2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)

Significance of Entropy Based Features For Dysarthric Severity Level Classification

Meghana Avula[1], Aditya Pusuluri[2]and Hemant A. Patil[2]
[1] Koneru Lakshmaiah Education Foundation, Hyderabad, Telangana
E-mail: 2210040091@klh.edu.in
[2] Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, Gujarat
E-mail: aditya_pss@daiict.ac.in, hemant_patil@daiict.ac.in

2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) | 979-8-3503-6733-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/APSIPAASC63619.2025.10848870

***Abstract*—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.**

Keywords: Dysarthria, Entropy, Gabor Filterbank

I. Introduction

Dysarthria is a speech disorder caused by impairments of neurons that affects the muscles necessary for speech production. These impairments can be caused by many conditions, such as stroke, traumatic brain injury, Parkinson's disease [1]. Dysarthria is manifested by a variety of speech abnormalities, such as slurred speech, decreased intelligibility, changes in speech rate, and changes in speech quality [2]. Given the diverse etiology and presentation of dysarthria, accurately classifying the severity of the disorder is essential for developing tailored treatment plans, monitoring patient progress, and optimizing clinical resource allocation. Traditional methods of assessing dysarthria severity often rely on subjective evaluations by clinicians , which can lead to variability and inconsistency. These assessments depend on the clinician's experience and judgment, which can cause differences in diagnosis and treatment outcomes. This demands the need for automatic speech recognizers (ASR), and dysarthria severity level classification can improve the performance of an ASR.

Earlier approaches focused on employing acoustic characteristics such as features derived from fundamental or pitch frequencies, formant frequencies and cepstral based techniques like Mel Frequency Cepstral Coefficients (MFCC), Linear Frequency Cepstral Coefficients (LFCC) for speech pathology classification task. The ability of fundamental frequency and formant frequency for dysarthria severity classification is showcased in [3]. Furthermore, in [4], [5], and [6] the cepstral based features such as MFCC, LFCC are explored which contains the ability to capture global spectral characteristics.

Entropy measures the randomness or complexity of a signal, making it useful for identifying the detailed variations in speech patterns seen in different severity levels of dysarthria. By looking at the entropy of speech signals, we can better understand the severity of the disorder as the entropy metric provides information related to amount of randomness present in the signal which can help us learn about the controllability of the voice production system for a dysarthric speaker and provide a more accurate classification. Ultimately, this approach aims to provide reliable and effective tools for assessing and managing dysarthria, improving the quality of life for individuals affected by the disorder. By leveraging the power of entropy-based features, we can achieve more accurate and consistent severity classification, leading to better treatment outcomes and enhanced communication abilities for those with dysarthria. Previously, entropy based features are used for various applications such as speech recognition [7], emotion recognition [8], speech/music segmentation [9]. Furthermore, [10] showcased the importance of entropy based features for detection of hypernasal speech emphasizing the usage of entropy based features for dysarthric speech. To the best of our knowledge, no prior research has explored the ability of multi band entropy for dysarthria severity classification.

The rest of the paper is organized as follows: Section 2 represents a brief technical details about the proposed zero mean entropy using gabor filter bank, Section 3 gives the details about the database used, classifiers used, and the baseline features considered for this work. Section 4 consists of the motivation for the application of entropy features for dysarthric severity-level classification task. Section 5 contains the experimental results and the work is concluded with results and conclusion along with potential future research direction.

II. Proposed Approach

A. Shannon Entropy

Entropy is used to separate the voiced and unvoiced regions of speech signal. The entropy is low for voiced region and high for unvoiced region. Since speech is a non stationary signal, the voiced and invoiced regions vary across each frame. This motivated us to use the multiband entropy for each frame as feature set instead of calculating the entropy across the entire signal. Entropy measures, such as Shannon entropy, quantify the unpredictability or randomness in the distribution of signal amplitudes over time [11]. In speech analysis, entropy reflects the variability and complexity of vocal characteristics, including pitch, intensity, and timing. Higher entropy values indicate greater disorder or irregularity in speech production, which can be indicative of dysarthric speech patterns characterized by inconsistent articulation and prosody. Entropy H(X), is mathematically expressed as:

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log_2 p(x)$$

where p(x) denotes the probability of the random variable X taking on the value x In the context of information theory and speech analysis, negative entropy values are not typically encountered, as they would imply a negative amount of uncertainty. Therefore, in practical applications like speech signal analysis, entropy values are non-negative.

