective fuzzy expectation-maximization phoneme prediction method in diffusion model-based dysarthria voice conversion (FEMPPDM-DVC) which is accessible to (i) training without parallel data (ii) converting a longer duration of the audio data (iii) preserves speaker identity. By integrating Fuzzy Expectation-Maximization (FEM) clustering and Diffusion model-based dysarthria voice conversion approach, the proposed method is able gradually generate normal utterances. Feature extraction is performed using Mel-frequency cepstral coefficients (MFCCs), a robust method in acoustic feature extraction in Text-to-Speech systems. Ensures the effectiveness and accuracy, through repeated parameter adjustment using the FEM clustering algorithm. The conversion network is a diffusion model-based structure, which can convert normal utterances to dysarthria utterances in the forward diffusion process. After forward diffusion process, the dysarthria utterance will go through a reverse diffusion process to convert dysarthria voice to normal utterance. Once converted, a GAN-based vocoder is applied to convert mel-frequency spectrogram to waveform. Objective evaluation is conducted on the Saarbrücken Voice Database (SVD) dataset show that FEMDDPM-DVC is able to improve the intelligibility and naturalness of dysarthria utterances in 15 kinds of dysarthria voice. The proposed method is compared with five other dysarthria voice conversion methods, show that the proposed method has the most performance among other method.

**Revolutionary Dysarthria Analysis Through Convolutional Neural Networks,**

Dysarthria occurs when there are impairments in the muscles necessary for speech, including those in the lips, tongue, voice box, and breathing muscles. This condition can result in unclear speech, even in individuals who are skilled communicators. Dysarthria can range from mild speech difficulties, where speech is hard to understand, to severe cases, where speech is completely unclear. A method for distinguishing between different types of speech involves using recordings and advanced computational techniques such as convolutional neural networks (CNNs). The strengths of CNNs are leveraged to construct a highly effective model for detecting dysarthria in speech samples. The dataset utilized includes recordings from individuals with and without dysarthria. The results indicate that the CNN model performs exceptionally well, achieving precision, recall, and F1-scores of 0.99 for both categories. With an accuracy of 0.99, the model demonstrates a high level of proficiency in differentiating between speech affected by dysarthria and normal speech. These findings highlight the effectiveness and utility of this CNN-based approach in diagnosing and assessing dysarthria. It offers valuable potential for assisting healthcare professionals and enhancing personalized speech therapy interventions.

**A Study of Speech Recognition Techniques for Dysarthria Speeches Based on Digit Recognition,**

With the introduction of speech recognition technology into people's lives, the application of which is no longer limited to ordinary people, but begins to target special populations, such as patients with dysarthria. It has become a major research challenge to build a high-performance speech recognition system for patients with dysarthria. To address this problem, in this paper the speech data of four patients with dysarthria and four healthy speakers on digital commands are firstly collected, secondly several speech recognition experiments are separately conducted by using Aishell-2 and Wenetspeech, two pre-trained models based on the WeNet open-source architecture, as well as the four major commercial recognition API systems, at the end a method to optimize the dysarthria data by using the so-vits-svc tool is proposed. The experiments show that the current state-of-the-art speech recognition model has a high recognition rate for speech data from the general population, but there is still room for improvement in the recognition performance of speech data from individuals with dysarthria. Later the optimization method for the dysarthria data proposed in this paper effectively improves the quality of speech data. For the two above open-source pre-trained models, the recognition rates have increased by 60% and 67.64 % compared to the original data, respectively.

**Fractal features for automatic detection of dysarthria,**

Amytrophic lateral sclerosis (ALS) is an incurable neurodegenerative disease. Difficulty articulating speech, dysarthria, is a common early symptom of ALS. Detecting dysarthria currently requires manual analysis of several different speech tasks by pathology experts. This is time consuming and can lead to misdiagnosis. Many existing automatic classification approaches require manually preprocessing recordings, separating individual spoken utterances from a repetitive task. In this paper, we propose a fully automated approach which does not rely on manual preprocessing. The proposed method uses novel features based on fractal analysis. Acoustic and associated articulatory recordings of a standard speech diagnostic task, the diadochokinetic test (DDK), are used for classification. This study's experiments show that this approach attains 90.2% accuracy with 94.2% sensitivity and 85.1% specificity.

**Assessing Dysarthria severity using global statistics and boosting,**

A new method for automatic assessment of Dysarthria severity is described. It uses the forward selection method (FSM) on global statistics of low-complexity features to find effective feature sets. FSM is embedded in a boosting algorithm that combines multiple weak classifiers to achieve a single strong classifier. Unlike standard boosting, this uses nonlinear class boundaries and unique feature sets per iteration. Results on a 39 speaker dysarthria database are described.

**An Intelligent System for Dysarthria Classification of Male and Female Processed Dataset using Sequential Model Parameters,**

There are several current studies including various methods for treating dysarthria in various developing countries. Research-based dysarthria treatments take a multidisciplinary approach that encompasses speech therapy, assistive technology, neurosurgery, pharmaceutical therapies, non-invasive brain stimulation, machine learning and AI methods. These methods seek to raise the general quality of life, communication and speaking abilities of dysarthria sufferers. With a 90% accuracy rate, the proposed sequential model was found to have good classification performance for identifying Dysarthria. Researchers have also been looking at the use of Machine Learning (ML) and Artificial Intelligence (AI) approaches to create tools and systems that can help people with dysarthria to speak and communicate for social welfare and provide access to best remedies.

**Dysarthria diagnosis via respiration and phonation,**

This report discusses the implementation of a computerized application - the Computerised Frenchay Dysarthria Assessment Procedure (CFDA) - which uses digital signal processing (DSP) techniques to objectively evaluate digitised speech recordings in order to detect any symptoms of dysarthria (a type of motor speech disorder). This investigation focuses specifically on two CFDA diagnostic sub-applications which assess a patient's ability to maintain a prolonged exhalation (the “Respiration at Rest” task) as well as execute a steady state phonation (i.e. the “Sustained Phonation” task). It is demonstrated that combining the functionality of these two sub-applications substantially increases their diagnostic effectiveness in the context of identifying a particular variety of dysarthria known as spastic dysarthria. It is further demonstrated that certain patterns of energy fluctuations are closely correlated with manifestations of spastic dysarthria. The CFDA is intended for use in real-world conditions to assist speech and language therapists when conducting clinical evaluations.

**Deep Learning Based Dysarthria Detection: A Comprehensive Approach,**

Making a model that can analyze numerous input features and predict whether a person will develop dysarthria is the first step in utilizing machine learning to forecast diseases like dysarthria. A motor speech disorder called dysarthria can be caused by a variety of underlying conditions, such as degenerative disorders, traumatic brain injuries, or neurological disorders. The authors used the 2000 audio signals from males and females with and without dysarthria from the TORGO data set. To extract the important features from the audio signals, the authors used the MFCC approach. To determine if dysarthria is present or absent, a machine learning model or algorithm will be fed with these MFCC coefficients.

**Brain Signals to Rescue Aphasia, Apraxia and Dysarthria Speech Recognition,**

In this paper, we propose a deep learning-based algorithm to improve the performance of automatic speech recognition (ASR) systems for aphasia, apraxia, and dysarthria speech by utilizing electroencephalography (EEG) features recorded synchronously with aphasia, apraxia, and dysarthria speech. We demonstrate a significant decoding performance improvement by more than 50% during test time for isolated speech recognition task and we also provide preliminary results indicating performance improvement for the more challenging continuous speech recognition task by utilizing EEG features. The results presented in this paper show the first step towards demonstrating the possibility of utilizing non-invasive neural signals to design a real-time robust speech prosthetic for stroke survivors recovering from aphasia, apraxia, and dysarthria. Our aphasia, apraxia, and dysarthria speech-EEG data set will be released to the public to help further advance this interesting and crucial research.

**Hmm-Based and Svm-Based Recognition of the Speech of Talkers With Spastic Dysarthria,**

This paper studies the speech of three talkers with spastic dysarthria caused by cerebral palsy. All three subjects share the symptom of low intelligibility, but causes differ. First, all subjects tend to reduce or delete word-initial consonants; one subject deletes all consonants. Second, one subject exhibits a painstaking stutter. Two algorithms were used to develop automatic isolated digit recognition systems for these subjects. HMM-based recognition was successful for two subjects, but failed for the subject who deletes all consonants. Conversely, digit recognition experiments assuming a fixed word length (using SVMs) were successful for two subjects, but failed for the subject with the stutter.

**Design of a dysarthria classifier using global statistics of speech features,**

Dysarthria is a neurological disorder in which the speech production system is impaired. There are five main types of dysarthrias depending on the location of the lesion in the nervous system. There is evidence suggesting a relationship between the location of the lesion and the resulting speech characteristics. This paper describes a non-intrusive classifier to identify the dysarthria type in a person using global statistics, e.g., mean, variance, etc., of speech features. A tree-based classifier was developed using multiple low-level maximum likelihood classifiers as inputs. An error of 10.5% was achieved in the classification of three types of dysarthrias.

**Dysarthria Speech Disorder Classification Using Traditional and Deep Learning Models,**

Dysarthria is a motor speech disorder that results in speech difficulties due to the weakness of associated muscles. This unclear speech makes it difficult for dysarthric patients to present himself understood. This neurological limitation is usually occurs due to damages to the brain or central nervous system. Speech therapy can be effectively employed to enhance the range and consistency of voice production and improve intelligibility and communicative effectiveness. Assessing the degree of severity of dysarthria provides vital information on the patient's progress which inturn assists pathologists in arriving at a treatment plan that includes developing automated voice recognition system suitable for dysarthria patients. This work performs an exhaustive study on dysarthria severity level classification using deep neural network (DNN) and convolution neural network (CNN) architectures. Mel Frequency Cepstral Coefficients (MFCCs) and their derivatives constitute feature vectors for classification. Using the UA-Speech database, the performance metrics of DNN/CNN based learning models have been compared to baseline classifiers like support vector machine (SVM) and Random Forest (RF). The highest classification accuracy of 97.6\% is reported for DNN under UA speech database. A detailed examination of the performance from the models discussed above reveal that appropriate choice of deep learning architecture ensures better results than traditional classifiers like SVM and Random Forest.

**Harnessing Deep Learning Techniques for Dysarthria Detection,**

Dysarthria, a motor speech disorder resulting from neurological impairments. This study explores various approaches for the automated detection of dysarthria, integrating both traditional and emerging technologies. Speech assessments by acoustic analysis, machine learning models, and deep learning techniques are considered. Machine learning models are trained on datasets containing normal and dysarthric speech for males and females, while deep learning methods, including convolutional and recurrent neural networks, are employed to automatically learn features from speech data. Our goal is to develop a reliable and accessible dysarthria detection system thatcomplements the expertise of healthcare professionals. Validation using diverse datasets is emphasized, acknowledging the importance of early detection and intervention in improving outcomes for individuals with dysarthria.

**Next-Gen Speech Disorder Diagnostics: CNN Methods for Dysarthria Classification,**

Neurological disorders such as amyotrophic lateral sclerosis (ALS) and cerebral palsy (CP) are common causes of dysarthria, the subject of this research article. Dysarthria has a significant impact on communication capacity, making early diagnosis essential for effective therapies and improved patient outcomes. The research used a Convolutional Neural Network (CNN) to categorise speech using the TORGO database, which includes 2000 samples of individuals with and without dysarthria, representing different genders. There are 500 samples from each gender in the collection, representing dysarthric and non-dysarthric people who were recorded during different sessions. Highlighting the promise of deep learning technology in aiding early identification of dysarthria and enabling rapid therapy for affected individuals, the CNN-based approach exhibits an amazing 96% accuracy in discriminating between dysarthric and non-dysarthric cases. This study offers a promising chance to enhance the quality of life for those with speech impairments while also making substantial contributions to the development of diagnostic capabilities.

**Advancing Speech Disorder Diagnostics: CNN Innovations in Dysarthria Recognition,**

Neurological disorders such as amyotrophic lateral sclerosis (ALS) and cerebral palsy (CP) are common causes of dysarthria, the subject of this research article. The capacity to communicate is greatly affected by dysarthria, thus early diagnosis is critical for effective therapies and improved patient outcomes. The study employed a Convolutional Neural Network (CNN) for speech classification using the TORGO database, which includes 2000 samples of individuals with and without dysarthria, representing different genders. With 500 samples each, collected over distinct sessions, the dataset contains both dysarthric and non-dysarthric people of both sexes. Distinguishing between dysarthric and non-dysarthric occurrences, the CNN-based approach achieves an astonishing 96% accuracy. This shows how deep learning technology may help with early dysarthria identification and enable rapid therapy for affected individuals. The findings of this study have important implications for the development of diagnostic tools and may lead to better living conditions for those who struggle with speaking.

**A preliminary study on self-care telemonitoring of dysarthria in spinal muscular atrophy,**

Spinal muscular atrophy (SMA) is a rare neuromuscular disease which may cause impairments in oro-facial musculature. Most of the individuals with SMA present bulbar signs such as flaccid dysarthria which mines their abilities to speak and, as consequence, their psychic balance. To support clinicians, recent work has demonstrated the feasibility of video-based techniques for assessing the oro-facial functions in patients with neurological disorders such as amyotrophic lateral sclerosis. However, no work has so far focused on automatic and quantitative monitoring of dysarthria in SMA. To overcome limitations this work’s aim is to propose a cloud-based store-and-forward telemonitoring system for automatic and quantitative evaluation of oro-facial muscles in individuals with SMA. The system integrates a convolutional neural network (CNN) aimed at identifying the position of facial landmarks from video recordings acquired via a web application by an SMA patient.Clinical relevance— The proposed work is in the preliminary stage, but it represents the first step towards a better understanding of the bulbar-functions’ evolution in patients with SMA.

**Automatic And Perceptual Discrimination Between Dysarthria, Apraxia of Speech, and Neurotypical Speech,**

Automatic techniques in the context of motor speech disorders (MSDs) are typically two-class techniques aiming to discriminate between dysarthria and neurotypical speech or between dysarthria and apraxia of speech (AoS). Further, although such techniques are proposed to support the perceptual assessment of clinicians, the automatic and perceptual classification accuracy has never been compared. In this paper, we investigate a three-class automatic technique and a set of handcrafted features for the discrimination of dysarthria, AoS and neurotypical speech. Instead of following the commonly used One-versus-One or One-versus-Rest approaches for multi-class classification, a hierarchical approach is proposed. Further, a perceptual study is conducted where speech and language pathologists are asked to listen to recordings of dysarthria, AoS, and neurotypical speech and decide which class the recordings belong to. The proposed automatic technique is evaluated on the same recordings and the automatic and perceptual classification performance are compared. The presented results show that the hierarchical classification approach yields a higher classification accuracy than baseline One-versus-One and One-versus-Rest approaches. Further, the presented results show that the automatic approach yields a higher classification accuracy than the perceptual assessment of speech and language pathologists, demonstrating the potential advantages of integrating automatic tools in clinical practice.

**Adversarial Auto-Encoders Based Model for Classification of Speech Dysarthria,**

Communication is effective based on various parameters, out of which phonetic or oral communication plays a vital role. Slurred speech or improper speech will lead to misunderstanding in speech, which could toss up any situation. There are many people, ranging from children to adults, who are affected with slurred speech, which is technically termed as Speech Dysarthria, a disease which tampers effective oral communication. Distinguishing between people affected with dysarthria and people with normal speech will be tedious process manually. Machine Learning (ML), and Artificial Intelligence (AI), can be pitched in to solve the problem. There are existing methodologies which classify people affected with speech dysarthria and people who communicate in a normal way. Some of the existing technologies used are Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and so on. This paper aims at distinguishing between people affected with speech dysarthria and people with normal speech, using Adversarial Auto Encoders (AAE), a model which has its roots from Variational Auto Encoders (VAE) and Generative Adversarial Networks (GAN). This paper brings out a good result and proves to be effective.

**Summary of Low-Resource Dysarthria Wake-Up Word Spotting Challenge,**

In recent years, the rapid advancement and widespread adoption of speech technology have made smart home systems a common feature in many households. However, individuals with dysarthria face difficulties using these technologies due to inconsistent speech patterns. This paper summarizes the Low-Resource Dysarthria Wake-Up Word Spotting (LRDWWS) Challenge at SLT 2024, which aimed to develop effective voice wake-up systems for individuals with dysarthria. The challenge attracted 25 teams from 4 countries, with 7 teams submitting results and 5 providing detailed system descriptions. This paper presents an overview of the dataset, evaluation metrics, and key innovations from participating teams. Our findings highlight the potential of these systems to enhance the accessibility and usability of smart home technologies for individuals with dysarthria. The challenge results underscore the importance of developing specialized solutions to meet the unique needs of this user group.

**Transformer-based Transfer Learning for Enhanced Speech Dysarthria Severity Assessment,**

Dysarthria, a neuromuscular communications disorder presented with impaired pronunciation, is a daunting task to identify and quantify the level of dysfunction. Through this paper, a critical study of automated dysarthria severity classification using transformer-based deep learning methods will be discussed. Transfer learning is facilitated by employing three variations of Vision Transformer (ViT) models: ViT-L-16, ViT-L-32, and ViT-B-16, on UA-Speech Corpus and TORGO datasets, achieving remarkable results with 99.01% accuracy for the UA-Speech dataset and 99.39% accuracy for the TORGO database. The study evaluates the models’ performance focusing on accuracy, precision, recall, and F1-scores. Experimental results emphasise the potential for automated diagnostic units in neurology clinical practice and establish a baseline for leading work in dysarthria severity classification with the selection of an efficient ViT model.

**Enhancing speech rate estimation techniques to improve dysarthria diagnosis,**

This report discusses the implementation of a computerized algorithm specifically designed to measure the syllables-per-minute rate of abnormal speech typically produced by persons suffering from an articulatory disorder known as dysarthria. This speech rate measurement application - which can also serve as a diagnostic tool in itself - has been integrated into the computerised Frenchay Dysarthria Assessment (CFDA) suite of diagnostic tests. It is demonstrated that, when processing dysarthric speech, syllables-per-minute measurements are more accurate when based on vowel transition detection techniques as opposed to using spectral moment computations.

**Personalizing TTS Voices for Progressive Dysarthria,**

Amyotrophic lateral sclerosis (ALS) patients experience progressive speech deterioration due to muscle paralysis, leading to eventual loss of verbal communication capability. Text-to-speech synthesis (TTS) is an important technology for speech generating devices, enabling users to communicate using generic electronic voices, but often without the vocal identity of the users. Our work is aimed at personalizing TTS voices for people with ALS induced dysarthria by integrating machine learning and speech processing techniques of voice conversion (VC) and TTS. This is challenging as only small quantities of dysarthric speech are available from individual patients. Our system includes both timbre and prosody conversion for VC, neural TTS to generate TTS speech, and neural feature converter to interface VC and TTS. We collected speech data from 4 ALS target speakers with mild to severe dysarthria. Subjective listening tests showed that on average, our approach improved speech intelligibility by about 72% over the target speakers’ speech, the converted voice was 2 to 3 times more similar to ALS targets than to TTS sources, and the converted speech quality was in the MOS scale of fair to good.

**A Phonological Control Method on A Speech Compensation System for Dysarthria Using A Standardized Space,**

We have developed a speech compensation system for dysarthria. The system aims at improving the phonological properties of vowels without losing speaker individuality. We propose a method for phonological control of vowels using a standardized space to control vowels in the normalized articulation space, normalized for speaker individuality. The method maps an original dysarthric speaker's normalized articulation space to a standardized space, then from the standardized space to the target speaker's normalized articulation space assuming normality to improve the phonological properties of vowels. We confirm phonological control of vowels by performing a processing simulation, comparison different target speakers and a processing simulation using a dummy original speaker as a dysarthria.

**Japanese Vowel-mora Visualization for Dysarthria Rehabilitation with Variational Autoencoder,**

This work proposes a Variational AutoEncoder (VAE)-based rehabilitation framework that visualizes the vowel-mora for Japanese dysarthria. Traditionally, Speech-Language Pathology (SLP) has shown the guideline of rehabilitation for dysarthria, but they should rely only on clinical experience and case-by-case adaption, highlighting the urgent necessity to push the boundary for showing a subjective guideline, which does not depend on the perspective of SLPs. The proposed framework takes advantage of two-dimensional latent representations of vowel-mora, which is assumed to be pre-processed by mel-spectrogram, via VAE. The experiments highlight the effectiveness of our proposed framework.

**Machine Learning Approaches for Automated Detection and Classification of Dysarthria Severity,**

Dysarthria, a speech disorder caused by neuro-motor problems resulting in impaired articulation, requires an assessment of its severity for diagnostic and monitoring purposes. Additionally, accurate severity classification facilitates the development of automated dysarthric speech detection and classification systems. This paper presents a comprehensive investigation into detecting dysarthric voices within a collection of normal voice samples, followed by the dysarthria severity classification utilizing neural network frameworks, specifically long short-term memory network (LSTM) and recurrent neural network (RNN). The study employs various features including Mel frequency cepstral coefficients (MFCC), formants, prosodic parameters, and voice quality. The performance of these models is evaluated against a baseline support vector machine (SVM) classifier using the Nemours corpus database. Remarkably, the highest classification accuracy achieved for this corpus is 99.69%. Detailed analysis demonstrates that selecting an appropriate neural network architecture yields superior performance compared to the conventional SVM classifier.

**The ISCSLP 2024 Multimodal Dysarthria Severity Assessment (MDSA) Challenge: Dataset, Tracts, Baseline and Results,**

To advance multimodal speech assessment and related research in developing objective diagnostic methods, we are launching the Multimodal Dysarthria Severity Assessment (MDSA) Challenge. This paper summarizes the outcomes from the ISCSLP 2024 MDSA Challenge. We first address the necessity of the challenge and then introduce the associated audio-video dataset selected from the MSDM database, including 62 subacute stroke patients and 25 normal controls. We then describe the challenge arrangement and the baseline system. Specifically, we set up a four-classification task for the severity of dysarthria (normal, mild, moderate and severe) with the aim of developing an objective and accurate automatic assessment method to assist in clinical diagnosis and treatment. Finally we summarize the challenge results and provide the major observations from the submitted systems. We hope the open data and challenge will serve as a benchmark and common test-bed for pathological speech assessment.

**An Analysis of Degenerating Speech Due to Progressive Dysarthria on ASR Performance,**

Although personalized automatic speech recognition (ASR) models have recently been improved to recognize even severely impaired speech, model performance may degrade over time for persons with degenerating speech. The aims of this study were to (1) analyze the change of performance of ASR over time in individuals with degrading speech, and (2) explore mitigation strategies to optimize recognition throughout disease progression. Speech was recorded by four individuals with degrading speech due to amyotrophic lateral sclerosis (ALS). Word error rates (WER) across recording sessions were computed for three ASR models: Unadapted Speaker Independent (U-SI), Adapted Speaker Independent (A-SI), and Adapted Speaker Dependent (A-SD or personalized). The performance of all models degraded significantly over time as speech became more impaired, but the A-SD model improved markedly when updated with recordings from the severe stages of speech progression. Recording additional utterances early in the disease before significant speech degradation did not improve the performance of A-SD models. This emphasizes the importance of continuous recording (and model retraining) when providing personalized models for individuals with progressive speech impairments.

**Dysarthria Voice Disorder Detection Using Mel Frequency Logarithmic Spectrogram and Deep Convolution Neural Network,**

Dysarthric speech recognition (DSR), often known as DSR, is an important tool that enables persons with vocal impairments to participate in voice-based automation systems and human-computer interaction. As a result of the poor intelligibility of handicapped speakers, the limited availability of datasets, and the low intra-class and inter-class variability in the speech samples, DSR is an essential component. This research provides a DSR based on the Mel Frequency Logarithmic Spectrogram (MFLS) and Deep Convolutional Neural Network (DCNN). The suggested MFLS+DCNN offers an improved representation of the voice signal in terms of its spectral and temporal characteristics. The results of the MFLS-DCNN scheme are validated using the UASpeech dataset, which was developed based on accuracy, recall, precision, and F1-score. With an accuracy of 96.83%, a recall performance of 0.97, a precision performance of 0.96, and an F1-score of 0.97, the scheme that is proposed has shown a considerable improvement above the conventional state-of-the-art.

**Wav2vec-Based Detection and Severity Level Classification of Dysarthria From Speech,**

Automatic detection and severity level classification of dysarthria directly from acoustic speech signals can be used as a tool in medical diagnosis. In this work, the pre-trained wav2vec 2.0 model is studied as a feature extractor to build detection and severity level classification systems for dysarthric speech. The experiments were carried out with the popularly used UA-speech database. In the detection experiments, the results revealed that the best performance was obtained using the embeddings from the first layer of the wav2vec model that yielded an absolute improvement of 1.23% in accuracy compared to the best performing baseline feature (spectrogram). In the studied severity level classification task, the results revealed that the embeddings from the final layer gave an absolute improvement of 10.62% in accuracy compared to the best baseline features (mel-frequency cepstral coefficients).

**Improved Speaker Independent Dysarthria Intelligibility Classification Using Deepspeech Posteriors,**

Individuals with dysarthria are unable to control rapid movement of the velum leading to reduction in intelligibility, audibility, naturalness and efficiency of vocal communication. Automatic intelligibility assessment of dysarthric patients allows clinicians diagnose the impact of therapy and medication and also to plan future course of action. Earlier works have concentrated on building speaker dependent machine learning systems for intelligibility assessment, due to limited availability of data. However, a speaker independent assessment system is of greater use by clinicians. Motivated by this observation, we propose a speaker independent intelligibility assessment system which relies on a novel set of features obtained by processing the output of DeepSpeech, an end to end Speech-to-Text engine. All experiments have been performed on the Universal Access Speech database. An accuracy of 53.9% was obtained using Support Vector Machine based four-class classification system for the speaker independent scenario while the accuracy obtained for the speaker dependent scenario is 97.4%.

**PB-LRDWWS System For the SLT 2024 Low-Resource Dysarthria Wake-Up Word Spotting Challenge,**

For the SLT 2024 Low-Resource Dysarthria Wake-Up Word Spotting (LRDWWS) Challenge, we introduce the PB-LRDWWS system. This system combines a dysarthric speech content feature extractor for prototype construction with a prototype-based classification method. The feature extractor is a fine-tuned HuBERT model obtained through a three-stage fine-tuning process using cross-entropy loss. This fine-tuned HuBERT extracts features from the target dysarthric speaker’s enrollment speech to build prototypes. Classification is achieved by calculating the cosine similarity between the HuBERT features of the target dysarthric speaker’s evaluation speech and prototypes. Despite its simplicity, our method demonstrates effectiveness through experimental results. Our system achieves second place in the final Test-B of the LRDWWS Challenge.

**On the use of array learners towards Automatic Speech Recognition for dysarthria,**

Providing Automatic Speech Recognition (ASR) systems for dysarthria is a challenging task since the normal and the disabled speech have different attributes; hence, using ASR systems designed and trained for normal speakers is not an effective approach. It is important to craft ASR technologies specifically for the speech disabled. Nonetheless, because of the complexity and variability of dysarthric speech, previous studies failed to achieve adequate performance. In this paper we investigated the applications of array learners towards dysarthric speech recognition. The array was implemented by several neural networks that configured to work in parallel. The proposed approach was verified by using the speech materials of seven dysarthric subjects with speech intelligibility from 2% to 86%. For comparison, the results were compared with a dysarthric ASR based on the legacy single-learner approach as the reference model. It is shown that the array learner-based dysarthric ASR improved the mean word recognition rate of 10.41% over the reference model, and decreased the error rate of 4.84%.

**Automated assessment and treatment of speech rate and intonation in dysarthria,**

Prosody assessment and treatment in dysarthria is clinically relevant, since prosodic impairment can have a negative impact on speech intelligibility and thus on participation in daily life conversation. We propose a speech-technology based software tool that provides automated numerical and visual feedback on two important aspects of prosody: speech rate and intonation. The tool includes speech rate and intonation algorithms, both specifically developed for the analysis of Dutch dysarthric speech. The tool enables speech-language pathologists to obtain objective measures of these prosodic aspects in a standardized and fast way, and enables dysarthric speakers to practise their prosodic skills intensively without the presence of a speech-language pathologist being required.

**Identification of hypokinetic dysarthria using acoustic analysis of poem recitation,**

Up to 90% of patients with Parkinson's disease (PD) suffer from hypokinetic dysarthria (HD). In this work, we analysed the power of conventional speech features quantifying imprecise articulation, dysprosody, speech dysfluency and speech quality deterioration extracted from a specialized poem recitation task to discriminate dysarthric and healthy speech. For this purpose, 152 speakers (53 healthy speakers, 99 PD patients) were examined. Only mildly strong correlation between speech features and clinical status of the speakers was observed. In case of univariate classification analysis, sensitivity of 62.63 % (imprecise articulation), 61.62% (dysprosody), 71.72% (speech dysfluency) and 59.60% (speech quality deterioration) was achieved. Multivariate classification analysis improved the classification performance. Sensitivity of 83.42% using only two features describing imprecise articulation and speech quality deterioration in HD was achieved. We showed the promising potential of the selected speech features and especially the use of poem recitation task to quantify and identify HD in PD.

**Respiratory and laryngeal influences on voice in post-stroke dysarthria: a pilot study,**

Dysarthria is a common disorder among stroke-patients that affects a wide range of speech and voice production processes. While voice disorders in post-stroke dysarthric patients have been well documented, few research has set out to investigate the complicated cause of the disorders, which potentially involve altered respiratory and laryngeal functions and their interaction. In this paper, we report a pilot study that preliminarily examined the respiratory and laryngeal influences on vocal performances in post-stroke dysarthric patients and healthy controls. Respiratory, laryngeal and vocal measures were collected in a maximum phonation time task and analyzed with linear mixed-effect regressions. The results suggested that pathways of influence may be established from respiratory to laryngeal functions and from laryngeal to vocal functions. Patients demonstrated increased laryngeal effort in response to reduced respiratory volumes and expiratory speed, which in turn may have contributed to a reduced voice quality and stability among patients. In addition, a few other findings suggested that reduced respiratory functions may also influence vocal performance through alternative pathways, although more work is needed to establish a clear chain of influence in these cases by unveiling the precise relations between the respiratory, laryngeal and vocal abnormalities in question.

**Assessing freezing of gait in parkinson's disease using analysis of hypokinetic dysarthria,**

Hypokinetic dysarthria (HD) and freezing of gait (FOG) are frequent symptoms of Parkinson's disease (PD). The aim of this work is to reveal pathological mechanisms common for HD and FOG, and use acoustic analysis of dysarthric speech to assess the gait difficulties in PD. We used a correlation analysis to investigate a relationship between speech features and FOG evaluated by freezing of gait questionnaire (FOG-Q). We found speech features quantifying reduced mobility of the articulatory organs significantly correlated with all parts of the questionnaire. Next, we built multivariate regression models to estimate the FOG-Q total score. With this approach, mean estimation error rate of 14.71% was achieved. We confirmed the previous findings of a close relationship between HD and FOG in PD. Furthermore, we showed it is possible to accurately (with the error of approximately 0.5 points) estimate FOG-Q using a reasonable number of conventional speech features.

**SARNet: Speaker-Attentive ResNet for Quantification of Dysarthria Severity,**

Analyzing differences in audio data of individuals with articulation disorders from the perspective of human speech and employing objective methods for automated dysarthria evaluation can significantly aid doctors in early patient screening and diagnosis. This proactive approach enables timely intervention and treatment during the initial stages of the condition. This paper presents a speech recognition network(named Speaker-Attentive ResNet (SARNet)) that aggregates and propagates features from different hierarchical levels using an attention-based statistical pooling module, built upon the ResNet architecture. This network aims to extract subtle features specific to individual speakers. The proposed method is evaluated using the TORGO dysarthric speech database, employing eight different acoustic features as well as features aggregated from these eight. Comparative experiments with two other models, Time-Delay Neural Network (TDNN) and Panns-CNN10(Large-Scale Pretrained Audio Neural Networks), demonstrate that due to the superior ability of MFCC(Mel-scaleFrequency Cepstral Coefficients) to emulate human auditory features, the proposed approach achieves a classification accuracy of $98 \%-99 \%$ in measuring speech intelligibility in healthy people, patients, and the severity of patient’s articulation disorders.

**Data Augmentation for Dysarthric Speech Recognition Based on Text-to-Speech Synthesis,**

In the field of automatic speech recognition (ASR) for people with dysarthria, it is problematic that not enough training speech data can be collected from people with dysarthria. To solve this problem, we propose a method of data augmentation using text-to-speech (TTS) synthesis. In the proposed data augmentation method, a deep neural network (DNN)-based TTS model is trained by utilizing speech data recorded from a speaker with dysarthria, and the trained TTS model is then used to generate the speaker’s speech data for training the ASR model for the speaker. The results of a speech recognition experiment on a person having spinal muscular atrophy (SMA) showed that the speech recognition error rate was improved by using the proposed data augmentation.

**Analysis and Classification of Dysarthric Speech,**

Classifying dysarthria using neural networks is challenging due to several factors inherent to the nature of dysarthria, the complexity of speech signals, and the requirements of effective machine learning models. Impaired speech classification is challenging for two main reasons: firstly, the data is scarce, and secondly, it is heterogeneous. In this paper, we have trained two different architectures on a dysarthric speech database. A comparison of results shows that according to precision, recall, f1-score, and accuracy, the Deep Neural Network (DNN) model outperforms the classical Convolutional Neural Network (CNN) model, even with a small database. For people with dysarthria, a DNN can improve the performance metrics during the classification of impaired speech by 99% compared to the classical architecture. This improvement is more than that provided by CNN. We have used the TORGO database for this work.

**Speech Recognition for a Person With Cerebral Palsy Using Whisper Fine-Tuned on Japanese and English Dysarthric Speech,**

People with cerebral palsy often have dysarthria, and this makes it hard for them to speak as they wish. In this paper, we present an automatic speech recognition (ASR) model for a person with cerebral palsy based on Whisper. Whisper is highly accurate when performing speech recognition for Japanese speech, but recognition accuracy for Japanese dysarthric speech tends to be low. One possible solution to this problem is to fine-tune Whisper using Japanese dysarthric speech. However, there is a problem in that it is difficult to record a large amount of speech for people with dysarthria. Therefore, in our proposed method, English dysarthric speech is utilized for training Whisper in addition to Japanese dysarthric speech. The results of our proposed method showed an improvement in the error rate of about 1% compared to using only Japanese dysarthric speech as training data.

**A joint-feature learning-based voice conversion system for dysarthric user based on deep learning technology,**

Dysarthria speakers suffer from poor communication, and voice conversion (VC) technology is a potential approach for improving their speech quality. This study presents a joint feature learning approach to improve a sub-band deep neural network-based VC system, termed J\_SBDNN. In this study, a listening test of speech intelligibility is used to confirm the benefits of the proposed J\_SBDNN VC system, with several well-known VC approaches being used for comparison. The results showed that the J\_SBDNN VC system provided a higher speech intelligibility performance than other VC approaches in most test conditions. It implies that the J\_SBDNN VC system could potentially be used as one of the electronic assistive technologies to improve the speech quality for a dysarthric speaker.

**A multi-smartwatch system for assessing speech characteristics of people with dysarthria in group settings,**

Speech-language pathologists (SLPs) frequently use vocal exercises in the treatment of patients with speech disorders. Patients receive treatment in a clinical setting and need to practice outside of the clinical setting to generalize speech goals to functional communication. In this paper, we describe the development of technology that captures mixed speech signals in a group setting and allows the SLP to analyze the speech signals relative to treatment goals. The mixed speech signals are blindly separated into individual signals that are preprocessed before computation of loudness, pitch, shimmer, jitter, semitone standard deviation and sharpness. The proposed method has been previously validated on data obtained from clinical trials of people with Parkinson disease and healthy controls.

**Weak Speech Supervision: A case study of Dysarthria Severity Classification,**

Machine Learning methodologies are making a remarkable contribution, and yielding state-of-the-art results in different speech domains. With this exceptionally significant achievement, a large amount of labeled data is the largest bottleneck in the deployment of these speech systems. To generate massive data, hand-labeling training data is an intensively laborious task. This is problematic for clinical applications where obtaining such data labeled by speech pathologists is expensive and time-consuming. To overcome these problems, we introduce a new paradigm called Weak Speech Supervision (WSS), a first-of-its-kind system that helps users to train state-of-the-art classification models without hand-labeling training data. Users can write labeling functions (i.e., weak rules) to generate weak data from the unlabeled training set. In this paper, we provide the efficiency of this methodology via showing the case study of the severity-based binary classification of dysarthric speech. In WSS, we train a classifier on trusted data (labeled with 100% accuracy) via utilizing the weak data (labeled using weak supervision) to make our classifier model more efficient. Analysis of the proposed methodology is performed on Universal Access (UA) corpus. We got on an average 35.68% and 43.83% relative improvement in terms of accuracy and F1-score w.r.t. baselines, respectively.

**End-to-end Dysarthric Speech Recognition Using Multiple Databases,**

We present in this paper an end-to-end automatic speech recognition (ASR) system for a person with an articulation disorder resulting from athetoid cerebral palsy. In the case of a person with this type of articulation disorder, the speech style is quite different from that of a physically unimpaired person, and the amount of their speech data available to train the model is limited because their burden is large due to strain on the speech muscles. Therefore, the performance of ASR systems for people with an articulation disorder degrades significantly. In this paper, we propose an end-to-end ASR framework trained by not only the speech data of a Japanese person with an articulation disorder but also the speech data of a physically unimpaired Japanese person and a non-Japanese person with an articulation disorder to relieve the lack of training data of a target speaker. An end-to-end ASR model encapsulates an acoustic and language model jointly. In our proposed model, an acoustic model portion is shared between persons with dysarthria, and a language model portion is assigned to each language regardless of dysarthria. Experimental results show the merit of our proposed approach of using multiple databases for speech recognition.

**On Using the UA-Speech and Torgo Databases to Validate Automatic Dysarthric Speech Classification Approaches,**

Although the UA-Speech and TORGO databases of control and dysarthric speech are invaluable resources made available to the research community with the objective of developing robust automatic speech recognition systems, they have also been used to validate a considerable number of automatic dysarthric speech classification approaches. Such approaches typically rely on the underlying assumption that recordings from control and dysarthric speakers are collected in the same noiseless environment using the same recording setup. In this paper, we show that this assumption is violated for the UA-Speech and TORGO databases. Using voice activity detection to extract speech and non-speech segments, we show that the majority of state-of-the-art dysarthria classification approaches achieve the same or a considerably better performance when using the non-speech segments of these databases than when using the speech segments. These results demonstrate that such approaches trained and validated on the UA-Speech and TORGO databases are potentially learning characteristics of the recording environment or setup rather than dysarthric speech characteristics. We hope that these results raise awareness in the research community about the importance of the quality of recordings when developing and evaluating automatic dysarthria classification approaches.

**Modification to correct distortions in stops of dysarthrie speech using TMS320C6713 DSK,**

Communication is a bridge between people which enables them to share facts, ideas, feelings etc. with each other. Speech communication is easier and simpler compared to other types of communication. But speech disorders affect one's ability to communicate. Dysarthria is a neuro-motor disorder, where one losses his/her ability to articulate words normally due to tongue/muscle weakness or stroke etc. It results in distorted speech which is hard to understand. Dysarthria affect more on the articulation of consonants (stops in particular) than on the articulation of vowels. Intelligibility varies greatly depending on the extent of neurological damage. As consonants are important than vowels in measuring the intelligibility, distortions in consonant production of dysarthric speech are studied and modifications are performed. This work aims at correcting devoicing of voiced stop (voiced stops are pronounced as unvoiced stops with the same place of articulation) by detecting important time instants such as glottal onset, glottal offset and burst onset thereby modifying the distortions in dysarthric speech using TMS320C6713 DSK with CC Studio 5.5. After the dysarthric speech has been recorded on the DSK 6713 processor modification is done to improve its intelligibility. Two databases were used in the experiment such as Universal Access (UA) database and Torgo database. Out of 65 bursts expected from UA database and 128 bursts expected from Torgo database, 84.6% and 82.81% bursts were detected correctly within 0.3 s intervals respectively.

**Prediction of Parkinson’s Disease Using Machine Learning Based on Vocal Frequency,**

Parkinson’s disease (PD) is a neurological disorder that affects most people after Alzheimer’s.The ageing neurodegenerative condition leads to PD that reduces dopamine levels in the brain are a defining feature. Tremor, stiffness, bradykinesia, and postural instability are the core characteristics of PD. PD patient’s(PWP) quality of life is impacted by both motor and non-motor symptoms, which may also have an indirect impact on family and caregivers. The patient’s quality of life can be improved and maintained with an early PD diagnosis. There is currently no cure, however there are therapies to control the illness, such as dopaminergic medications. Speech problems associated with PD are categorized under the term hypokinetic dysarthria for the diagnosis of PD.The traditional approach is based on their clinical history and also with their physical examination. With the help of automatic analysis tools, practitioners may diagnose patients, monitor their progress, and conduct regular, economical assessments that are objective. This study uses a machine learning approach to conduct a pilot experiment to identify the existence of dysarthria in speech and gauge its severity. With the help of SVM Algorithm and HyperTunningdisplays 93% accuracy on KAGGLE database test samples.

**Acoustic Space in Motor Disorders of Speech: Two Case Studies,**

Studies on acoustic space have strengthened the view that vowels are acoustically and perceptually defined in terms of their relative positioning in vowel space. Every speaker identifies an optimal vowel space within which perceptual, phonological contrast is maintained. This is an interdisciplinary study involving speech pathology, physics of speech and neurology of speech. Two case studies of dysarthria presented in this paper are -- one Parkinson's disease and one case of acute ischemic stroke with age-gender-language matched controls. A detailed acoustic analysis shows how acoustic space gets considerably reduced, in both PD and stroke, and in these two very different kinds of dysarthrias the acoustic space is also modified very differently. The study also examines the third formant to show that the higher formants are consistently lowered in both PD and stroke. Hypokinetic speech production in these cases is reflected in lower intensity. The results have significant applications in clinical acoustics and in the theoretical fields of neurology of speech, linguistics and phonology.

**Robust Assessment of Dysarthrophonic Voice with RASTA-PLP Features: A Nonlinear Spectral Measures,**

This paper presents an artificial intelligence based speech signal processing technique to identify dysarthrophonic voice with relative spectral-perceptual linear prediction (RASTA-PLP) features. Dysarthria is a neural motor speech disorder caused by muscular weakness. Voice analysis of dysarthrophonic patients is challenging as this disease has multidimensional effects on the human voice generation system. Conventional spectral analysis is unable to accurately characterize the pathology associated with nonlinear dynamicity of human voice. This work investigates the suitability of RASTA-PLP features excerpted from speech signals to identify dysarthrophonic patients. The speech samples of healthy and dysarthrophonic patients are collected from the Saarbrücken Voice Database (SVD). Several machine learning and Artificial neural network (ANN) based algorithms are developed to evaluate the classification performance of the proposed system. The designed system can achieve excellent performance in terms of accuracy (100%) considering female and male subjects separately.

**Comparison of English and Chinese Speech Recognition Using High-Density Electromyography,**

Speaking different languages requires different ways of pronunciation, and the muscular activities associated with phonation show different articulation styles. Therefore, clarifying the contributions of the articulatory muscles in different regions, such as the face and neck, is helpful for automatic speech recognition. However, it remains unclear how the articulatory muscles at different positions affect the classification accuracies of speech recognition across different languages. In this study, the technique of high-density surface electromyography (HD sEMG) was proposed to investigate the role of different articulatory muscles in classifying English and Chinese speaking tasks, respectively. The HD sEMG signals were recorded by 120 electrodes evenly placed on the facial and neck muscles across six subjects while they were speaking five English and Chinese daily words. Four time-domain features were extracted from sEMG recordings and used to construct a linear-discriminant-analysis classifier for speech recognition. The results showed that the classification accuracies of using neck sEMG were higher than that of using facial sEMG in both English and Chinese recognition tasks. The accuracies for Chinese speaking tasks were significantly higher than that for English when using facial sEMG only. Moreover, there was no significant difference in accuracies between the two types of languages when using neck sEMG. This study might provide useful information about the contributions of different articulatory muscles, and pave the way for automatic speech recognition across different languages for patients with dysarthria.

**Hybrid CNN-LSTM network to detect Dysarthria using Mel-Frequency Cepstral Coefficients,**

Dysarthria is a speech problem acquired at birth due to cerebral palsy (CP) or developed after severe brain damage. Dysarthria affects more than 70% of Parkinson's patients and 10% to 65% of people with traumatic brain injury. It is critical to detect dysarthria and other voice speech difficulties early to diagnose the underlying cause. Intelligent systems capable of identifying dysarthria with incredible precision have been developed using audio processing techniques and various deep learning models. This paper presents a hybrid CNN-LSTM model for classifying patients with dysarthria using audio recordings. The CNN-LSTM combination helps capture spatial and temporal information where CNN acts as a feature extractor while LSTM functions as a classifier. The proposed model was trained on the publicly available 9184 audio recordings from the TORGO dataset, and various audio augmentation techniques were employed to generate synthetic data. A total of 128 features were extracted using Mel Frequency Cepstral Coefficients (MFCC) and fed into the architecture as inputs. The K-fold cross-validation technique was used to avoid overfitting and increase the generalization capability of the model. The proposed architecture achieved a state-of-the-art 99.59% accuracy on the dataset. The presented work will minimize the workload of speech pathologists and help them detect dysarthria precisely and effectively.

**Optimizing Dysarthria Wake-Up Word Spotting: an End-to-End Approach For SLT 2024 LRDWWS Challenge,**

Speech has emerged as a widely embraced user interface across diverse applications. However, for individuals with dysarthria, the inherent variability in their speech poses significant challenges. This paper presents an end-to-end Pretrain-based Dual-filter Dysarthria Wake-up word Spotting (PD-DWS) system for the SLT 2024 Low-Resource Dysarthria Wake-Up Word Spotting Challenge. Specifically, our system improves performance from two key perspectives: audio modeling and dual-filter strategy. For audio modeling, we propose an innovative 2 branch- d 2 v 2 model based on the pre-trained data2vec $2(\mathrm{~d} 2 \mathrm{v} 2)$, which can simultaneously model automatic speech recognition (ASR) and wake-up word spotting (WWS) tasks through a unified multi-task finetuning paradigm. Additionally, a dual-filter strategy is introduced to reduce the false accept rate (FAR) while maintaining the same false reject rate (FRR). Experimental results demonstrate that our PD-DWS system achieves an FAR of 0.00321 and an FRR of 0.005, with a total score of 0.00821 on the test-B eval set, securing first place in the challenge.

**Enhancing Dysarthria Diagnosis With Deep Learning Techniques,**

Dysarthria is characterized by delayed, slurred speech that can be challenging to comprehend. This condition is frequently brought on by nerve injury that affects the muscles used to produce speech. Depending on which muscles are affected and the underlying cause, symptoms can differ greatly. Talk therapy can help with early detection and intervention, which can enhance treatment outcomes. Convolutional neural networks (CNNs) are the method suggested here for detecting dysarthria from audio data. The method highlights the distinctions in speech patterns between people with and without dysarthria by using feature extraction, notably Mel-frequency cepstral coefficients (MFCC), and audio visualization approaches. These properties are used in the development and training of several neural network models, such as CNN, Long Short Term Memory (LSTM), Gated Recurrent Units (GRU), Bidirectional LSTM, SimpleRNN, and Deep Neural Networks (DNN). These models’ performance is assessed with the use of confusion matrices and classification reports. This all-inclusive dysarthria detection pipeline includes advanced deep learning approaches for evaluation, model training, and data preprocessing. The objective is to develop a trustworthy dysarthria detection system that will help healthcare professionals identify and treat the ailment early on.

**Using Novel Hybrid Convolutional Neural Network for Dysarthria Diagnosis,**

Dysarthria is a motor speech disorder characterized by articulation and phonation difficulties resulting from speech muscle weakness, paralysis, or incoordination. A precise and timely diagnosis of dysarthria is essential for effective treatment and management of the condition, as it may deteriorate over time or be a precursor to a much more serious disease. On the other hand, this is becoming a severe problem in recent times owing to the rising ageing population and the prevalence of neurological disorders among such people. Hence early detection of dysarthria is deemed essential for the timely management of the disease. In recent years, Artificial Intelligence (AI) applications have shown promising results in various audio processing tasks and incorporated into pathological voice analysis for disease diagnosis. The majority of previous studies on dysarthria detection employed Machine Learning (ML) and Deep Learning (DL) models as the disease classification models. In light of this, this study presents a novel hybrid approach for classifying dysarthria based on audio data using Convolutional Neural Network (CNN) and Support Vector Machine (SVM). According to the experimental results, the proposed classification schema achieves an accuracy of 98.25 % compared to previous research work. Overall, our proposed method aims to automate the classification process, enabling faster and more reliable diagnoses.

**Breathiness Indices for Classification of Dysarthria Based on Type and Speech Intelligibility,**

Dysarthria classification based on intelligibility level is useful for speech pathologists for deciding the therapy. However, intelligibility assessment also depends on perceptual attributes like hypernasality, breathiness, slow rate, short pauses etc. These perceptual attributes vary depending on the cause for dysarthria giving rise to different types of dysarthria. In this work, we explore the use of breathiness features for intelligibility assessment of dysarthria and for distinguishing type of dysarthria. Voiced segments from two controlled speakers, two dysarthric speakers with low and mid intelligibility level each from UA database are used in the work. Features were analysed for use in dysarthria intelligibility assessment vs. distinguishing type of dysarthria.

**Cross-lingual Dysarthria Severity Classification for English, Korean, and Tamil,**

Data scarcity hinders research on dysarthria severity classification due to the limited size of datasets. While the crosslingual approach has been applied to alleviate the problem, the roles of language-specific features have been underestimated. This paper proposes a cross-lingual classification method for English, Korean, and Tamil, which employs both language-independent features and language-unique features. First, we extract thirty-nine features from diverse speech dimensions such as voice quality, pronunciation, and prosody. Second, feature selections are applied to identify the optimal feature set for each language. A set of shared features and a set of distinctive features are distinguished by comparing the feature selection results of the three languages. Lastly, automatic severity classification is performed, utilizing the two feature sets. Notably, the proposed method removes different features by languages to prevent the negative effect of unique features for other languages. Accordingly, eXtreme Gradient Boosting (XGBoost) algorithm is employed for classification, due to its strength in imputing missing data. In order to validate the effectiveness of our proposed method, two baseline experiments are conducted: experiments using the intersection set of mono-lingual feature sets (Intersection) and experiments using the union set of monolingual feature sets (Union). According to the experimental results, our method achieves better performance with a 67.14% F1 score, compared to 64.52% for the Intersection experiment and 66.74% for the Union experiment. Further, the proposed method attains better performances than mono-lingual classifications for all three languages, achieving 17.67%, 2.28%, 7.79% relative percentage increases for English, Korean, and Tamil, respectively. The result specifies that commonly shared features and language-specific features must be considered separately for cross-language dysarthria severity classification.

**The Open-Access Mandarin Subacute Stroke Dysarthria Multimodal (MSDM) Database for Intelligent Assessment,**

Early objective identification and assessment of dysarthria due to neurological deficits are essential for neurorehabilitation. Developing a system to achieve this requires a large-scale database of pathological information with detailed labeling. In the present study, a high-quality Chinese multimodal audio-visual database, consisting of 64 subacute stroke patients and 25 healthy participants, named the “Mandarin Subacute Stroke Dysarthria Multimodal (MSDM) database”, was established. The materials of MSDM include a series of speech tasks such as syllables, characters, words, sentences, and spontaneous speech. All audio-visual data in this database were manually annotated and simultaneously verified by experienced researchers. Additionally, comprehensive clinical assessments of speech-motor function (e.g., Frenchay Dysarthria Assessment) and cognitive function (e.g., Montreal Cognitive Assessment) for each individual were included in the database. In conclusion, the MSDM database is believed to provide sufficient data resources for developing automatic assessment and speech recognition methods and contribute to understanding the pathological mechanisms of dysarthria.

**Significance of Filterbank Structure for Capturing Dysarthric Information through Cepstral Coefficients,**

The short-term Fourier transform magnitude spectra (STFT-MS) computed from the dysarthric speech deviates nonlinearly from the normal speech in different frequency bands depending on underlying sound units. This discriminating information can be captured by segmenting the STFT-MS into different frequency bands following the power spectra of board categories of sound units. Motivated by this observation in this study, we have computed the cepstral coefficients by analyzing the STFT-MS in 0–500 Hz, 500–2000 Hz, 2000–4000 Hz, and 4000 – 8000Hz, respectively for 16 kHz sampled speech data. Each of the selected frequency bands is analyzed by using a 30 channel Mel filterbank. The log filterbank energies computed for each sub-band are then polled together and discrete cosine transform (DCT) is applied to compute the cepstral coefficients, here termed as sub-band enhanced Mel frequency cepstral coefficients (SE-MFCC). The i-vector based dysarthric intelligibility assessment system reported in this study shows that the SEMFCC outperforms the conventional Mel frequency cepstral coefficients (MFCC), and the cepstral coefficients computed using inverse-Mel filterbank (IMFCC), and linear filterbank (LFCC). The score level combination of SE-MFCC with the MFCC further improves the overall performance.

**Data Augmentation Based on Frequency Warping for Recognition of Cleft Palate Speech,**

In this paper, we present an automatic speech recognition (ASR) system for the speech of a person with a cleft lip and palate (CLP). The accuracy of speech recognition for a person with CLP is lower than that of a physically-unimpaired (PU) person because the CLP speech has characteristics that differ from those of a PU person; moreover, the amount of available training data is quite limited. In the field of ASR for PU people, data augmentation and self-supervised learning have been studied to tackle this problem of data scarcity. In this paper, we evaluate the effectiveness of those approaches on CLP speech recognition, and propose a data augmentation technique based on frequency warping. The formant of CLP speech tends to fluctuate compared to that of PU people. In order to compensate for the large variety of formant components, our data augmentation method stretches or contracts the spectrogram through the frequency axis. The experimental results on an ASR task with two CLP subjects showed that both data augmentation and self-supervised learning were effective for CLP speech recognition, and our proposed method further improved the performance of those two approaches based on conventional SpecAugment techniques.

**Improving Pronunciation Clarity of Dysarthric Speech Using CycleGAN with Multiple Speakers,**

In this paper, we propose a method that improves pronunciation clarity of dysarthric speech using CycleGAN based non-parallel voice conversion. This method converts dysarthric speech into healthy speech using CycleGAN. We considered the use of single and multiple speakers as healthy speech. The subjective evaluations showed the effectiveness of using multiple speakers as healthy speech.

**Adaptation of a Pronunciation Dictionary for Dysarthric Speech Recognition,**

In the general framework of an automatic speech recognition system, a pronunciation dictionary, that is a mapping table from a phoneme sequence to a word, is used both in the processes of training and recognition. However, this pronunciation dictionary is not always adequate in the case of dysarthric speech recognition because dysarthric people often have difficulty pronouncing words in the same way they are pronounced in the dictionary. In this paper, we investigate the adaptation of a pronunciation dictionary to an individual dysarthric person and evaluate the effectiveness of adapting the dictionary using a dysarthric speech recognition task. In the proposed method, in order to find rules we can use to modify a dictionary, we analyze the pattern of mis-recognition in the phoneme recognition results. By following the extracted rules, we add pronunciations to the dictionary for the target dysarthric person. We evaluate the effectiveness of the adapted pronunciation dictionary on a continuous speech recognition task and demonstrate that the adapted dictionary can decrease the word error rate.

**Comparative Analysis of Glottal and Vocal Tract Features in Dysarthria,**

Dysarthria is a neurological disorder associated with the muscles in the vocal tract system that is caused by problems with coordination. It affects the vocal tract system and the glottis with different levels of severity. This research deals with the role of vocal tract (i.e., filter) and glottal excitation source in dysarthric speech, using Glottal Flow Model with Iterative Adaptive Inverse Filtering (GFM-IAIF). By decomposing the vocal tract and glottal source into two separate components, we were able to identify which component (source or filter) is affected the most. Our research showed that the vocal tract system is the most affected part, determining a better classification of features the magnitude spectrum-based Mel frequency cepstral coefficient (MFCC) resulted in 95.75% accuracy for vocal tract components and lower accuracy of 86.5% for glottal source components. In the same way, Modified Group Delay Cepstral Coefficients (MGDCC) correlated with a test accuracy of 94.43% of the vocal tract in comparison with 88.70% of the glottal source. These outcomes reveal that the vocal tract gets most damaged due to dysarthria and thus, emphasis the use of specific diagnostic and therapy interventions targeted at this area. The research not only explores feature fusion but also recommends further efforts to refine dysarthria diagnosis and treatment.

**A Fuzzy Cognitive Map Hierarchical Model for Differential Diagnosis of Dysarthrias and Apraxia of Speech,**

This paper presents a novel soft computing system for differential diagnosis of the dysarthrias and apraxia of speech based on well accepted dysarthrias' classification system used by speech and language pathologists. The dysarthrias and apraxia are complex disorders of speech because they represent a variety of neurological disturbances that can potentially affect every component of speech production. Since an accurate diagnosis is a very challenging task for the clinician, the under development system based on hierarchical fuzzy cognitive maps (FCMs) will be used as a "second opinion" or training system. The hierarchical FCM differential diagnosis system is capable of differentiating between the six types of dysarthria as well as apraxia. The system was tested using published case studies and real patients and examples are presented here

**A Novel Gamified Approach for Collecting Speech Data from Young Children with Dysarthria: Feasibility and Positive Engagement Evaluation,**

Dysarthria is a common and treatable speech problem in children, and computer-assisted speech therapy is a promising way for children's speech therapy. However, data collection poses a significant challenge for computer-assisted therapy, especially when it comes to gathering speech data from young children, particularly those with dysarthria. Finding a better way to collect young children's speech data is, therefore, an urgent need. This paper prompted a gamified speech collection method and carried out an experiment to compare the participation time, error rate, and collection efficiency with the gamified method and with a traditional method where adults are imitated. Moreover, we also explore whether the gamified collection methods increase the children's positive engagement. A feasibility study including 10 children with dysarthria and 10 children without speech problems was conducted. Their participation duration, number of spoken utterances, number of mispronunciation utterances, and a questionnaire about children's engagement attitude were recorded. The findings indicate that the gamified collecting method reduces pronunciation mistake rates in children with dysarthria while also increasing their engagement to participate.

**Edge Computing Solutions Supporting Voice Recognition Services for Speakers with Dysarthria,**

In the framework of Automatic Speech Recognition (ASR), the synergism between edge computing and artificial intelligence has led to the development of intelligent objects that process and respond to human speech. This acts as a key enabler for multiple application scenarios, such as smart home automation, where the user’s voice is an interface for interacting with appliances and computer systems. However, for millions of speakers with dysarthria worldwide, such a voice interaction is impossible because nowadays ASR technologies are not robust to their atypical speech commands. So these people, who also live with severe motor disabilities, are unable to benefit from many voice assistant services that might support their everyday life. To cope with the above challenges, this paper proposes a deep learning approach to isolated word recognition in the presence of dysarthria conditions, along with the deployment of customized ASR models on machine learning powered edge computing nodes. In this way, we work toward a low-cost, portable solution with the potential to operate next to the user with a disability, e.g., in a wheelchair or beside a bed, in an always active mode. Finally, experiments show the goodness (in terms of word error rate) of our speech recognition solution in comparison with other studies on isolated word recognition for impaired speech.

**Dysarthria Diagnosis and Dysarthric Speaker Identification Using Raw Speech Model,**

Dysarthria is a medical condition that causes difficulty in producing coherent speech due to muscle paralysis or weakness. This article presents a unique approach to identifying dysarthric speakers using a deep learning model that works directly with unprocessed speech waveforms. By eliminating the need for feature extractions, the model's resistance to noise and voice variability is increased. The proposed approach utilizes a SincNet layer model with multiple initializations including Mel, Erb, and Bark scales for dysarthria detection (DD) and dysarthric speaker identification (DSI). Bark scaling, aligning better with human auditory perception and capturing distinctive acoustic features, notably outperforms other initialization methods. When Bark scaling was employed, the study's results demonstrated outstanding accuracy rates of 97.0% for DD and 88.0% for DSI. The results demonstrated exceptional performance, that surpassing existing literature benchmarks.

**Acoustic and Kinematic Examination of Dysarthria in Cantonese Patients of Parkinson’s Disease,**

Hypokinetic dysarthria is one of the core symptoms of Parkinson's disease, characterized by reduced loudness, slurred speech and distorted consonant productions. Dopaminergic medication for Parkinson's disease (PD) has been proven to be effective in treating limb and gross motor movement problems. However, the literature sees contradictory findings regarding dopaminergic effects on speech problems associated with PD. Previous perceptual and acoustic studies of PD hypokinetic dysarthria mostly involved heterogeneous population and variations of speech tasks which complicated data interpretation. Also, the lack of kinematic data limited our understanding of the details of articulation errors associated with PD, as well as the medication effects. Electromagnetic articulography (EMA) enabled the examination of PD articulatory patterns with high degree of accuracy and safety. The aim of the present study was to address these inconsistencies by providing an integrative description of basic kinematic and acoustic parameters of speech production about the dopaminergic effect on early PD speech by using EMA. The results revealed a significant improvement of articulatory function on dopaminergic effects for PD speech, evidenced by increased vowel space, the velocity and accelerated velocity initiation and coordination of articulation during bilabial or alveolar productions.

**In-Domain Data Augmentation to Enhance Severity Level Classification of Dysarthria from Speech,**

In this paper, we present our endeavor to construct an automatic dysarthria severity level classification system tai-lored for low-resource dysarthric speech datasets. The scarcity of available speech data from dysarthric speakers poses a significant challenge to training an effective classification system. Addressing this challenge, we devised a robust baseline system by blending a distinctive set of features, encompassing temporal, prosodic, and spectral information, with the traditional MFCC features. This amalgamation aptly captures the nuanced characteristics of dysarthric speech, facilitating efficient model training. To tackle the constraints of low-resource conditions, we explored four prominent augmentation techniques: Speaking Rate, Pitch, Formant, and Vocal Tract Length Perturbation (VTLP) modification based data augmentation for the task of severity classification. The explored data augmentation gives a significant reduction in the classification error rate (CER), and VTLP-based data augmentation is superior among others. Further, we also investigated combinations of explored data augmentation methods, fortifying the reliability of our dysarthria severity classification system. The combined novel augmentation gives a noteworthy relative improvement of 42.86% over the baseline on the dysarthric severity classification.

**A modern approach to dysarthria classification,**

This work deals with the assessment of neurological diseases known as dysarthrias, using a novel approach based on objective and perceptual features extracted from pathological speech signals. A methodology for the classification of dysarthria is developed in which digital signal processing algorithms are used to appraise the severity of those features less reliably judged by the clinicians, while the others are taken directly from perceptual judgments or medical records. The assessment process evaluates the performance of two different classifiers and compares them with the traditional assessment system. The first approach is based on the lineal discriminant analysis and the second is a non-lineal technique based on self-organizing maps. The non-lineal classifier provided the highest percent of correct classification and the most accurate information on the relevance of the features in the classifier decision. It also provided a bi-dimensional representation of de data that allows a better understanding of the correspondence between the speech deviations and the location of the damage in the peripheral or central nervous system.

**Observations from a Simple Vocal-Tract-Model's Behaviour for PD-Dysarthric Speech: Applicability,**

The uniform-element tube model of a speaker's vocal tract is a by-product from LPC speech analysis. It is used here to observe a speaker's articulation. The aim is an insight into possible articulatory weaknesses of PD (Parkinson-Disease) patients who suffer from dysarthria, i.e., speaking problems. Approved auditive methods exist for evaluation of the patients' speech handicaps, applying also instrumental signal features. But a direct view on vocal-tract movements can be an additional diagnostic aid. Due to its real-time potential, the simple estimation of vocal-tract areas from LPC analysis is regarded here, despite its known impreciseness. In a first measurement set, it is checked with fluent speech of healthy persons, whether any reaction appears on articulatory changes between clear and intentionally mumbled speech. Then, with sustained vowels of healthy and slightly as well as strongly handicapped speakers, the potential of the approach to display such differences is examined.

**A Multi-modal Approach to Dysarthria Detection and Severity Assessment Using Speech and Text Information,**

Automatic detection and severity assessment of dysarthria are crucial for delivering targeted therapeutic interventions to patients. While most existing research focuses primarily on speech modality, this study introduces a novel approach that leverages both speech and text modalities. By employing cross-attention mechanism, our method learns the acoustic and linguistic similarities between speech and text representations. This approach assesses specifically the pronunciation deviations across different severity levels, thereby enhancing the accuracy of dysarthric detection and severity assessment. All the experiments have been performed using UA-Speech dysarthric database. Improved accuracies of 99.53% and 93.20% in detection, and 98.12% and 51.97% for severity assessment have been achieved when speaker-dependent and speaker-independent, unseen and seen words settings are used. These findings suggest that by integrating text information, which provides a reference linguistic knowledge, a more robust framework has been developed for dysarthric detection and assessment, thereby potentially leading to more effective diagnoses.

**Processing of pathological changes in speech caused by dysarthria,**

Computer analysis of voice isolated sounds may lead to identification of parameters correlated with neurological diseases. This paper presents results of preliminary research of voice pathological changes caused by dysarthria. The selection of linguistic material was characterized according to the place and manner of articulation in the phonetic system of Polish. Results of clinical examination allowed to determine simple markers of neurodegenerative diseases, which serves as a basis for construction of objective examination model.

**Significance of Entropy Based Features For Dysarthric Severity Level Classification,**

Dysarthria is a motor speech disorder arising from impairment of muscles that makes difficult to form or pronounce words while speaking. In this paper, we introduce an approach of multiband entropy based features extracted from dysarthric speech signals for dysarthric severity level classification. Generally, entropy is measured as the number of bits of information contained in each message signal. The information content of these signal measures how much randomness or uncertainity contains in a signal. Extending this, we use a frame-wise processing technique to divide the speech signal into short frames which allows for detailed analysis of different characteristics of speech signal. Furthermore we divide frames into equal sub bands and compute the entropy and zero mean entropy in each sub band using Gabor filterbank. It is expected that the mean entropy of very low dysarthric severity is lower compared to high dysarthric severity level indicating that randomness is increasing as the severity level of dysarthria is increasing. Experimental analysis were conducted on extensively used dataset namely UA Speech. Results were carried out by Convolutional Neural Network (CNN), along with 5-cross validation. The results are compared against standard MFCC, LFCC, and glottal source based LFRCC. It was observed that the addition of entropy information boosted the performance of MFCC by 4.38%, LFCC by 2.54%, and LFRCC by 1.88% compared to their traditional techniques indicating the crucial information captured by the entropy for dysarthria severity classification.

**Automated Acoustic Analysis in Parkinson’s Disease Using a Smartphone\*,**

Dysarthria is a common speech disorder in Parkinson’s Disease (PD). The Dysarthria Analyzer software has emerged as a viable tool for automatic speech analysis in PD and quantification of dysarthria severity. However, most studies use the Dysarthria Analyzer with recordings obtained under tightly controlled conditions and high-quality microphones, and the utility of the Dysarthria Analyzer when used with recordings acquired under non-ideal conditions, such as in busy clinical settings, remains unexplored. This study investigates the Dysarthria Analyzer’s performance in a setting more akin to a clinical environment using a smartphone. We obtained data from three groups, including healthy controls (HC), PD patients with their deep brain stimulation on (ON-DBS), and PD patients with their DBS off (OFF-DBS). We found a significant decrease in pitch variability and an increase in speech rate for the OFF-DBS group compared to the HC. Furthermore, most of the estimated values for the speech markers fall within the reported values in the literature. Our findings demonstrate that the Dysarthria Analyzer effectively extracts relevant speech markers even when used with recordings obtained under non-ideal conditions, emphasizing its potential for widespread clinical adoption.Clinical Relevance— Our findings demonstrate the potential of using smartphone recordings obtained in clinical environments for automatic objective speech analysis. These findings are relevant for developing a clinical tool that can be widely accessible and easily implemented during routine clinical visits of PD to improve the assessment of dysarthria in PD.

**An Optimal Speech Recognition Module for Patient's Voice Monitoring System in Smart Healthcare Applications,**

During recent years, health care domain has rapidly developed in which patients and medical resources are directly connected with the smart way that enables Smart Health Care. The growth in design and development of a speech automated system will provide a life assistant service in smart health care environment. In automating the speech system, speech recognition is one of the basic steps to understand the human recognition and their behaviors. These speech recognition systems will be very much accessible for speakers who suffer from dysarthria, a neurological disability that damages the control of motor speech articulators. In this paper, the main objective is to develop an efficient speech recognition module based on the Voice Input Voice Output Communication Aid (VIVOCA) architecture that can device a support aid to the people with DYSARTHRIA. Totally there are seven features extracted from each noise eliminated real time bilingual isolated word speech signal data uttered by a speaker both in Tamil and English languages. Vector Quantization based Genetic Algorithm codebook is created for the recognition modeling. Optimization of Hidden Markov Model (HMM) is done based on Particle Swarm Optimization (PSO) method to improve the recognition accuracy compared to the conventional HMM and also experiment results of the proposed module shows 95% of accuracy. The proposed module will be very much useful for developing a speech recognition system that facilitates the patients and persons with special needs for communication. The proposed module is also evaluated for its complexity which will be therefore efficient for low consumption of energy.

**Optical force and distance sensing in intraoral devices for stroke rehabilitation: a distance calibration and force classification approach,**

Stroke survivors often suffer from oro-facial impairments, affecting swallowing function and speech production. Measuring tongue pressure and position intraorally can help to improve therapy for both symptoms, but space inside the oral cavity is extremely limited and such devices can easily be prohibitively large and obstructive if too many sensors are needed. In this work, we present our efforts to sense the force of the tongue exerted against the hard palate and the tongue-palate distance, using only optical proximity sensors. To explore the feasibility and accuracy of this approach and to evaluate the selected sensor, we conducted a study with 10 subjects and measured the sensor's response to 10 discrete distances ranging from 0mm to 30mmbetween tongue and sensor, and to a continuously increasing tongue force against the sensor from 0.1N to 8N. For distance measurements, an existing in-situ calibration method was applied and verified that yielded errors of less than 2mm for the estimated distances in nearly every case. For force measurements, a Bayesian classification approach was adopted to map sensor data to two force regions (below and above a certain boundary value), where up to 84.1% (average: 71.7 %) of ADC values were classified correctly within-sample.

**Abstract of Presentations on March 9th,**

Presents abstracts for the articles comprising the conference proceedings.

**Speech task based automatic classification of ALS and Parkinson’s Disease and their severity using log Mel spectrograms,**

We consider the task of speech based classification of patients with amyotrophic lateral sclerosis (ALS), Parkinson's disease (PD) and healthy controls (HC). Recent work in convolutional neural networks (CNN) to solve image classification problems raises the possibility of utilizing spectral representation of speech for detection of neurological diseases. In this paper, a spectrogram based approach is used. Feeding overlapping windows to the CNN makes sure that the temporal aspects are considered by using short signal segments or wide analysis filters. A three class (ALS, PD or HC) dysarthria classification is performed. In addition, we perform two severity classification experiments for ALS (5 class) and PD (3 class) respectively. Experiments are conducted on both baseline MFCC data [1] and log Mel spectrograms. Classification results show that for several audio lengths, models trained on log Mel spectrograms consistently outperform those of MFCC's. The ability of the network to accurately classify different classes is evaluated via the area under receiver operating characteristic curve [2],[3]. The findings from this study could aid in better detection and monitoring of ALS and PD diseases.

**A Review and Classification of Amyotrophic Lateral Sclerosis with Speech as a Biomarker,**

Amyotrophic Lateral Sclerosis (ALS) is a motor system neurodegenerative disease that affects speech impairment, spinal, respiratory and swallowing difficulties in patients. It has gradually increased in elderly people in recent years and is not easy to diagnose. The ALS bulbar form system is based on detecting dysarthria speech classification in discriminating healthy subjects from ALS patients. To construct the classification model by using various machine learning techniques, the studies used datasets related to speech impairment recordings of ALS patients and healthy subjects. Early diagnosis of ALS can somewhat improve the patient’s quality of life to an extent. Sustained vowel phonation is very useful for classifying ALS and healthy control (HC). In this study, jitter and shimmer were used to extract features. A support vector machine classifier was applied, and it classified ALS/HC with an accuracy of 98.5%. This study presents a detailed review of various machine learning techniques applied to the speech signal for the diagnosis of ALS and their impact on future research in this direction.

**Teager Energy Cepstral Coefficients For Classification of Dysarthric Speech Severity-Level,**

Dysarthria is a neuro-motor speech impairment that renders speech unintelligibility, which is generally imperceptible to humans w.r.t severity-levels. Dysarthric speech classification acts as a diagnostic tool for evaluating the advancement in a patient's severity condition and also aids in automatic dysarthric speech recognition systems (an important assistive speech technology). This study investigates the significance of Teager Energy Cepstral Coefficients (TECC) in dysarthric speech classification using three deep learning architectures, namely, Convolutional Neural Network (CNN), Light-CNN (LCNN), and Residual Networks (ResNet). The performance of TECC is compared with state-of-the-art features, such as Short-Time Fourier Transform (STFT), Mel Frequency Cepstral Coefficients (MFCC), and Linear Frequency Cepstral Coefficients (LFCC). In addition, this study also investigate the effectiveness of cepstral features over the spectral features for this problem. The highest classification accuracy achieved using UA-Speech corpus is 97.18%, 94.63%, and 98.02% (i.e., absolute improvement of 1.98%, 1.41%, and 1.69%) with CNN, LCNN, and ResNet, respectively, as compared to the MFCC. Further, we evaluate feature discriminative capability using

**Sequence-to-Sequence Models in Italian Atypical Speech Recognition,**

In the domain of automatic speech recognition (ASR), we explore the usage of a state-of-the-art transformer-based sequence-to-sequence model to build a speaker-dependent isolated word recognizer for native Italian speakers with a speech disorder, such as dysarthria. In particular, this paper is concerned with a self-supervised learning approach, where the Wav2Vec2 has been fine-tuned on our private Italian corpus containing a total of 41 hours of speech contributions authored by 191 individuals with a disability and atypical speech. The discussed approach has been also evaluated thanks to collaboration of sixteen speakers with diverse degrees of speech disorders (mild, moderate, severe), and our analysis has shown a remarkable performance of our ASR system, with an overall word recognition accuracy of 97.4%.

**Monitoring Progress of Parkinson's Disease Based on Changes in Phonation: a Pilot Study,**

Hypokinetic dysarthria (HD) is a frequent symptom of idiopathic Parkinson's disease (PD). Although it is hypothesized its progress is tightly linked with changes in other motor/non-motor features of PD, it has not been proved yet. The aim of this work is to employ acoustic analysis of sustained phonation in order to identify significant correlates between phonatory measures and motor/non-motor deficits in a two-year follow-up study. For this purpose, we repeatedly quantified a sustained vowel/a/ in 51 PD patients who were also assessed by 5 common clinical scales. In addition, a multivariate regression model was trained to predict the motor/non-motor deficits in the horizon of two years. Results suggest that mainly instability in vocal folds oscillation increases with the progress of PD and with overall cognitive decline. Based on the acoustic analysis, the change in clinical scores could be predicted with the error in the range of 11.83-19.60 %.

**An automatic diagnosis and assessment of dysarthric speech using speech disorder specific prosodic features,**

To diagnose and classify the dysarthric speech, speech language pathologist (SLP) conducts a listening test. On the basis of the scores given by listeners the dysarthria is diagnosed and assessed. The above mentioned method is costly, time consuming and not very accurate. Unlike the traditional method, this research proposes an automatic diagnosis and assessment of dysarthria. The aim of this paper is to diagnose and classify the severity of dysarthria. The speech disorder specific prosodic features are selected by using genetic algorithm. The diagnosis and assessment of dysarthric speech is done by support vector machines. During diagnosis the classification accuracy of 98% has been achieved. And 87% of the dysarthric speech utterances are correctly classified. The standard UASPEECH database has been used in this work.

**Signal Analysis for Voice Evaluation in Parkinson’s Disease,**

Parkinson's Disease (PD) is a neurodegenerative disorder that is frequently correlated with vowel articulation difficulties. The phonation problem arises in patients affected by PD is commonly known as Parkinsonian Dysarthria and identifiedby vocal signal analysis. The analysis supporte physicians and specialists in early detection and monitoring of dysarthria aiming, to increase patients life quality and to evaluate the efficacy of treatments. We investigate on vocal signal analysis correlation with speech patterns related to PD. Vowel parameters are considered as discriminant elements among PD patients and healthy subjects. Aim of this work is to define possible indicators for dysarthria in PD patients.

**Two-stage and Self-supervised Voice Conversion for Zero-Shot Dysarthric Speech Reconstruction,**

Dysarthria is a motor speech disorder commonly associated with conditions such as cerebral palsy, Parkinson’s disease, amyotrophic lateral sclerosis, and stroke. Individuals with dysarthria typically exhibit significant speech difficulties, including imprecise articulation, lack of fluency, slow speech rate, and decreased volume and clarity, which can hinder their ability to communicate effectively with others. We propose a two-stage Voice Conversion method to enhance the reconstruction of dysarthric speech. In the first stage, we develop a KNN-VC approach based on a same-gender-retrieval strategy to preliminarily repair the dysarthric speech. In this stage, we match the dysarthric speech only with normal speech of the same gender. In the second stage, we adapt so-vits-svc to restore the speaker’s timbre and improve the sound quality of the speech repaired in the first stage. Both objective and subjective evaluations were conducted on the dataset of the Low Resource Dysarthria Wake-Up Word Spotting Challenge (LRDWWS Challenge) shows that the proposed approach can achieve some improvements in terms of speaker similarity, speech intelligibility and naturalness for unknown speakers, and these evaluations also show our method has a good Zero-shot performance. Our audio samples can be accessed online 1.

**Computer aided methods for diagnosis and therapy of speech breathing disorders,**

Computer-aided methods for the diagnosis and therapy of speech breathing disorders caused by brain damage (dysarthrias) are described. These methods make it possible to evaluate a speech disorder for diagnostic purposes using objective parameters. Special therapeutic tasks which proceed in accordance with the principle of biofeedback make it possible for the patient to carry out self-therapy through individualized exercise packages. Pathological speech patterns can be readily recognized, classified and quantified using the methods presented. It is concluded that these methods are clinically practical and show promise of becoming a useful clinical tool for the treatment of dysarthrias.<>

**Automatic detection of voice onset time in dysarthric speech,**

Although a number of speech disorders reflect varying involvement of brain areas, recently published automatic speech analyses have primarily been limited to hypokinetic dysarthria in Parkinson's disease (PD). Therefore, the aim of the present study was to provide an automatic algorithm suitable for the assessment of voice onset time (VOT) in various dysarthria types. Twenty-four PD participants with hypokinetic dysarthria and 40 Huntington's disease (HD) subjects with hyperkinetic dysarthria were included. These two types of dysarthria were selected in the design of a robust algorithm as they contain most of the dysarthric patterns found among all dysarthria subtypes. For a 10 ms threshold, the proposed algorithm reached approximately 90% accuracy in PD speakers and 80% accuracy in HD speakers. The accuracy of 80% obtained in HD was superior to the performance of 55% achieved by a previous algorithm designed particularly for hypokinetic dysarthria in PD.

**Hybrid CNN-GRU Model for Predicting Dysarthric Speech Using Deep Learning Approaches,**

This research study presents the novel method for dysarthria speech detection and classification. In order to create systems that are capable of accurately categorizing speech patterns that are impacted by dysarthria, it is vital to have speech detection and classification capabilities for dysarthria. The condition known as dysarthria, which is characterized by difficulties in articulation, phonation, resonance, and prosody, is a disorder that leads to a reduction in the intelligibility of speech. This method of classification necessitates the development of machine learning or deep learning models, more specifically a combination of Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU), in order to accomplish the objective of automatically classifying dysarthric speech into categories such as severity or presence/absence of dysarthria. The primary purpose of dysarthria speech categorization is to bring about the development of a reliable instrument that can be used for identifying, categorizing, and managing dysarthric speech. As an added benefit, the classification of dysarthric speech allows for the adaptation of treatment methods and the acquisition of information about the underlying causes of the condition. The individuals who are affected by this speech difficulty are the target audience for this strategy, which ultimately aims to enhance the quality of life of those individual. The study addresses the challenge of classifying dysarthric speech caused by neurological impairments using a Hybrid CNN-GRU model, achieving accuracies of 70.3% (high-severity), 38% (severe-moderate), 34.5% (moderate), and 10.3% (mild dysarthria).