Entropy is calculated on speech signals by filtering them through a Gabor filter bank, extracting the linear predictive residual, and then computing the entropy of the filtered signal. *Zero-mean entropy is chosen in order to emphasize the variability of entropy values around the mean value.* It is calculated by subtracting the mean entropy of each frame from its individual entropy values, providing a normalized measure of entropy variation across frames [12]. It can be calculated by,

$$ZeroMeanEntropy = H(X) - \frac{1}{n}\sum_{i=1}^{n} H(x_i)$$

In this process of entropy calculation, Gabor filter banks are used for their ability to capture both frequency and temporal information from speech signals, which is crucial for analyzing dysarthric speech [13]. They decompose the signal into various subbands, allowing the extraction of fine-grained spectral features. *This multi-resolution analysis helps in capturing the small patterns and variations in the speech signal, making it easier to identify and quantify the irregularities associated with dysarthria.* By using a Gabor filter bank, the analysis becomes more robust and comprehensive, as it effectively highlights the relevant spectro-temporal characteristics that are essential for accurate entropy measurement and subsequent classification of dysarthric severity levels [14]. Furthermore, fusion of traditional baseline features and zero mean multi band entropy based features are performed. Fig. 1 contains the block diagram of the proposed feature set.

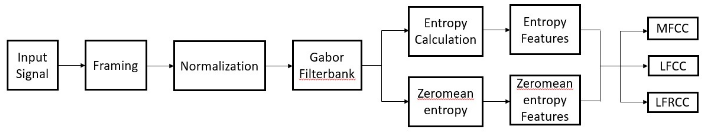

Fig. 1. Block Diagram of Proposed Architecture

The entropy and zero mean entropy features are extracted using 40 Gabor filterbanks while segmenting the speech signal using 20 ms window length with 10 ms hop length. These overlapping windows ensured effective temporal feature capture. Entropy values were computed for each window, normalized to achieve zero-mean entropy, and then arranged in descending order to prioritize the most significant features, enhancing the representation for subsequent processing or classification. Furthermore, the top 20 and bottom 20 features from the ordered sub bands are analyzed separately for entropy and zero mean entropy.

By integrating entropy features with MFCC, LFCC, and LFRCC, we enhance the analysis of dysarthric speech. Entropy measures the unpredictability or variability in the speech signal, which, when combined with these cepstral coefficients, improves the sensitivity to different spectral characteristics. This integration enables a more accurate assessment of speech impairments, making it possible to better differentiate and analyze variations in dysarthric speech.

III. Experimental Setup

A. Dataset Used

In this study, we use two well-known datasets of dysarthric speech: the Universal Access Dysarthria Speech Corpus (UA-Speech) and the TORGO corpus. These datasets mainly feature speech from individuals with spastic dysarthria, characterized by symptoms such as breathiness, hypernasality, a harsh voice, and incorrect articulation, making the speech difficult to understand. The UA-Speech corpus includes recordings from eight speakers (four males: M01, M05, M07, and M09; and four females: F02, F03, F04, and F05). Each speaker contributes 465 utterances, selected from a total of 735 available utterances per speaker. The UA corpus, contains 3537 speech samples categorized into four severity levels: 930 samples of very low severity, 926 samples of low severity, 930 samples of medium severity and 751 samples of high severity. The TORGO corpus, on the other hand, contains 1982 speech samples categorized into three severity levels: 671 samples of very low severity, 627 samples of low severity, and 684 samples of medium severity. For both datasets, we use 80% of the data for training and 20% of the data for testing. The training and testing sets are carefully split to ensure both contain a variety of words, non-words, and sentences, providing a comprehensive assessment of the model's performance.We employ a 5-fold cross-validation (CV) approach in our experiments. This involves dividing the training data into five subsets, training the model on four of these subsets, and validating it on the remaining subset. This process is repeated five times, with each subset used once for validation.