**Utterance Verification-Based Dysarthric Speech Intelligibility Assessment Using Phonetic Posterior Features,**

r

**Smart Voice Assistance for Speech disabled and Paralyzed People,**

People who are paralyzed, confront numerous challenges in meeting their basic necessities on a daily basis. It is very difficult to understand the speech of people with dysarthria, amyotrophic lateral sclerosis (ALS) and similar conditions. Automatic speech command recognition system will enhance the lifestyle of people with voice disorder like dysarthria and paraplegics. The proposed work will convert the speech command of paralyzed people into text and send it to the care taker's mobile with the help of Twilio message services. Algorithms like Support Vector Machine (SVM) and Convolutional Neural network (CNN) model is used for speech command identification and speech to text conversion. CNN model yields an accuracy of 90.62%, whereas the SVM algorithm gives a very low accuracy. The developed TensorFlow model is deployed in the flask server.

**A Tool for Training Speech Imitation Accuracy,**

Dysarthria is a neurological motor speech disorder that commonly results in reduced intelligibility. Communication partners can learn to better understand the speech of someone with dysarthria through perceptual training. Vocal imitation of the degraded speech during perceptual training has been shown to elevate this learning. A tool that provides the learner with real-time feedback regarding the accuracy of their imitation attempts during training may further enhance this learning. We describe a training tool that compares dysarthric speech productions with the imitation attempts of healthy subjects, using a two-level dynamic warp that accounts for both spectral and temporal degradation. Feature vectors derived from both the spectrogram and LPC are examined.

**Leveraging OpenAI Whisper Model to Improve Speech Recognition for Dysarthric Individuals,**

Automatic Speech Recognition (ASR) systems are pivotal in facilitating human-technology interactions through voice commands. However, individuals with dysarthria face significant challenges in benefiting from these technologies due to their speech disorder. This paper proposes finetuning the Whisper model for Dysarthric Speech Recognition (DSR) by incorporating additional features extracted from Mel-frequency cepstral coefficients (MFCCs). By combining spectrograms and MFCCs within an attention mechanism, the model creates a richer feature representation, with spectrograms providing broader context and MFCCs highlighting crucial formant frequencies. The attention mechanism dynamically weighs the importance of each feature based on specific speech segments and dysarthric speech characteristics. Furthermore, a hierarchical attention approach is adopted, which encompasses a two-stage attention mechanism. This mechanism directs attention at both local and global levels, facilitating the capture of both fine-grained details and broader contextual information within the speech signal. This study involved the development and training of 45 speaker-adaptive dysarthric ASR systems. The proposed model achieves an average Word Recognition Accuracy (WRA) of 74.08%, showing a notable enhancement compared to the benchmark of 69.23%. The findings underscore the efficacy of the proposed approach in addressing dysarthria-related challenges in ASR systems.

**Empirical Analysis of Machine Learning Models on Parkinson’s Speech Dataset,**

Parkinson’s disease (PD) is a chronic and progressive neurodegenerative disorder that worsens over time. Diagnosing PD primarily relies on clinical assessments, which can be costly, time-consuming, and invasive. These evaluations may also be subjective and vulnerable to inaccuracy. Dysarthria, a common condition characterised by delayed and distorted speech, frequently coexists with PD. This opens up the possibility of using speech features for diagnostic reasons. This research paper explores different machine learning models trained on numerical data of changes in speech patterns due to Dysarthria. These models are based on classifiers such as Artificial Neural Networks (ANN), Multi-Layer Perceptron (MLP), Random Forests, and Decision Trees. Additionally, we compare the performance of a newly introduced HyperTab classifier with the existing models. Our findings demonstrate the significant potential of machine learning in diagnosing PD based on speech analysis. This progress holds the promise of creating a cost-effective tool to expedite disease detection. Furthermore, this research is of utmost importance in offering essential support to regions with limited access to specialized medical facilities.

**Investigation of Cross Modality Feature Fusion for Audio-Visual Dysarthric Speech Assessment,**

Dysarthria, a speech disorder resulting from neurological conditions, presents significant obstacles to speech intelligibility and daily communication. Automatic dysarthria assessment has the capability to provide low-cost diagnosis and treatment assistant support for such diseases as Parkinson's disease, Alzheimer's disease, and stroke. This study investigates the efficacy of cross-modality feature fusion using audio-visual data for the automatic assessment of dysarthric speech. Leveraging advanced self-supervised learning models, AV-HuBERT and Wav2Vec 2.0, we develop a multimodal system to enhance dysarthria severity classification. Utilizing the Mandarin Subacute Stroke Dysarthria Multimodal (MSDM) dataset, which includes synchronized audio and lip movement video recordings, our system achieves promising performance. Experimental results demonstrate that our back-end fusion and feature fusion approaches both outperform traditional single-modality methods, with the best back-end fusion system achieving a speaker-level F1 score of 0.841 while the best feature-level fusion system achieving a speaker-level F1 score of 0.772. This study marks the first application of pre-trained self-supervised learning models for multimodal dysarthria assessment, highlighting the potential for the assistance of diagnosis and treatment.

**Experimental Investigation on STFT Phase Representations for Deep Learning-Based Dysarthric Speech Detection,**

Mainstream deep learning-based dysarthric speech detection approaches typically rely on processing the magnitude spectrum of the short-time Fourier transform of input signals, while ignoring the phase spectrum. Although considerable insight about the structure of a signal can be obtained from the magnitude spectrum, the phase spectrum also contains inherent structures which are not immediately apparent due to phase discontinuity. To reveal meaningful phase structures, alternative phase representations such as the modified group delay (MGD) and instantaneous frequency (IF) spectra have been investigated in several applications. The objective of this paper is to investigate the applicability of the unprocessed phase, MGD, and IF spectra for dysarthric speech detection. Experimental results show that dysarthric cues are present in all considered phase representations. Further, it is shown that using phase representations as complementary features to the magnitude spectrum is beneficial for deep learning-based dysarthric speech detection, with the combination of magnitude and IF spectra yielding a high performance. The presented results should raise awareness in the research community about the potential of the phase spectrum for dysarthric speech detection and motivate research into novel architectures which optimally exploit magnitude and phase information.

**Speech Recognition-Based Feature Extraction For Enhanced Automatic Severity Classification in Dysarthric Speech,**

Due to the subjective nature of current clinical evaluation, the need for automatic severity evaluation in dysarthric speech has emerged. DNN models outperform ML models but lack user-friendly explainability. ML models offer explainable results at a feature level, but their performance is comparatively lower. Current ML models extract various features from raw waveforms to predict severity. However, existing methods do not encompass all dysarthric features used in clinical evaluation. To address this gap, we propose a feature extraction method that minimizes information loss. We introduce an ASR transcription as a novel feature extraction source. We finetune the ASR model for dysarthric speech, then use this model to transcribe dysarthric speech and extract word segment boundary information. It enables capturing finer pronunciation and broader prosodic features. These features demonstrated an improved severity prediction performance to existing features: balanced accuracy of 83.72%.

**The Nemours database of dysarthric speech,**

The Nemours database is a collection of 814 short nonsense sentences; 74 sentences spoken by each of 11 male speakers with varying degrees of dysarthria. Additionally, the database contains two connected-speech paragraphs produced by each of the 11 speakers. The database was designed to test the intelligibility of dysarthric speech before and after enhancement by various signal processing methods, and is available on CD-ROM. It can also be used to investigate general characteristics of dysarthric speech such as production error patterns. The entire database has been marked at the word level and sentences for 10 of the 11 talkers have been marked at the phoneme level as well. The paper describes the database structure and techniques adopted to improve the performance of a Discrete Hidden Markov Model (DHMM) labeler used to assign initial phoneme labels to the elements of the database. These techniques may be useful in the design of automatic recognition systems for persons with speech disorders, especially when limited amounts of training data are available.

**Corpus Design and Automatic Speech Recognition for Deaf and Hard-of-Hearing People,**

This study describes automatic speech recognition (ASR) for the deaf and hard-of-hearing people. In the relevant literature, ASR for the deaf has been studied in a manner similar to the recognition of speech by people with dysarthria. However, prior studies have been conducted over a small number of deaf speakers. Therefore, to date, it remains unclear how the performance of ASR varies with different speakers. We conducted phoneme recognition experiments using speech from 12 deaf students to obtain analytical results from the perspective of ASR performance.

**Inappropriate Pause Detection in Dysarthric Speech Using Large-Scale Speech Recognition,**

Dysarthria, a common issue among stroke patients, severely impacts speech intelligibility. Inappropriate pauses are crucial indicators in severity assessment and speech-language therapy. We propose to extend a large-scale speech recognition model for inappropriate pause detection in dysarthric speech. To this end, we propose task design, labeling strategy, and a speech recognition model with an inappropriate pause prediction layer. First, we treat pause detection as speech recognition, using an automatic speech recognition (ASR) model to convert speech into text with pause tags. According to the newly designed task, we label pause locations at the text level and their appropriateness. We collaborate with speech-language pathologists to establish labeling criteria, ensuring high-quality annotated data. Finally, we extend the ASR model with an inappropriate pause prediction layer for end-to-end inappropriate pause detection. Moreover, we propose a task-tailored metric for evaluating inappropriate pause detection independent of ASR performance. Our experiments show that the proposed method better detects inappropriate pauses in dysarthric speech than baselines. (Inappropriate Pause Error Rate: 14.47%)

**Using Automatic Speech Recognition to Measure the Intelligibility of Speech Synthesized From Brain Signals,**

Brain-computer interfaces (BCIs) can potentially restore lost function in patients with neurological injury. A promising new application of BCI technology has focused on speech restoration. One approach is to synthesize speech from the neural correlates of a person who cannot speak, as they attempt to do so. However, there is no established gold-standard for quantifying the quality of BCI-synthesized speech. Quantitative metrics, such as applying correlation coefficients between true and decoded speech, are not applicable to anarthric users and fail to capture intelligibility by actual human listeners; by contrast, methods involving people completing forced-choice multiple-choice questionnaires are imprecise, not practical at scale, and cannot be used as cost functions for improving speech decoding algorithms. Here, we present a deep learning-based “AI Listener” that can be used to evaluate BCI speech intelligibility objectively, rapidly, and automatically. We begin by adapting several leading Automatic Speech Recognition (ASR) deep learning models - Deepspeech, Wav2vec 2.0, and Kaldi - to suit our application. We then evaluate the performance of these ASRs on multiple speech datasets with varying levels of intelligibility, including: healthy speech, speech from people with dysarthria, and synthesized BCI speech. Our results demonstrate that the multiple-language ASR model XLSR-Wav2vec 2.0, trained to output phonemes, yields superior performance in terms of speech transcription accuracy. Notably, the AI Listener reports that several previously published BCI output datasets are not intelligible, which is consistent with human listeners.

**Contemporary speech/speaker recognition with speech from impaired vocal apparatus,**

Speech is the effective form of communication between human and its environment. Speech also has potential of being important mode of interaction with computer. This review paper deals with both speech and speaker recognition of persons with speech motor disorders. Normally speaker recognition consists of speaker verification and speaker identification. Speaker identification is the process of determining which registered speaker provides a given input sample. Speaker verification is the process of accepting or rejecting the identity claim of a speaker. On the other hand, the speech recognition system deals with the following challenges such as speech representation, feature extraction techniques, speech classifiers, databases and performance evaluation. Motor speech disorders are a class of speech disorder that disturbs the body's natural ability to speak. These disturbances vary in their etiology based on the integrity and integration of cognitive, neuromuscular, and musculoskeletal activities. There are various types of speech disorders like Apraxia, Cluttering (similar to stuttering), Dyspraxia, Dysarthria, Dysprosody and so on. The main objective of this review paper is to summarize and compare the well known methods used in various stages of speech and speaker recognition system.

**Speech training system based on resonant frequencies of vocal tract,**

Speech sounds are air pressure vibrations produced by air exhaled from the lungs and modulated and shaped by the vibrations of the glottal cords and the vocal tract as it is pushed out through the lips and nose. Speech signals, in addition to communicating the linguistic information, convey a multitude of other information including gender, age, accent, intent, emotion, humor and the state of health of the speaker. There are several neurological (e.g., aphasia, dysarthria, and apraxia) or anatomical (e.g., cleft lip and palate) factors that could affect the intelligibility and audibility of the human speech. In this paper we present our work in developing a training system based on the resonant frequencies (widely known as formants) of the vocal tract to help a human subject with speech impairment train himself or herself to improve the intelligibility and audibility of his or her speech.

**Analysis of adverse effects of stimulation during DBS surgery by patient-specific FEM simulations,**

Deep brain stimulation (DBS) represents today a well-established treatment for movement disorders. Nevertheless the exact mechanism of action of DBS remains incompletely known. During surgery, numerous stimulation tests are frequently performed in order to evaluate therapeutic and adverse effects before choosing the optimal implantation site for the DBS lead. Anatomical structures responsible for the induced adverse effects have been investigated previously, but only based on stimulation data obtained with the implanted DBS lead. The present study introduces a methodology to identify these anatomical structures during intraoperative stimulation tests based on patient-specific electric field simulations and visualization on the patient specific anatomy. The application to 4 patients undergoing DBS surgery and presenting dysarthria, paresthesia or pyramidal effects shows the different anatomical structures, which might be responsible for the adverse effects. Several of the identified structures have been previously described in the literature. To draw any statistically significant conclusions, the methodology has to be applied to further patients. Together with the visualization of the therapeutic effects, this new approach could assist the neurosurgeons in the future in choosing the optimal implant position.

**Targeting Friedreich Ataxia: A Sustainable Path to Safer and Smarter Therapeutics Through Integrated Docking and Toxicology,**

Friedreich's ataxia (FA) is an autosomal recessive disorder affecting the nervous and cardiovascular systems, characterized by progressive ataxia, dysarthria, and muscle weakness. It results from a GAA trinucleotide repeat expansion in the FXN gene, leading to epigenetic silencing via heterochromatin formation, which reduces frataxin production-a mitochondrial protein essential for iron-sulphur cluster biogenesis and cellular energy production. Frataxin deficiency causes oxidative stress and mitochondrial dysfunction. Current treatments offer limited effectiveness, primarily addressing symptoms. Our study aims to develop novel therapeutic candidates by leveraging computational protein-ligand docking through the Galaxy EU platform, targeting frataxin deficiencies. High-throughput docking and toxicology studies have identified several promising compounds with potential therapeutic efficacy, offering hope for more effective and sustainable treatments for FA.

**Check Your Audio Data: Nkululeko for Bias Detection,**

We present a new release of the software tool Nkululeko. New additions enable users to automatically perform sanity checks, data cleaning, and bias detection in the data based on machine learning predictions. Two open-source databases from the medical domain are investigated: the Androids de-pression corpus and the UASpeech dysarthria corpus. Results show that both databases have some bias, but not in a severe manner.

**Extraction of patients subpopulations with psychiatric symptoms using a transformer architecture,**

In this paper, we demonstrate a novel pipeline for identifying and extracting patient subpopulations from unstructured physician’s notes. We validate the method by extracting patients with psychiatric issues from a general patient population. This method first uses a clinical metathesaurus to select terms of interest from reports, then vectorizes the terms using a transformer model. These vectors’ dimensions are reduced using Uniform Manifold Approximation and Projection (UMAP), and the results grouped by optimal cluster selection methods. We demonstrate this technique on a freely-available collection of deidentified patient notes (MIMIC IV), extracting and clustering “mental or behavioral dysfunctions”. Our results show that it is possible to select user-defined groups of patients from unstructured text with minimal model oversight to group patients with similar profiles. In our study cohort, the models automatically segmented the patients into two groups: patients with more physical symptoms (alcohol/drug abuse, dysarthria, tongue-biting, eating disorders) and patients with mental/emotional symptoms. By detecting the underlying similarities in patient profiles, we believe this method can be utilized for symptom prediction tasks, as well as curating treatment plans based on their cluster profiles. Such a system can assist in clinical decision-making without the need for individually-created NLP models.

**Signal Recognition for Parkinson's Disease Diagnosis Based on Multi-Source Deep Domain Generalization,**

Parkinson's disease (PD) is an incurable neurodegenerative disorder. Dysarthria, stemming from PD, in speech offers an accessible and non-invasive diagnostic indicator. However, PD speech data, with its limited sample size and high aliasing, presents challenges for machine learning. Thought multi-source deep transfer learning methods show promise in PD speech recognition, they typically rely on labelled target domain data. To overcome this limitation, our study proposes an unsupervised transfer learning approach, namely multi-source deep domain generalization (MDDG). MDDG comprises four key modules: feature extraction, inter-domain and classes adversarial learning, and classifier training solely on multisource datasets. Leveraging adversarial networks, MDDG extracts invariant features from source domains, minimizing both inter-domain distribution discrepancies and intra-class difference while maximizing the distance between different classes from multi-source domains. Experimental results demonstrate the superiority of the MDDG over recent methodologies, where MDDG achieves remarkable metrics: highest accuracy (75.80%), precision (66.48%), and specificity (87.61%), all with the lowest standard error, underscoring the effectiveness and stability of MDDG offering valuable support to medical professionals in efficient PD diagnosis and monitoring.

**Notice of Violation of IEEE Publication Principles: A voice-input voice-output communication assists in favor of people with harsh verbal communication destruction,**

An actual VOICE INPUT VOICE OUTPUT Correspondence Help identifies the actual disordered talk from the person as well as develops communications that are changed into artificial talk. Tests demonstrated this technique works within producing great acknowledgement overall performance (mean precision 98 percentage) upon extremely disordered talk, even if acknowledgement perplexity is actually elevated. The actual VOICE INPUT VOICE OUTPUT Communication Assists (VIVOCA) had been examined inside an area test through people with reasonable in order to serious dysarthria as well as verified that they'll utilize the gadget to create intelligible talk result through disordered talk enter. The actual test outlined a few problems that restrict the actual overall performance as well as user friendliness from the gadget whenever used within actual utilization circumstances, along with imply acknowledgement precision associated with 85 percentages within these types of conditions. Once after receiving the clear speech in English as input, the same can be translated to any other language as an output, with the same meaning as it is in input. These types of restrictions are going to be tackled within long term function.

**A Wearable System for Monitoring Neurological Disorder Events with Multi-Class Classification Model in Daily Life,**

Dysphagia and dysarthria are the prominent sequelae of neurological disorders. Treatment and rehabilitation of these impairments necessitate continuously monitoring symptoms related to swallowing and speaking. However, current medical technologies require large and diverse equipment to record these symptoms, which are predominantly limited to clinical environments. In this study, we propose an innovative wearable system for distinguishing neurological disorder events using a mechano-acoustic (MA) sensor and multi-class ensemble classification model. The MA sensor exhibits a high sensitivity to neck vibration without any interference from ambient sounds. A multi-class classification model was also developed to discern the symptoms from the recorded signals accurately. The proposed classification model is an ensemble neural network trained on waveforms and mel spectrograms. As a result, we achieve a high classification accuracy of 91.94%, surpassing the performance of previous single neural networks.

**Tongue-Trackpad: Tongue Rehabilitation Through Quantitative Personalized Visual-Feedback Intervention,**

Effective rehabilitation of tongue movement relies on delivering personalized interventions and quantitatively assessing the intervention's impact. Current approaches primarily focus on enhancing tongue movement control and strength through oromotor exercises and audio feedback. Here, we introduce the Tongue- Trackpad a novel solution for tongue movement rehabilitation and progress tracking. The Tongue- Trackpad is a wireless intra-oral device that provides real-time visual feedback of tongue movement while quantifying the movement, thus enabling personalized interventions. In this preliminary feasibility study, a stroke survivor diag-nosed with dysarthria participated in eighteen sessions of per-sonalized visual-feedback pursuit intervention using the Tongue-Trackpad. The intervention was customized to the participant's tongue movement deficit areas identified during the initial evaluation. Despite the participant's severe speech disorder, characterized by a Diadochokinetic rate of zero, the results indicated a modest positive change in tongue movement both within a single session and over the intervention period. The results showed an average increase of 3.5

**Classifying Speech Disorders Using Voice Signals and Machine Learning,**

This study explores the application of machine learning (ML) and advanced voice signal analysis to classify speech disorders, including vocal tremors, dysarthria, and stuttering. These disorders pose significant challenges to communication and quality of life, requiring precise and reliable diagnostic tools. To address the inherent challenges of limited data availability in this domain, the study employs synthetic voice data generated using mathematical models to augment real-world recordings. This hybrid approach ensures a more comprehensive dataset, enabling robust training and evaluation of machine learning models.Key machine learning algorithms such as Support Vector Machines (SVMs), Random Forest, and Gradient Boosting are utilized to extract and analyze a wide range of acoustic features from speech samples. These methods are selected for their ability to handle complex, nonlinear patterns and to identify subtle distinctions between normal and disordered speech. By leveraging these techniques, the study investigates their performance in accurately classifying speech disorders, emphasizing their potential in improving diagnostic outcomes.The findings highlight the transformative potential of machine learning in the field of speech pathology. ML models demonstrated notable improvements in the precision, efficiency, and consistency of diagnosing speech disorders compared to traditional methods. This breakthrough not only enhances diagnostic reliability but also equips clinicians with objective tools that reduce subjectivity in assessments. Moreover, the integration of these technologies promotes earlier detection of disorders, enabling timely and tailored therapeutic interventions.By facilitating more accurate and accessible diagnostic practices, this work paves the way for significant advancements in patient care. The use of ML-powered tools can bridge the gap between clinical expertise and technological innovation, empowering healthcare professionals to offer personalized treatment plans. Ultimately, this approach has the potential to improve patient outcomes, reduce the stigma associated with speech disorders, and ensure more equitable access to high-quality care for individuals worldwide.

**Speech Formation Objectivisation Methods Development and Analysis for Correct Phonation Diagnostics,**

In connection with various infringements speech production people have requirement for development of methods of an objective estimation of the vocal apparatus functional condition. Methods should be directed on revealing of the attributes forming articulation-acoustic characteristics of correct harmonization of themes. The given problem is especially important for children for whom increase speech pathologies (legasthenia, dysgraphia, dysarthria, phonasthenia, hoarseness, etc.) is observed. As results of the carried out research, it is possible to come to a conclusion that for the analysis of voice pathologies it is expedient to use the offered method.

**Synthesizing Dysarthric Speech Using Multi-Speaker Tts For Dysarthric Speech Recognition,**

Dysarthria is a motor speech disorder often characterized by reduced speech intelligibility through slow, uncoordinated control of speech production muscles. Automatic Speech recognition (ASR) systems may help dysarthric talkers communicate more effectively. To have robust dysarthria-specific ASR, sufficient training speech is required, which is not readily available. Recent advances in Text-To-Speech (TTS) synthesis multi-speaker end-to-end systems suggest the possibility of using synthesis for data augmentation. In this paper, we aim to improve multi-speaker end-to-end TTS systems to synthesize dysarthric speech for improved training of a dysarthria-specific DNN-HMM ASR. In the synthesized speech, we add dysarthria severity level and pause insertion mechanisms to other control parameters such as pitch, energy, and duration. Results show that a DNN-HMM model trained on additional synthetic dysarthric speech achieves WER improvement of 12.2% compared to the baseline, the addition of the severity level and pause insertion controls decrease WER by 6.5%, showing the effectiveness of adding these parameters. Audio samples are available at https://mohammadelc.github.io/SpeechGroupUKY/

**The Change of Vocal Tract Length in People with Parkinson’s Disease,**

Hypokinetic dysarthria is one of the early symptoms of Parkinson’s disease (PD) and has been proposed for early detection and also for monitoring of the progression of the disease. PD reduces the control of vocal tract muscles such as the tongue and lips and, therefore the length of the active vocal tract is altered. However, the change in the vocal tract length due to the disease has not been investigated. The aim of this study was to determine the difference in the apparent vocal tract length (AVTL) between people with PD and age-matched control healthy people. The phoneme, /a/ from the UCI Parkinson’s Disease Classification Dataset and the Italian Parkinson’s Voice and Speech Dataset were used and AVTL was calculated based on the first four formants of the sustained phoneme (F1-F4). The results show a correlation between Parkinson’s disease and an increase in vocal tract length. The most sensitive feature was the AVTL calculated using the first formants of sustained phonemes (F1). The other significant finding reported in this article is that the difference is significant and only appeared in the male participants. However, the size of the database is not sufficiently large to identify the possible confounding factors such as the severity and duration of the disease, medication, age, and comorbidity factors.Clinical relevance—The outcomes of this research have the potential to improve the identification of early Parkinsonian dysarthria and monitor PD progression.

**Weak-Supervised Dysarthria-Invariant Features for Spoken Language Understanding Using an Fhvae and Adversarial Training,**

The scarcity of training data and the large speaker variation in dysarthric speech lead to poor accuracy and poor speaker generalization of spoken language understanding systems for dysarthric speech. Through work on the speech features, we focus on improving the model generalization ability with limited dysarthric data. Factorized Hierarchical Variational Auto-Encoders (FHVAE) trained unsupervisedly have shown their advantage in disentangling content and speaker representations. Earlier work showed that the dysarthria shows in both feature vectors. Here, we add adversarial training to bridge the gap between the control and dysarthric speech data domains. We extract dysarthric and speaker invariant features using weak supervision. The extracted features are evaluated on a Spoken Language Understanding task and yield a higher accuracy on unseen speakers with more severe dysarthria compared to features from the basic FHVAE model or plain filterbanks.

**Dysarthric Speech Detection Using Hybrid Models,**

Dysarthria is a speech disorder caused by weak or poorly coordinated speech-related muscles. Dysarthria can be caused by various factors such as stroke, multiple sclerosis, or cerebral palsy, as well as brain injury or certain medications. The ultimate goal of our paper is to make a suitable and accurate tool for detecting dysarthric speech that can be used in clinical settings for early diagnosis, treatment planning for dysarthric individuals. In this work, the machine learning models namely CNN (Convolutional Neural Network), and hybrid models, such as CNN, combined with LSTM (Long Short-Term Memory), and CNN combined with GRU (Gated Recurrent Unit) are trained on TORGO dataset, to classify the dysarthric speech from non-dysarthric speech. The model's performance is evaluated on the testing data and all of these models produce more than 95% accuracy rate. From this study, we understood that this methodology can be used in developing Automatic Speech Recognition(ASR) for people with dysarthria, which would become an essential technology for many applications.

**Towards Improving the Performance of Dysarthric Speech Severity Assessment System,**

Dysarthria is a kind of speech disorder caused by difficulties in muscular control of the speech mechanism, owing to the impairment of the central or peripheral nervous systems. Dysarthria is a speaking impairment due to the weak muscles of the patient due to brain injury. Face, throat muscles, lips are important to generate intelligible speech in humans. The severity of the disorder decides the understandability of the dysarthric speech, as the speech can be slow and unclear. This affects dysarthric people's life. This creates a communication gap between the common people and dysarthric patients. This work proposes a prototype that will improve the intelligibility of dysarthric speech. The main objective of the prototype is to enhance the rehabilitation of dysarthric people's livelihood. From the severity assessment, the clinician and caretakers can provide an improved speech treatment for dysarthric patients. Audio features like MFCC (Mel Frequency Spectral Coefficients), ZCR (Zero Crossing Rate), Spectral centroid, spectral rolloff, Mel spectrogram are used for feature extraction process. The KNN (K-nearest neighbor) and SVM (Support Vector Machine) classifiers are used to categorize the severity level of dysarthria based upon the features obtained.

**Effects of acoustic features modifications on the perception of dysarthric speech — Preliminary study (Pitch, intensity and duration modifications),**

Marking stress is important in conveying meaning and drawing listener's attention to specific parts of a message. Extensive research has shown that healthy speakers mark stress using three main acoustic cues; pitch, intensity, and duration. The relationship between acoustic and perception cues is vital in the development of a computer-based tool that aids the therapists in providing effective treatment to people with Dysarthria. It is, therefore, important to investigate the acoustic cues deficiency in dysarthric speech and the potential compensatory techniques needed for effective treatment. In this paper, the relationship between acoustic and perceptive cues in dysarthric speech are investigated. This is achieved by modifying stress marked sentences from 10 speakers with Ataxic dysarthria. Each speaker produced 30 sentences using the 10 SubjectVerb-Object-Adjective (SVOA) structured sentences across three stress conditions. These stress conditions are stress on the initial (S), medial (O) and final (A) target words respectively. To effectively measure the deficiencies in Dysarthria speech, the acoustic features (pitch, intensity, and duration) are modified incrementally. The paper presents the techniques involved in the modification of these acoustic features. The effects of these modifications are analysed based on steps of 25% increments in pitch, intensity and duration. For robustness and validation, 50 untrained listeners participated in the listening experiment. The results and the relationship between acoustic modifications (what is measured) and perception (what is heard) in Dysarthric speech are discussed.

**FBSE-FTFCWT-Based Novel Automated Framework for Dysarthric Speech Detection,**

Neurological injuries or neurodegenerative diseases can lead to dysarthria, a condition that impairs speech intelligibility. Accurate detection of dysarthria and its severity from speech signals are crucial for advancing smart healthcare solutions. This study presents an automated system for dysarthria detection and severity classification, using a Fourier-Bessel series expansion-based flexible time-frequency coverage wavelet transform (FBSE-FTFCWT) and an autoencoder. Initially, FBSE-FTFCWT decomposes the speech signal into 16 sub-band signals, which are used as an input in the form of tensor for autoencoders to generate latent representation. This latent representation is subsequently used for dysarthric speech and its severity level detection. The proposed framework outperformed the current state-of-the-art in classifying dysarthric and normal speech on the UA-speech dataset, achieving 3.28% higher accuracy. Additionally, for the dysarthric severity detection task using speech signals from the same dataset, it showed 3.1% improvement in accuracy over existing methods.

**Study of the Automatic Detection of Parkison’s Disease Based on Speaker Recognition Technologies and Allophonic Distillation,**

The use of new tools to detect Parkinson's Disease (PD) from speech articulatory movements can have a considerable impact in the diagnosis of patients. In this study, a novel approach involving speaker recognition techniques with allophonic distillation is proposed and tested separately in four parkinsonian speech databases (205 patients and 186 controls in total). This new scheme provides values between 72% and 94% of accuracy in the automatic detection of PD, depending on the database, and improvements up to 9% respect to baseline techniques. Results not only point towards the importance of the segmentation of the speech for the differentiation of parkinsonian and control speakers but confirm previous findings about the relevance of plosives and fricatives in the detection of parkinsonian dysarthria.

**Monitoring the Effect of Levodopa Using Sustained Phonemes in Parkinson’s Disease Patients,**

Parkinson's disease (PD) is a neurological disease identified by multiple symptoms, and levodopa is one of the most effective medications for treating the disease. To determine the dosage of levodopa, it is necessary to meet on a regular basis and observe motor function. The early detection and progression of the disease have been proposed using hypokinetic dysarthria. However, previous studies have not examined the effects of levodopa on speech rigorously and have provided inconsistent results. In this study, three sustained phonemes of PD patients were investigated for the effect of medication. A set of features characterizing vocal fold dynamics as well as the vocal tract coordinators were extracted from the sustained phonemes /of 28 PD patients during levodopa medication off and on states. All the features were statistically investigated and classified using a linear discriminant analysis (LDA) classifier. LDA classifier identified medication on from medication off based on the combined features from phoneme /a/, /o/ and /m/ with the accuracy=82.75% and F1-score=82.18%. Voice recording of PD patients during sustained phonemes /a/, /o/ and /m/ has the potential for identifying whether the patients are in On state or Off state of medication.Clinical Relevance— The outcomes of this study have the potential to monitor the effect and progress of levodopa on PD patients.

**Adapting acoustic and lexical models to dysarthric speech,**

Dysarthria is a motor speech disorder resulting from neurological damage to the part of the brain that controls the physical production of speech. It is, in part, characterized by pronunciation errors that include deletions, substitutions, insertions, and distortions of phonemes. These errors follow consistent intra-speaker patterns that we exploit through acoustic and lexical model adaptation to improve automatic speech recognition (ASR) on dysarthric speech. We show that acoustic model adaptation yields an average relative word error rate (WER) reduction of 36.99% and that pronunciation lexicon adaptation (PLA) further reduces the relative WER by an average of 8.29% on a large vocabulary task of over 1500 words for six speakers with severe to moderate dysarthria. PLA also shows an average relative WER reduction of 7.11% on speaker-dependent models evaluated using 5-fold cross-validation.

**Improving the intelligibility of dysarthric speech towards enhancing the effectiveness of speech therapy,**

Dysarthria is a neuro-motor disorder in which the muscles used for speech production and articulation are severely affected. Dysarthric patients are characterized by slow or slurred speech that is difficult to understand. This work aims at enhancing the intelligibility of dysarthric speech towards developing an effective speech therapy tool. In this therapy tool, enhanced speech is used for providing auditory feedback with a delay to instill confidence in the patients, so that they can improve their speech intelligibility gradually through relearning. Feature level transformation techniques based on linear predictive coding (LPC) coefficient mapping and frequency warping of LPC poles are experimented in this work. Speech utterances from Nemours dataset with mild and moderate dysarthria are used to study the effectiveness of the proposed algorithms. The quality of the transformed speech is evaluated using subjective and objective measures. A significant improvement in the intelligibility of speech was observed. Our method henceforth could be used to enhance the effectiveness of speech therapy, by encouraging the dysarthric patients talk more, thus helping in their fast rehabilitation.

**A Novel Approach for Intelligibility Assessment in Dysarthric Subjects,**

Dysarthria is a motor speech impairment caused by muscle weakness. Individuals, with this condition, are unable to control rapid movement of the velum leading to reduction in intelligibility, audibility, naturalness and efficiency of vocal communication. Systems that can assess intelligibility of dysarthric speech can help clinicians diagnose the impact of therapy and medication. In the paper, we propose a usable novel method to assess intelligibility of dysarthric speakers. The approach is based on the observation that the performance of a speech recognition engine deteriorates with increase in severity of the disorder. The mismatch between the original word and the recognized string is exploited to compute the dysarthria intelligibility score. Experiments on UA speech corpus show that the computed intelligibility score exhibits a significant correlation with perceptually assessed intelligibility scores. We further show that a small set of words spoken by the dysarthric subject is sufficient to assess the speech intelligibility reliably.

**Intelligibility modification of dysarthric speech using HMM-based adaptive synthesis system,**

Dysarthria is a manifestation of an inability to control and coordinate on one or more articulatory subsystems, which results in poorly articulated, slurred, and unintelligible speech. In order to enable a dysarthric speaker to communicate more efficiently with others, a text-to-speech synthesis system that generates speech in his voice, but without the errors he makes would be desirable. In this regard, the current work proposes a system, where the dysarthric speech is first recognized by an HMM-based speech recognition system. A sentence-level network is used to ensure 100% recognition accuracy. The recognized text is then synthesized by a speech synthesis system adapted to the dysarthric speaker's voice. This system replaces the sound units wrongly uttered by the dysarthric speaker, thereby improving intelligibility. The rate of synthesized speech is quite low for speakers with moderate and severe dysarthria. Therefore, the speech rate is modified using time-domain pitch synchronous overlap add (TD-PSOLA) technique. Degradation mean opinion score (DMOS) is used to prove that wrongly uttered sound units are replaced by correct sound units and that the synthetic speech is made more intelligible with the speaker's identity.

**Enhancement of dysarthric speech for developing an effective speech therapy tool,**

Dysarthria is a neuromuscular disorder that results from weakened movement of muscles used in speech production. This results in poor articulation causing the dysarthric speech to be slurred and difficult to understand. The natural auditory feedback makes the patients understand that their speech is of low quality, and this lowers their self confidence and they become more and more introverted causing the disorder to aggravate. In this work, we enhance the dysarthric speech and provide the enhanced speech to the patient through auditory feedback. This helps the patients to feel comfortable with their speech and gradually develop confidence to speak more and hence achieve a speedy rehabilitation. The utterances are analyzed using linear predictive coding (LPC). The LPC features in the acoustic space of dysarthric speaker are mapped to the feature space of the normative population using constrained maximum likelihood linear regression (CMLLR) before they are used for re-synthesising the enhanced speech. We then evaluated the quality of the enhanced speech using subjective and objective measures, DMOS and PESQ, and obtained an improvement of 63% and 43.4% for DMOS and PESQ measures respectively. Clinical trials are being pursued at Amrita Institute of Medical Sciences, Kochi on patients with dysarthria to evaluate the effectiveness of the proposed approach for faster rehabilitation of the patients. The results of these clinical trials will be reported in due course of time.

**Cochlear Filter-Based Cepstral Features for Dysarthric Severity-Level Classification,**

Severity-level classification of dysarthria helps in diagnosing a patient and choosing an appropriate course of treatment. This would also aid in redirecting the speech to an appropriate dysarthric Automatic Speech Recognition (ASR), as traditional ASR does not perform well on dysarthric speech. In the recent past, several approaches have been used to study the severity-level classification of dysarthria using state-of-the-art features, such as Short-Time Fourier Transform (STFT) and Mel Frequency Cepstral Coefficients (MFCC). This study investigates novel auditory transform-based Cochlear Filter Cepstral Coefficients (CFCC) features for dysarthric severity-level classification. Three DNN-based classifiers, namely, Convolutional Neural Network (CNN), Light-CNN (LCNN), and Residual Neural Network (ResNet) were employed on UA-Speech Corpus and TORGO corpus. Our proposed CFCC feature set yields an improved classification accuracy of 97.46% (98.99%), 94.92% (94.97%), and 96.66% (98.93%) on UA (Torgo)-corpus using CNN, LCNN and ResNet classifiers respectively. Furthermore, performance metrics, such as the Jaccard index, Matthew's Correlation Coefficient (MCC),

**Dysarthric Speech Recognition using Depthwise Separable Convolutions: Preliminary Study,**

As a neurological disability that affects muscles involved in articulation, dysarthria is a speech impairment that leads to reduced speech intelligibility. In severe cases, these individuals could also be handicapped and unable to interact with digital devices. For such individuals, Automatic Speech Recognition (ASR) technologies could be life changing by enabling them to communicate with others as well as computing devices via voice commands. Nonetheless, ASR systems designed to recognize healthy speech have shown very poor performance to transcribe dysarthric speech, signaling the need to design ASR specifically tailored for dysarthria. Dysarthric Speech Recognition (DRS) research has progressed gradually because of the challenges the research community faces such as the scarcity of dysarthric speech that does not allow the researchers to design deeper acoustic models needed to better learn dysarthric speech variations. In this paper we report on our preliminary findings to improve our previous DSR called Speech Vision and study the effects of Separable Convolutional neurons to improve its acoustic model. Speech Vision is a novel Dysarthric Speech Recognition system that learns to recognize the shape of the words uttered by dysarthric speakers instead of recognizing phone sequences and then mapping them to words. Experiments conducted on the utterances provided by all UA-Speech dysarthric speakers indicate the proposed Depthwise separable architecture provided better word recognition accuracies compared to the original Speech Vision’s architecture across all dysarthric speech intelligibility classes.

**Feature selection in Parkinson's disease: A rough sets approach,**

Parkinson's disease is a neurodegenerative disorder with a long time course and a significant prevalence, which increases significantly with age. Although the etiology is currently unknown, the disease presents with neurodegeneration of regions of the basal ganglia. the onset occurs later in life, and the disease progresses slowly. The disease is diagnosed clinically, requiring the identification of several factors such as distal resting tremor, rigidity, and bradykinesia. The common thread throughout the range of symptoms is motor dysfunction, and recent reports have focused on dysphonia, the impairment in voice production as a diagnostic measure. In this paper, a number of features associated with speech have been collected through clinical studies from both healthy and people with Parkinson's (PWP) and analysed in order to determine if one or more of them can be used to diagnose PWP. The feature set is analysed using the rough sets paradigm, which maps feature vectors associated with objects onto decision classes. The results from applying rough sets is a set of rules that map features via rules into a decision support system - performing classification of objects. the results FOM this study indicate that a subset of typical voice derived features is adequate to differentiate healthy from PWP with 100% accuracy. These result are important in that they imply that a diagnosis can be automated and performed remotely. This work will be extended to determine if this approach can be utilised with the same effectiveness for the diagnosis of parkinsonism disorders - a collection-diseases with Parkinson's like symptoms.

**Universal Paralinguistic Speech Representations Using self-Supervised Conformers,**

Many speech applications require understanding aspects beyond the words being spoken, such as recognizing emotion, detecting whether the speaker is wearing a mask, or distinguishing real from synthetic speech. In this work, we introduce a new state-of-the-art paralinguistic representation derived from large-scale, fully self-supervised training of a 600M+ parameter Conformer-based architecture. We benchmark on a diverse set of speech tasks and demonstrate that simple linear classifiers trained on top of our time-averaged representation outperform nearly all previous results, in some cases by large margins. Our analyses of context-window size demonstrate that, surprisingly, 2 second context-windows achieve 96% the performance of the Conformers that use the full long-term context on 7 out of 9 tasks. Furthermore, while the best per-task representations are extracted internally in the network, stable performance across several layers allows a single universal representation to reach near optimal performance on all tasks.