B. Classifier Used

The experiments used a Convolutional Neural Network (CNN) classifier due to its effectiveness in capturing essential spectro-temporal patterns in speech signals and its ability to maintain translation invariance. The model was trained with a stratified 5-fold cross-validation (CV) approach. This method ensures that each fold has a data distribution similar to the entire dataset, enhancing the robustness and reliability of the training process. The data was split into 80% for training and 20% for validation, and the model was optimized using the Adam optimizer, with categorical cross-entropy as the loss function, and accuracy as the evaluation metric. The CNN used two activation functions: ReLU and softmax. ReLU was selected to improve learning speed and reduce computational cost, while softmax was used in the final layer for multiclass classification. To prevent overfitting, normalization layers and dropout layers were added after each convolutional layer. These layers help the model generalize better by reducing the risk of becoming too specific to the training data, thereby improving its performance on unseen data.

C. Baseline Features

In this study, MFCC, LFCC and LFRCC features are used as baselines. MFCCs are widely used in speech processing and are derived from the short-term power spectrum of a speech signal. They transform the frequency components of the signal to the Mel frequency scale, which is more aligned with human auditory perception. This transformation captures important aspects of speech that are relevant to human hearing, allowing for effective representation of the spectral characteristics of speech.

LFCCs, on the other hand, are computed using a linear frequency scale rather than the Mel scale. This approach is advantageous for capturing detailed frequency variations in the speech signal. LFCCs especially in cases of speech disorders like dysarthria is crucial due to its linear scale.

LFRCCs are derived from the residual signal obtained after applying pre-emphasis and frame-wise processing [15]. The process starts with pre-emphasizing the residual signal to enhance high-frequency components, followed by dividing the signal into frames and applying a windowing function to each frame. The Fourier Transform is then used to compute the magnitude spectrum of each frame. This spectrum is passed through a linear filter bank, and the resulting energies are logarithmically transformed. Finally, the Discrete Cosine Transform (DCT) is applied to these log energies to obtain the cepstral coefficients [16]. This method captures the detailed spectral characteristics of the speech signal on a linear frequency scale, providing a comprehensive representation of the residual speech signal.

To ensure a fair comparison, 20-dimensional feature vectors were extracted from the speech data using the librosa toolkit, with a window length of 25 ms and a hop length of 10 ms for all feature sets. These features are commonly used in speech processing because they effectively represent the speech signal's spectral characteristics, making them suitable for analyzing dysarthric speech patterns [17]. By maintaining consistent extraction parameters across all feature sets, the study ensures that the comparative analysis of different features is reliable and unbiased.

IV. Spectrographic Analysis

A. Effect of entropy

Figure 2 illustrates the distribution of entropy across various subbands and frames for the word "November." Each subplot, labeled from (a) to (d), demonstrates how entropy varies, with subbands on the x-axis and frames on the y-axis. The z-axis represents entropy values, using a color gradient from blue (indicating lower entropy) to red (indicating higher entropy) to make differences easily visible. Subplots (a) and (b) display considerable fluctuations with notable peaks and valleys, which reflect a high degree of entropy variability. In contrast, subplots (c) and (d) have smoother, more uniform surfaces, indicating more consistent entropy values. The spectrogram reveals that as dysarthric severity increases, the overall entropy also increases. This pattern is evident in the transition from low to high severity levels, where the entropy values become more pronounced. This correlation between rising entropy levels and increased dysarthric severity underscores the utility of entropy as a metric for assessing the severity of dysarthria. The visual representation provided by these plots offers valuable insights into how entropy distribution changes with the progression of the condition.

From Fig 3 the provided bar graphs illustrate the zero-mean entropy of each subband for varying degrees of dysarthric severity. Each subplot, from (a) to (d), represents different severity levels, with the x-axis indicating the subband index and the y-axis showing the zero-mean entropy values. Subplots (a) indicates values are relatively low and stable indicating quite regular and predictable character of speech. This minimal variation around the mean can suggest controlled phonation. Subplot (b) demonstrate that when compared to very low severity, the variations around the mean value of entropy increases indicating the phonatory instability, reflecting high variability in entropy. Subplots (c) and (d) shows as the severity increases, the fluctuations ncreases and the entropy value increases indicating the irregularity and unpredictability of the speech signal resulting in source phonation instability. The higher entropy values can indicate severe noise in the speech production system. The overall pattern reveals that as the severity of dysarthria increases, the entropy values also increase, highlighting a direct correlation between higher entropy levels and greater dysarthric severity. This consistent rise