**Feature extraction using pre-trained convolutive bottleneck nets for dysarthric speech recognition,**

In this paper, we investigate the recognition of speech uttered by a person with an articulation disorder resulting from athetoid cerebral palsy based on a robust feature extraction method using pre-trained convolutive bottleneck networks (CBN). Generally speaking, the amount of speech data obtained from a person with an articulation disorder is limited because their burden is large due to strain on the speech muscles. Therefore, a trained CBN tends toward overfitting for a small corpus of training data. In our previous work, the experimental results showed speech recognition using features extracted from CBNs outperformed conventional features. However, the recognition accuracy strongly depends on the initial values of the convolution kernels. To prevent overfitting in the networks, we introduce in this paper a pre-training technique using a convolutional restricted Boltzmann machine (CRBM). Through word-recognition experiments, we confirmed its superiority in comparison to convolutional networks without pre-training.

**Modeling pathological speech perception from data with similarity labels,**

The current state of the art in judging pathological speech intelligibility is subjective assessment performed by trained speech pathologists (SLP). These tests, however, are inconsistent, costly and, oftentimes suffer from poor intra- and inter-judge reliability. As such, consistent, reliable, and perceptually-relevant objective evaluations of pathological speech are critical. Here, we propose a data-driven approach to this problem. We propose new cost functions for examining data from a series of experiments, whereby we ask certified SLPs to rate pathological speech along the perceptual dimensions that contribute to decreased intelligibility. We consider qualitative feedback from SLPs in the form of comparisons similar to statements “Is Speaker A's rhythm more similar to Speaker B or Speaker C?” Data of this form is common in behavioral research, but is different from the traditional data structures expected in supervised (data matrix + class labels) or unsupervised (data matrix) machine learning. The proposed method identifies relevant acoustic features that correlate with the ordinal data collected during the experiment. Using these features, we show that we are able to develop objective measures of the speech signal degradation that correlate well with SLP responses.

**Towards a clinical tool for automatic intelligibility assessment,**

An important, yet under-explored, problem in speech processing is the automatic assessment of intelligibility for pathological speech. In practice, intelligibility assessment is often done through subjective tests administered by speech pathologists; however research has shown that these tests are inconsistent, costly, and exhibit poor reliability. Although some automatic methods for intelligibility assessment for telecommunications exist, research specific to pathological speech has been limited. Here, we propose an algorithm that captures important multi-scale perceptual cues shown to correlate well with intelligibility. Nonlinear classifiers are trained at each time scale and a final intelligibility decision is made using ensemble learning methods from machine learning. Preliminary results indicate a marked improvement in intelligibility assessment over published baseline results.

**Semantic Analysis of NIH Stroke Scale using Machine Learning Techniques,**

In particular, stroke is a major disease leading to death in adults and elderly people, as well as disability. Rapid detection of stroke is very difficult because the cause and cause of the onset are different for each individual. In this paper, we design and implement a system for semantic analysis of early detection of stroke and recurrence of stroke in Koreans over 65 years old, based on the National Institutes of Health (NIH) Stroke Scale. Using C4.5 of the decision tree series represented by the analytics algorithm of machine learning technique, we conduct a semantic interpretation that analyzes and extracts the semantic rules of the execution mechanism that are additionally provided by C4.5. The C4.5 algorithm is used to construct a classification and prediction model using the information gain of the NIH stroke scale features, and to obtain additional NIH Stroke Scale feature reduction effects.

**Online speaking rate estimation using recurrent neural networks,**

A reliable online speaking rate estimation tool is useful in many domains, including speech recognition, speech therapy intervention, speaker identification, etc. This paper proposes an online speaking rate estimation model based on recurrent neural networks (RNNs). Speaking rate is a long-term feature of speech, which depends on how many syllables were spoken over an extended time window (seconds). We posit that since RNNs can capture long-term dependencies through the memory of previous hidden states, they are a good match for the speaking rate estimation task. Here we train a long short-term memory (LSTM) RNN on a set of speech features that are known to correlate with speech rhythm. An evaluation on spontaneous speech shows that the method yields a higher correlation between the estimated rate and the ground-truth rate when compared to the state-of-the-art alternatives. The evaluation on longitudinal pathological speech shows that the proposed method can capture long-term and short-term changes in speaking rate.

**Breath to speech communication with fall detection for elder/patient with take care analytics,**

People suffering from Developmental-Disabilities are almost entirely paralyzed disabling them to communicate in any way except using an Augmentative and Alternative Communication device. Survey analysis tells that 1.4% of globe's population suffers from speech disorder which is more than the Karnataka's population. Looking into the elderly group it was analyzed that the fall events cannot be predicted and might be an unsafe event. Estimates tell that 33.33% of 65 and above aged people fall every year. It can be seen that out of these falls 55% occur at home and 23% occur near the home. Hence, a dependable fall detection system has to be developed, and commercially be used all over the globe among the elderly. Depending on fast detection and delivering signals, the cost of the system can be reduced which is interconnected to the reaction and saving time. An enhanced breathe to speech communication and fall detection system for elderly people and also monitoring through a take care analytics is suggested that are based on intelligent sensors that are put by the person using that device.

**Raw Source and Filter Modelling for Dysarthric Speech Recognition,**

Acoustic modelling for automatic dysarthric speech recognition (ADSR) is a challenging task. Data deficiency is a major problem and substantial differences between the typical and dysarthric speech complicates transfer learning. In this paper, we build acoustic models using the raw magnitude spectra of the source and filter components. The proposed multi-stream model consists of convolutional and recurrent layers. It allows for fusing the vocal tract and excitation components at different levels of abstraction and after per-stream pre-processing. We show that such a multi-stream processing leverages these two information streams and helps s model towards normalising the speaker attributes and speaking style. This potentially leads to better handling of the dysarthric speech with a large inter-speaker and intra-speaker variability. We compare the proposed system with various features, study the training dynamics, explore usefulness of the data augmentation and provide interpretation for the learned convolutional filters. On the widely used TORGO dysarthric speech corpus, the proposed approach results in up to 1.7% absolute WER reduction for dysarthric speech compared with the MFCC base-line. Our best model reaches up to 40.6% and 11.8% WER for dysarthric and typical speech, respectively.

**Glottal signal parameters as features set for neurological voice disorders diagnosis using K-Nearest Neighbors (KNN),**

Disorders affecting nervous system can affect the voice in different ways. Different neurological disorders may lead to speech problems; this may modify the articulatory characteristics related to vocal folds function, which provide important information for detecting certain neurological diseases. In order to improve the diagnosis of Neurological Voice Disorders (NVD) an objective technique based on articulatory evaluations of vocal folds vibration, based on an estimation of a Glottic Signal (GS) extracted from Speech signal. In this work, we propose a method based on parameters extracted from GS obtained by an inverse filtering algorithm for automatic classification and diagnosis of NVD using K-Nearest Neighbors (KNN). Our work is developed around Saarbrucken Voice Database it is an open German database containing deferent samples, words, sentences of normal and pathological voices. We have selected three groups of subjects: persons with normal voices, which considered as reference, persons having suffered Parkinson disease (PD) and persons with spasmodic dysphonia.

**A Kepstrum based approach for enhancement of dysarthric speech,**

A novel speech processing algorithm based on Kepstrum analysis procedure is proposed in this paper, which provides very good speech enhancement for Dysarthric speech. Kepstrum approach has so far been used in communication applications like two microphone noise cancellation. The other applications are derivation of Kalman filter and wiener filter equations. So an attempt to use kepstrum approach to enhance the dysarthric speech is made in this paper. The algorithm is tested on various monosyllabic and bisyllabic (Consonant-Vowel pattern and Consonant-Vowel-Consonant-Vowel pattern) dysarthric speech samples of cerebral palsy patients between the age group of 40–60 years and it was found that there was considerable formant shift and modification in the energy of the output signal. Also the results obtained by kepstrum approach is compared with the results obtained by Linear Prediction Coefficients (LPC) method and it is found that kepstrum approach gives better results.

**A 40nm CMOS SoC for Real-Time Dysarthric Voice Conversion of Stroke Patients,**

This paper presents the first dysarthric voice conversion SoC, which can translate stroke patients' voice into more intelligible and clearer speech in real time. The SoC is composed of a RISC-V MPU and a compact DNN engine with a single 16-bit multiply-accumulator, which improves 12x performance and > 100x energy efficiency, and has been implemented in 40nm CMOS. The silicon area is 0.68×0.79mm2, and the measured power is 18.4mW for converting 3-sec dysarthric voice within 0.5 sec (at 200MHz and 0.8V) and 4.8mW for conversion < 1 sec (at 100MHz and 0.6V).

**Speech recognition of deaf and hard of hearing people using hybrid neural network,**

This paper describes isolated word recognition of deaf students by unsupervised and supervised neural network. Compared to normal speech, there is high variability in deaf speech and by hearing once we couldn't understand it. By the use of proposed method deaf people can make use of all voice operated devices. In this paper we use combination of SOFM and BPN neural network for recognition. Initially the input is sampled, filtered, windowed and Perceptual Linear Predictive Coefficients are determined for each frame. These coefficients are applied as input to the SOFM neural network. The output of this network is given to BPN neural network comprising of 3 layers for learning. The network has been trained with five words uttered by five different deaf persons in the age group of 5-10 years. Another set of same five words uttered by same five deaf persons were used for test purposes. The recognition results for the word one, three, four is 50 to 60% and for five is 50%...But the recognition results for word two are only10% since the variability is high for two. The results can be improved by varying the parameters of the hybrid neural network.

**ASR for electro-laryngeal speech,**

The electro-larynx device (EL) offers the possibility to re-obtain speech when the larynx is removed after a total laryngectomy. Speech produced with an EL suffers from inadequate speech sound quality, therefore there is a strong need to enhance EL speech. When disordered speech is applied to Automatic Speech Recognition (ASR) systems, the performance will significantly decrease. ASR systems are increasingly part of daily life and therefore, the word accuracy rate of disordered speech should be reasonably high in order to be able to make ASR technologies accessible for patients suffering from speech disorders. Moreover, ASR is a method to get an objective rating for the intelligibility of disordered speech. In this paper we apply disordered speech, namely speech produced by an EL, on an ASR system which was designed for normal, healthy speech and evaluate its performance with different types of adaptation. Furthermore, we show that two approaches to reduce the directly radiated EL (DREL) noise from the device itself are able to increase the word accuracy rate compared to the unprocessed EL speech.

**Navigo--Accessibility Solutions for Cerebral Palsy Affected,**

This paper presents the architecture of an integrated teachable interface, designed and implemented to provide accessibility solutions to the cerebral palsy patient in the virtual and urban space. It enables multimodal interaction between the palsy user and computers, handhelds and remote controlled appliances for enhanced usability. The idea is to redesign the system by changing the way we interpret the input, to suit the user's needs. The USP of the solution is that it is extremely dynamic, flexible and customizable to the entire range and levels of a complex condition as cerebral palsy. It works on the basis of allowing an individual to define simple variable inputs for himself and the system enabling him using the same by various specially designed tools, interfaces and feedback mechanisms.

**A study of pronunciation verification in a speech therapy application,**

Techniques are presented for detecting phoneme level mispronunciations in utterances obtained from a population of impaired children speakers. The intended application of these approaches is to use the resulting confidence measures to provide feedback to patients concerning the quality of pronunciations in utterances arising within interactive speech therapy sessions. The pronunciation verification scenario involves presenting utterances of known words to a phonetic decoder and generating confusion networks from the resulting phone lattices. Confidence measures are derived from the posterior probabilities obtained from the confusion networks. Phoneme level mispronunciation detection performance was significantly improved with respect to a baseline system by optimizing acoustic models and pronunciation models in the phonetic decoder and applying a nonlinear mapping to the confusion network posteriors.

**Building a Truly Inclusive Protocol for Students with Disabilities from an Experience in STEM areas,**

Three of the most reported strategies for students with disabilities (SWD) along their stay in higher education institutions include adapting changes in management and facilities through protocols that focused on Admission, Retention and Graduation. These strategies are important, although they do not guarantee full inclusivity during the teaching, learning and evaluation process. In this work, through a qualitatively methodology, we describe the perceptions of inclusivity of five lecturers, eight undergraduates and one SWD in science courses for Computer Systems Engineering. Furthermore, we report one successful experience to design an inclusive evaluation in a Mathematics course. The evidence found in this work suggests that the center of every protocol for SWD should be focused on Attention, as a new and longer stage. This stage includes e.g. teacher training, inclusive curriculum (for teaching and evaluation), student service and university extension programs, and inclusivity-focused research. Our purpose is to address this experience to the engineering community and promote the establishment of policies towards a more inclusive higher education system.

**Fuzzy inference system & fuzzy cognitive maps based classification,**

Fuzzy classification is very necessary because it has the ability to use interpretable rules. It has got control over the limitations of crisp rule based classifiers. This paper mainly deals with classification on the basis of soft computing techniques fuzzy cognitive maps and fuzzy inference system on the lenses dataset. The results obtained with FIS shows 100% accuracy. Sometimes the data available for classification contain missing or ambiguous data so Neutrosophic logic is used for classification to deal with indeterminacy.

**Evaluating Methods for Ground-Truth-Free Foreign Accent Conversion,**

Foreign accent conversion (FAC) is a special application of voice conversion (VC) which aims to convert the accented speech of a non-native speaker to a native-sounding speech with the same speaker identity. FAC is difficult since the native speech from the desired non-native speaker to be used as the training target is impossible to collect. In this work, we evaluate three recently proposed methods for ground-truth-free FAC, where all of them aim to harness the power of sequence-to-sequence (seq2seq) and non-parallel VC models to properly convert the accent and control the speaker identity. Our experimental evaluation results show that no single method was significantly better than the others in all evaluation axes, which is in contrast to conclusions drawn in previous studies. We also explain the effectiveness of these methods with the training input and output of the seq2seq model and examine the design choice of the non-parallel VC model, and show that intelligibility measures such as word error rates do not correlate well with subjective accentedness. Finally, our implementation is open-sourced to promote reproducible research and help future researchers improve upon the compared systems.

**Using Acoustic Deep Neural Network Embeddings to Detect Multiple Sclerosis From Speech,**

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system. It affects cognitive and motor functions, and the limitation of executive functions can also manifest itself in speech production. Due to this, automatic speech analysis might serve as an effective technique for assessing MS, or for monitoring the status of the patient. However, choosing the features to be extracted from the recordings is not straightforward. In the past few years, general feature extractors such as i-vectors, d-vectors and x-vectors have found their way into automatic speech analysis. In this study we show that there is no need to employ a special neural network architecture such as x-vectors to calculate effective features, but (even more) indicative features can be derived on the basis of a standard Deep Neural Network acoustic model. From our results, these features could effectively be used to distinguish MS subjects from healthy controls, as we measured AUC scores up to 0.935. We found that classification performance depended only slightly on the choice of the hid-den layer used to extract our features, but the speech task per-formed by the subject turned out to be an important factor.

**Speech Therapy Applications based on Speech Technologies,**

The proposed paper focuses on the area of ICTbased supportive tools for speech therapy in children with speech and hearing disorders in Slovak language. The main idea was to design a concept of the supportive tool, where speech technologies can be used in a new, modern way. The web application was designed to help a child to train a correct pronunciation of particular sounds. To evaluate the similarity of the spoken word, Dynamic Time Warping algorithm was implemented, which measure distance between spoken word and the pattern in the database. Obtained distance help a caregivers or therapist to evaluate the need of continuation in speech therapy with a particular sound, or it is possible to move to another one. We also proved the concept of using automatic speech recognition in the speech therapy to create modern game-like speech therapy tools. The proposed paper brings the first observations from preliminary tests and discuss advantages and drawbacks of designed tools.

**Automatic Screening Of Children With Speech Sound Disorders Using Paralinguistic Features,**

Subjective screening of children with speech disorders is costly, time consuming and infeasible due to the limited availability of Speech and Language Pathologists (SLPs). Therefore, there is an increasing interest in automatic speech analysis of children with speech disorders as it can offer a practical alternative to human assessment. Paralinguistic features are a set of low-level descriptors commonly used in speech emotion recognition. However, they have not yet been examined with childhood speech sound disorders such as, apraxia-of-speech and phonological and articulation disorders. In this paper, we investigated the effectiveness of paralinguistic features in discriminating between typically developing children and those who suffer from different types of speech sound disorders. Two types of standard paralinguistic features were explored, the Geneva Minimalistic Acoustic Parameter Set (GeMAPS) and its extended version, (eGeMAPS) feature sets. We applied feature selection to find the most discriminant set of features and employed binary classification using a support vector machine (SVM) to discriminate between the two groups. The method was tested on a recently-released public speech corpus collected from typically developing children and children with various types of speech sound disorders. The system achieved segment-level and subject-level unweighted average recall (UAR) of around 78% and 87% respectively.

**Dysarthric vocal interfaces with minimal training data,**

Over the past decade, several speech-based electronic assistive technologies (EATs) have been developed that target users with dysarthric speech. These EATs include vocal command & control systems, but also voice-input voice-output communication aids (VIVOCAs). In these systems, the vocal interfaces are based on automatic speech recognition systems (ASR), but this approach requires much training data and detailed annotation. In this work we evaluate an alternative approach, which works by mining utterance-based representations of speech for recurrent acoustic patterns, with the goal of achieving usable recognition accuracies with less speaker-specific training data. Comparisons with a conventional ASR system on dysarthric speech databases show that the proposed approach offers a substantial reduction in the amount of training data needed to achieve the same recognition accuracies.

**Interaction of speech disorders with speech coders: effects on speech intelligibility,**

Modern speech coding schemes have been developed to address the demand for economical spoken language telecommunication of acceptable quality. A variety of speech coding algorithms have been described, which compress speech to facilitate efficient transmission of spoken language over communication networks ((J.R. Deller Jr., 1993; P.E. Papamichalis, 1987). Most such speech coding algorithms are lossy in the sense that the processed speech is not identical to the original speech. As a result, some distortion is invariably introduced with any lossy speech coding strategy. For this reason, candidate coders undergo detailed evaluation to ensure that the associated speech output is of acceptable quality (S.R. Quackenbush et al., 1988). Three different coding algorithms were investigated relative to unprocessed speech: the Codebook Excited Linear Prediction (CELP), the Global System for Mobile Communications (GSM) algorithm which is a standardized speech coding algorithm in Europe, and the Linear Predictive Coding (LPC) algorithm. The specific coding schemes evaluated were MatLab implementations of NSA FS-1015 LPC-l0e; NSA FS-1016 CELP-v3.2; and ETSI GSM (A. Spanias, 1995). One of the goals of this study was to quantify the coding distortion using objective measures and to correlate these measures with speech intelligibility and subjective quality data, in the hope of identifying one or more measures that can predict the subjective results.

**Models for objective evaluation of dysarthric speech from data annotated by multiple listeners,**

In subjective evaluation of dysarthric speech, the inter-rater agreement between clinicians can be low. Disagreement among clinicians results from differences in their perceptual assessment abilities, familiarization with a client, clinical experiences, etc. Recently, there has been interest in developing signal processing and machine learning models for objective evaluation of subjective speech quality. In this paper, we propose a new method to address this problem by collecting subjective ratings from multiple evaluators and modeling the reliability of each annotator within a machine learning framework. In contrast to previous work, our model explicitly models the dependence of the speaker on an evaluators reliability. We evaluate the model on a series of experiments on a dysarthric speech database and show that our method outperforms other similar approaches.

**Evaluation of the word fluency in parkinson's disease patients treated with deep brain stimulation-a pilot study-,**

Parkinson's disease is a chronic, progressive and common disease of neurological disorders. Motor complication in parkinson's disease (PD) is resting tremor, slow movement, rigidity and postural instability. The motor symptoms of PD respond well to bilateral deep brain stimulation (DBS). Recent study, there are also reports of worsened verbal fluency, executive dysfunction, and processing speed with DBS. Whether subthalamic nuclei (STN) stimulation worsens there are under debate. The aim of this study was to explore the effects of STN stimulation verbal fluency as assessed with clinical neuropsychological tests. Eight patients treated with deep brain stimulation were enrolled, and some of the patients continued anti-PD medications. Assessments were done both with the STN stimulation turned OFF and ON. In both test conditions, the following were assessed: speech , word fluency A, and B. The score of the word fluency test of all patients have undergone DBS surgery significantly worsened as compared with before surgery. Five patients speech ware worsened, but three patients were improved when the STN stimulation was turned OFF. On the other hand, five patients were reduced the word fluency's total score when the STN stimulation was turned OFF. In this sample, STN stimulation significantly worsened the result of the word fluency test. When the STN stimulation was turned OFF, it was reduced. These finding suggests that STN-DBS might be worse speech conditions and verbal fluency.

**Compression of acoustic inventories using asynchronous interpolation,**

A compression method is proposed that takes advantage of a powerful property of acoustic unit inventories: In the appropriate acoustic space, units that share a (context-dependent or -independent) phoneme label must be close to a vector phoneme template associated with the phoneme. The method approximates units by interpolation between templates. The interpolation operation involves two asynchronous weight functions operating on the template. One is associated with spectral peak locations, the second with spectral balance. This enables approximating transitions such as [i:]/spl rarr/[v], in which formant movement precedes frication onset. The algorithm guarantees smooth concatenation points.

**Development of Speech Therapy Mobile Application for Speech Disorder Post-Stroke Patients,**

In Malaysia, stroke is the third cause of death and disability. Stroke cause significant injury to the brain that may result in long-term problems such as communication, concentration, memory, and executive functions. About one-third of post-stroke patients have speech and communication problems that require to undergo series of one-to-one speech therapy sessions. However, there are only 300 speech therapists in Malaysia which limit the recovery and may not reach to the needed patient. More importantly, frequent therapy conducted could fasten the recovery of the patients' speech. Therefore, this research develops a mobile application to be used as an alternative for speech therapy session. Although there are mobile applications for speech therapy, none of them are in Bahasa Melayu. The mobile application implements an automatic speech recognition technology that accepts the vowel speech sound from a post-stroke patient in Bahasa Melayu. The mobile application will process the sound, evaluate, and provide feedback score for the vowel sound in an accuracy percentage. It is expected that the ASR speech therapy mobile application could help speech disorder post-stroke patients to practice their speech ability at their own time without attending speech therapy sessions.

**Exploratory analysis of speech features related to depression in adults with Aphasia,**

Aphasia is an acquired communication disorder resulting from brain damage and impairs an individual's ability to use, produce, and comprehend language. Loss of communication skills can be stressful and may result in depression, yet most depression diagnostic tools are designed for adults without aphasia. This paper discusses preliminary results from a research effort to examine acoustic profiles of adults with aphasia who have been assessed as having possible depression versus those who assessment suggests they are not depressed based on tools completed by their caretakers. This study analyzes prosodic and spectral features in 14 participants (7 assessed as having possible depression and 7 whose assessment does not suggest depression). The results showed using Cepstral Peak Prominence provided the best overall performance in separating depressed and non-depressed speech among adults with aphasia.

**Speech enhancement for pathological voice using time-frequency trajectory excitation modeling,**

This paper proposes a speech enhancement algorithm for pathological voices using a time-frequency trajectory excitation (TFTE) modeling. The TFTE model has a capability of delicately controlling the periodic and non-periodic excitation components by taking a single pitch based decomposition process. By investigating the difference of frequency characteristics between pathological and normal voices, this paper proposes an enhancement algorithm which can efficiently reduce the breathiness of the pathological voice while maintaining the identity of the speaker. Subjective test results are presented to verify the effectiveness of the proposed algorithm.

**Comparative Analysis of Machine Learning and Ensemble Learning Classifiers for Parkinson’s Disease Detection,**

A neurological illness called Parkinson's disease (PD) commonly appears between the ages of 55 and 65. Moreover, a patient's entire quality of life is significantly impacted by the progressive development of motor as well as non-motor symptoms due to this disease. There is no known cure for PD, although a number of therapies have been created to assist control its symptoms. Therefore, the management of PD is a field that is expanding, and there is a need to develop a comprehensive framework for the timely detection and classification of PD. In this paper, we developed six machine learning-based and five ensemble learning-based classification models to forecast PD. The six base classifiers which are used in the current study are Support Vector Machine (SVM), Decision Trees (DTs), Random Forest (RF), K-Nearest Neighbor (KNN), Logistic Regression (LR), Naive Bayes (NB); and five ensemble classifiers named XGBoost, Gradient Boost, Bagging, CatBoost and Light Gradient Boosted Machine (LGBM) respectively, are then carefully compared. To improve the performance of the classifiers and to reduce the problem of overfitting, a feature selection method named Principal Component Analysis (PCA) and various preprocessing techniques are applied. Further, this study uses the voice samples dataset from the UCI repository having 188 PD and 64 normal patients. Overall, our findings revealed that, when compared to the other five base classifiers, the RF model offered the best classification performance with an accuracy of 82.37%, and the ensemble classifier named LGBM shows best results when compared with base as well as ensemble classifiers having an accuracy of 85.90%.

**Dysarthric Speech Recognition using Multi-Taper Mel Frequency Cepstrum Coefficients,**

Vast industrial growth has increased the demand of automatic speech recognition for various automation and human machine interaction application. Performance of various artificial intelligence based approaches is limited because of the speech disability caused due to communication disorders, neurogenic speech disorder or psychological speech disorders. The dysarthric disorder is neurogenic speech disorder that limits the human voice articulation capability. This paper presents, dysarthric speech detection using Multi-Taper Mel Frequency Cepstral coefficients (MTMFCC) that is capable to smallest variation over the dysarthric speech. The efficiency of the proposed algorithm is estimated using the K-Nearest Neighbor (KNN) classifier and support vector machine (SVM) based on accuracy, sensitivity and specificity. The system has shown 99.04 % and 96.00 % accuracy for the MTMFCC+KNN and MTMFCC+SVM which is superior to traditional MFCC.

**Significance of Feature Selection for Acoustic Modeling in Dysarthric Speech Recognition,**

In this paper, a comparative study of various feature extraction methods is carried out on dysarthric speech. Dysarthric speech is difficult to recognize and thus pose challenges that normal speech does not. Since various features can be used to model phonemes in hidden Markov model (HMM) based recognition system, which feature is suitable for the task specified is a topic to be addressed.Dysarthric speech becomes unintelligible due to the improper coordination of articulators. In this paper, recognition results are compared using mel-frequency cepstral coefficients (MFCC), perceptual linear prediction (PLP), filter bank and reflection coefficients feature sets. The performance is analyzed using TORGO database. Phonemes are grouped for the analysis. Our study shows that MFCC and PLP gave better results than filter bank and reflection coefficients for dysarthric speech analysis.

**An Unsupervised Approach to Speech Segmentation using Auto Resonance Networks,**

Automatic speech segmentation finds application in speech recognition, speaker diarization etc. However, it is challenging as it demands huge labelled data. Unsupervised approach can overcome the laborious labeling task. Auto resonance networks are one choice for unsupervised approach. Recently auto resonance networks have been used in classification problems. This work proposes an unsupervised approach for speech segmentation using auto resonance networks. The system is tested using the standard TIMIT database and a few pathological speech utterances. Results show that the speech frames belonging to the same broad class are assigned to the same cluster. The results are compared with another neural network approach, namely, self organizing maps.

**A Noval Text To Speech Conversion Using Hierarchical Neural Network,**

The creation of a realistic voice from written text is revolutionized by the state-of-the-art paradigm in text-to-speech (TTS) systems, the Hierarchical Neural Network Architecture (HNNA). An extensive analysis of the important role the HNNA plays in identifying hierarchical patterns in textual input a function that enables the generation of expressive and contextually relevant speech is done in this abstract. The HNNA captures language hierarchies via hierarchical representations. It starts with word embedding analysis of textual inputs and moves on to sentence and document abstractions. Modern models like RNN are used in this architecture to generate high-quality speech signals, and transformer-based encoders decode hierarchical text representations. The integration of linguistic patterns and emphasis on context-aware processing make the HNNA noteworthy. Because of this, speech produced by contemporary TTS systems can sound more intonated, realistic, and understandable.

**Automatic assessment of articulation disorders using confident unit-based model adaptation,**

This paper presents an approach to automatic assessment on articulation disorders using unsupervised acoustic model adaptation. Prior knowledge is obtained via the phonological analysis of the speech data from 453 articulation disordered children. A confusion matrix of the recognition units for a specific subject is re-estimated based on the prior knowledge and the recognition results to choose the confident units for adaptation. The adapted acoustic models can effectively improve the recognition performance of the disordered speech and thus used for articulation assessment. In the experiments, the proposed unsupervised adaptation method achieved a significant performance improvement of 9.1% for disordered speech on syllable recognition rate. Automatic assessment also shows encouraging consistency to the assessment from the therapist.

**Features dimensionality reduction and multi-dimensional voice processing program to Parkinson disease discrimination,**

Parkinson's disease is a pathology that involves characteristic perturbations in patients' voices. This paper describes a proposed method that aims to diagnose persons with Parkinson (PWP) by analyzing on line their voices signals. First, Thresholds signals alterations are determined by the Multi-Dimensional Voice Program (MDVP). Principal Analysis (PCA) is exploited to select the main voice principal components that are significantly affected in a patient. The decision phase is realized by a Multinomial Bayes (MNB) Classifier that categorizes an analyzed voice in one of the two resulting classes: healthy or PWP. The prediction accuracy achieved reaching 98.8% is very promising.

**Production of Consonants /s/ and /t/ in the Malay Language of Subjects with Paralysis,**

Paralysis causes affected person having difficulty in moving their bodies partly or wholly. Paralysis may occur due to several factors including stroke and spinal cord injury (SCI). The sign and symptoms of paralysis depends on its aetiology and affected body parts. One of the indirect effects of paralysis is slurred speech and difficulty in speaking. This study investigated the contact pattern of three paralysed subjects during speech production of consonants /s/ and /t/ in the Malay Language. The subjects had paralyses due to different aetiologies and with different medical history background. All participants were required to produce two single consonants; /s/ and /t/ in a studio laboratory with a soundproof system. Electropalatography (EPG) was selected as the device used for recording the tongue-hard palate contact throughout this study. All subjects were required to wear the Reading EPG palate consisting of 62 electrodes which matched the tongue-palate contact. The speech data were analysed using Articulate Assist 1.18. Then, the patterns from three paralysed subjects were compared with the average contact pattern of Malay speakers obtained in a separate study. Results showed differences in consonant /s/ and /t/ production between paralysed and normal subjects. Hence, data presented can, therefore, be used in the treatment planning of paralysed subjects during speech therapy.

**Spectrogram and Mel-Spectrogram Based Dysphonic Voice Detection Using Convolutional Neural Network,**

A non-invasive pathological voice detection system has been presented in this paper. This work considers two spectral images of voice signals namely spectrogram and Mel-spectrogram to detect dysphonic voices. The spectrogram is a convenient representation of voice signals on a time-frequency scale and has been popularly investigated in pathological voice detection algorithms. However, the spectrogram uses a linear frequency scaling and hence, does not consider the resolution of the human auditory model. The Mel-spectrogram overcomes this limitation by using a quasi-logarithmic frequency spacing. This work uses vowel samples available in the Saarbrucken voice database. The simulation results show that the Melspectrogram outperforms the spectrogram in terms of classification accuracy.

**Data processing for Parkinson's disease: Tremor, speech and gait signal analysis,**

The Parkinson has been placed on the second position of the most frequent neurodegenerative illnesses list, after Alzheimer, consisting in slowly and progressive neurons damage. Parkinson's disease is associated with motor symptoms, including tremor, postural instability, rigidity, bradykinesia and dysphonia. The time series specific to tremor, speech and gait Parkinson's Disease signals were firstly analyzed, by tools derived from chaotic analysis, such as: correlation dimension, recurrence plot, recurrence quantification analysis or Lyapunov exponent, as well. From the tremor management point of view and based on the results, this paper emphasizes the importance of the non-linear dynamics specific parameters in Parkinson tremor analysis. Currently, there hasn't been underlined any “gold standard” method, by which quantitative and qualitative evaluation of symptoms' gravity for patients with Parkinson are observed. The goal of the research carried out in this way consists in finding a screening test, in order to identify early the Parkinson's disease. By analyzing the tremor, gait and speech signals for patients with Parkinson's disease, this paper brings an approach over the most important information that can be used within a knowledge based system, specific to Parkinson's disease diagnosis.

**Automatic Parkinson’s disease detection from speech: Layer selection vs adaptation of foundation models,**

In this work, we investigate Speech Foundation Models (SFMs) for Parkinson’s Disease (PD) detection. We explore two main approaches: (1) using SFMs as frozen feature extractors and, (2) fine-tuning/adapting SFMs for PD detection. We propose a cross-validation-based layer selection methodology to identify the layer effective for PD detection. Additionally, we compare the performance of the layer selection scheme with full fine-tuning and, parameter-efficient fine-tuning (PEFT) using Low-Rank Adaptation (LoRA). Our results show that layer selection and LoRA-based fine-tuning can perform on par with full fine-tuning, providing a more parameter-efficient alternative. The highest accuracy was achieved by fine-tuning Whisper using LoRA.

**Cross-lingual Evaluation Of Hypernasality Using Wav2Vec2 Features,**

Hypernasality, a speech resonance disorder characterized by excessive nasal airflow, presents challenges in accurate detection across languages. Traditional assessments of hypernasality include perceptual evaluation and nasometry. More recently, objective acoustic measures based on formant analysis and other acoustic features have been proposed as proxies for hypernasality; however, these acoustic measures exhibit considerable variability, especially in multilingual contexts. In this study, we utilize the wav2vec2-large-xlsr-53 model, a cross-lingual speech representation framework, and evaluate hypernasality-related features within its transformer layers. We extracted features from each layer of the model’s transformer and trained a machine learning model to predict hypernasality ratings across three datasets: Americleft (English), New Mexico Cleft Palate Center (English), and All India Institute of Speech and Hearing (Kannada). Our analysis reveals that the 11th and 12th layer contextualized embeddings effectively model hypernasality cross-lingually, demonstrating significant within-language average correlations (0.78 and 0.78) and cross-lingual average correlations (0.68 and 0.67) between predicted and perceptual ratings. These findings suggest that the wav2vec2-large-xlsr-53 model’s intermediate layers effectively capture hypernasality cross-lingually.

**Dysarthric Speech Recognition Using Deep Convolution Neural Network,**

Dysarthric speech recognition (DSR) plays imperative role in assisting vocally impaired people to indulge in voice based automation systems and human computer interaction. DSR is exigent due to low intelligibility of impaired speaker, less dataset availability, low intra-class and inter-class variability in the speech samples. This paper presents DSR based on Deep Convolution Neural Network (DCNN and Mel Frequency Logarithmic Spectrogram (MFLS). The proposed MFLS+DCNN provides superior spectral-temporal representation of the speech signal. The results of the MFLS-DCNN scheme are validated on UASpeech dataset based on accuracy, recall, precision and F1-score. The proposed scheme provides accuracy of 96.83%, recall of 0.97, precision of 0.96, and F1-score of 0.97 which has shown significant improvement over traditional state of arts.

**SSDCVA: Support system to the diagnostic of cerebral vascular accident for physiotherapy students,**

The aim of this paper is to provide an intelligent pedagogical agents based on cases to aid in the diagnosis and treatment of patients with neurological disorders. The study proposes an architecture which facilitates the decision making process of students in the healthcare field, in order to advice adequate treatment for patients suffering a cerebral vascular accident (CVA).

**The role of the Lombard reflex in parkinson's disease,**

Parkinson's disease (PD) is a severe disease with many symptoms, including speech disorders. Although many methods exist to treat some of PD's symptoms, therapies for speech impairment are not effective and satisfactory, resulting in an open area of research. The current project aims at taking advantage of the Lombard reflex to improve the speech loudness of PD patients. As a first step, the experience of the Lombard reflex by Japanese PD people was confirmed, and the perception of PD patients' speech was evaluated by several subjects. In a following step, methods based on masking sound will be used for intensive training and for self-training of PD patients. However, after intensive training, PD patients may be able to talk louder even without masking noise. In addition, the design and the development of a device based on masking sound that can be used by PD patients while using phone is under consideration.

**Functional MRI changes affected by horticultural therapy for cerebrovascular disorders,**

The aim of this paper is to assess whether horticultural therapy (HT) is effective for improvement of daily activity and brain functional activity in brain-damaged patients from a medical point of view. Five patients with cerebrovascular disorders were invited to participate in horticultural therapy for a month in Ishikawa Hospital. The Independence Scale in Activities of Daily Living (ISADL) and the Mini Mental State Examination (MMSE) were performed before and after a whole series of horticultural therapy program to assess a patient’s activities of daily life and to determine the patient’s cognitive function, respectively. Functional magnetic resonance imaging (fMRI) under recognition tasks were measured before and after the horticultural therapy. Independency of the daily activities and cognitive function tended to improve after the horticlutural therapy. fMRI examinations showed that the supramarginal gyrus (SMG), the motor area, the supplementary motor area (SMA), the sensory area, the visual area, the middle temporal gyrus, and the frontopolar area were activated after the therapy in most patients.

**Repetition detection in dysarthric speech,**

Repetition detection is an important pre-processing step in application such as speech to text alignment, voice based interactive system etc. It is very challenging to detect the repeated words because a speaker may utter the repeated words partially or may miss some words in between as it is more often, in the case of Dysarthric utterances. To address these issues, we propose an approach for repetition detection and tested on Dysarthric utterances by extracting features such as MFCC and formants. For calculating similarity scores between the words, we employed two approaches: Dynamic time warping (DTW) and polynomial curve fitting (PCF). Finally, we compared the results of both the approaches by taking each feature independently. DTW based approach found to be more accurate exemplified by the experimental results.

**A Novel Multi-task Learning based Automatic Speech Impairment Assessment Algorithm,**

Speech impairment assessment is crucial to the treatment evaluation of speech therapy. The current evaluation methods mainly depend on speech-language pathologists (SLPs). Automatic speech impairment assessment (ASIA), especially the regression of severity scores, has not received enough attention. So we present a novel ASIA algorithm which based on the multi-task leaning with joint severity level classification and score regression. Owing to the auxiliary classification task, the precision of the severity score prediction can be improved effectively. In addition, the residual network (ResNet) and the long short-term memory (LSTM) are cascaded as the backbone. The performance of the model is demonstrated on the Mandarin AphasiaBank dataset and the experiments show that the algorithm achieves promising performance.

**Coomputer-based auditory analysis of neurogenic speech disorders,**

Abstract:

**Data Selection Based on Phoneme Affinity Matrix for Electrolarynx Speech Recognition,**

Electrolarynx (EL) is a communicative aid for the patient after laryngectomy to generate communicable speech. Since EL speech exhibits low speech intelligibility and produces loud noise, understanding the content of the speech remains challenging for listeners, even if the patient is proficient in using the EL device. Accordingly, it is important to develop the tools that offer additional communication methods. Automatic speech recognition (ASR) of EL speech emerges as a method worth considering in this regard. However, the problem of under-resourced data dramatically degrades the recognition performance of EL speech. Data augmentation is one of the viable solutions for addressing the issue of under-resourced speech data. However, even with an increased health training corpus, the improvement in EL speech recognition may not be satisfactory. Because the characteristics of the EL speech still differ significantly from those of health speech. This paper proposes a data selection method using the phoneme affinity matrix to prioritize the selection of health speech that closely resembles EL speech for data augmentation. The affinity between two phonemes is defined as the similarity of the Phone Posteriorgrams(PPGs) of the two phonemes, considering the phoneme models. The experimental results demonstrate that the approach utilizing data selection based on the phoneme affinity matrix yields superior results compared to both the baseline and the method employing random sampling to select the augmented health speech corpus.

**The self-taught vocal interface,**

We describe a demonstration of a novel vocal user interface (VUI), ALADIN, which is trained through usage, by demonstrating spoken commands with manual controls. It works by mining the vocal commands to find recurrent acoustic patterns corresponding to words or phrases that constitute elements of the user's commands. The demonstration consists of the VUI controlling a 3D environment with home automation devices such as lights and doors, as well as an actual home automation setup. Additionally, a tablet interface is provided for feedback and manual control.

**Measurement of Formant Frequency in /hVd/ Words of Distorted Speech in Adult New Zealanders,**

This paper investigates the acoustic features of distorted speech in adult New Zealanders. There are total of sixteen subjects in this study: eight healthy individuals as control group and eight patients, four of them suffer from non-neurological dysphonia hence, their voice is hoarse but they have not undergone any surgery whereas the other four have gone through total laryngectomy and use a voice prosthesis (TEP) to speak. Building upon our previous work on temporal acoustic characteristics of distorted speech, analysis of the two formant frequencies of vowel articulation in /hVd/ words are presented in this paper. The results of both TEP speech and dysphonic speech are compared to the samples collected from healthy control group.Independent sample t-tests have been carried out once between TEP speech and normal speech, and once between dysphonic and normal speech to show how significant the differences in acoustic measurements are. The results show that formant frequencies in TEP speech are significantly different from normal speech whereas in patients with dysphonia these variabilities are not as notable.

**Automatic Estimation of the Triangular Vowel Space Area from i-Vectors,**

Parkinson's Disease (PD) is a neurodegenerative disorder which gradually effects the neurological condition of the patient. In many cases the disease impairs the reliability of the articulatory system and the ability to pronounce vowels normally. One prominent way to measure the degree of the functioning of the articulatory system is the Vowel Space Area (VSA). However, the typical way to measure it, is to manually annotate sustained vowel recordings or phonetically annotated speech utterances of a speaker and then analyze the signals. However, it is often desirable to measure the VSA directly from unlabeled natural speech. Therefore an automatic model-based system is proposed in this paper to estimate the triangular Vowels Space Area (tVSA) and the underlying corner vowel formant frequencies directly from unlabeled natural speech. The proposed algorithm is able to estimate the tVSA automatically from the speech signals without the need of phonetical or vowel transcriptions. The i-Vectors are extracted from the signals as the speaker's characteristic representation, from which the speaker's corner vowel formant frequencies are estimated by regression classifiers. Two regression classifiers, namely Deep Neural Networks (DNN) and Support Vector Regression (SVR), are investigated in this work. The proposed configuration employs the SVR classifier, which is able to predict the corner vowel formant frequencies of the test speakers with R(exp 2) up to 0.56719 and rho up to 0.76485.

**A personal voice analyzer and trainer,**

This paper presents a personal voice analyzer and trainer that allow the user to perform four daily exercises to improve the voice capacity. The system grades how well the user is performing the exercises by analyzing the duration, the intensity and the pitch of the user's voice.

**Hybrid Voice Spectrogram-Chromogram Based Deep Learning (HVSC-DL) Model for the Detection of Parkinson’s Disease,**

Parkinson’s disease is one of the serious neurological disorders that restricts the life quality of individuals significantly. The changes in sound signals contain important clues for detecting the disease at an early stage. In this study, a newly collected Parkinson’s voice dataset is introduced, and preliminary results with classical machine learning and the proposed three alternative deep learning models are presented comparatively. We observed that our proposed two-channel Hybrid Voice Spectrogram-Chromogram based Deep Learning Model (HVSC-DL) with the patient-healthy classification accuracy rates of 96.6%, 92.9% and 94.5% on /a/, /o/ and /i/ sounds respectively, showed superior performance compared to pure tone chromogram and spectrogram based models.

**Improving Recognition for Disordered Speech in Indonesian Language: a Data Augmentation approach,**

For individuals with speech disorders, speech recognition serves as a crucial communication tool. But most speech recognition systems are trained on normal speech data, worsen by limited disordered speech data which mostly available in English language. Data augmentation is one possible approach, where we can apply signal processing such as speed perturbation to transform normal speech into disordered speech. In this work, we studied how augmented data composition in a dataset affects the system performance on recognizing disordered speech. Using QuartzNet CNN as the acoustic model, we evaluated the speech recognition system on an augmented dataset built based on an Indonesian language speech database from Mozilla Corpus. Using speed perturbation to generate disordered speech, the initial results show that augmenting normal speech dataset with 25-50% more disordered speech data could help improve the Word Error Rate (WER) of the model in recognizing disordered speech. This result, however, is limited as we use the same speed perturbation method for training and testing datasets.

**Phonatory Analysis on Parkinson's Disease Patients Attending Singing and Discussion Therapy (Parkinsonics) using Signal Processing Techniques,**

Persons with Parkinson's Disease (PD) frequently have speech and voice disorders. Regular speech therapy with a speech-language pathologist is essential to mitigate progressive symptom deterioration. Speech-related therapies, such choral singing groups are alternative approaches designed to be more naturalistic and enhance participant enjoyment. It is important to measure and quantify the effects of these therapies on the vocal features of PD patients to determine efficacy. We performed a prospective crossover study of 25 PD patients attending discussion or choral-singing groups for 12 weeks each (Parkinsonics NCT02753621). Every six weeks, each participant produced several recordings of the sustained vowels /a:/ and /e:/ at ‘normal’ and ‘maximum’ loudness. The goal was to identify if there are signal-processing-based features that can help track changes in the voice of PD patients over time. Voice features were extracted from these recordings using the Automatic Voice Condition Analysis (AVCA) library and were compared using non-parametric statistical tests. Results suggest that neither therapy caused any significant improvements in the analyzed phonatory aspects of the patients' voices. Future work should require use of connected speech to analyze articulation and comparison with a control group of participants with PD not attending any therapy to evaluate if therapy can mitigate the progressive effects of PD on the voice of patients.

**Creating Personalized Synthetic Voices from Articulation Impaired Speech Using Augmented Reconstruction Loss,**

This research is about the creation of personalized synthetic voices for head and neck cancer survivors. It is focused particularly on tongue cancer patients whose speech might exhibit severe articulation impairment. Our goal is to restore normal articulation in the synthesized speech, while maximally preserving the target speaker’s individuality in terms of both the voice timbre and speaking style. This is formulated as a task of learning from noisy labels. We propose to augment the commonly used speech reconstruction loss with two additional terms. The first term constitutes a regularization loss that mitigates the impact of distorted articulation in the training speech. The second term is a consistency loss that encourages correct articulation in the generated speech. These additional loss terms are obtained from frame-level articulation scores of original and generated speech, which are derived using a separately trained phone classifier. Experimental results on a real case of tongue cancer patient confirm that the synthetic voice achieves comparable articulation quality to unimpaired natural speech, while effectively maintaining the target speaker’s individuality. Audio samples are available at https://myspeechproject.github.io/ArticulationRepair/.

**A Study in Comparing Acoustic Space: Korean and Hindi Vowels,**

The present paper is an empirical study on acoustic space of five peripheral vowels of Korean and Hindi. The study attempts to explore the difference between acoustic space Hindi and Korean that will have significant implications for any language teaching program, as well as in other fields such as speech pathology and forensic acoustics for speaker identification studies. Data from acoustic analysis would also be of great help to the teacher who is teaching foreign language in understanding and explaining the precise differences between L1 & L2. We also wish to examine the Dispersion Theory (D.T. hypothesis) and the Quantal Theory of Speech (QTS) in the light of the findings of the study.

**Speech-Based Screening of Multiple Sclerosis By Features Derived from Self-Supervised Models,**

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system. Since it, among other symptoms, adversely affects the speech of the subject, automatic speech analysis might offer a simple, inexpensive and remote tool for MS screening or monitoring the progression of the disease. We employ ten different wav2vec 2.0 models as the base of feature extraction and compare the performance with pre-trained and custom x-vector models. Based on our results, cross-lingual models perform better than the base wav2vec 2.0 networks, but the model size is crucial as the best results were obtained with a model having one billion trainable parameters. We found fine-tuning the application language to be beneficial to the classification performance, but for other languages, it did not improve the AUC scores. Surprisingly, though, we did not outperform standard x-vectors, which might be due to the standard, but perhaps too simple aggregation strategy of the frame-level embeddings.

**Corpus Construction for Deaf Speakers and Analysis by Automatic Speech Recognition,**

This study explores automatic speech recognition (ASR) for the deaf and hard-of-hearing. Despite the recent progress in ASR for dysarthric speakers, existing research primarily focuses on people with motor speech disorders. Thus, the effect of speech diversity on the performance of ASR is not considered for ambiguous deaf speech owing to a lack of auditory feedback. Therefore, we compiled a corpus of speech of many profoundly deaf speakers to compare the ASR performance with that of normal-hearing speakers. The performance analysis is reported through a set of phoneme recognition experiments. Furthermore, we show that additional phonological features that reflect deaf speakers’ articulation can improve performance in phoneme recognition for deaf speech.

**Cognitive Constraint Rules In Design Of Multi-mode Computer Interfaces For Disabled Users,**

Abstract:

**Study Of Muscle Activation During The Articulation Of Phoneme Using Features Extracted From Electromyography Signals,**

Humans express their emotions mostly through speech. People can speak effectively through the active coordination of facial as well as vocal muscles. The proposed study attempts to analyze the muscle activation during the articulation of phonemes, particularly ‘അ’ (a) and ‘ആ’ (aa), which are commonly used in most Indian languages. In this work, the Malayalam language is selected for the study. Surface Electromyography signals are recorded from three facial muscles (Zygomaticus Major, Depressor Anguli Oris, and Mentalis) and one neck muscle (Anterior Belly of Digastric) of healthy female subjects using standard protocols. Root Mean Square (RMS) and mean frequency are extracted from the recorded signals. Further, classifiers namely Decision Tree, Random Forest, and KNN are applied to differentiate the phonemes ‘അ’ (a) and ‘ആ’ (aa). The results show that the mean RMS of the signals from Mentalis is greater during the articulation of phonemes ‘അ and ‘ആ’, compared to other muscles. In addition, KNN showed a maximum classification accuracy of 87 % while differentiating the phonemes with the extracted features. This study may be helpful for developing assistive devices for people suffering from speech disorders.

**SAY & SEE articulation therapy software,**

SAY & SEE is a software tool for speech therapy that displays an animated mid-sagittal view of the vocal tract in real-time on Macintosh computers in response to speech input with an ordinary microphone. The system permits users to see in a graphic and easily comprehensible way how closely they approach a targeted speech pattern. Trials of SAY & SEE at multiple sites around the country where it was used by hearing-impaired children, speech-impaired children with normal hearing, and adults with normal hearing who had suffered strokes or head injuries, have shown that SAY & SEE provides valuable feedback on the modifications to articulation needed to improve speech production. The same results have been found with students learning English as a second language (ESL), since they must struggle to pronounce phonemes that do not appear in their native languages.<>

**Exploratory data analysis on nuclei in cantonese dysarthric speech,**

Phone substitutions, distortions, deletions, and insertions are some of the main problems in dysarthric speech. This work aims to explore the articulatory error patterns in dysarthric speech, which provides insights for the improvement of automatic dysarthric speech analysis technologies. A set of dysarthric speech is collected and phonetically transcribed manually by different transcribers. Transcriptions are mapped into distinctive feature values. Error rates for each value are extracted. Substitutions and distortions are found to be the major errors in dysarthric speech. Their error patterns are analyzed in this paper. The analysis may provide guidance for labelling dysarthric speech errors, which will be useful to future development of technologies to achieve automated analysis of dysarthric speech.

**Optimized spectral floor in multi-band spectral subtraction for dysarthric speech recognition,**

The advancements in automatic speech recognition through pre-processing of speech is an alternative way to improve word recognition rate. In this work, we explored the speech enhancement techniques for improving the performance of automatic speech recognition for degraded dysarthric speech. The spectral floor parameter is optimized in the multi-band spectral subtraction technique as it controls the amount of residual noise estimation to avoid crossing the lower bound. The enhanced dysarthric speech is compared with standard enhancement techniques such as Boll's spectral subtraction, multi-band subtraction and Kalman filtering using segmental signal-to-noise ratio, log spectral distortion, mel cepstral distortion and perceptual scores. Further, the improved speech is represented using complex cepstrum based features, and trained using general regression based classier. The recognition system is tested and evaluated for word accuracy. The proposed pre-processed speech enhances the speech recognition accuracy in comparison to baseline models along with no enhancement scheme.

**End-to-end acoustic-articulatory dysarthric speech recognition leveraging large-scale pretrained acoustic features,**

Automatic dysarthric speech recognition (ADSR) remains challenging due to the irregularities in speech caused by motor control impairments and the limited availability of dysarthric speech data. This paper explores the integration of articulatory features, captured using Electromagnetic Articulography (EMA), with both conventional acoustic features and those extracted from large-scale pretrained models including Whisper and XLSR-53 as well as the fine-tuned Whisper model. We propose end-to-end (E2E) Conformer-based acoustic-articulatory models for ADSR and compare their performance against the corresponding hybrid TDNNF models. The experimental results show that using the fine-tuned Whisper features (Whisper-FT) fused with articulatory features achieves the lowest (10.5%) word error rate (WER) on dysarthric speech, with particularly significant improvements for severely dysarthric speech, reaching a WER of 20.8%.

**The Inclusion of Students from high Level Education with Disabilities through ICT's,**

Nowadays, the challenges that teachers face on different levels of education are varied and some are more complex than others, such is the case of the inclusion of students who present some disability in high level education with the rest of the class. The role of teacher is of the utmost importance, as this is the first contact with the student, it should be the support and bridge which will best manage the relationships between students, as well as finding the best path that allows all students to advance at the same pace and above all to make possible a better development in the classroom by the students. The Information and Communication Technologies (ICT's) play an important role in education, since they're one of the primary technological tools which the students count with to comprehend, communicate and exchange information between them, socialize and develop better and better as much as outside and inside the classroom.

**Normal versus pathology voice-an analysis,**

This paper proposes a novel method for diagnosis of dysathria using acoustic voice analysis. It is an effective and on-invasive method that could be used for confirmation of initial diagnosis and it would provide an objective determination of the vocal impairment. The acoustic parameters pitch and formants are extracted which gives differentiation of normal and pathological voices. Support Vector Machine is used to classify normal and disordered voice, the proposed algorithm offers least computational complexity with an accuracy of 80 to 90%.

**Two-Step Acoustic Model Adaptation for Dysarthric Speech Recognition,**

This paper introduces a model adaptation approach for a speaker-dependent dysarthric speech recognition system. The dysarthria we focus on in this paper is caused by athetoid cerebral palsy, which causes involuntary muscle movements in those with the disease. For this reason, the dysarthric people's speech is often unstable and difficult for conventional automatic speech recognition (ASR) systems to recognize. A model-adaptation approach, which adapts an ASR model to dysarthric speech, is one possible solution. However, because the difference in speaking styles between dysarthric and non-dysarthric people is so significant, the conventional adaptation method is not able to sufficiently adapt the model to the dysarthric speech. In our proposed two-step model-adaptation approach, an ASR model is first adapted to the general speaking style of multiple dysarthric speakers, and then the adapted model is further adapted for the target speaker. From our experiments on an ASR task, our two-step adaptation approach showed better performance than a conventional one-step adaptation approach.

**Multi-view representation learning via gcca for multimodal analysis of Parkinson's disease,**

Information from different bio-signals such as speech, handwriting, and gait have been used to monitor the state of Parkinson's disease (PD) patients, however, all the multimodal bio-signals may not always be available. We propose a method based on multi-view representation learning via generalized canonical correlation analysis (GCCA) for learning a representation of features extracted from handwriting and gait that can be used as a complement to speech-based features. Three different problems are addressed: classification of PD patients vs. healthy controls, prediction of the neurological state of PD patients according to the UPDRS score, and the prediction of a modified version of the Frenchay dysarthria assessment (m-FDA). According to the results, the proposed approach is suitable to improve the results in the addressed problems, specially in the prediction of the UPDRS, and m-FDA scores.

**Phonological features in discriminative classification of dysarthric speech,**

In an attempt to overcome problems associated with articulatory limitations and generative models, this work considers the use of phonological features in discriminative models for disabled speech. Specifically, we train feed-forward and recurrent neural networks, and radial basis and sequence-kernel support vector machines to abstractions of the vocal tract, and apply these models to phone recognition on dysarthric speech. The results show relative error reduction of between 1.5% and 10.9% with this approach against standard hidden Markov modeling, and increases in accuracy with speaker intelligibility across all classifiers. This work may be applied within components of assistive software for speakers with dysarthria.

**Objective Measures of Plosive Nasalization in Hypernasal Speech,**

Hypernasal speech is a common symptom across several neurological disorders; however it has a variable acoustic signature, making it difficult to quantify acoustically or perceptually. In this paper, we propose the nasal cognate distinctiveness features as an objective proxy for hypernasal speech. Our method is motivated by the observation that incomplete velopharyngeal closure changes the acoustics of the resultant speech such that alveolar stops /t/ and /d/ map to the alveolar nasal /n/ and bilabial stops /b/ and /p/ map to bilabial nasal /m/. We propose a new family of features based on likelihood ratios between the plosives and their respective nasal cognates. These features are based on an acoustic model that is trained only on healthy speech, and evaluated on a set of 75 speakers diagnosed with different dysarthria subtypes and exhibiting varying levels of hypernasality. Our results show that the family of features compares favorably with the clinical perception of speech-language pathologists subjectively evaluating hypernasality.

**Quantifying perturbations in temporal dynamics for automated assessment of spastic dysarthric speech intelligibility,**

Spastic dysarthric speech is often associated with imprecise placement of articulators which, in turn, cause perturbations in speech temporal dynamics, such as unclear distinctions between adjacent phonemes. While these perturbations can lead to a significant reduction in intelligibility, measures to objectively assess their detrimental effect on intelligibility are lacking. In this paper, short- and long-term temporal dynamics measures are proposed and evaluated as correlates of subjective intelligibility. The former is based on log-energy temporal dynamics information, whereas the latter is based on an auditory-inspired modulation spectral signal representation. A composite measure is also developed based on linearly combining the proposed measures with a tone-unit duration parameter. Experiments with the publicly-available 'Universal Access' database of spastic dysarthric speech show that the proposed composite measure can achieve rank correlations with subjective ratings as high as 0.87, thus providing a tool to automatically diagnose speech disorder severity and to evaluate dysarthria treatment outcomes.

**Improving the intelligibility of dysarthric speech by modifying system parameters, retaining speaker's identity,**

Dysarthria is a neuromotor impairment of speech that affects one or more subsystems involved in speech production. Such impairment is reflected in the acoustic characteristics of phonemes uttered by a dysarthric speaker. If such a speaker suffers from laryngeal dysfunction and improper articulation, then he/she may not be able to utter some/most of the phonemes properly. In our work, from the utterance of a dysarthric speaker, the poorly uttered phonemes are located and replaced with that of the normal speaker's speech signal. However, the resultant speech signal after concatenation doesn't sound natural due to the discontinuities, at the concatenation points in short-term energy, pitch period, and formant contour. In our work, the discontinuity at the concatenation point, in the short-term energy function is handled by smoothening the short-term energy of few frames before and after the concatenation point. Since, the pitch period in the replaced segment (phoneme) is considerably different from the dysarthric speaker's pitch period, the pitch period is adjusted to resemble the dysarthric speaker. The quality and naturalness of the utterance, after pitch modification, are considerably increased. The discontinuity in the formant contour is due to the reason that the co-articulation effect is absent since the replaced unit is taken from a different context. From the linear prediction analysis, the pole locations and their corresponding radii are adjusted based on the pole locations of adjacent phonemes. The quality and naturalness of speech signal, after all the three modifications, are found to be very close to the natural speech.

**Landmark based modification to correct distortions in dysarthric speech,**

Dysarthria is a motor speech disorder in which the person lacks the control over articulators used for speech production. This results in distorted speech, whose intelligibility is poor compared to the normal speakers. The distortion depends on the type of articulator and severity of the disease, because of which the problems tend to be person specific. Phonemic distortions in the dysarthric speech database for universal access research are analyzed in this work and the problem of devoicing in voiced stops is addressed. In this work, segments for modification are identified by automatically detecting the landmarks. Devoicing in voiced stops is rectified by introducing voice bar of low frequency before the burst onset. Intelligibility assessment is carried out by subjective test where, listeners assessed the intelligibility. Large improvement in intelligibility is found after modification.

**Improved Silence-Unvoiced-Voiced (SUV) Segmentation for Dysarthric Speech Signals using Linear Prediction Error Variance,**

A novel algorithm for the segmentation of dysarthric speech into silence, unvoiced and voiced (SUV) segments is presented. The proposed algorithm is based on the combination of short-time energy (STE), zero-crossing rate (ZCR) and linear prediction error variance (LPEV) or the segmentation problem. Extending the previous work in this field, the proposed method will address the difficulties in distinguishing between voiced and unvoiced segments in dysarthric speech. More precisely, the error variance of the linear prediction coefficients will be used to design a three-fold decision matrix that can accommodate the high variability in loudness experienced in dysarthric speech. In addition, a moving average threshold approach will be proposed in order to provide an “as-fit” segmentation technique that is fully automated and that will be able to handle highly severe dysarthric speech with varying loudness and ZCRs. The ability of the proposed fully-automated algorithm will be validated using real speech samples from healthy speakers, and speakers with ataxic dysarthria. The results of the proposed approach are compared with known methods using STE and ZCR. It is observed that the proposed classification method does not only show an improvement in segmentation performance but also provides consistent results in low signal energy situations.

**Sensorimotor adaptation of speech using real-time articulatory resynthesis,**

Sensorimotor adaptation is an important focus in the study of motor learning for non-disordered speech, but has yet to be studied substantially for speech rehabilitation. Speech adaptation is typically elicited experimentally using LPC resynthesis to modify the sounds that a speaker hears himself producing. This method requires that the participant be able to produce a robust speech-acoustic signal and is therefore not well-suited for talkers with dysarthria. We have developed a novel technique using electromagnetic articulography (EMA) to drive an articulatory synthesizer. The acoustic output of the articulatory synthesizer can be perturbed experimentally to study auditory feedback effects on sensorimotor learning. This work aims to compare sensorimotor adaptation effects using our articulatory resynthesis method with effects from an established, acoustic-only method. Results suggest that the articulatory resynthesis method can elicit speech adaptation, but that the articulatory effects of the two methods differ.

**Exploring articulatory characteristics of Cantonese dysarthric speech using distinctive features,**

Dysarthria is a kind of motor speech disorder due to neurological deficits. Understanding the articulatory problems of dysarthric speakers may help to design suitable intervention strategies to improve their speech intelligibility. We have developed an automatic articulatory characteristics analysis framework based on a distinctive feature (DF) recognition. We recruited 16 Cantonese dysarthric subjects with spinocerebellar ataxia (SCA) or cerebral palsy (CP) to support our research. To the best of our knowledge, this is among the first efforts in collecting and automatically analyzing Cantonese dysarthric speech. The framework shows a close Pearson correlation to manual annotation of the subjects in most DFs and also in the average DF error rates. It indicates a potential way to describe articulatory characteristics of dysarthric speech and automatically assess it.

**Vowel Formant Transformation Techniques for Dysarthric Speech,**

Speech communication is the fundamental of an individual's participation in society. But this type of communicating medium is often disrupted by various types of physical disorder. Dysarthria is one such disorder that encapsulates various neuro-motor disorders that disrupt or impair the physical production of speech. Speech of dysarthric speaker is less intelligible when compared to normal speaker, which creates difficulty during communication. One way to improve the intelligibility of dysarthric speech is to modify the formants of dysarthric speech and resynthesize the speech using modified formants. In this work, we modify the formants of dysarthric vowels by using three different formant transformation techniques available. When tested on the same data with root mean square as evaluation parameter, it is found that use of probabilistic least square regression results in smaller RMSE compared to formant transformation using target features and joint density estimation.

**A Novel Dysarthric Speech Synthesis system using Tacotron2 for specific and OOV words,**

Among various speech disorders, dysarthria presents a unique challenge when it comes to end-to-end speech synthesis. Its severity and complexity further escalate if appropriate treatment is not taken. In this paper, we propose a speaker-adaptive dysarthric speech synthesis technique using the Tacotron2 model. We used this model to generate dysarthric speech utterances that already exist in the UASpeech database and for Out-Of-Vocabulary (OOV) words as well to expand the vocabulary size. By generating dysarthric speech from textual input, we achieved favorable Mean Opinion Score (MOS) ratings for both known and OOV words. This approach was successful in adapting to the inherent variability and intelligibility differences among speakers. To ensure the fidelity of our generated speech, we incorporated Dynamic Time Warping (DTW) to demonstrate the similarity between the original and generated waveforms. Additionally, we applied Waveform similarity-based Synchronized OverLap-Add (WSOLA) on the OOV words, ensuring a flawless evaluation process. This approach allows us to address the challenge of limited data availability by enhancing database and vocabulary size. This paper aims to generate synthetic dysarthric speech to expand the pool of available UASpeech database, facilitating advancements in automatic recognition of dysarthric speech. Audio samples for all generated words for 6 dysarthric speakers are available at https://github.com/samiulhaq4424/MOS

**Fusion of Multiple Audio Descriptors for the Recognition of Dysarthric Speech,**

Dysarthria is a motor speech disorder majorly caused from brain stroke or severe accidents affecting the left hemisphere of the brain. Due to the injury, speech muscles and tongue movement become weak, leading to highly unintelligible speech. In the past decade, the performance of Automatic Speech Recognition (ASR) systems have improved signif-icantly for normal speech, but when it comes to dysarthric speech, performance is far from satisfactory. In this pa-per we implement dysarthric speech recognition system in a speaker-adaptive manner, leveraging additional acoustic features to help the model effectively learn and adapt to each speaker's unique speech patterns. Feature selection is based on MFCC combined with different variants of Jitter and Shimmer. Despite the challenges posed by limited data and considerable variability among the speakers, our system achieves an average Word Recognition Accuracy (WRA) of 83.11 % on the UASpeech database, which is a significant improvement over previous studies. In addition to ASR, we also analyze dysarthric speakers with very low intelligibility rates and found some prominent insights into their speech characteristics.

**Comparison in Suprasegmental Characteristics between Typical and Dysarthric Talkers at Varying Severity Levels,**

Dysarthria is a speech disorder often characterized by slow speech with reduced intelligibility. This preliminary study investigates suprasegmental characteristics between typical and dysarthric speakers at varying severity levels, with the long-term goal of improving methods for dysarthric speech synthesis/augmentation and enhancement. First, we aim to analyze phonemes, speaking rate and pause characteristics of typical and dysarthric speech using the phoneme- and word-level alignment information extracted by Montreal Forced Aligner (MFA). Then, pitch and intensity declination trends and range analysis are conducted. The pitch and intensity declination are measured by fitting a regression line. These analyses are conducted on dysarthric speech in TORGO, containing 8 dysarthric speakers involved with cerebral palsy or amyotrophic lateral sclerosis and 7 age- and gender-matched typical speakers. These results are important for the development of dysarthric speech synthesis, augmentation to statistically model and evaluate characteristics such as pause, speaking rate, pitch, and intensity.

**A Situational Semantic Projection Model for Ontology Completion in Dysarthric Speech Using Emotion and Dialogue Acts,**

Enhancing semantic understanding of dysarthric speech via ontology completion faces challenges due to variability and ambiguity of dysarthric speech. This study introduces a novel situational semantic projection model that employs emotions and dialogue acts to enhance ontology completion accuracy. By incorporating emotional cues and dialogue act information, the model robustly maps dysarthric speech to relevant semantic categories. Data from 320 speech prompts of dysarthria patients from the DEED database, supplemented by 550 prompts from the TORGO database, were utilized. Convolutional neural networks extracted speech dialogue acts, while dense neural networks generated emotions. Ontology triples were extracted using Spacy’s engine, with cosine similarity-based estimation of missing triple elements. Following an accuracy of 94% in triple extraction, the situational semantic projection model achieved a 74% hit for a cosine closeness of 0.001 and a 55% correlation between the triple classes and the corresponding emotional dialogue acts.

**Exploring Speech Profiles in Parkinson's Disease Patients on Long - Term Dopaminergic Medication,**

Hypokinetic dysarthria (HD), prevalent in Parkin-son's disease (PD), is characterized by reduced motor movement, leading to speech and voice impairments. The effectiveness of dopamine therapy varies, and symptoms may change over time. To explore the HD in patients on medication, we conducted hierarchical clustering to categorize 83 PD patients using the acoustic features extracted from recorded text reading, sustained phonation, and diadochokinetic task. We then compared these clusters to a group of 44 healthy controls. Regardless of the medication dosage and PD duration, patients displayed three distinct speech profiles: articulatory; respiratory-prosodic; and another more complex/severe regarding the number of HD do-mains affected. The newly identified profiles exhibit distinctions from previously defined speech subtypes in de novo PD patients. No differences in clinical scores were observed among the newly identified profiles, suggesting uniformity in clinical presentation despite distinct speech characteristics.

**A New Method for Predicting Severity Level of Dysarthric Speech Based on Joint Feature-Sample Selection using Audio-Visual Data,**

Automatic objective assessment of dysarthria is valuable and crucial. Most previous studies focus on using audio-only data, ignoring the complementary of other modal data. In addition, traditional methods ignore the relationship between the pre-defined features and different pronunciations, reducing the performance of the automatic assessment system. To address these issues, this paper proposes a joint feature-sample selection (JFSS) based dysarthria severity level regression model using audio-visual data. In the proposed framework, relevant pronunciation samples and features are simultaneously obtained and unreliable noisy samples are discarded by the JFSS method. On the Mandarin Subacute Stroke Dysarthria Multimodal (MSDM) Database, the proposed regression model outperformed several baseline models. By using acoustic-visual features, the root mean square error (RMSE) of 13.78 and fitting coefficient R-square of 0.77 computed between the automatically predicted and perceptual evaluation metrics (i.e. Frenchay Dysarthria Assessment) were obtained, which confirmed the capacity of the proposed JFSS-based regression method in predicting dysarthria severity level.

**Brain Computer Interaction Framework for Speech and Motor Impairment Using Deep Learning,**

Some people may have trouble communicating with others because of motor and speech difficulties brought on by accidents, strokes, or diseases. Paralyzed people often struggle to communicate their demands, which can make even the simplest tasks challenging. People with disorders like dysarthria and amyotrophic lateral sclerosis may have trouble following conversations. The proposed method for automatic detection of everyday basic desires will help those with dysarthria and quadriplegic paralysis has more fulfilling lives. The device accomplishes this by monitoring brain impulses, which it then converts into either perceptible voice commands or messages that can be delivered to a healthcare provider's smart devices, depending on the user's preferences. The suggested method randomly displays an image of one of six basic requirements while using event-related-potentials(ERPs) detected from the Electroencephalogram (EEG) data to determine which need to fulfill. The dataset utilized for train, test, &validation was built using input from 10 participants. The proposed method had a 96.3|% success rate.

**Advancing Parkinson's Disease Detection: An Empirical Evaluation of Machine Learning Models Based on Speech Analysis,**

Parkinson's disease (PD) is a neurodegenerative disorder that adversely impacts a considerable portion of the population, especially senior citizens. Currently, the basis of PD diagnosis is clinical assessments, which can be costly, time-consuming, and invasive, that can be subjective or prone to errors. Dysarthria, a prevalent condition characterized by slow and distorted speech, often coexists with PD, presenting an opportunity to leverage speech features for diagnostic purposes. In this research paper, the altered speech features caused by Dysarthria are harnessed to train a diverse set of cutting-edge machine learning algorithms, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forests, and Decision Trees. A thorough examination of the latest advancements in the diagnosis of Parkinson's disease (PD) using speech analysis is provided. Additionally, a comparative analysis between the performance of a newly introduced classifier, HyperTab, and existing models is conducted. The findings reveal the remarkable promise of machine learning in speech-based PD diagnosis, paving the way for developing a cost-effective tool for accelerating disease detection. Moreover, this contribution is crucial in providing vital support to those areas with limited access to specialized medical facilities, significantly improving the quality of life and outcomes for PD patients in rural regions.

**Speech Synthesis Using Ambiguous Inputs From Wearable Keyboards,**

This paper proposes a new application in speech communication using text-to-speech (TTS), and the goal is to enable dysarthria, articulation disorder, or persons who have difficulty in speaking to communicate anywhere and anytime using speech to express their thoughts and feelings. To achieve this goal, an input method is required. Thus, we propose a new text-entry method based on three concepts. First, from an easy-to-carry perspective, we used a wearable keyboard that inputs digits from 0 to 9 in decimal notation according to 10-finger movements. Second, from a no-training perspective, users input sentences in a way of touch typing using the wearable keyboard. Following this method, we obtained a sequence of numbers corresponding to the sentence. Third, a neural machine translation (NMT) method is applied to estimate texts from the sequence of numbers. The NMT was trained using two datasets; one is a Japanese-English parallel corpus containing 2.8 million pairs of sentences, which were extracted from TV and movie subtitles, while the other is a Japanese text dataset containing 32 million sentences, which were extracted from a question-and-answer platform. Using the model, phonemes and accent symbols were estimated from a sequence of numbers. Thus, the result accuracy in symbol levels was 91.48% and 43.45% of all the sentences were completely estimated with no errors. To subjectively evaluate feasibility of the NMT model, a two-person word association game was conducted; one gave hints using synthesized speech that is generated from symbols estimated by NMT, while the other guessed answers. As a result, 67.95% of all the quizzes were correctly answered, and experiment results show that the proposed method has the potential for dysarthria to communicate with TTS using a wearable keyboard.

**Assessment of dysarthric speech and an analysis on velopharyngeal incompetence,**

In the present work the role of a speech recognition system in the assessment of dysarthric speech is studied. Initially a continuous speech recognition system is developed for the assessment of dysarthric speech. The performance of the continuous speech recognition system on the assessment of dysarthric speech is found to be dis-satisfactory due to greater number of insertions. Analysis conducted on duration of phoneme and speech rate confirms the fact that the more number of insertions in the recognizer output is due to lower speech rate of dysarthric speakers. To overcome the problem with the continuous speech recognition system an isolated-style speech recognition system is developed. The performance of this system on the assessment is compared with the Frenchay dysarthric assessment (FDA) scores provided with the Nemours speech corpus. From the performance of the isolated-style speech recognition system it is observed that apart from the articulatory problems, some of the speakers are affected with velopharyngeal incompetence also and is analyzed with group delay function based acoustic measure for the detection of hypernasality on dysarthric speech

**1. State-of-the-art speaker recognition methods applied to speakers with dysarthria,**

Speech-based biometrics is one of the most effective ways for identity management and one of the preferred methods by users and companies given its flexibility, speed and reduced cost. Current state-of-the-art speaker recognition systems are known to be strongly dependent on the condition of the speech material provided as input and can be affected by unexpected variability presented during testing, such as environmental noise, changes in vocal effort or pathological speech due to speech and/or voice disorders. In this chapter, we are particularly interested in understanding the effects of dysarthric speech on automatic speaker identification performance. We explore several state-of-theart feature representations, including i-vectors, bottleneck neural-networkbased features, as well as a covariance-based feature representation. High-level features, such as i-vectors and covariance-based features, are built on top of four different low-level presentations of dysarthric/controlled speech signal. When evaluated on TORGO and NEMOURS databases, our best single system accuracy was 98.7%, thus outperforming results previously reported for these databases.

**Dysarthric Speech Enhancement Based on Convolution Neural Network,**

Generally, those patients with dysarthria utter a distorted sound and the restrained intelligibility of a speech for both human and machine. To enhance the intelligibility of dysarthric speech, we applied a deep learning-based speech enhancement (SE) system in this task. Conventional SE approaches are used for shrinking noise components from the noise-corrupted input, and thus improve the sound quality and intelligibility simultaneously. In this study, we are focusing on reconstructing the severely distorted signal from the dysarthric speech for improving intelligibility. The proposed SE system prepares a convolutional neural network (CNN) model in the training phase, which is then used to process the dysarthric speech in the testing phase. During training, paired dysarthric-normal speech utterances are required. We adopt a dynamic time warping technique to align the dysarthric-normal utter-ances. The gained training data are used to train a CNN - based SE model. The proposed SE system is evaluated on the Google automatic speech recognition (ASR) system and a subjective listening test. The results showed that the proposed method could notably enhance the recognition performance for more than 10% in each of ASR and human recognitions from the unprocessed dysarthric speech. Clinical Relevance— This study enhances the intelligibility and ASR accuracy from a dysarthria speech to more than 10%

**A Few-Shot Approach to Dysarthric Speech Intelligibility Level Classification Using Transformers,**

Dysarthria is a speech disorder that hinders communication due to difficulties in articulating words. Detection of dysarthria is important for several reasons as it can be used to develop a treatment plan and help improve a person’s quality of life and ability to communicate effectively. Much of the literature focused on improving ASR systems for dysarthric speech. The objective of the current work is to develop models that can accurately classify the presence of dysarthria and also give information about the intelligibility level using limited data by employing a few-shot approach using a transformer model. This work also aims to tackle the data leakage that is present in previous studies. Our whisper-large-v2 transformer model trained on a subset of the UASpeech dataset containing medium intelligibility level patients achieved an accuracy of 85%, precision of 0.92, recall of 0.8 F1-score of 0.85, and specificity of 0.91. Experimental results also demonstrate that the model trained using the ’words’ dataset performed better compared to the model trained on the ’letters’ and ’digits’ dataset. Moreover, the multiclass model achieved an accuracy of 67%.

**Exploring AI-based Speaker Dependent Methods in Dysarthric Speech Recognition,**

In this paper, we present our recent improvements within the CapisciAMe project, an Italian initiative aimed at investigating the usage of deep learning strategies for automatic speech recognition in presence of speech disorders, such as dysarthria. Our research activity is focused on isolated word recognition by exploiting a convolutional neural network (CNN) architecture to predict the presence of a reduced number of speech commands within an atypical speech. Currently, by following speaker-dependent approaches, our speech models have been trained on a 21K speech dataset consisting of voice contributions, i.e., single speech data recordings, from 156 Italian users with neuromotor disabilities and dysarthria. Having a large number of repetitions (into the thousands) for each word on which to train our deep learning model is of crucial importance for our project. Nevertheless, people with impaired speech are generally weak in repetitive vocalization tasks, so producing a large number of speech samples for each work is a complex operation to be accomplished. To mitigate this difficulty, we investigate possible relationships between the number of samples for each word and the accuracy of automatic speech recognition. This study plays a critical role in our research, allowing us to minimize the amount of speech data samples required for each work from people with dysarthria to train the automatic speech recognition system.

**Smart fog: Fog computing framework for unsupervised clustering analytics in wearable Internet of Things,**

The increasing use of wearables in smart telehealth system led to the generation of large medical big data. Cloud and fog services leverage these data for assisting clinical procedures. IoT Healthcare has been benefited from this large pool of generated data. This paper suggests the use of low-resource machine learning on Fog devices kept close to wearables for smart telehealth. For traditional telecare systems, the signal processing and machine learning modules are deployed in the cloud that processes physiological data. This paper presents a Fog architecture that relied on unsupervised machine learning big data analysis for discovering patterns in physiological data. We developed a prototype using Intel Edison and Raspberry Pi that was tested on real-world pathological speech data from telemonitoring of patients with Parkinson's disease (PD). Proposed architecture employed machine learning for analysis of pathological speech data obtained from smart watches worn by the patients with PD. Results show that proposed architecture is promising for low-resource machine learning. It could be useful for other applications within wearable IoT for smart telehealth scenarios by translating machine learning approaches from the cloud backend to edge computing devices such as Fog.

**Perceptually Enhanced Single Frequency Filtering for Dysarthric Speech Detection and Intelligibility Assessment,**

This paper proposes a new speech feature representation that improves the intelligibility assessment of dysarthric speech. The formulation of the feature set is motivated from the human auditory perception and high time-frequency resolution property of single frequency filtering (SFF) technique. The proposed features are named as perceptually enhanced single frequency cepstral coefficients (PE-SFCC). As a part of SFF technique implementation, speech signal passed through a single pole complex bandpass filter bank to obtain high-resolution time-frequency distribution. Then, the distribution is enhanced by using a set of auditory perceptual operators. Lastly, traditional homomorphic analysis has been carried out on the resulting signal to obtain PE-SFCC feature vector. The performance of proposed features in dysarthric speech detection and its intelligibility assessment has been reported on UASPEECH database. The PE-SFCC features outperformed the state-of-the-art features in dysarthric speech detection and intelligibility assessment.

**Dysarthric Speech Recognition with Lattice-Free MMI,**

Recognising dysarthric speech is a challenging problem as it differs in many aspects from typical speech, such as speaking rate and pronunciation. In the literature the focus so far has largely been on handling these variabilities in the framework of HMM/GMM and cross-entropy based HMM/DNN systems. This paper focuses on the use of state-of-the-art sequence-discriminative training, in particular lattice-free maximum mutual information (LF-MMI), for improving dysarthric speech recognition. Through a systematic investigation on the Torgo corpus we demonstrate that LF-MMI performs well on such atypical data and compensates much better for the low speaking rates of dysarthric speakers than conventionally trained systems. This can be attributed to inherent aspects of current speech recognition training regimes, like frame subsampling and speed perturbation, which obviate the need for some techniques previously adopted specifically for dysarthric speech.

**Degree of Parkinson's disease severity estimation based on speech signal processing,**

This paper deals with Parkinson's disease (PD) severity estimation according to the Unified Parkinson's Disease Rating Scale: motor subscale (UPDRS III), which quantifies the hallmark symptoms of PD, using an acoustic analysis of speech signals. Experimental dataset comprised 42 speech tasks acquired from 50 PD patients (UPDRS in ranged from 6 to 92). It was divided into subsets: words, sentences, reading text, monologue and diadochokinetic tasks. We performed a parametrization of the whole corpus and these groups separately using a wide range of conventional and novel speech features. We used guided regularized random forest algorithm to select features with maximum clinical information and performed random forests regression to estimate PD severity. According to significant correlations between true UPDRS in scores and scores predicted by the proposed methodology it was shown that information extracted through variety of speech tasks can be used to estimate the degree of PD severity.

**Design and creation of Dysarthric Speech Database for development of QoLT software technology,**

In this paper we will introduce the work of creation of a speech database to develop speech technology for disabled persons, which has been done as part of a national program to help better life for Korean people. We will report about the creation of speech database of a total of 160 persons: prompting items, designs, etc. for the creation of a database which is needed to develop an embedded key-word spotting speech recognition system tailored for the persons disabled in articulation. The created database is being used by the technology development team in the national program to study the phonetic characteristics of the different types of disabled persons, develop the automatic method to assess degrees of disability, investigate the phonetic features of speech of the disabled, and design and implement the software prototype for personal embedded speech recognition systems adapted to the disabled persons.

**Speech assistive technology to improve the interaction of dysarthric speakers with machines,**

In this paper, we propose a communication scheme using automatic speech recognition and speech synthesis in order to assist dysarthric speakers. This system aims at improving recognition rate and intelligibility of dysarthric speech. An HMM-based recognizer using variable duration of Hamming window permits to raise the recognition rate of dysarthric speech up to 80 %. In order to improve the intelligibility of synthetic speech while keeping the naturalness close to the voice of dysarthric speaker, we synthesize the recognized text using new basic unit segmenter, a new concatenating algorithm and a grafting technique to correct the bad pronounced phonemes. The NEMOURS dysarthric database is used to evaluate the proposed assistive communication system. Results show that a rate of 65 % to 80% of correct word recognition and a Mean Opinion Score (MOS) of 4 were obtained.

**Improving dysarthric speech intelligibility through re-synthesized and grafted units,**

In this paper, an assistive speech-enabled system is proposed to help dysarthric speakers. This system aims at improving the intelligibility of dysarthric speech making it as natural as possible and close to the original voice of the dysarthric speaker. The resynthesized utterances use new basic units, a new concatenating algorithm and a grafting technique to correct the bad pronounced phonemes. The Nemours database is used to evaluate the proposed assistive communication system. The results of subjective evaluations show that the proposed technique reaches the very satisfactory MOS (mean opinion score) of 4.

**Dysarthric speech corpus in Tamil for rehabilitation research,**

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.

**Nonverbal Sound Detection for Disordered Speech,**

Voice assistants have become an essential tool for people with various disabilities because they enable complex phone-or tablet-based interactions without the need for fine-grained motor control, such as with touchscreens. However, these systems are not tuned for the unique characteristics of individuals with speech disorders, including many of those who have a motor-speech disorder, are deaf or hard of hearing, have a severe stutter, or are minimally verbal. We introduce an alternative voice-based input system which relies on sound event detection using fifteen nonverbal mouth sounds like "pop", "click", or "eh." This system was designed to work regardless of ones’ speech abilities and allows full access to existing technology. In this paper, we describe the design of a dataset, model considerations for real-world deployment, and efforts towards model personalization. Our fully-supervised model achieves segment-level precision and recall of 88.6% and 88.4% on an internal dataset of 710 adults, while achieving 0.31 false positives per hour on aggressors such as speech. Five-shot personalization enables satisfactory performance in 84.5% of cases where the generic model fails.

**DNN acoustic models for dysarthric speech,**

In this paper, we investigate various training methods for building deep neural network (DNN) based acoustic models for dysarthric speech data. Methods like multitask learning, knowledge distillation and model adaptation, which overcome data sparsity and model over-fitting problems are employed to study the merits of each method. In Knowledge distillation framework, some privilege information in addition to featurelabels pairs available only during training, is exploited to help the model learn better without using such previleged information during testing [1]; knowledge from one model can be distilled to another and thereby guiding it in learning better [2]. In this work, a DNN acoustic model trained using data pooled from Dysarthric speech data and parallel un-impaired data is used as the intelligent teacher while the student DNN model is trained using only Dysarthric speech. The target label for training the student model is a combination of hard aligned labels and those obtained from forward-pass through the teacher model. In addition to this technique, other knowledge sharing techniques like multitask learning were explored for Dysarthric speech data and have found to show a relative improvement of 11% over the corresponding baseline models.

**Objective Assessment of Vocal Tremor,**

Detecting early signs of neurodegeneration is vital for planning treatments for neurological diseases. Speech plays an important role in this context because it has been shown to be a promising early indicator of neurological decline, and because it can be acquired remotely without the need for specialized hardware. Typically, symptoms are characterized by clinicians using subjective and discrete scales. The poor resolution and subjectivity of these scales can make the earliest speech changes hard to detect. In this paper, we propose an algorithm for the objective assessment of vocal tremor, a phenomenon associated with many neurological disorders. The algorithm extracts and aggregates a feature set from the average spectra of the energy and fundamental frequency profiles of a sustained phonation. We show that the resultant low-dimensional feature set reliably classifies healthy controls and patients with amyotrophic lateral sclerosis perceptually rated for tremor by speech language pathologists.

**A NLP-based Approach to Improve Speech Recognition Services for People with Speech Disorders,**

Current speech recognition services are not suitable for people with speech disorders, which present difficulties in coordinating muscles and articulating words and sentences. In this case, a speaker-dependent approach is strongly required in order to address the specific vocal disarticulation. Several Deep learning approaches have been proposed in the literature to address this problem. However, they require many voice samples of people to properly work, and this is not practical. In this paper, we present an innovative Automatic Speech Recognition (ASR) system which is able to correct failures of deep learning based solution adopting Natural Language Processing (NLP) techniques. The proposed solution can perform both single word and whole sentence corrections by analyzing the speech context. We evaluated the solution in a home automation case study and proved the good accuracy of our model.

**Dysarthric Speech Augmentation Using Prosodic Transformation and Masking for Subword End-to-end ASR,**

End-to-end speech recognition systems are effective, but in order to train an end-to-end model, a large amount of training data is needed. For applications such as dysarthric speech recognition, we do not have sufficient data. In this paper, we propose a specialized data augmentation approach to enhance the performance of an end-to-end dysarthric ASR based on sub-word models. The proposed approach contains two methods, including prosodic transformation and time-feature masking. Prosodic transformation modifies the speaking rate and pitch of normal speech to control prosodic characteristics such as loudness, intonation, and rhythm. Using time and feature masking, we apply a mask to the Mel Frequency Cepstral Coefficients (MFCC) for robustness-focused augmentation. Results show that augmenting normal speech with prosodic transformation plus masking decreases CER by 5.4% and WER by 5.6%, and the further addition of dysarthric speech masking decreases CER by 11.3% and WER by 11.4%.

**ParkinSense: A Novel Approach to Remote Idiopathic Parkinson’s Disease Diagnosis, Severity Profiling, and Telemonitoring via Ensemble Learning and Multimodal Data Fusion on Webcam-Derived Digital Biomarkers,**

Despite being the fastest growing Neurodegenerative disease in the world, with over ten million cases worldwide, there is no definitive diagnosis method for Parkinson’s Disease currently. Current diagnosis technologies misdiagnose one in four patients, relying on tests and technologies that are subjective to the administrating clinicians. Furthermore, they are inaccessible to billions due to immobility, geographic barriers, or associated costs. With early and accurate diagnosis being vital to effective treatment, a tremendous issue emerges. ParkinSense addresses these issues, serving as an automated web application that contactlessly analyzes three cardinal symptoms of Parkinson’s Disease over a standard webcam and microphone: Hypomimia, Dysarthria, and Bradykinesia. With Ensemble Learning coupled with various classifiers, ParkinSense was able to achieve an accuracy rate of 99.72% when tested on synthetic patients. By examining and combining multiple modalities, ParkinSense boasts accuracy rates higher than many advanced unimodal technologies, as patients often do not substantially display all the symptoms of Parkinson’s Disease, leading to misdiagnosis when relying on one symptom for diagnosis. Furthermore, ParkinSense is able to accurately estimate disease severity for positively diagnosed patients, with a Mean Absolute Error of 9.7 on the Unified Parkinson’s Disease Rating Scale. ParkinSense strongly suggests that an entirely contactless diagnosis of Parkinson’s Disease through digital biomarkers can be effective, and web applications can be utilized for rapid, automated, and remote diagnosis and severity analysis, which can be crucial for those affected by barriers that prevent access to healthcare.

**Phase-Based Feature Representations for Improving Recognition of Dysarthric Speech,**

Dysarthria is a neurological speech impairment, which usually results in the loss of motor speech control due to muscular atrophy and incoordination of the articulators. As a result the speech becomes less intelligible and difficult to model by machine learning algorithms due to inconsistencies in the acoustic signal and data sparseness. This paper presents phase-based feature representations for dysarthric speech that are exploited in the group delay spectrum. Such representations are found to be better suited to characterising the resonances of the vocal tract, exhibit better phone discrimination capabilities in dysarthric signals and consequently improve ASR performance. All the experiments were conducted using the UASPEECH corpus and significant ASR gains are reported using phase-based cepstral features in comparison to the standard MFCCs irrespective of the severity of the condition.

**MFCC-Based Analysis of Vibratory Anomalies in Parkinson's Disease Detection using Sustained Vowels,**

Parkinson's disease (PD) is a neurological disorder resulting from the degeneration of dopaminergic neurons, which leads to speech production disorders, particularly at the level of the vibratory and articulatory aspects. Early clinical observation of voice disorders can be difficult, hence the growing interest in voice analysis-based diagnoses using machine learning methods. Mel frequency cepstral coefficients (MFCC) are widely used in speech and audio processing as they capture valuable insight about the shape of the vocal tract. The objective of this study is to investigate the relevance of MFCC coefficients as descriptors to identify abnormal vibratory patterns indicative of Parkinson's disease during sustained vowel production. To do this, we evaluate the performance of MFCC in discriminating voices of patients with PD from healthy voices, without combining them with any other speech features. We also perform a comparative study of different classification techniques, such as KNN, SVM, DT, and RF.

**An Investigation of End-to-End Speech Recognition Using Model Adaptation for Dysarthric Speakers,**

In this paper, we present an end-to-end automatic speech recognition (ASR) system for dysarthric speech. Because the speaking style of a person suffering from an articulation disorder is quite different from that of a physically unimpaired person, speech recognition systems for such persons need to be constructed in such a way that they specialize in meeting the needs of such dysarthric people. However, the amount of training data that can be collected from dysarthric people is limited because of their large burden. Therefore, it is a challenge to effectively train an ASR model for dysarthric people. In this paper, we introduce a model adaptation approach to train a more accurate model with limited training data, which adapts an ASR model trained by non-dysarthric speech samples for dysarthric speech recognition. From our experiments on an ASR task with two dysarthric subjects, the model adaptation approach with non-dysarthric speech showed better performance than training from scratch.

**MiMi: Sinhala Language Speech Assistive Learning Bot to Support Children with Stuttering,**

This research paper presents “MiMi”, a Sinhala Language voice assistive gamified solution that is designed to address stuttering in children aged between three and fourteen. Speech disorders occur when the regular flow of communication is disrupted. Stuttering, Lisps, Dysarthria, and Apraxia are some variations of speech impairments. Stuttering can be caused by a variety of factors including physical weaknesses, inherited diseases, Autism, and accidents. The risk of continuing to stutter into adulthood is highest in children between the ages of three to fourteen. It is recognized that stuttering therapy activities were less effective in managing stuttering after this age. Stuttering treatments comprise speech therapy with speech-language therapists, which requires in-person sessions that can be challenging and expensive in some circumstances. A parent’s financial ability, their busy schedules, the state of the economy in the nation, and the feasibility of physically seeing therapists and enduring treatments are all factors that might encourage or demotivate participation in therapy sessions. The development in technology and technical approaches have revolutionized the medical field and several studies have been conducted regarding communication disorders in recent years. The application can be used to practice a child’s needed speech therapy virtually and can also be used to aid speech therapy sessions done by speech therapists. The main aim of the system is to provide a customized, engaging, and innovative therapeutic strategy for children to manage stuttering.

**A CNN Approach to Detect Parkinson's Disease using T1-Weighted, T2-Weighted, and Flair MRI,**

Parkinson's disease (PD) has always been under the category of incurable disease to date. This neurodegenerative disease is generally caused by the deficiency of dopamine-developing nerve cells in the brain. The primary symptom of PD is a tremor, or shaking, in one or both hands. Other symptoms may include harshness of the limbs and trunk, slow movement (bradykinesia), poor balance and coordination, and a change in speech or writing (dysarthria or micrographia). As the senior citizen population is mounting, the patients of Parkinson's are also increasing. Unfortunately, there is absolutely no effective treatment for Parkinson's disease. Therefore, the task at hand is to enhance the timely identification of the illness. This research study has developed a novel 2D convolutional neural network to learn the intricate patterns in three different types of MR Images (MRI) to detect Parkinson's disease, which is thought to be brought on by minor, undetected strokes in the brain or the severity of white matter lesions, using a sizable dataset available on Kaggle. Through this model, a high accuracy of 94% has been achieved. Next, transfer learning is applied by using InceptionV3. The proposed model demonstrates an overall accuracy of 95.29% by leveraging a pre-trained model. Moreover, the precision, recall, and F1-score stand at 95%, 100%, and 97% correspondingly and are higher for the large dataset.

**Improved Recognition of the Speech of People with Parkinson’s Who Stutter,**

Stuttering is a speech disorder often associated with neurological conditions, including Parkinson’s disease (PD). Despite advancements in modern automatic speech recognition (ASR) technologies, today’s systems still face challenges in accurately recognizing dysarthric speech, particularly when stuttering is present. In this study, we propose a novel stuttered speech data augmentation approach to improve dysarthric speech recognition. We utilize typical speech data from LibriSpeech to generate artificial stuttered speech by applying Voice Activity Detection and Forced Alignment techniques to accurately identify word boundaries, and integrating an adaptive stuttering filter to simulate severe stuttering patterns. Additionally, dysarthric speech data from individuals with PD, collected by the Speech Accessibility Project (SAP), is integrated into the model. Our experimental results demonstrate that the proposed augmentation approach outperforms existing methods in enhancing the recognition of stuttered speech. Furthermore, fine-tuning the ASR systems with SAP data yields additional performance improvements for both stuttering and non-stuttering individuals with PD.

**Graph Neural Networks for Parkinson’s Disease Detection,**

Despite the promising performance of state-of-the-art approaches for Parkinson’s Disease (PD) detection, these approaches often analyze individual speech segments in isolation, which can lead to sub-optimal results. Dysarthric cues that characterize speech impairments from PD patients are expected to be related across segments from different speakers. Isolated segment analysis fails to exploit these inter-segment relationships. Additionally, not all speech segments from PD patients exhibit clear dysarthric symptoms, introducing label noise that can negatively affect the performance and generalizability of current approaches. To address these challenges, we propose a novel PD detection framework utilizing Graph Convolutional Networks (GCNs). By representing speech segments as nodes and capturing the similarity between segments through edges, our GCN model facilitates the aggregation of dysarthric cues across the graph, effectively exploiting segment relationships and mitigating the impact of label noise. Experimental results demonstrate the advantages of the proposed GCN model for PD detection and provide insights into its underlying mechanisms.

**Vsameter: Evaluation of a New Open-Source Tool to Measure Vowel Space Area and Related Metrics,**

Vowel space area (VSA) is an applicable metric for studying speech production deficits and intelligibility. Previous works suggest that the VSA accounts for almost 50% of the intelligibility variance, being an essential component of global intelligibility estimates. However, almost no study publishes a tool to estimate VSA automatically with publicly available codes. In this paper, we propose an open-source tool called VSAmeter to measure VSA and vowel articulation index (VAI) automatically and validate it with the VSA and VAI obtained from a dataset in which the formants and phone segments have been annotated manually. The results show that VSA and VAI values obtained by our proposed method strongly correlate with those generated by manually extracted F1 and F2 and alignments. Such a method can be utilized in speech applications, e.g., the automatic measurement of VAI for the evaluation of speakers with dysarthria.

**Vision Transformer for Parkinson’s Disease Classification using Multilingual Sustained Vowel Recordings,**

Parkinson’s disease (PD) is the 2nd most prevalent neurodegenerative disease in the world. Thus, the early detection of PD has recently been the subject of several scientific and commercial studies. In this paper, we propose a pipeline using Vision Transformer applied to mel-spectrograms for PD classification using multilingual sustained vowel recordings. Furthermore, our proposed transformed-based model shows a great potential to use voice as a single modality biomarker for automatic PD detection without language restrictions, a wide range of vowels, with an F1-score equal to 0.78. The results of our study fall within the range of the estimated prevalence of voice and speech disorders in Parkinson’s disease, which ranges from 70-90%. Our study demonstrates a high potential for adaptation in clinical decision-making, allowing for increasingly systematic and fast diagnosis of PD with the potential for use in telemedicine.Clinical relevance— There is an urgent need to develop non invasive biomarker of Parkinson’s disease effective enough to detect the onset of the disease to introduce neuroprotective treatment at the earliest stage possible and to follow the results of that intervention. Voice disorders in PD are very frequent and are expected to be utilized as an early diagnostic biomarker. The voice analysis using deep neural networks open new opportunities to assess neurodegenerative diseases’ symptoms, for fast diagnosis-making, to guide treatment initiation, and risk prediction. The detection accuracy for voice biomarkers according to our method reached close to the maximum achievable value.

**Cross-Corpus Disparity of Parkinson's Voice Datasets Observed on Control Group Distribution,**

Parkinson'$s$ disease (PD) is one of the most common neurodegenerative disorders. PD has been the fastest growth in prevalence, and it has become the leading cause of disability. The severity or progression of PD can be reduced if diagnosed at the early stages. It is therefore necessary to develop rapid and simple screening methods or tools to diagnose PD. Speech impairment is one of the early symptoms of PD which is commonly termed Parkinsonian hypokinetic dysarthria. Many researchers have developed a computerized method to identify of diagnosing PD based on voice features. However, the inaccuracy of the developed models was inconsistent especially when being tested on different datasets. The possible cause is the unwanted variability and biases between datasets. This study investigates the possible inconsistencies between Parkinson's voice datasets. The inconsistencies were investigated in the statistical distribution of voice parameters of the healthy-control (HC) group. This work observes the statistical distribution of sustained phoneme parameters extracted from the healthy-control (HC) group of five datasets using ANOVA and the Post-Hoc Turkey-Cramer test. The result suggests that the diversity in language and ethnicity were not contributing significantly to any biases between databases. The other result confirms that noises in the recording contribute to the biases in the extracted voice features, especially the harmonic features

**Speech Feature-Based Machine Learning Model and Smart Devices for Stroke Early Recognition,**

Dysarthria is one of the early symptoms of stroke, and the traditional diagnosis method relies on the doctors listening to the patient's pronunciation, which suffers from low accuracy. We constructed a multi-feature fusion feature engineering and combined it with machine learning methods for early recognition of stroke. According to the specificity of stroke pathological speech, different features will imply different information, which will lead to different recognition results. Therefore, we extracted features with different classes of information of the stroke pathological speech and combined them with different machine learning models for comparison experiments. Finally, the random forest model with a combination of prosodic features and spectral features achieved an accuracy of 94.33%, a precision of 91.89%, a recall of 99.14%, and an f1-score of 95.17%. The model has been deployed into an application running on smart devices. Stroke patients can use this for early recognition at home to save valuable time for early warning and intervention treatment.

**Phonetic Analysis of Dysarthric Speech Tempo and Applications to Robust Personalised Dysarthric Speech Recognition,**

Improving the accuracy of personalised speech recognition for speakers with dysarthria is a challenging research field. In this paper, we explore an approach that non-linearly modifies speech tempo to reduce mismatch between typical and atypical speech. Speech tempo analysis at the phonetic level is accomplished using a forced-alignment process from traditional GMM-HMM in automatic speech recognition (ASR). Estimated tempo adjustments are applied directly to the acoustic features rather than to the time-domain signals. Two approaches are considered: i) adjusting dysarthric speech towards typical speech for input into ASR systems trained with typical speech, and ii) adjusting typical speech towards dysarthric speech for data augmentation in personalised dysarthric ASR training. Experimental results show that the latter strategy with data augmentation is more effective, resulting in a nearly 7% absolute improvement in comparison to baseline speaker-dependent trained system evaluated using UASpeech corpus. Consistent recognition performance improvements are observed across speakers, with greatest benefit in cases of moderate and severe dysarthria.

**Comparison of Noise Reduction Techniques for Dysarthric Speech Recognition,**

The paper investigates the impact of denoising techniques on a deep learning recognition system for speak-ers with dysarthria, i.e., a neuromotor speech disorder which compromises speech intelligibility and that affects approximately 46 million of people worldwide. In particular, we compare a manual noise reduction techniques with automatic approaches based on classical signal processing techniques, i.e. filtering and spectral analysis, as well as more recent deep learning techniques based on recurrent neural network models. Comparison results reported in this paper are based on a dataset with more than 21K audio files collected with the collaboration of 156 Italian native speakers with different disabilities that cause dysarthria speech impairment. Therefore, different diseases and dysarthric severity levels have been taken into account. Moreover, differently from several other studies related to automatic recognition systems, audio files considered in our analysis have been collected in real environments, with a very limited supervision and simply using users' smartphones. Our analysis shows that, in this context, the effectiveness of automatic denoising tools is quite limited, particularly for dysarthric speakers with severe grades of disorder. However, comparisons with the proposed manual denoising intervention provide new and interesting insights which can be effectively and easily exploited with the aim of empowering actual automatic dysarthric speech recognition systems and that could drive future research in this field.

**Exploring the Role of Fricatives in Classifying Healthy Subjects and Patients with Amyotrophic Lateral Sclerosis and Parkinson’s Disease,**

Dysarthria due to Amyotrophic Lateral Sclerosis (ALS) and Parkinson’s Disease (PD) impairs sustained phoneme productions. Vowels and fricatives get affected differently owing to the differences in their production mechanisms. This paper examines three sustained voiceless fricatives - /s/, /sh/ and /f/, as compared to three sustained vowels - /a/, /i/ and /o/, for classifying patients with ALS/PD and Healthy Controls (HC). Fricatives are found to achieve higher classification accuracies than /a/ and /o/, though /i/ outperforms all. Patients seem to find it difficult to form constrictions while producing fricatives, or to proximally position the tongue and palate while uttering /i/, due to dysarthria. Unwanted voicing added to voiceless fricatives by the patient population further contributes towards the discrimination. Both source (related to vocal cord) and filter (related to vocal tract) cues of fricatives, on average, outperform those of vowels. Lastly, decision-level fusion of /i/-/s/-/sh/, with a pooled classifier for these three phonemes, achieves the highest mean ALS vs. HC classification accuracy of 83.35%, although in PD vs. HC case, fusion of multiple /i/ utterances performs the best with an accuracy of 80.03%.

**A Fine-Tuned Transformer Model for Dysarthric Speech with Spectrograms,**

This study examines the enhancement of communication for individuals with dysarthria speech disorder that impairs control of speech muscles through Automated Speech Recognition (ASR) technologies. While transformer and neural attention-based sequence-to-sequence models have achieved state-of-the-art results in typical speech-to-text applications, their adaptation for dysarthric speech recognition presents significant challenges. These include the unique articulatory variation and a lack of substantial training data characteristic of dysarthric speech. To bridge this gap, our research introduces a novel Dysarthric Speech Transformer, which incorporates a deep transformer architecture specifically designed to accommodate the complexities of dysarthric speech patterns. To overcome the hurdle of data scarcity, we employ a two-phase transfer learning strategy that leverage high-quality clear speech data, alongside innovative neural freezing configurations, to enhance model training and effectiveness. Our methodology is validated through rigorous training on the UASpeech(M12) and TORGO datasets, which are tailored for speaker-specific dysarthric ASR. The results indicate a substantial improvement in recognition accuracy, underscoring the efficacy of our specialized transformer model. Further, our study explores the synergistic potential of combining signal processing techniques with deep learning, suggesting that this approach could lead to breakthroughs in ASR system performance for individuals with severe dysarthria. This integration promises not only to refine the accuracy of dysarthric speech recognition, but also to extend the accessibility and usability of ASR technologies for those most in need.

**An ontology-based expert system to generate therapy plans for children with disabilities and communication disorders,**

Nowadays there are no precise estimates about the number of persons (especially children) living with disabilities and communication disorders in the world. This situation becomes more complex in developing countries, where the World Health Organization (WHO) claims that only between 5 and 15% of children and adults with disabilities have the opportunity to access to assistive technologies. In the same way, a Speech-Language Pathologist (SLP) has a work overload, and must carry out several activities related with monitoring patients, designing therapy plans, preparing reports, providing couseling services, among others. On those grounds, in this paper we present an expert system able to automatically infer general intervention guidelines for children with disabilities and communication disorders. The system relies on ontologies and implements a semantic web environment to provide several services related with information querying, reports generation, inference of intervention strategies, etc. In order to populate the ontology and validate the system, we have used the clinical information of 152 real cases of patients of Cuenca, Ecuador and 1,005 speech-language therapy information elements.

**Exploring Self-Supervised Pre-Trained ASR Models for Dysarthric and Elderly Speech Recognition,**

Automatic recognition of disordered and elderly speech remains a highly challenging task to date due to the difficulty in collecting such data in large quantities. This paper explores a series of approaches to integrate domain adapted Self-Supervised Learning (SSL) pre-trained models into TDNN and Conformer ASR systems for dysarthric and elderly speech recognition: a) input feature fusion between standard acoustic frontends and domain adapted wav2vec2.0 speech representations; b) frame-level joint decoding of TDNN systems separately trained using standard acoustic features alone and with additional wav2vec2.0 features; and c) multi-pass decoding involving the TDNN/Conformer system outputs to be rescored using domain adapted wav2vec2.0 models. In addition, domain adapted wav2vec2.0 representations are utilized in acoustic-to-articulatory (A2A) inversion to construct multi-modal dysarthric and elderly speech recognition systems. Experiments conducted on the UASpeech dysarthric and DementiaBank Pitt elderly speech corpora suggest TDNN and Conformer ASR systems integrated domain adapted wav2vec2.0 models consistently outperform the standalone wav2vec2.0 models by statistically significant WER reductions of 8.22% and 3.43% absolute (26.71% and 15.88% relative) on the two tasks respectively. The lowest published WERs of 22.56% (52.53% on very low intelligibility, 39.09% on unseen words) and 18.17% are obtained on the UASpeech test set of 16 dysarthric speakers, and the DementiaBank Pitt test set respectively.

**Classification-based screening of Parkinson’s disease patients through voice signal,**

In this paper a classification algorithm for Parkinson's Disease screening is proposed. Code executes the processing of specific voice signals recorded by healthy and ill subjects. In the direction of a future implementation and validation in a home telemonitoring system, the algorithm has been built with the objective to serve as a screening tool for the precocious directing of subjects with high risk of neurological diseases to instrumental exams. In fact, in several neurological disorders, such as Parkinson's disease, motor impairments of vocal apparatus arise earlier than postural and ambulatory symptoms. In a home telemonitoring system, in which hardware would consist in a voice recorder (that could be a simple smartphone) and a server for the web platform, data would be acquired and instantly stored on a platform for their processing through machine learning algorithms and to be viewed by specialists. For this purpose, a fully automatic process is needed. Therefore, in this work, audio-preprocessing and features computation are completely performed automatically, using Matlab. Final models have been trained in Matlab environments from Weka's libraries. The family of developed models are trained with different type of phonations, from simple vowels to complex sounds, for a wider and more efficient analysis of vocal apparatus motor impairments. Moreover, dataset was 612 observation large, that is significantly above the mean size of similar works using simple phonations only. For a deeper analysis, different groups of parameters have been tested and cepstral features have been found to be optimal for classification and made up the big part of final algorithm. Developed models are part of the K-Nearest Neighbor family, thus, available for implementation in web platform. Finally, obtained models have shown high accuracies on the whole dataset, reaching values comparable with the literature but with more stability (standard deviation less than 1%). These results have been confirmed in the last validation session in which models have been exported and validated with 25% of data, reaching a best performance with a true positive rate of 98% and a true negative rate of 87%.

**An educative environment based on ontologies and e-learning for training on design of speech-language therapy plans for children with disabilities and communication disorders,**

Nowadays several developing countries must to face several problems related with the lack of personnel, resources or adequate health-care structures to attend children with disabilities and communication disorders. Some studies of the World Health Organization (WHO) and other organizations point that in some countries of Africa exists one Speech-Language Pathologist (SLP) per each 2 to 4 million people, whereas in developed countries like United States or Australia the ratio is much higher: one SLP per each 2500 to 4700 people. For these reasons, in this paper we propose an intelligent educative environment aimed to support the training and development of practice skills of the future SLPs. Our proposal relies on ontologies and several e-learning modules and at the moment is able to support the learning process of designing therapy plans. Currently, the ontology that we have designed contains more than 1,005 instances and two tests to evaluate the articulation, and hearing skills on children.

**Parkinson’s Disease Detection based on Changes of Emotions during Speech,**

Parkinson’s disease (PD) is the neurodegenerative disease which affects 2-3 % of the population beyond 65 years of age in EU. When PD treatment is administered early, it is significantly more effective. Unfortunately, it is quite challenging to detect this disease at its early stage and when the symptoms can be recognized it is usually quite late. For this reason there is big motivation for development more accessible and accurate solutions for the detection of PD. One of the early symptoms is so-called hypomimia. This paper introduces an automatic method, which can objectively detect PD. The method is based on analysis of emotion changes during pronunciation defined speech exercises. We achieved balanced accuracy 69 % using XGBoost algorithm. As the exercise we proposed to use a Czech tongue twister - the difficult to pronounce sentence. The features can be explained and thus it can be used in clinical practice. We identified that the most valuable emotion for PD detection in this case is fear.

**A robotic assistant to support the development of communication skills of children with disabilities,**

Currently, the robotic assistants, as well as the intelligent ICT (Information and Communication Technologies) tools, play an important role in different stages of evaluation, therapy, or education of People With Communication Disorders (PWCD). However, not all of these tools can easily incorporate new functionalities, exercises, or activities for therapy sessions. Likewise, in developing countries, a Speech-Language Pathologist (SLP) is in charge to perform several activities related with design and planning of therapies, patient's monitoring, counseling services, among many others. On those grounds, in this paper we present a robotic assistant that has multiple costumes, is able to provide support during therapy, can manage the information, among others functionalities. This robotic assistant is a part of a complete intelligent-system for supporting the Speech-Language Therapy (SLT) process from a viewpoint of domain knowledge modelling. With the aim of analyzing the feasibility of our approach, we have carried out two tests, one to determine the robot's response on controlled environments, and another to analyze the children's response to the robot.

**An automated speech-language therapy tool with interactive virtual agent and peer-to-peer feedback,**

Widespread use of technology has brought about revolutionary changes in health domain. Treatment of speech and language disorders is one such field where computer-based techniques have the potential to provide readily available therapy at low cost by reducing the reliance on therapists and clinicians. Although significant research has been performed on computer-aided speech and language therapy, existing approaches do not focus on the nonverbal behavioral cues of a patient in a natural conversational scenario. In this paper, we propose an automated speech and language therapy tool with an intelligent and interactive virtual agent playing the role of a therapist. Besides giving feedback to the patient's response on various therapeutic tasks, the virtual therapist has the ability to carry out a conversation with the patient and give feedback on their nonverbal behavioral cues to help them improve their communicating ability. Additionally, our proposed system supports computerized peer-to-peer conversation in real-life scenarios which is also underrepresented in existing research concerning computer-assisted speech language therapy.

**Detection of voice impairment for parkinson's disease using machine learning tools,**

Parkinson's disease is a progressive nervous system disorder that affects movement, and patients need periodic monitoring which is difficult for them and costs a lot. In recent years, there has been much research about the link between Parkinson's disease (PD) and speech impairment in order to provide an early diagnosis of the disease and create a system for remote monitoring of patients as well. Many studies have used signal and speech processing techniques to convert acoustic signals into vectors of features which are then mapped into different machine learning algorithms. The results obtained in PD telemedicine studies have shown that the choice of feature extraction techniques and classification algorithms directly influence the accuracy and reliability of the proposed system. This article provides a system to assess the speech disorders in the context of PD using features extracted from three domains (time/frequency, cepstral, and wavelet domain) and machine learning tools. Our goal is to assess the ability of each individual to distinguish those with Parkinson's disease from healthy people. The results suggest that cepstral domain gives the most reliable parametrization comparable to time/frequency and wavelet domain with a high accuracy using Support Vector Machine classifier.

**Speech disabilities in adults and the suitable speech recognition software tools - a review,**

Speech impairment, though not a major obstacle, is still a problem for people who suffer from it, while they are making public presentations. This paper describes the different speech disabilities in adults and reviews the available software and other computer based tools that facilitate better communication for people with speech impairment. The motivation for this writing has been the fact that stuttering, one of the types of speech disability has affected about 1 percentage of the people worldwide. This fact was provided by the Stuttering Foundation of America, a Non-profit Organization, functioning since 1947. A solution to stuttering is expected to benefit a considerable population. Speech recognition software tools help people with disabilities use their computers and other hand held devices to satisfy their day-to-day needs which otherwise, require dedicated domestic help and also question the person's ability to be independent. ASR (Automatic Speech Recognition) systems are popular among the common people and people with motor disabilities, while using these techniques for the treatment of speech correction is a current research field and is of interest to SLPs/SLTs (Speech Language Pathologist / Speech Language Therapist). On-going research also includes development of ASR based software to facilitate comfortable oral communication with people suffering from speech dysfunctions, i.e., in the domain of AAC (Augmentative and Alternative Communication).

**Articulation acoustic kinematics in ALS speech,**

Patients affected by Amyotrophic Lateral Sclerosis (ALS) show specific dysarthric clues in speech. These marks could be used to detect early symptoms and monitor the evolution of the disease in time. Classically articulation marks have been mainly based on static premises. Articulation Kinematics from acoustic correlates may help in producing measurements based on the dynamic behavior of speech. Specifically, distribution functions from the absolute kinematic velocity estimated by a simplified articulation model can be used in establishing distances based on Information Theory concepts between running speech segments from patients and controls. As an example, a longitudinal case of ALS has been studied using this methodology. It shows that the performance of dynamic articulation quality correlates may be more sensitive and robust than static ones. Conclusions foresee the use of speech as a valuable monitoring methodology for ALS timely evolution.

**The Possibilities of Smart Clothing in Adult Speech Therapy: Speech Therapists' Visions for the Future,**

The potential of technology in healthcare has been closely explored in recent years. Increasingly more innovative technology-assisted rehabilitation methods for various customer groups are constantly being developed. However, the possibilities of smart clothing in adult speech rehabilitation have not been previously studied. The purpose of this study was to discover speech therapists' visions about the possibilities of smart clothing in adult rehabilitation. We organized an ideation workshop in December 2020 with four speech therapists who had worked in adult rehabilitation for at least five years. The workshop was held online on the Zoom platform. In the workshop we presented three questions for the speech therapists: 1) Which adult speech therapy clients could benefit from smart clothing? 2) What could smart clothing be used for in speech therapy rehabilitation for adults? and 3) How could smart clothing be used in speech therapy rehabilitation for adults? Qualitative data from this research was analyzed by thematic analysis. The main results of this study were that patients with dysphagia and patients with voice disorders were seen as the groups with the greatest potential use smart clothing, and continuous registration of various physiological functions of voice and swallowing were voted as the most usable applications of smart clothing. The most discussed topics were using smart clothing to monitor rehabilitation and using the clothing to activate and motivate the client by giving feedback. And finally, the easiest ways to control smart clothing were seen to be body movements, gestures, and touch.

**The Effect of Vibration Therapy on Spasticity of Children Suffering from Cerebral Palsy: A Systematic Review,**

Background: Rehabilitation is one of the primary concerns for children with Cerebral Palsy. Recently in non-pharmacological treatments, vibration therapy is increasingly used. Objective: To systematically identify and evaluate studies that applied vibration therapy on children having spasticity due to cerebral palsy. Methods: For the search process, we searched three databases Medline, Cochrane library, and Google scholar from Jan 2000 to May 2021. For the selection process, our inclusion criteria were the English language published articles that discussed I) vibration therapy II) Children III) Spasticity and IV) Patients of Cerebral palsy. Results: Results showed eight Quality studies and among those, only two showed an effect of the decrease in spasticity in the intervention group. Conclusion: Due to the limited quality studies we can say that there is no strong evidential base for vibration therapy action so this area requires more experiments with a healthy number of participants. Standardization of some parameters in protocol will be an important step towards the evidential research of the therapy. Moreover, the use of cognitive neuroscience techniques will help in making the therapy evidence.

**Efficient Recognition and Classification of Stuttered Word from Speech Signal using Deep Learning Technique,**

Fluency is a metric that assesses how well a speaker communicates with another person while presenting the information. Stuttering is one of the fluency problems that have a significant impact on speech recognition. The fluency of a speech is disrupted by involuntary word repetitions and prolongations, as well as external and internal noises. The objective of this study is to improve stuttered speech and create a better speech recognition system that decimates involuntary prolongations of sounds and repetitions of syllables or words. To get a good-quality speech signal, we propose a method in which a stuttered voice signal is analyzed using the classification algorithm called Convolutional Neural Network (CNN). For conversion of data into recognized speech, the approach is to save the input audio (speech signal of a person) with help of a microphone, then eradicate the external noises and stammers, extract features, and finally classify the speech data. The algorithm’s performance is compared using several filters such as Median Filter, Gaussian Filter, Gabor Filter, and Kalman Filter with the measures such as Mean Square Error (MSE), Signal to Noise ratio (SNR), Cross-correlation (CC), Mean Absolute Error (MAE), and Peak Signal to Noise ratio (PSNR). As per the experimental observations, the proposed scheme outperforms the established methods in terms of maintaining the overall speech signal intelligibility of the stuttered speech signal by identifying the stuttered word and removing the repetitions or prolongations. The Kalman filter performs better when compared to other used filters for analysis in terms of pre-processing level.

**Machine learning methods in assistive technologies,**

Assistive devices are essential in enhancing the quality of life for individuals who have severe disabilities, such as quadriplegia and amyotrophic lateral sclerosis, or who have had massive brainstem strokes. However, the effectiveness of these systems is dependent on preserved residual movements or speech. In the absence of means to repair the damaged nervous system, three options exist for restoring function: 1) augmenting the capabilities of remaining pathways, 2) detouring around points of damage, or 3) providing the brain with new channels for communication and control. The paper reviews the use of machine learning methods for development of assistive technology. Three projects are described, representing the three options listed above. In each of them machine learning methods are employed to help with pattern recognition and classification. The three projects are: automatic speech recognition of dysarthric speech; control strategies for FES-assisted locomotion (functional electrical stimulation); and an EEG-based computer access. Although these three projects may look very different from each other, the structure of their experimental set-ups, and their potential for application in assistive devices are very similar. All experimental set-ups consist of sensory signal acquisition, signal processing for feature extraction, and data processing by machine learning techniques for pattern recognition and classification. In addition, all three projects deal with digital signal processing and machine learning method applications in development of man-machine interfaces.

**Deep Transfer Learning Approach for Facial and Verbal Disease Detection,**

Millions of people have been subjected to different kind of acute diseases, some of them are eye diseases, facial skin diseases, tongue diseases and voice abnormalities. Most of eye diseases cause fully or partial blindness. Skin and tongue complications can be signs of cancers. Voice abnormalities can be cured at initial stages. Well-practiced medical practitioners have the ability of diagnose these diseases, but due to the pandemic situations and high consultation costs people do not tend to consult doctors. This research is predominantly focused on development of an application for automatic detection of eye, skin, tongue and verbal diseases using transfer learning (TL) based deep learning (DL) approach. Deep learning is a part of machine learning (ML) which has been used in most computer vision approaches. Transfer learning has been used to rebuild the existing convolutional neural network (CNN) models and used in disease detection. DenseNet121, MobileNetV2, RestNet152V2, models have been used to detect eye, skin and tongue diseases respectively and a new model has been used to detect voice abnormalities. CNN models are capable of automatically extracting features from the given images and voice data. All the trained models have been given accuracy rate of 80%-95%.

**Multi-objective Association Analysis of Parkinson Disease with Intelligent Optimization Algorithms,**

Parkinson's disease is a neurological disorder that has significant social and economic impacts affecting the patient's quality of life. Combined Parkinson's disease assessment scale is used together with clinical observations and evaluations in the diagnosis of the disease. However, this method may be insufficient especially at the beginning of the disease. Using data mining for analyzing of Parkinson's disease data can lead to the identification of similar patients, with the aim to assist the clinicians to respond more promptly and in a more personalized fashion to the changes of the patients' status. The majority of the studies within Parkinson disease are related to the classification task of data mining. However, association rules mining is one of the most common data mining problems used to find interesting and valuable associations that often occur in large data sets. There are very few studies on association analysis of Parkinson disease. To the best of our knowledge, there is not any study about multi-objective optimization for association analysis in real Parkinson data sets. In this study, multi-objective association analysis of Parkinson disease with MOPNAR and QAR\_CIP\_NSGAII that aims to automatically find the association rules with related intervals by optimizing many conflicting objectives such as support, confidence, lift, certainty factor, netconf, yulesQ, and coverage simultaneously.

**Towards Automatic Data Augmentation for Disordered Speech Recognition,**

Automatic recognition of disordered speech remains a highly challenging task to date due to data scarcity. This paper presents a reinforcement learning (RL) based on-the-fly data augmentation approach for training state-of-the-art PyChain TDNN and end-to-end Conformer ASR systems on such data. The handcrafted temporal and spectral mask operations in the standard SpecAugment method that are task and system dependent, together with additionally introduced minimum and maximum cut-offs of these masks, are now automatically learned using an RNN-based policy controller and tightly integrated with ASR system training. Experiments on the UASpeech corpus suggest the proposed RL-based data augmentation consistently produced performance superior or comparable to that obtained using expert or handcrafted SpecAugment policies. Our RL auto-augmented PyChain TDNN system produced an overall WER of 28.79% on the UASpeech test set of 16 dysarthric speakers.

**Towards Data-Driven Cognitive Rehabilitation for Speech Disorder in Hybrid Sensor Architecture,**

Recently, research in human-focused robot technologies has aimed to bring closer social-emotional intelligence, which interconnects with human beings' lifestyles doubtlessly. Among several major areas that can benefit from it, the healthcare sector consisting of many possible areas of therapeutic interposition is the most vital one. Recent developments in machine learning have broadened the scope of this work. The present research aims to achieve cognitive rehabilitation for speech disorders using a machine learning approach. Data collected from Electroencephalogram and Kinect sensor has been passed through Chebyshev filter for noise removal, and Autoencoder has been applied to extract features. Finally, Convolutional Neural Network-based transfer learning has shown an accuracy of 96% in classifying speech disorders. The performance of our proposed framework has been quite promising for applying it in real-time scenarios and achieving a better quality of life.

**An Evaluation of Neural Vocoder-Based Voice Cloning System for Dysphonia Speech Disorder,**

Dysphonia is a voice disorder affecting voice quality, quantity, and intensity, occurring at various ages and diverse backgrounds. Dysphonia impacts the difficulty of communication, thereby reducing the overall quality of life. Medical solutions have been proposed to improve the speech quality of individuals with dysphonia. However, these solutions are often limited by considerable expenses and time-consuming procedures. Therefore, alternative solutions are needed to enhance speech quality. The widespread development of technology in various domains can be proposed as an alternative solution. One is speech processing technology using text-to-speech (TTS) with voice cloning techniques. Our work presents the impact of the vocoder in a voice cloning system on the quality of synthesized speech for dysphonia speakers. We compare selected vocoder models based on architecture and performance. Furthermore, we explore the effect of using Speaker Conditionals on the vocoder. We perform an objective evaluation for each vocoder to measure the quality of the models.

**VoiceBank-2023: A Multi-Speaker Mandarin Speech Corpus for Constructing Personalized TTS Systems for the Speech Impaired,**

Services of personalized TTS systems for the Mandarin-speaking speech impaired are rarely mentioned. Taiwan started the VoiceBanking project in 2020, aiming to build a complete set of services to deliver personalized Mandarin TTS systems to amyotrophic lateral sclerosis patients. This paper reports the corpus design, corpus recording, data purging and correction for the corpus, and evaluations of the developed personalized TTS systems, for the VoiceBanking project. The developed corpus is named after the VoiceBank-2023 speech corpus because of its release year. The corpus contains 29.78 hours of utterances with prompts of short paragraphs and common phrases spoken by 111 native Mandarin speakers. The corpus is labeled with information about gender, degree of speech impairment, types of users, transcription, SNRs, and speaking rates. The VoiceBank-2023 is available by request for noncommercial use and welcomes all parties to join the VoiceBanking project to improve the services for the speech impaired.

**Benchmarking Automatic Speech Recognition Technology for Natural Language Samples of Children With and Without Developmental Delays,**

Natural language sampling (NLS) offers rich insights into real-world speech and language usage across diverse groups; yet, human transcription is time-consuming and costly. Automatic speech recognition (ASR) technology has the potential to revolutionize NLS research. However, its performance in clinical-research settings with young children and those with developmental delays remains unknown. This study evaluates the OpenAI Whisper ASR model on n=34 NLS sessions of toddlers with and without language delays. Manual comparison of ASR to human transcriptions of children with Down Syndrome (DS; n=19; 2-5 years old) and typically-developing children (TD; n=15; 2-3 years old) revealed ASR accurately captured 50% of words spoken by TD children but only 14% for those with DS. About 20% of words were missed in both groups, and 21% (TD) and 6% (DS) of words were replaced. ASR also struggled with developmentally informative sounds, such as non-speech vocalizations, missing almost 50% in the DS data and misinterpreting most of the rest. While ASR shows potential in reducing transcription time, its limitations underscore the need for human-in-the-loop clinical machine learning systems, especially for underrepresented groups.

**Voice Disorders Assessment of Chinese Pronunciation Parkinson's Disease Dataset,**

People with Parkinson's disease (PD) have significant voice disorders and can be diagnosed by sustained vowels. There are few studies of PD with sustained vowels in Chinese phonatory. This paper aims to study the differences among different vowels in Chinese. We collect a sustained vowel dataset of patients with PD in China with multiple vowel types, and use SVM and KNN algorithms to compare and analyze Chinese dataset and other language dataset. The experimental results show differences between Chinese and other languages in PD detection with the same classifiers and sustained vowels, and the best detected vowel in Chinese pronunciation is the vowel /u/. In this study, two main conclusions have been drawn. Firstly, different sustained vowels have different effects on the detection of PD in Chinese. Secondly, different language pronunciations of the same sustained vowel have varied effects on the detection of PD.

**Detection of Parkinson's Disease using Extreme Gradient Boosting,**

Parkinson's disease is a brain-related disease that is common in every person mainly persons above age 45 years. This disease causes numbness in muscles, swallowing problems, bending of the back, shivering in hands, smell dysfunction, speaking problem, Hearing problem, and many more. Parkinson's disease has to be diagnosed as early as possible since the clinical tests, which take hours to detect, may cost a loss of time and money. An automated model for detecting Parkinson's disease in a person with greater accuracy is proposed in this paper. While several models for detecting Parkinson's disease have been established, they are all less reliable and precise. Our model is created using the gradient boosted decision tree, which not only reliably predicts Parkinson's disease in a human, but also predicts it quickly. The feature set contains 22 parameters of the voice signal, which are given to the XGBoost classifier. The developed model predicts Parkinson's disease with 96.6% of accuracy, 95.6% of sensitivity, 100% of specificity, 100% of Precision, F-Score 97.7%.

**Automated Dynamic Stuttered Speech Recognition System and Conversion System using Mel Filter And Enhanced Logistic Regression Model,**

Confidence and communication are both negatively impacted by stuttering, a speech condition that affects more than 70 million people globally. To improve fluency and intelligibility, the Stuttered Speech Recognition (SSR) system uses cutting-edge technologies such as the Weighted Mel Frequency Cepstral Coefficient algorithm and Machine Learning (ML) models to detect and fix speech defects. It can transform text into speech for assured communication and includes noise reduction for clarity. This goes beyond the scope of conventional speech recognition and holds great promise for helping stuttering people lead more fulfilling lives and be more integrated into society at large. The characteristic features of stuttering speech are an abundance of initial repetitions and prolongations. As a result, the stuttering episodes can be categorized using sound analysis. Finding specific stuttering occurrences, such as prolongations and repeats, in stuttered speech using a feature extraction technique is described in this study. To determine how effectively it can identify stuttering prolongations and repetitions, we apply the famous Mel Frequency Cepstral Coefficient (MFCC) feature extraction. Stuttering is a disease of neurodevelopment characterized by a lack of fluency in conventional speech patterns. Prolonged or repeated sounds, syllables, sentences, or phrases are examples of disfluency. Stuttering detection is challenging and complex since it involves pathology, psychology, acoustics, and signal processing, all of which make it an intriguing interdisciplinary domain study challenge. A thorough analysis of stuttering and disfluency classification methods is presented based on acoustic data as well as statistical and machine learning and deep learning approaches. This research is helpful for scholars, academicians and medical experts to identify the limitations in traditional models and to design efficient solutions for stuttered speech classification.

**Detecting Speech Abnormalities With a Perceiver-Based Sequence Classifier that Leverages a Universal Speech Model,**

We propose a Perceiver-based sequence classifier to detect abnormalities in speech reflective of several neurological disorders. We combine this classifier with a Universal Speech Model (USM) that is trained on 12 million hours of diverse audio recordings. Our model compresses long sequences into a small set of class-specific latent representations and a factorized projection is used to predict different attributes of the disordered input speech. The benefit of our approach is that it allows us to model different regions of the input for different classes and is at the same time data efficient. We evaluated the proposed model extensively on a curated corpus from the Mayo Clinic. Our model outperforms standard transformer (80.9%) and perceiver (81.8%) models and achieves an average accuracy of 83.1%. With limited task-specific data, we find that pretraining is important and surprisingly pretraining with the un-related automatic speech recognition (ASR) task is also beneficial. Encodings from the middle layers provide a mix of both acoustic and phonetic information and achieve best prediction results compared to just using the final layer encodings (83.1% vs 79.6%). The results are promising and with further refinements may help clinicians detect speech abnormalities without needing access to highly specialized speech-language pathologists.

**Parkinson’s Disease Detection Using Machine Learning Algorithms,**

In order to come up with an efficient method for the discovery of Parkinson’s Disease (PD), various research papers were taken into consideration that are dependent on Parkinson’s Detection using various Machine Learning algorithm (MLA). Patients with Parkinson’s disease (PWP) have movement issues as well as communication impairments, making physical visits for medication and monitoring challenges. Early identification of Parkinson’s disease allows people to live a normal life. The global aging population has brought about an increased focus on the importance of early, remote, and accurate detection of Parkinson’s disease (PD). This paper utilized the various techniques of MLA and it was also observed that the suggested model is more effective and provides higher precision. Here, the research has been carried out on the biological voice data of 31 individuals among which 23 suffered from PD. The potential outcomes of the paper is based on the detection accuracy using six different classifiers of ML. The accuracy output includes Support Vector Machine (SVM) 85%, Decision Tree 85%, Logistic Regression 85%, Random Forest 85%, Gradient Boost 87% and XG Boost 85%. From the above findings we propmote the incorporation of machine learning methodologies into clinical decision-making processes to enhance the accuracy and systematization of Parkinson’s disease diagnosis.

**LSTM Siamese Network for Parkinson’s Disease Detection from Speech,**

Parkinson's disease (PD) is a neurodegenerative disorder that severely affects motor functions. Symptoms include dysarthria and this fact has been the basis for PD detection from speech in several works. Machine learning-based technologies have made significant strides in automatic speech recognition, but their use is fairly limited for clinical diagnosis due to the lack of large corpora containing parkinsonian speech. In this work, we propose a two-step strategy to use machine learning methods for PD detection. In the first step, we use Long Short-Term Memory (LSTM)-based siamese networks to learn feature representations that highlight the information related to speech articulation and prosody relevant for PD detection. Siamese networks are trained on data pairs employing a Spanish corpus containing 52 patients and 56 control subjects. In the second step, we train a classifier to make decisions about the presence or absence of PD employing the features provided by the LSTM networks. We achieve an EER of 1.9% in the detection by combining the scores of different text-dependent models. Preliminary experiments show the efficacy of the proposed method and prove the usefulness of LSTM for PD detection from speech.

**Automatic Dysarthric Speech Detection Exploiting Pairwise Distance-Based Convolutional Neural Networks,**

Automatic dysarthric speech detection can provide reliable and cost-effective computer-aided tools to assist the clinical diagnosis and management of dysarthria. In this paper we propose a novel automatic dysarthric speech detection approach based on analyses of pairwise distance matrices using convolutional neural networks (CNNs). We represent utterances through articulatory posteriors and consider pairs of phonetically-balanced representations, with one representation from a healthy speaker (i.e., the reference representation) and the other representation from the test speaker (i.e., test representation). Given such pairs of reference and test representations, features are first extracted using a feature extraction front-end, a frame-level distance matrix is computed, and the obtained distance matrix is considered as an image by a CNN-based binary classifier. The feature extraction, distance matrix computation, and CNN-based classifier are jointly optimized in an end-to-end framework. Experimental results on two databases of healthy and dysarthric speakers for different languages and pathologies show that the proposed approach yields a high dysarthric speech detection performance, outperforming other CNN-based baseline approaches.

**Selection of optimal parameters for automatic analysis of speech disorders in Parkinson's disease,**

Patients with Parkinson's disease (PD) usually suffer from hypokinetic dysarthria (HD), which involves impairment of phonation, articulation, prosody, and speech fluency. Our paper deals with parameters that can be used for the evaluation of motor aspects of speech and relevant methods of data acquisition and analysis. A review of specific parameters of HD and methods used for their evaluation may from the practical point of view contribute both to the diagnostic approaches to HD and to the development of suitable measures for assessment of its progression. The paper gives a description of the most frequently used parameters and their optimization to enable the best possible automatic classification of the various stages of Parkinson's disease.

**High-Intelligibility Speech Synthesis for Dysarthric Speakers with LPCNet-Based TTS and CycleVAE-Based VC,**

This paper presents a high-intelligibility speech synthesis method for persons with dysarthria caused by athetoid cerebral palsy. The muscular control of such speakers is unstable because of their athetoid symptoms, and their pronunciation is unclear, which makes it difficult for them to communicate. In this paper, we present a method for generating highly intelligible speech that preserves the individuality of dysarthric speakers by combining Transformer-TTS, CycleVAE-VC, and a LPCNet vocoder. Rather than repairing prosody from the dysarthric speech, this method transfers the dysarthric speaker’s individuality to the speech of a healthy person generated by TTS synthesis. This task is both important and challenging. From the results of our evaluation experiments, we confirmed that the proposed method can partially transfer the individuality of the target dysarthric speaker while maintaining the intelligibility of the source speech.

**Multilingual Analysis of Speech and Voice Disorders in Patients with Parkinson’s Disease,**

Parkinson’s disease (PD) is associated with several speech/voice disorders collectively referred to as hypokinetic dysarthria (HD). The main goal of this study is to identify acoustic features that support the diagnosis of PD while being independent of the language of a speaker. We recorded seven speech (e.g. monologue) and voice (e.g. sustained phonation) tasks in a cohort of 59 PD patients and 44 age- and gender-matched healthy controls (HC) speaking Czech or US English. A non-parametric test revealed that the best discrimination power has a measure quantifying the number of interword pauses per minute. In a consequent classification analysis, utilising logistic regression, we observed a drop in the classification accuracy from 72–73% to 67%, when moving from single-language modelling to the multilingual one. The results of this study suggest that especially the prosodic (pause-based) features could play a significant role in the automatic language-independent diagnosis of PD.

**A Review of Automated Intelligibility Assessment for Dysarthric Speakers,**

Automated dysarthria intelligibility assessment offers the opportunity to develop reliable, low-cost, and scalable tools, which help to solve current shortcomings of manual and subjective intelligibility assessments. This paper reviews the literature regarding automated intelligibility assessment, identifying the highest performing published models and concluding on promising avenues for further research. Our review shows that most of the existing work were able to achieve very high accuracies. However, we have found that most of these studies validated their models using speech samples of the same speakers used in training, making their results less generalizable. Furthermore, there is a lack of study on how well these models perform on speakers from different datasets or different microphone setups. This lack of generalizability has implications to the real-life application of these models. Future research directions could include the use of more robust methods of validation such as using unseen speakers, as well as incorporating speakers from different datasets. This would provide confidence that the models are generalized and therefore allow them to be used in real-world clinical practice.

**IDEA: An Italian Dysarthric Speech Database,**

This paper describes IDEA a database of Italian dysarthric speech produced by 45 speakers affected by 8 different pathologies. Neurologic diagnoses were collected from the subjects' medical records, while dysarthria assessment was conducted by a speech language pathologist and neurologist. The total number of records is 16794. The speech material consists of 211 isolated common words recorded by a single condenser microphone. The words that refer to an ambient assisted living scenario, have been selected to cover as widely as possible all Italian phonemes.The recordings, supervised by a speech pathologist, were recorded through the RECORDIA software that was developed specifically for this task. It allows multiple recording procedures depending on the patient severity and it includes an electronic record for storing patients' clinical data. All the recordings in IDEA are annotated with a TextGrid file which defines the boundaries of the speech within the wave file and other types of notes about the record.This paper also includes preliminary experiments on the recorded data to train an automatic speech recognition system from a baseline Kaldi recipe. We trained HMM and DNN models and the results shows 11.75% and 14.99% of WER respectively.

**Region Based Prediction and Score Combination for Automatic Intelligibility Assessment of Dysarthric Speech,**

The intelligibility assessment of dysarthric speech is essential for planning therapy. Time-frequency representations have been used for automatic intelligibility assessment of dysarthria. These representations have been derived from the utterance as a whole. As voiced and unvoiced components have different characteristics; in this work, we use different time-frequency representations for voiced and unvoiced segments and use them for intelligibility assessment with CNN classifier. Finally, we combine the scores obtained by the two systems to assign an intelligibility level. The combined system's performance is found to be superior to the systems that used both voiced and unvoiced components separately or together as one utterance. The utterances of the TORGO database are used in the intelligibility assessment. Automatic assessment of speech intelligibility reduces speech-language pathologists' time and effort in assisting diagnosis and treatment design.

**A Preliminary Study on Constructing Mandarin Personalized Speech Recognition Systems for the Speech Impaired,**

This study explores personalized Mandarin ASR for the speech impaired. Sentence and short phrase tasks selected from the VoiceBank-2023 corpus are examined for amyotrophic lat-eral sclerosis and esophageal-speech speakers. Unadapted speaker-independent ASR model is trained with the NER-Pro corpus to serve as a pretrained model for the follow-up speaker-independent and dependent models fine-tuned with the CDSD and VoiceBank-2023 corpora. Several fine-tuning strategies are investigated. Experimental results show that the adapted speaker-dependent models attain average character error rates (CERs) ranging from 6.8% (normal) to 11.9% (mod-erate dysarthria) and 5.9% (normal) to 12.9% (moderate) for sentence and short phrase tasks, respectively. Last, we have implemented the fine-tuning strategies on a large corpus of a hereditary spastic paraplegia patient to evaluate frequently used sentence tasks under open-set and close-set conditions, resulting in CERs of 10.8% and 3.8%, respectively.

**Dysarthric Speech Conformer: Adaptation for Sequence-to-Sequence Dysarthric Speech Recognition,**

Automatic Speech Recognition (ASR) holds immense potential to provide an effective interface for assistive technologies, but its performance remains unsatisfactory for people with speech impairments such as dysarthria. Existing ASR systems struggle to accurately recognize dysarthric speech due to the significant speaker variability in dysarthric speech and the scarcity of dysarthric datasets. In this study, we propose a two-phase adaptation pipeline based on the Conformer architecture that leverages typical speech to transfer to individualized ASR models for dysarthric speakers. ASR performance is evaluated for isolated words and continuous sentences, yielding an average Word Error Rate of 21.5% on the UASpeech dataset and 12.7% on the TORGO dataset. Selectively freezing decoder layers was more often successful than selectively freezing encoder layers, suggesting that optimal performance is achieved by focusing the adaptation on the acoustic information contained in the encoder.

**Robust Cross-Etiology and Speaker-Independent Dysarthric Speech Recognition,**

In this paper, we present a speaker-independent dysarthric speech recognition system, with a focus on evaluating the recently released Speech Accessibility Project (SAP-1005) dataset, which includes speech data from individuals with Parkinson’s disease (PD). Despite the growing body of research in dysarthric speech recognition, many existing systems are speaker-dependent and adaptive, limiting their generalizability across different speakers and etiologies. Our primary objective is to develop a robust speaker-independent model capable of accurately recognizing dysarthric speech, irrespective of the speaker. Additionally, as a secondary objective, we aim to test the cross-etiology performance of our model by evaluating it on the TORGO dataset, which contains speech samples from individuals with cerebral palsy (CP) and amyotrophic lateral sclerosis (ALS). By leveraging the Whisper model, our speaker-independent system achieved a CER of 6.99% and a WER of 10.71% on the SAP-1005 dataset. Further, in cross-etiology settings, we achieved a CER of 25.08% and a WER of 39.56% on the TORGO dataset. These results highlight the potential of our approach to generalize across unseen speakers and different etiologies of dysarthria.

**Dysarthric speaker identification with constrained training durations,**

Dysarthria is a neurological speech disorder that induces badly or no pronunciation of phonemes. In order to promote biometric identification of dysarthic speakers under constrained training scenario, we propose in this paper a recognition framework based on the score level fusion of two systems: The first is based on the classical Mel Frequency Cepstral Coefficients (MFCCs) while the second system uses Auditory Cues (ACs) which simulate the external, middle and inner parts of the ear. A simple energy based voice activity detector (VAD) is incorporated in both systems and its impact on performance is evaluated. The experimental investigations are accomplished using Nemours database and Torgo database and Gaussian Mixture Models (GMMs) for speaker modeling. The experimental results demonstrate the effectiveness of the energy based VAD, especially for the MFCC-based system. Moreover, the complementarity of the two features is manifested by a significant gain in identification performance of the fused system under different training durations. Interestingly, the proposed system surpasses the state of the art results and achieves 100% correct speaker identification under long duration training scenario.

**Dysarthric Speech Recognition: A Comparative Study,**

Automatic Speech Recognition (ASR) systems are designed to convert spoken words into written text. These systems have the potential to greatly benefit individuals with speech impairments like dysarthria, improving their overall quality of life. However, there are significant challenges that researchers must address in order to develop an effective ASR system for atypical speech, specifically dysarthric speech. Among these challenges are the limited availability of dysarthric speech data and the need for an accurate interpretation model. In this study, we explored the development of dysarthric ASR systems, starting from traditional approaches that were specifically designed to recognize dysarthric speech. We then delve into deep-learning-based ASR systems and modern ASR engines that utilize transformers. Additionally, we highlighted the notable dysarthric speech collections and identified the key factors that impede the performance of dysarthric ASR systems. The findings indicate that although researchers have achieved gradual improvements in performance, further research is required to achieve practical performance levels. Likewise, it is important to establish guidelines and standards to accurately measure and benchmark the performance of dysarthric ASR systems in order to ensure their usefulness and effectiveness in real-world applications.

**Electromagnetic articulograph sensor-to-sound unit mapping-based intelligibility assessment of dysarthric speech,**

Dysarthria is a traumatic neuromotor disorder that affects the physical production of speech. It reduces the function of primary articulators that are involved in speech. Recent research has presented that the use of articulatory data provide a better assessment towards identifying the speech intelligibility of dysarthric speakers over acoustic models that are modelled based on the biological perception of speech through Mel scale than directly involving the anatomy of speech production. Articulatory data of speech, include positional data of the articulators obtained from 12 Electromagnetic Articulograph (EMA) sensor channels with each sensor channel containing 6 parameters namely x, y, z, Φ, θ, and rms. While identifying the phone intelligibility deficit for the dysarthric speakers, articulatory data of all 12 sensor combinations each with 6 parameters may create turbulence in the intelligibility assessment. Hence, an appropriate mapping of the sound units with the sensor combinations is vital. The current work, focuses on mapping the phones with the EMA sensor channels based on their place of articulation. The parameters are also grouped based on the 3D-spherical co-ordinate distribution of the articulator as (x, y, θ) and (y, z, Φ). The mapping of the EMA sensor channels with the sound units are further validated through HMM-based acoustic only models and FDA scores. The phones are trained through their corresponding articulatory sensor combination information using a 5-class support vector machine and tested through a 10-fold cross validation technique for both the parameter groups. The 5 classes include mild, moderate, moderate-to-severe, severe and normal. Each phone can be classified to any of the 5 class in both the parameter groups based on their speech intelligibility. The results show that the EMA sensor combinations with (y, z, Φ) parameter group for each phone coincide well with the acoustic and FDA scores.

**Introducing microphone calibration techniques for reliable loudness estimation in the CFDA tests,**

This study reports on the implementation of low-cost microphone calibration procedures to ensure that volume levels of digital audio recordings are of similar intensity (loudness) to the actual speech utterances that they represent. In the context of this investigation, these utterances are produced by patients as they attempt the various articulation exercises of the Computerised Frenchay Dysarthria Assessment (CFDA) diagnostic test suite. In fact, these calibration procedures are designed to normalize the signal intensity of CFDA-recorded audio samples so that two audio recordings which produce the same volume readings will actually be of the same perceived loudness even though these recordings may have been made on different computers with different microphone and sound card configurations. It is demonstrated that the aforementioned calibration procedures are fit for purpose in the context of pure tones, but less so for speech-like noise which is more complex in its tonal qualities and composition. Nonetheless, the proposed calibration procedures have consistently produced a close similarity in signal energy between “live” utterances and their corresponding digital recordings. This consistent similarity in energy has been deemed by a panel of expert clinicians to be a substantial improvement over the previous CFDA calibration protocols.

**Selecting disorder-specific features for speech pathology fingerprinting,**

The general aim of this work is to learn a unique statistical signature for the state of a particular speech pathology. We pose this as a speaker identification problem for dysarthric individuals. To that end, we propose a novel algorithm for feature selection that aims to minimize the effects of speaker-specific features (e.g., fundamental frequency) and maximize the effects of pathology-specific features (e.g., vocal tract distortions and speech rhythm). We derive a cost function for optimizing feature selection that simultaneously trades off between these two competing criteria. Furthermore, we develop an efficient algorithm that optimizes this cost function and test the algorithm on a set of 34 dysarthric and 13 healthy speakers. Results show that the proposed method yields a set of features related to the speech disorder and not an individual's speaking style. When compared to other feature-selection algorithms, the proposed approach results in an improvement in a disorder fingerprinting task by selecting features that are specific to the disorder.

**Machine Learning-Assisted Diagnosis of Speech Disorders: A Review of Dysarthric Speech,**

Speech disorders are commonly associated with neurodegenerative impairments in the motor system of the brain. Challenges to speech and communication are often experienced by those with conditions including Amyotrophic Lateral Sclerosis (ALS), Cerebral Palsy (CP), and Parkinson's Disease (PD). This review focuses on the machine learning applications for diagnosing speech disorders, with a particular emphasis on dysarthric speech. ALS, in particular, presents difficulties in speech production, spinal functions, respiration, and swallowing, making early diagnosis crucial for improving the patient's quality of life. Sustained vowel phonation emerges as a promising technique for distinguishing among ALS, PD, CP, and individuals in good health. This comprehensive review explores the various machine learning approaches utilized for diagnosing ALS, PD, and CP through speech analysis, highlighting their potential impact on future research.

**Acoustic-to-Articulatory Inversion for Dysarthric Speech: are Pre-Trained Self-Supervised Representations Favorable?,**

Acoustic-to-articulatory inversion (AAI) involves mapping from the acoustic to the articulatory space. Signal-processing features like the MFCCs, have been widely used for the AAI task. For subjects with dysarthric speech, AAI is challenging because of an imprecise and indistinct pronunciation. In this work, we perform AAI for dysarthric speech using representations from pre-trained self-supervised learning (SSL) models. We demonstrate the impact of different pre-trained features on this challenging AAI task, at low-resource conditions. In addition, we also condition x-vectors to the extracted SSL features to train a BLSTM network. In the seen case, we experiment with three AAI training schemes (subject-specific, pooled, and fine-tuned). The results, consistent across training schemes, reveal that DeCoAR, in the fine-tuned scheme, achieves a relative improvement of the Pearson Correlation Coefficient (CC) by ∼1.81% and ∼4.56% for healthy controls and patients, respectively, over MFCCs. We observe similar average trends for different SSL features in the unseen case. Overall, SSL networks like wav2vec, APC, and DeCoAR, trained with feature reconstruction or future timestep prediction tasks, perform well in predicting dysarthric articulatory trajectories.

**Multilingual Analysis Of Intelligibility Classification Using English, Korean, And Tamil Dysarthric Speech Datasets,**

This paper analyzes dysarthric speech datasets from three languages with different prosodic systems: English, Korean, and Tamil. We inspect 39 acoustic measurements which reflect three speech dimensions including voice quality, pronunciation, and prosody. As multilingual analysis, examination on the mean values of acoustic measurements by intelligibility levels is conducted. Further, automatic intelligibility classification is performed to scrutinize the optimal feature set by languages. Analyses suggest pronunciation features, such as Percentage of Correct Consonants, Percentage of Correct Vowels, and Percentage of Correct Phonemes to be language-independent measurements. Voice quality and prosody features, however, generally present different aspects by languages. Experimental results additionally show that different speech dimension play a greater role for different languages: prosody for English, pronunciation for Korean, both prosody and pronunciation for Tamil. This paper contributes to speech pathology in that it differentiates between language-independent and language-dependent measurements in intelligibility classification for English, Korean, and Tamil dysarthric speech.

**3. Assessment and intelligibility modification for dysarthric speech,**

Dysarthria is a motor speech disorder that is often associated with irregular phonation and amplitude, incoordination and restricted movement of articulators. This condition is caused by cerebral palsy, degenerative neurological disease and so on. The pattern of speech impairment can be determined by the amount of compromise detected in the muscle groups. That is, the dysarthrias have global effect rather than focal effects on speech production systems of phonation, articulation and resonance. Clinically, assessment of dysarthria is carried out using perceptual judgment by experienced listeners. One of the limitations of perceptual assessment is that it can be difficult even for highly trained listeners to differentiate the multiple dimensions of dysarthric speech, as dysarthria has multisystem dysregulation. Although many researchers are involved in developing assistive devices, acoustic analyses are carried out on each of the subsystems independent of each other. As dysarthria affects the speech system globally, a multidimensional approach is required for the assessment and an associated intelligibility improvement system to develop an assistive device. This chapter will describe the significance and methods to develop a detection and assessment system by analyzing the problems related to laryngeal, velopharyngeal and articulatory subsystems for dysarthric speakers, using a speech recognition system and relevant signal-processing-based techniques. The observations from the assessment system are used to correct and resynthesize the dysarthric speech, conserving the speaker’s identity, thereby improving the intelligibility. The complete system can detect the multisystem dysregulation in dysarthria, correct the text and resynthesize the speech, thus improving the lifestyle of the dysarthric speaker by giving them the freedom to communicate easily with the society without any human assistance.

**Experimental Results of CNN and RNN Models for Identification of Parkinson’s Disease,**

There are many people with neurological disorders, and these disorders lead to dysarthria. Dysarthria, makes it more difficult for croakers to control the issue if not addressed promptly. The psyche and physiology of the patient will be affected by the symptoms. The majority of early studies on diagnosis dysarthria employed deep or machine learning methods to categorize these symptoms. In order to describe about dysarthria, this study suggests a CNN-LSTM model using convolutional neural networks and reopened intermittent units. According to experimental researches the CNN-LSTM model performs better in detecting dysarthria than the previous models. The study's suggested model has a delicacy of 98.38. Compared to past exploration models, this is more precise.

**Synthesis of New Words for Improved Dysarthric Speech Recognition on an Expanded Vocabulary,**

Dysarthria is a condition where people experience a reduction in speech intelligibility due to a neuromotor disorder. Previous works in dysarthric speech recognition have focused on accurate recognition of words encountered in training data. Due to the rarity of dysarthria in the general population, a relatively small amount of publicly-available training data exists for dysarthric speech. The number of unique words in these datasets is small, so ASR systems trained with existing dysarthric speech data are limited to recognition of those words. In this paper, we propose a data augmentation method using voice conversion that allows dysarthric ASR systems to accurately recognize words outside of the training set vocabulary. We demonstrate that a small amount of dysarthric speech data can be used to capture the relevant vocal characteristics of a speaker with dysarthria through a parallel voice conversion system. We show that it’s possible to synthesize utterances of new words that were never recorded by speakers with dysarthria, and that these synthesized utterances can be used to train a dysarthric ASR system.

**Verbal and visuo-spatial working memory in cerebral palsy patients,**

To investigate the differences in verbal and visuo-spatial memory abilities in cerebral palsy patients. Twenty-seven cerebral palsy patients with dysarthria, 29 cerebral palsy patients without dysarthria and 32 mental age-matched normally developed controls were compared using a Digit span test, a Verbal learning test, the Corsi Tapping test and a Figure memory test. We found that Verbal and visual working memory functioning is selectively impaired in CP patients. Dysarthria in CP patients may play an important role both in verbal and visuo-spatial working memory. The two components of working memory, PL and VSSP, work in synergy and especially in high point tasks.

**Source Domain Data Selection for Improved Transfer Learning Targeting Dysarthric Speech Recognition,**

This paper presents an improved transfer learning framework applied to robust personalised speech recognition models for speakers with dysarthria. As the baseline of transfer learning, a state-of-the-art CNN-TDNN-F ASR acoustic model trained solely on source domain data is adapted onto the target domain via neural network weight adaptation with the limited available data from target dysarthric speakers. Results show that linear weights in neural layers play the most important role for an improved modelling of dysarthric speech evaluated using UASpeech corpus, achieving averaged 11.6% and 7.6% relative recognition improvement in comparison to the conventional speaker-dependent training and data combination, respectively. To further improve the transferability towards target domain, we propose an utterance-based data selection of the source domain data based on the entropy of posterior probability, which is analysed to statistically obey a Gaussian distribution. Compared to a speaker-based data selection via dysarthria similarity measure, this allows for a more accurate selection of the potentially beneficial source domain data for either increasing the target domain training pool or constructing an intermediate domain for incremental transfer learning, resulting in a further absolute recognition performance improvement of nearly 2% added to transfer learning baseline for speakers with moderate to severe dysarthria.

**A Survey on ASR Systems for Dysarthric Speech,**

Recently Automatic Speech Recognition (ASR) has been widely overblown with many applications and assistance but orally challenged people, such as people with disordered speech, can’t get much benefits. Speech technologies are very useful on a daily basis to assist people with speech disorders. Dysarthria is a neurological speech disorder caused by significant injury in the left hemisphere of the brain. Dysarthric people have difficulty in the movement of speech-related muscles. As a result of strain on their speech muscles, individuals with dysarthria are able to generate limited speech data for analysis.In order to recognize speech of dysarthria sufferers, a robust technique is needed that can cope with extreme irregularity and narrow training data. This survey details a brief understanding of dysarthric speech characteristics and behavior. It also presents several attempts that have been made to make robust ASR systems for dysarthric speech.

**Multi-Task Transformer with Input Feature Reconstruction for Dysarthric Speech Recognition,**

Dysarthria is a motor speech disorder caused by damage to the part of the nervous system that controls the physical production of speech. It poses great challenges in building robust dysarthric speech recognition (DSR) due to the high inter- and intra-speaker variability. To this end, we propose a multi-task Transformer with input feature reconstruction as an auxiliary task, where the main task of DSR and the auxiliary reconstruction task share the same encoder network. The auxiliary task aims to reconstruct clear speech features from corrupted speech of healthy speakers (intra-domain) or dysarthric speakers (cross-domain). Further, to alleviate the imbalanced distribution of dysarthria data sets, we devise an adaptive rebalance sampling scheme to improve the utterance sampling frequency of dysarthric speech. Experimental results show that the proposed model considerably outperforms other baselines across speakers with varying severity of dysarthria.

**A Study on The Impact of Self-Supervised Learning on Automatic Dysarthric Speech Assessment,**

Automating dysarthria assessments offers the opportunity to develop practical, low-cost tools that address the current limitations of manual and subjective assessments. Nonetheless, the small size of most dysarthria datasets makes it challenging to develop automated assessment. Recent research showed that speech representations from models pre-trained on large unlabelled data can enhance Automatic Speech Recognition (ASR) performance for dysarthric speech. We are the first to evaluate the representations from pre-trained state-of-the-art Self-Supervised models across three downstream tasks on dysarthric speech: disease classification, word recognition and intelligibility classification, and under three noise scenarios on the UA-Speech dataset. We show that HuBERT is the most versatile feature extractor across dysarthria classification, word recognition, and intelligibility classification, achieving respectively +24.7%, +61%, and + 7.2% accuracy compared to classical acoustic features.

**Lip Contour Identification in Texture Data of 3D Face Mesh Sequences,**

Parkinson's disease (PD) is a degenerative neurological disease affecting motor, cognitive and autonomic function with symptoms including a resting tremor, rigidity and reduced overall movement. Sufferers often experience communication changes, such as reduced vocal loudness and imprecise articulation, as a result of motor speech disorder. The extent of motor speech disorder (dysarthria) and the efficacy of treatment is currently subjectively assessed by speech and language therapists using standardised tests such as the Frenchay Dysarthria Assessment which assesses speech characteristics and oro-motor control. The outcome of these subjective assessments can be adversely affected by external factors such as the experience of the therapist, therefore there is a recognised need for more objective assessment methods. 3D dynamic computer modelling can be used to capture the movements of the human face in real time to produce a sequence of 3D face meshes over time. This dynamic 3D modelling can be used to objectively assess the extent of facial mobility inParkinson's disease patients with dysarthria. An important step in measuring facial mobility from a sequence of 3D meshes is to identify, in each mesh of the sequence, the location of facial features such as the lip outline. We describe an approach for identifying the lip outline using the mesh data and texture data of a face mesh using integral projection models and an active contour model.

**2. Enhancement of continuous dysarthric speech,**

Dysarthria is a set of motor-speech disorders resulting due to neurological injuries. It affects the motor component of the motor-speech system. Disruption in muscular control makes the speech imperfect. As a result, dysarthric speech is not as comprehensible as normal speech. Most often, people with dysarthria have problems with communicating and this inhibits their social participation. Hence, for effective communication, it is extremely vital that we develop assistive speech technologies for people with dysarthria. The aim is to improve the naturalness and intelligibility of dysarthric speech while retaining the characteristics of the speaker. For this purpose, durational attributes across dysarthric and normal speech utterances are first studied. An automatic technique is developed to correct dysarthric speech to bring it closer to normal speech. The performance of this technique is compared with that of two other techniques available in the literature - a formant resynthesis technique and a hidden Markov model based adaptive speech synthesis technique. Subjective evaluations show a preference for dysarthric speech modified using the proposed approach over existing approaches.

**Exploring Appropriate Acoustic and Language Modelling Choices for Continuous Dysarthric Speech Recognition,**

There has been much recent interest in building continuous speech recognition systems for people with severe speech impairments, e.g., dysarthria. However, the datasets that are commonly used are typically designed for tasks other than ASR development, or they contain only isolated words. As such, they contain much overlap in the prompts read by the speakers. Previous ASR evaluations have often neglected this, using language models (LMs) trained on non-disjoint training and test data, potentially producing unrealistically optimistic results. In this paper, we investigate the impact of LM design using the widely used TORGO database. We combine state-of-the-art acoustic models with LMs trained with data originating from LibriSpeech. Using LMs with varying vocabulary size, we examine the trade-off between the out-of-vocabulary rate and recognition confusions for speakers with varying degrees of dysarthria. It is found that the optimal LM complexity is highly speaker dependent, highlighting the need to design speaker-dependent LMs alongside speaker-dependent acoustic models when considering atypical speech.

**Multi-Modal Acoustic-Articulatory Feature Fusion For Dysarthric Speech Recognition,**

Building automatic speech recognition (ASR) systems for speakers with dysarthria is a very challenging task. Although multi-modal ASR has received increasing attention recently, incorporating real articulatory data with acoustic features has not been widely explored in the dysarthric speech community. This paper investigates the effectiveness of multi-modal acoustic modelling for dysarthric speech recognition using acoustic features along with articulatory information. The proposed multi-stream architectures consist of convolutional, recurrent and fully-connected layers allowing for bespoke per-stream pre-processing, fusion at the optimal level of abstraction and post-processing. We study the optimal fusion level/scheme as well as training dynamics in terms of cross-entropy and WER using the popular TORGO dysarthric speech database. Experimental results show that fusing the acoustic and articulatory features at the empirically found optimal level of abstraction achieves a remarkable performance gain, leading to up to 4.6% absolute (9.6% relative) WER reduction for speakers with dysarthria.

**A Sequential Contrastive Learning Framework for Robust Dysarthric Speech Recognition,**

Dysarthria is a manifestation of disruption in the neuromuscular physiology resulting in uneven, slow, slurred, harsh, or quiet speech. Despite the remarkable progress of automatic speech recognition (ASR), it poses great challenges in developing stable ASR for dysarthric individuals due to the high intra- and inter-speaker variations and data deficiency. In this paper, we propose a contrastive learning framework for robust dysarthric speech recognition (DSR) by capturing the dysarthric speech variability. Several speech data augmentation strategies are explored to form two branches of the framework, meanwhile alleviating the scarcity of dysarthria data. We also develop an efficient projection head acting on a sequence of learned hidden representations for defining contrastive loss. Experiment results on DSR demonstrate that the model is better than or comparable to the supervised baseline.

**Intelligibility Improvement of Dysarthric Speech using MMSE DiscoGAN,**

Dysarthria is a manifestation of the disordering in articulatory parts that are used during speech production, which results in uneven, slow, slurred, monotone speech or speech in an abnormal rhythm. People with dysarthria produce less intelligible speech. Improving the intelligibility of dysarthric speech is challenging because unlike normal speech, there is less amount of data for dysarthric speech. It is a known fact that dysarthric speech and normal speech are different in speech production-perception perspectives. Recently, Generative Adversarial Network (GAN)-based architectures have become more popular to learn such kind of cross-domain relationships efficiently. In this paper, we propose to use Discover GAN (DiscoGAN) along with Mean Square Error (MSE) regularization (i.e., MMSE DiscoGAN) for Dysarthric-to-Normal speech conversion. In particular, a direct feature-based mapping technique is used to train all the models. In the end, we use the Automatic Speech Recognition (ASR) to measure the Phoneme Error Rate (PER) for a particular speaker. Proposed method is compared with baseline Deep Neural Network (DNN)-based system. Training of both the architectures and the evaluations were carried out on UA corpus. By analyzing the results, we observed that MMSE DiscoGAN outperforms DNN by 13.16% and 9.64% for male and female, respectively. Moreover, proposed GAN-based frameworks efficiently improve the intelligibility of dysarthric speech, and generate more naturalsounding speech compared to the DNN-based models.

**Optimization of dysarthric speech recognition,**

Explores the residual vocal ability of people who have severe motor impairments accompanied with severe dysarthria, and develops methods for improving the performance of automatic speech recognition (ASR) of dysarthric speech. The target applications for this technology are in the development of communication and control devices for these people. In our speech recognition system, we developed an adaptive word detection algorithm to detect words in highly irregular dysarthric speech. We also implemented perceptually-based mel frequency cepstrum coefficients (MFCC) for the parametric representation of the speech signal, and we adopted the left-to-right discrete hidden Markov model (DHMM) for speech pattern recognition. The system was tested with one person who has cerebral palsy and dysarthria, reducing the intelligibility of her speech to less than 15%. Our initial results on a word set consisting of ten digits demonstrated that recognition rates above 90% can be achieved if more than ten repetitions are used for training.

**HMM-based speech recognition system for the dysarthric speech evaluation of articulatory subsystem,**

Dysarthria is a neuromotor impairment of speech that affects one or more of the speech subsystems, but is often associated with irregular co-ordination of articulators and restricted movement of articulators among other problems. It is reflected in the acoustic characteristics of the phonemes as deviations from their healthy counterparts. To capture these deviations, in this work, isolated-style, phoneme recognition system is developed using monophone as the sub word unit. The performance of this phoneme recognition system for a dysarthric speaker can be directly related to the severity of the problem. To train the sub word unit models, speech data is collected from seven normal speakers. Time-aligned phonetic transcriptions are derived using forced Viterbi alignment procedure. Using this data, hidden Markov models for the required phonemes are trained. Nemours database contains time-aligned phonetic transcriptions for all the ten dysarthric speakers. Using these transcriptions, phonetic inventory is created for each of the dysarthric speakers separately. These phoneme segments are tested with the phoneme models trained using the normal speakers' data. The performance of this speech recognition system is analyzed after phoneme grouping, based on the place of articulation, for the assessment of the articulatory subsystem of the dysarthric speech. The analysis output correlates well with the Frenchey dysarthria assessment (FDA) scores provided with the database.

**Exploiting Audio-Visual Features with Pretrained AV-HuBERT for Multi-Modal Dysarthric Speech Reconstruction,**

Dysarthric speech reconstruction (DSR) aims to transform dysarthric speech into normal speech by improving the intelligibility and naturalness. This is a challenging task especially for patients with severe dysarthria and speaking in complex, noisy acoustic environments. To address these challenges, we propose a novel multi-modal framework to utilize visual information, e.g., lip movements, in DSR as extra clues for reconstructing the highly abnormal pronunciations. The multi-modal framework consists of: (i) a multi-modal encoder to extract robust phoneme embeddings from dysarthric speech with auxiliary visual features; (ii) a variance adaptor to infer the normal phoneme duration and pitch contour from the extracted phoneme embeddings; (iii) a speaker encoder to encode the speaker’s voice characteristics; and (iv) a mel-decoder to generate the reconstructed mel-spectrogram based on the extracted phoneme embeddings, prosodic features and speaker embeddings. Both objective and subjective evaluations conducted on the commonly used UASpeech corpus show that our proposed approach can achieve significant improvements over baseline systems in terms of speech intelligibility and naturalness, especially for the speakers with more severe symptoms. Compared with original dysarthric speech, the reconstructed speech achieves 42.1% absolute word error rate reduction for patients with more severe dysarthria levels. 1

**Real time detection of fricative landmarks to modify distortion in dysarthric speech using TMS320C6713 DSK,**

Speech is the most preferred way of communication. However, people with communication disorders find it difficult to communicate through speech. Dysarthria is one such disorder which involves poor articulation because of lack of control over articulatory muscles resulting in poor intelligibility. It is important to increase the intelligibility of dysarthric speech. One method is to identify the region of interest and modify the signal in that duration where accurate identification of the region is important. Fricatives are a class of consonants which are found to be distorted in dysarthria speech. This work aims at detection of fricative onset and offset to modify the distortions in dysarthric speech using TMS320C6713 DSK implemented on CC Studio 5.5. Among the 203 utterances tested, 71% of fricative onsets and 85% of fricative offsets were identified within ±0.3 s interval of the actual fricative onsets and offsets respectively.

**Dysarthric Speech Enhancement using Hybrid model involving Dynamic Time Warping and Feed Forward Neural Network,**

Any neurological impairment to the speech production system results in a range of motor speech disorders known as dysarthria. Speech quality is degraded in dysarthric speech, which is characterized by poor speech articulation. Therefore, it is crucial to enhance or correct dysarthric speech in order to help those who have the condition communicate more effectively. The purpose of this work is to enhance the continuous speech of many dysarthria sufferers. To enhance the intelligibility of speech, Dynamic Time Warping (DTW) with Artificial Neural Network (DTWANN) Speech Enhancement (SE) system has been used. Here, reconstructing the dysarthric signal from the speech signal has been focused on improving intelligibility of the speech. In order to process the dysarthric speech signal in the testing phase, the proposed SE system first trains an Artificial Neural Network (ANN) model. Pairs of normal speech and dysarthric speech utterances is used for training. The outcomes demonstrated that the suggested strategy improves PESQ score to 2.550 which shows significant improvement over PESQ of 1.002 before enhancement.

**Improving Acoustic Models for Dysarthric Speech Recognition using Time Delay Neural Networks,**

Recently, deep learning approaches have been widely used to solve problems in the pattern recognition area, especially speech recognition. The deep structures of neural networks have made the system gain impressive performance for the normal speaker speech acoustic model. However, there has remained a challenge to build a speech recognition model for dysarthric speakers. This paper investigates the performance of speech recognition models for dysarthric speakers using time delay deep neural networks. Moreover, we also explore the model performance by combining dysarthria and normal speech corpus. Finally, well-tuned hyperparameters of deep neural network structures give promising results on Mandarin and English dysarthria speech.

**Mora-Based Evaluation of Naturalness for Japanese Dysarthric Speech,**

This study aims to develop an automatic evaluation system for speech naturalness in dysarthric speech, focusing on temporal aspects less affected by noise. Speech samples from 19 young, 20 healthy older adults, and 21 dysarthria patients were analyzed for articulation rate, pause frequency, and mora length variability. Dysarthria patients exhibited slower articulation rates, more pauses, and greater mora length variability. These findings indicate that temporal features can be effectively utilized for the objective assessment of speech naturalness.

**Parkinson-Speech Analysis: Methods and Aims,**

Parkinson's Disease (PD) is one of the most frequent neurodegenerative diseases worldwide. Besides motor disorders, patients affected by this disease mostly suffer from a speech disorder named dysarthria. It is treated by different speech therapies. As a part of that, it is practice to observe the progress of the patients' dysarthria by classifying its severity by one of different possible scales like the National Technical Institute of the Deaf (NTID) scale. These ratings are currently done by auditive assessments, which are very time consuming and expensive. This work presents the conception of a tool for acoustical analysis of Parkinson speech and classification of the severity of the speech disorder, with the goal to improve the success of the speech therapies and simplify the classification.

**Efficacy of Current Dysarthric Speech Recognition Techniques,**

Speech Recognition system development is critical to supporting good communication. Furthermore, since speech and language technology progress on a daily basis, advancements in Dysarthric Speech Recognition systems are advantageous to those with dysarthria. In this work, we look in depth at the primary techniques employed in Dysarthric Speech Recognition. It starts with a quick description of dysarthria and how it affects speech patterns. It then analyses several ways to character extraction. Some of these methods are Mel Frequency Cepstral Coefficients, Linear Predictive Coding, Perceptual Linear Prediction, and Gammatone Frequency Cepstral Coefficients. Various ways for determining differentiating aspects of dysarthric speech have also been developed. These include techniques such as formant re-synthesis and acoustic space manipulation. The goal of these strategies is to improve speech recognition and intelligibility by enhancing various acoustic elements of speech. This study examines current advancements and problems in Dysarthric Speech Recognition in depth. Deep learning approaches and improvements to speaker and lexical models are being researched as strategies to increase DSR system efficiency and accuracy. Despite advancements, there are still a number of hurdles in the field of dysarthric speech recognition. Gathering data, coping with noise interference, building specialist models, and reacting to variances in speech are all examples of challenges faced in DSR model. This research also considers potential future directions for DSR, such as the utilisation of assistive technology and multimodal approaches. It also discusses advanced methodologies for feature extraction and user accessibility.

**Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease,**

Parkinson's disease (PD) is a neurodegenerative disorder of unknown etiology. PD patients suffer from hypokinetic dysarthria, which manifests on all aspects of voice production, respiration, phonation, articulation, nasality and prosody. To evaluate these disorders, clinicians have adopted perceptual methods, based on acoustic cues, to distinguish the different disease states. To develop the assessment of voice disorders for detecting patients with Parkinson's disease (PD), we have used a PD dataset of 34 sustained vowel / a /, from 34 people including 17 PD patients. We then extracted from 1 to 20 coefficients of the Mel Frequency Cepstral Coefficients from each person. To extract the voiceprint from each voice sample, we compressed the frames by calculating their average value. For classification, we used Leave-One-Subject-Out validation-scheme along with the Support Vector Machines with its different types of kernels. The best classification accuracy achieved was 91.17% using the first 12 coefficients of the MFCC by Linear kernels SVM.

**Intelligibility of modifications to dysarthric speech,**

Dysarthria is a motor speech impairment affecting millions of people. Dysarthric speech can be far less intelligible than that of non-dysarthric speakers, causing significant communication difficulties. The goal of our work is to understand the effect that certain modifications have on the intelligibility of dysarthric speech. These modifications are designed to identify aspects of the speech signal or signal processing that may be especially relevant to the effectiveness of a system that transforms dysarthric speech to improve its intelligibility. A result of this study is that dysarthric speech can, in the best case, be modified only at the short-term spectral level to improve intelligibility from 68% to 87%. A baseline transformation system using standard technology, however, does not show improvement in intelligibility. Prosody also has a significant (p<0.05) effect on intelligibility.

**Towards the improvement of automatic recognition of dysarthric speech,**

Dysarthria is a motor speech disorder that is often associated with irregular phonation (e.g. vocal fry) and amplitude, in coordination of articulators, and restricted movement of articulators, among other problems. The aim of this study is to raise dysarthic speech recognition rate through producing intelligibility enhanced speech using a procedure in which formants and energies are estimated from dysarthic speech and modified to more closely approximately desired normal targets. The modified parameters are taken to formant synthesizer to get final transformed speech, tested through perceptual tests to ensure quality and intelligibility. Then, we passed the modified dysarthric speech through an automatic speech recognition engine based on the HTK hidden Markov model toolkit. Speech recognition tests results indicate that the applied conversion algorithm raises the recognition rate of the dysarthric speech from 28% to 71.4%.

**Analysis of phonation in patients with Parkinson's disease using empirical mode decomposition,**

This paper deals with an acoustic analysis of hypokinetic dysarthria in patients with Parkinson's disease (PD). The analysis is based on parametrization of five basic Czech vowels using conventional features and parameters based on empirical mode decomposition (EMD). Experimental dataset consists of 84 PD patients with different disease progress and 49 healthy controls. From the single-vowel-analysis point of view we observed that sustained vowels pronounced with minimum intensity (not whispering) outperformed the other vowels' realization (including the most popular sustained vowel [a] pronounced with normal intensity). Then we employed a classification along with feature selection and again obtained the best results in the case of silent sustained vowels (accuracy ACC = 84 %, sensitivity SEN = 86% and specificity SPE = 82 %). Finally we considered classification of PD using different vowels' realization and reached accuracy = 94 %, sensitivity = 96% and specificity = 90 %. Features based on EMD significantly improved the results.

**Improved acoustic modeling for automatic dysarthric speech recognition,**

Dysarthria is a neuromuscular disorder, occurs due to improper coordination of speech musculature. In order to improve the quality of life of people with speech disorder, assistive technology using automatic speech recognition (ASR) systems are gaining importance. Since it is difficult for dysarthric speakers to provide sufficient data, data insufficiency is one of the major problems in building an efficient dysarthric ASR system. In this paper, we focus on handling this issue by pooling data from unimpaired speech database. Then feature space maximum likelihood linear regression (fMLLR) transformation is applied on pooled data and dysarthric data to normalize the effect of inter-speaker variability. The acoustic model built using the combined features (acoustically transformed dysarthric + pooled features) gives an relative improvement of 18.09% and 50.00% over baseline system for Nemours database and Universal Access speech (digit set) database.

**Deep Learning Based Prediction of Hypernasality for Clinical Applications,**

Hypernasality refers to the perception of excessive nasal resonance during the production of oral sounds. Existing methods for automatic assessment of hypernasality from speech are based on machine learning models trained on disordered speech databases rated by speech-language pathologists. However, the performance of such systems critically depends on the availability of hypernasal speech samples and the reliability of clinical ratings. In this paper, we propose a new approach that uses the speech samples from healthy controls to model the acoustic characteristics of nasalized speech. Using healthy speech samples, we develop a 4-class deep neural network classifier for the classification of nasal consonants, oral consonants, nasalized vowels, and oral vowels. We use the classifier to compute nasalization scores for clinical speech samples and show that the resulting scores correlate with clinical perception of hypernasality. The proposed approach is evaluated on the speech samples of speakers with dysarthria and cleft lip and palate speakers.

**Noise robust dysarthric speech classification using domain adaptation,**

This paper will investigate viability of a screening application that could be used to identify individuals with Dysarthria from among a larger population using sentence-level speech data. This task presents a number of challenged particularly if we aim to identify the disorder in the earlier stages before the more significant symptoms have begun to manifest themselves. A principal challenge in this task is acheiving robustness to the large number of confounding variables such as gender, age, accent, speaking style, and recording conditions. All of these variables will affect an individuals speech in a manner unrelated to the disorder, and identifying what information is relevant to the disorder amongst these confounding variables given the limited amount of data that is available in this regime presents a major engineering challenge. In this paper we will focus on achieving robustness to different types and levels of noise by employing a feature selection algorithm that attempts to minimize a non-parametric upper bound of the error in the noisy condition. This is a crucial problem to solve as the clean recording conditions used in data collection are typically a poor reflection of the type of data that will be encountered upon deployment.

**Development of Communication System from EMG of Suprahyoid Muscles Using Deep Learning,**

A certain number of people have difficulty communicating because of quadriparesis or dysarthria. Various human-machine interfaces (HMIs) have been proposed for such individuals, and silent speech interfaces using EMG have also been studied. For Japanese people, however, while the accuracy of vowel recognition is high, the accuracy of consonant recognition is low. Therefore, we are still in the development stage of practical application as an HMI. Therefore, instead of identifying consonants, we can expect to register words as commands and identify the registered words to communicate intentions. In recent years, deep neural networks (DNN) have made remarkable achievements in various fields. Convolutional Neural Net-work (CNN) has achieved high accuracy in many tasks. In addition, Long Short-Term Memory (LSTM), a type of Recurrent Neural Networks (RNNs), has also achieved high accuracy in estimating time series data such as speech, language, and video images. Therefore, we propose a classifier based on DNN that predicts intended words using surface electromyogram (sEMG) of the suprahyoid muscles. Experimental results showed CNN classifier is the best classifier.

**Disorder in the vocal tract and its detection from speech signal analysis,**

The results of analyzing the speech signals of two patients with dysarthria are presented. The magnitude of misarticulation is quantized, and the pitch period variability is calculated using a Fourier search filter. The latter is a fast Fourier transform algorithm which searches for a narrow range of frequencies in the spectrum. An attempt is made to identify the malfunctioning muscles/nerves in the vocal tract.<>

**Oriented IC Design Approach for Small-Volume, Performance-Demanded Edge-AI ASICs,**

The first-tier CMOS process for the design of integrated circuits (ICs) is often unavailable for small-volume edge-artificial intelligence (AI) application-specific integrated circuits (ASICs), highlighting demanded design goals in either speed, power, or leakage. The urgency in entering the market may force the designer to consider using the second-tier process with iterative designs to meet the demanded design goals. The overall paid prices are a long time to market, a long time to production, and a high development cost. This work proposes an oriented IC design approach for edge-AI ASICs to quickly meet uniquely demanded design goals and achieve a short time to market and production with low cost. We implement a test chip in 22nm CMOS for concept proving, where a deep neural network accelerator for dysarthria speech conversion is the vehicle to illustrate the proposed design approach in increasing the speed with a lower area cost and power consumption.

**AI and augmentative communication: where are we and where to go?,**

The author discusses the motivations behind the application of artificial intelligence (AI) techniques in augmentative communication, describes the representative research activities in augmentative communication, and explores the potential research areas which merge AI and augmentative communication. Current research on augmentative communication consists of three areas of concentration: clinical study, device assessment, and design methodologies. Clinical study focuses on identifying the types and levels of dysarthria that can be matched to the use of certain 'assistive' technologies, for example, use of speech synthesis and recognition to assist voice communication. The other two areas focus on the assessment of existing augmentative communication devices and design methodologies for new augmentative communication devices, respectively.<>

**Eye position feedback in a model of the vestibulo-ocular reflex for spino-cerebellar ataxia 6,**

The autosomal dominant spinocerebellar ataxias (SCAs) are a group of neurodegenerative diseases characterized by progressive instability of posture and gait, incoordination, ocular motor dysfunction, and dysarthria due to degeneration of cerebellar and brainstem neurons. Recent studies have established that there are more than 16 genetically distinct subtypes. Clinical observations suggest that eye movements and postural stability are universally but differentially impaired in the SCAs. The aim of the present work was to study the horizontal vestibulo-ocular reflex (VOR) in SCA6 patients to understand the pathophysiology of the VOR due to cerebellar Purkinje cell degeneration. The VOR was recorded in patients with genetically defined SCA6 during rotation in the dark. Severely affected subjects had an intact VOR, but there were quantitative differences in the gain and dynamics compared to normal controls. During angular velocity ramp rotations, there was a reversal in the direction of the VOR that was more pronounced in SCA6 compared to controls. Modeling studies indicate that abnormal feedback of an eye position signal into the velocity storage network can account for this reversal. These and other results will help to identify features that are diagnostic for SCA subtypes and provide new information about selective vulnerability of neurons controlling vestibular reflexes.

**Low Complexity Model with Single Dimensional Feature for Speech Based Classification of Amyotrophic Lateral Sclerosis Patients and Healthy Individuals,**

Lightweight automatic diagnostic tools for Amy-otrophic Lateral Sclerosis (ALS) and the associated dysarthria are essential for deployment in resource-limited platforms like mobile phones or general purpose computers. This study performs speech-based low-complexity classification of ALS and healthy subjects by cutting down (1) model complexity and (2) input feature dimensionality. Low complexity Dense Neural Network (DNN) models with 2 or less hidden layers are explored in comparison with the highly complex state-of-the-art Convolutional Neural Network (CNN) with Bidirectional Long Short-Term Memory (BiLSTM) architecture. On the other hand, various temporal statistics (standard deviation, autocorrelation at varying lags) obtained from the commonly used Mel-Frequency Cepstral Coefficients (MFCC) or its individual coefficients are investigated as the low dimensional features. Experiments with 72 ALS and 55 healthy subjects using Spontaneous Speech (SPON) and Diadochokinetic Rate (DIDK) tasks indicate the following. Model complexity reduction with DNN architectures gives comparable, or in some cases better performance, w.r.t. the CNN-BiLSTM model. DNN architectures, with lag 1 au-tocorrelation of MFCC (along with its delta and double delta coefficients) as the input feature vector for SPON task and standard deviation of the same for DIDK task, can respectively achieve 5.67% and 6.59% higher mean classification accuracies than the CNN-BiLSTM model with entire MFCC sequence as input while causing 99.99% reduction in the model parameter count. Moreover, using single dimensional standard deviation feature of the first delta coefficient for SPON and that of the second delta coefficient for DIDK, together with the DNN models, achieve 94.59% further reduction in the model parameter count while incurring only 1.76% and 5.17% further decrease, respectively, in the classification performance.

**Exploring Alternative Data Augmentation Methods in Dysarthric Automatic Speech Recognition,**

Patients with dysarthria face challenges in verbal communication, which affects their interaction with speech-activated devices. Intelligent systems that can interpret dysarthric speech could significantly enhance their quality of life. Neural Networks (NN) and Convolutional Neural Networks (CNN) have been used for sparse word classification in dysarthric speech, achieving an average accuracy of 64.1%. Spatial Convolutional Neural Networks (SCNN) and Multi-Head Attention Transformers (MHAT) have improved this accuracy by 20%. However, these methods have been tested on limited databases and yield specific results, making their application in more natural speech environments challenging. To address the lack of dysarthric speech data, some researchers have used speech synthesis techniques, but they require extensive training and careful database structuring. To alleviate this problem, we explore data augmentation methods based on the spectral characteristics of dysarthric speech, which focus on transforming spectrographic images extracted from available audio files. Two such methods were implemented and achieved results similar to complex state-of-the-art solutions.

**Assessing progress of Parkinson's disease using acoustic analysis of phonation,**

This paper deals with a complex acoustic analysis of phonation in patients with Parkinson's disease (PD) with a special focus on estimation of disease progress that is described by 7 different clinical scales (e. g. Unified Parkinson's disease rating scale or Beck depression inventory). The analysis is based on parametrization of 5 Czech vowels pronounced by 84 PD patients. Using classification and regression trees we estimated all clinical scores with maximal error lower or equal to 13 %. Best estimation was observed in the case of Mini-mental state examination (MAE = 0.77, estimation error 5.50 %). Finally, we proposed a binary classification based on random forests that is able to identify Parkinson's disease with sensitivity SEN = 92.86% (SPE = 85.71 %). The parametrization process was based on extraction of 107 speech features quantifying different clinical signs of hypokinetic dysarthria present in PD.

**Extracting cues from speech for predicting severity of Parkinson'S disease,**

Speech pathologists often describe voice quality in hypokinetic dysarthria or Parkinsonism as harsh or breathy, which has been largely attributed to incomplete closure of vocal folds. Exploiting its harmonic nature, we separate voiced portion of the speech to obtain an objective estimate of this quality. The utility of the proposed approach was evaluated on predicting 116 clinical ratings of Parkinson's disease on 82 subjects. Our results show that the information extracted from speech, elicited through 3 tasks, can predict the motor subscore (range 0 to 108) of the clinical measure, the Unified Parkinson's Disease Rating Scale, within a mean absolute error of 5.7 and a standard deviation of about 2.0. While still preliminary, our results are significant and demonstrate that the proposed computational approach has promising real-world applications such as in home-based assessment or in telemonitoring of Parkinson's disease.

**Super-gaussianity of Speech Spectral Coefficients as a Potential Biomarker for Dysarthric Speech Detection,**

Parkinson's disease (PD) and Amyotrophic Lateral Sclerosis (ALS) are progressive neurodegenerative diseases which, among other symptoms, cause dysarthria of speech. To assist the clinical diagnosis and treatment of neurological diseases, several studies have addressed the characterization and classification of healthy and dysarthric speech. However, most contributions deal with PD speech, with significantly fewer results presented for ALS speech. The objective of this paper is to show that ALS speech has a similar statistical distribution as PD speech, with the complex spectral coefficients being significantly less super-Gaussian than healthy speech spectral coefficients. In addition, a method to exploit the super-Gaussianity of speech signals as a feature to classify healthy and dysarthric speech is presented and evaluated. The proposed approach is evaluated on a French database of healthy and dysarthric (PD and ALS) speech. Experimental results show that the use of the super-Gaussianity of speech signals yields a significantly higher classification accuracy than state-of-the-art features such as fundamental frequency, jitter, shimmer, harmonics-to-noise ratio, or Mel frequency cepstral coefficients.

**Deep Learning of Articulatory-Based Representations and Applications for Improving Dysarthric Speech Recognition,**

Improving the accuracy of dysarthric speech recognition is a challenging research field due to the high inter- and intra-speaker variability in disordered speech. In this work, we propose to use estimated articulatory-based representations to augment the conventional acoustic features for better modeling of the dysarthric speech variability in automatic speech recognition. To obtain the articulatory information, long short-time memory recurrent neural networks are employed to learn the acoustic-to-articulatory inverse mapping based on a simulated articulatory database. Experimental results show that the estimated articulatory features can provide consistent improvement for dysarthric speech recognition with more improvement observed for speakers with moderate and moderate-severe dysarthria.

**Effect of Noise and Model Complexity on Detection of Amyotrophic Lateral Sclerosis and Parkinson’s Disease Using Pitch and MFCC,**

Dysarthria due to Amyotrophic Lateral Sclerosis (ALS) and Parkinson’s disease (PD) impacts both articulation and prosody in an individual’s speech. Complex deep neural networks exploit these cues for detection of ALS and PD. These are typically done using recordings in laboratory condition. This study aims to examine the robustness of these cues against background noise and model complexity, which has not been investigated before. We perform classification experiments with pitch and Mel-frequency cepstral coefficients (MFCC) using models of three different complexities and additive white Gaussian noise in four signal-to-noise-ratio (SNR) conditions. The findings are as follows: 1) In clean condition, pitch performs similar to MFCC across most model complexities considered, suggesting that one-dimensional pitch pattern provides discriminative cues for the classification to an extent equal to that of multi-dimensional MFCC, 2) Similar trend is observed in noisy cases when classifiers are trained and tested in matched noise and SNR conditions, 3) When the classifiers trained on clean data are applied in noisy cases, pitch based average classification accuracies are found to be 20.09% and 24.73% higher than those using MFCC for ALS vs. healthy and PD vs. healthy, respectively, suggesting robustness of pitch based classifier against noise and model complexity.

**Waveform Analysis and Feature Extraction from Speech Data of Dysarthric Persons,**

Speech recognition systems provide a natural way of interacting with computers and serve as an alternative to the more popular but less intuitive peripherals (input / output devices). Tools employing the techniques of Automatic Speech Recognition (ASR) can be extended to serve people with speech disabilities so that they can overcome the difficulties faced in their interaction with general public. An attempt is made here to achieve this goal by mapping the distorted speech signals of people with severe levels of dysarthria to that of a normal speech and/or less severe dysarthric speech. The analysis is carried out by comparing the speech waveforms of the people with and without communication disorders and then extracting the features from the audio files. The differences in time, duration, frequency and PSD are used to facilitate the mapping of unintelligible speech data to intelligible ones. When reasonable accuracy levels are achieved in this mapping, the normal voice can be used as the substitute / surrogate of the original distorted voice.

**Speaker Identity Preservation in Dysarthric Speech Reconstruction by Adversarial Speaker Adaptation,**

Dysarthric speech reconstruction (DSR), which aims to improve the quality of dysarthric speech, remains a challenge, not only because we need to restore the speech to be normal, but also must preserve the speaker’s identity. The speaker representation extracted by the speaker encoder (SE) optimized for speaker verification has been explored to control the speaker identity. However, the SE may not be able to fully capture the characteristics of dysarthric speakers that are previously unseen. To address this research problem, we propose a novel multi-task learning strategy, i.e., adversarial speaker adaptation (ASA). The primary task of ASA fine-tunes the SE with the speech of the target dysarthric speaker to effectively capture identityrelated information, and the secondary task applies adversarial training to avoid the incorporation of abnormal speaking patterns into the reconstructed speech, by regularizing the distribution of reconstructed speech to be close to that of reference speech with high quality. Experiments show that the proposed approach can achieve enhanced speaker similarity and comparable speech naturalness with a strong baseline approach. Compared with dysarthric speech, the reconstructed speech achieves 22.3% and 31.5% absolute word error rate reduction for speakers with moderate and moderate-severe dysarthria respectively. Our demo page is released here1.

**A solution to overcome speech disorder of patients using Brain Neuron EEG Signals,**

Speech disorders are neurodevelopmental disorders such as Stuttering, Dysarthria, Dysphonia and Aphasia associated with left inferior frontal structural anomalies that involve repeating or prolonging a word, syllable or phrase, or stopping during speech and making no sound for certain syllables. Most of the people who are suffering from speech disorders encounter difficulties in professional communication. Since people are busy with their day to day life, it is not practical to spend more time in consulting a doctor or do speech therapies for their medical issues. The speech therapist generally charges a significantly much higher rate for a single speech therapy practice, which the patient needs to practice at least twice or more for a week to get a better result. In an economy like Sri Lanka, people with average income cannot afford such an amount of money. Therefore, an innovative desktop application for speech disorder patients to overcome this problem has arisen. The main aim of this application is to reduce the speech imperative percentage of speech disorder patients via capturing the electroencephalogram feed of speech motor (Broca's area) using brain neuron O1, O2, C3, C4, F3, F4, F7, F8 electrodes and analyzing it to identify speech imperative issues. This system identifies the current impact on the left hemisphere of the brain (Broca's area) using EEG neurofeedback. Using speech voice analysis, the system provides the user to measure the articulation interference of the speech process. Self-Learning video tutorials are available for the clinical practices and treatments are available as prolong, relaxing, and humming exercises. Patients can track down the improvements daily or monthly by the rating system which makes the system unique among all other systems and the result can be directly sent to the desired consultant/neurophysiologist by the system itself. Patients can save time and the total cost of a therapy fee by using this system.

**Using StarGANv2 Voice Conversion to Enhance the Quality of Dysarthric Speech,**

In this paper, we propose to use StarGAN, a pow-erful Generative Adversarial Network (GAN), to improve the quality of dysarthric speech. Through extensive experiments, we demonstrate the effectiveness of StarGANv2-VC in converting dysarthric speech and significantly improving its intelligibility and naturalness. In addition, this research contributes to the field by conducting a comparative study between StarGANv2-VC and MaskCycleGAN-VC, another well-established GAN architecture, recently used in dysarthric speech conversion tasks. The results show that StarGANv2-VC performs the best, making it a promising solution for improving the speech quality of people suffering from dysarthria.

**Detecting Parkinson's Disease with Prediction: A Novel SVM Approach,**

In recent years, the prevalence of neurological diseases has been on the rise due to an increasing population and aging. Conditions like Parkinson's disease, stroke, cerebral palsy, and others often lead to dysarthria among patients. Detecting and treating Parkinson's disease promptly is crucial as it can pose challenges to managing the disease's progression and have adverse effects on a patient's mental and physical well-being when symptoms worsen. Previous research has been predominantly using the concept of machine learning for Parkinson's exposure. This study introduces a novel integrated Support Vector Machine archetypal for detecting the disease by gathering data, organizing it followed by training and testing, finally assessing its performance thereby decreasing the causalities. The experimental findings indicate the use of SVM model proposed in the research achieves an accurateness rate of 87 percentages and f1-score of 0.87%.

**Trusted Detection for Parkinson's Disease Based on Multi-Type Speech Fusion,**

Most patients with Parkinson's disease (PD) suffer from varying degrees of dysarthria. Therefore, speech can be exploited as an effective source of diagnostic information for PD. Different types of speech tasks are designed to evaluate subjects' verbal ability. Currently, machine learning methods for the detection of PD mostly concentrate on a single type of speech data. To make full use of the information from multiple sources, multimodal learning has been proposed and developed rapidly in recent years. However, most multimodal frameworks fall short of the reliability requirements of medical diagnosis. To solve this problem, a trustworthy model based on multi-type acoustic materials is proposed in this paper. The framework consists of three major components, i.e., pseudo-type generation, decision-making, and opinion combination. The objective of this endeavor is to offer accurate and reliable PD detection, aiding in the diagnostic process. Experimental results demonstrate the advantage of the proposed model over state-of-the-art fusion methods and highlight the necessity of each component in the proposed framework.

**Static and Dynamic Source and Filter Cues for Classification of Amyotrophic Lateral Sclerosis Patients and Healthy Subjects,**

Dysarthria due to Amyotrophic Lateral Sclerosis (ALS) affects speech production. Even the elementary sustained vowel utterances get impaired. For these, the impairments can be in achieving vowel-specific articulatory configurations, reflected in static acoustic cues, and/or in sustaining a configuration for a prolonged duration, reflected in dynamic cues. Such cues can further be attributed to the vocal cord (source) and vocal tract (filter) involved in speech production. This paper analyzes the relative contributions of these static (captured through average spectral characteristics) and dynamic (captured through spectral variations over time) source and filter cues toward automatic classification of ALS patients and healthy subjects using sustained utterances of /a/, /i/, /o/ and /u/. Experiments with 80 ALS patients and 80 healthy subjects suggest that the source cues (static/dynamic) are not the primary discriminators. For /i/, the static filter cues achieve the highest mean classification accuracy of 76.66%, whereas, for /a/, /o/ and /u/, the dynamic filter attributes contribute the most attaining average accuracies of 66.29%, 73.03% and 70.27%, respectively. Hence, ALS patients seem to face difficulties in forming the front closed vocal tract structure of /i/, whereas, holding the target vocal tract shape for long appears to be the primary challenge in case of /a/, /o/ and /u/.

**The Sophisticated Prognostication of Migraine Aura Using Machine Learning,**

Migraine is one of the most disabling diseases in the world and impacting more than one billion individuals. The symptoms such as intensity, Nausea, Vomit, Phonophobia, Photophobia, Visual, Dysphasia, Dysarthria, Vertigo, Sensory, intense to sound, occurs before the migraine. Migraine drains the quality of persons' life. Furthermore, the study of this research was to train machine learning model on migraine aura dataset using different modern approaches that could assist medical patients before occurring migraine and could also give the indications related to migraine. Likewise, by applying random forest algorithm, we got 99.5 percent accuracy using google colab which will be enough to deploy on application for future Project. Moreover, we leveraged traditional framework to complete the research incorporating, Data collection (Online free Kaggle), Preprocessing, Classification, Model training and Comparison of models. As a result, our peak algorithm precision was including, Naive Bayes 94.52%, decision tree 98.3%, K-nearest neighbours 98.6%, random forest 99.5%, Similarly, Rapid miner and Google Colab software are used for the comparison of algorithms and best one is chosen.

**Speaker recognition system of flexible throat microphone using contrastive learning,**

Recently, Flexible pressure sensor-based Throat Microphones (FTM) have attracted more attention in noise-robust speaker recognition and are promising for helping people with specific dysarthria to complete speaker recognition. FTM has outstanding flexibility compared with Hard Throat Microphones (HTM) and noise-robustness compared with Close-talk microphones (CM). However, speaker recognition for FTM is still an open task awaiting exploration since FTM has degradation problems and a lack of data sets. To tackle these two obstacles, referring to feature mapping methods for HTM, we introduce an FTM-oriented supervised contrastive learning (FTMSCL) method. An FTM speech data set is collected, then a contrastive loss function is designed to avoid the feature mapping methods' problems and effectively leverage label information from this data set. Furthermore, a critical parameter margin in this loss and several data augmentations for FTM are investigated. Experimental results show that, with no need for CM data, FTMSCL can achieve a False Acceptance Rate (FAR) of 2.97% and a False Rejection Rate (FRR) of 2.83%, which outperforms a conventional End-to-End one and an advanced feature mapping one significantly. Moreover, the best FAR and FRR of our FTMSCL method are only 0.86% and 0.83% higher than the best one using clean CM data.

**Design and Development of Computational Methodologies for Predicting Parkinson's Disease with Artificial Intelligence,**

Parkinson's disease is a degenerative ailment that affects the central nervous system, hinders movement, and produces tremors and rigidity. Its four stages have an impact on more than 1 million people in India each year. For this persistent illness, there is no known cure. It can only be treated if it is discovered early. Scientists have long theorized that voice impairment is one of the initial signs of Parkinson's disease. Vocal changes in Parkinson's disease patients could be recognized, enabling earlier intervention before physically incapacitating symptoms manifest. With the use of a number of speeches signaling techniques, the proposed work intends to develop a model that examines a person's voice and forecasts their likelihood of developing an illness. Parkinson's disease is a brain disorder where a part of your brain gradually deteriorates, leading to ever-worse symptoms. Numerous symptoms, including stiffness, bradykinesia, difficulties with coordination, and voice irregularities, are brought on by this neurodegeneration. Dysarthria is another symptom of Parkinson's disease. The primary flaws of PD speaking are decrease in intensity, monotony of pitch and loudness, inappropriate silences, and a breathy voice. Clinical signs, such as the description of a variety of movement symptoms and medical observations, are frequently used for treatment of Parkinson's disease (PD). On the other hand, conventional diagnostic techniques may be subjectivity-prone since they rely on the interpretation of motions that may be challenging to identify because they are occasionally imperceptible to human sight.

**Review on Dysarthric Speech Severity Level Classification Frameworks,**

Dysarthria is a motor speech disorder that results from malfunctions in the mechanisms and systems which regulate the movements required for speech production. It is challenging to organise, programme, manage, coordinate, and carry out speech productions due to these neurological limitations. Assessing the degree of dysarthria severity can provide information about the patient’s progress, assist pathologists in planning a treatment plan, and make automated voice recognition software for dysarthria more useful. This paper provides a thorough analysis of fresh and cutting-edge machine learning and deep learning methods and algorithms for dysarthric speech severity level classification. Depending on the degree of the impairment, various acoustic features and feature selection techniques are considered for the precise classification of dysarthric speech. This study is focused on various deep learning methodologies as well as the comparison of outcomes with various machine learning algorithms. An overview of the data sets and evaluation metrics frequently utilised in the classification of dysarthric severity levels is also included in this review. Finally, an audio-video cross model approach is reviewed for the comparison of severity level classification based on different features using different classifiers.