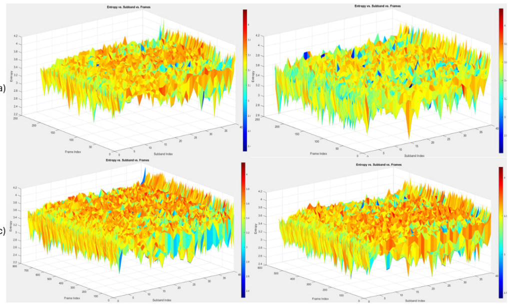

Fig. 2. Spectrogram of multi band entropy for each frame of the word "November" for (a)very low, (b)low, (c)medium, and (d)high severity-level dysarthric signal, respectively. X-axis represents the sub-band index, Y-axis represents frame index, and Z-axis represents the entropy value, respectively.

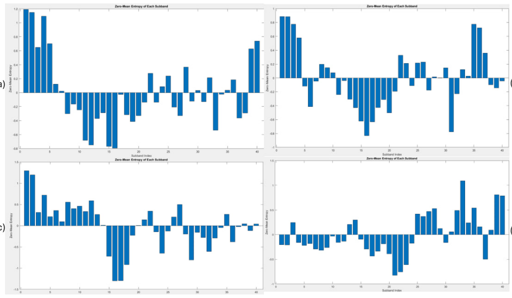

Fig. 3. Effect of entropy for the word "November" for (a)very low, (b)low, (c)medium, and (d)high severity-level dysarthria speaker. X-axis represents the sub-band index and Y-axis represents the zero mean entropy, respectively.

in entropy with increasing severity offers valuable insights into the relationship between entropy distribution and dysarthric speech impairment.

Together, these analyses clearly demonstrate that as dysarthria severity increases, speech becomes more chaotic and disordered, with higher and more variable entropy values.

V. Experimental Results

Table I indicates the classification accuracy of baseline features using cnn classifier.

TABLE I
Fold, Test, Precision (P), Recall (R), F1-Score for baseline features using CNN classifier on UA-Speech and TORGO.

Data	Features	Fold Acc.	Acc.	P	R	F1
UA-Speech	MFCC	93.09	93.36	93.89	93.38	93.56
	LFCC	94.11	95.45	95.63	95.60	95.57
	LFRCC	92.01	93.78	93.55	94.06	93.71
TORGO	MFCC	85.71	88.62	89.40	88.64	89.01
	LFCC	91.58	91.82	91.51	91.02	91.26
	LFRCC	92.43	94.20	93.96	94.19	94.07

Table 1 contains the fold test accuracies of baseline features. For AU Speech dataset, LFCC outperforms both MFCC and LFRCC baseline features by a fold (test) accuracy margin of 1.02% (2.09%), 2.1% (1.67%) and for TORGO dataset, LFRCC outperforms both MFCC and LFCC baseline features by a fold (test) accuracy margin of 6.72% (5.58%), 0.85% (2.38%). This might be because dysarthric speech is seen to have turbulent noise in a higher frequency range and thus, the linear frequency scale in LFCC and LFRCC offers relatively better frequency resolution than its MFCC counterpart to capture the spectral details at higher frequency region.

TABLE II
Fold, test, precision (P), recall (R), F1-score for 40 sub band features along with first 20 and last 20 of decreasing arranged sub band entropy and zero mean entropy features using CNN classifier on UA-Speech.

Features	Fold Acc.	Acc.	P	R	F1
Entropy	67.77	77.11	76.86	77.13	76.65
Entropy (first 20)	57.71	59.46	59.71	59.46	59.56
Entropy (last 20)	50.95	58.47	58.06	59.26	56.62
Zero mean entropy	**82.75**	**86.15**	**86.65**	**85.87**	**86**
Zero mean entropy (first 20)	79.86	80.36	80.40	80.04	80.17
Zero mean entropy (last 20)	68.86	71.89	71.64	72.57	70.66

TABLE III
Fold, test, precision (P), recall (R), F1-score for 40 sub band features along with first 20 and last 20 of decreasing arranged sub band entropy and zero mean entropy features using CNN classifier on Torgo.

Features	Fold Acc.	Acc.	P	R	F1
Entropy	71.36	74.81	77	75.53	74.86
Entropy (first 20)	65.55	70.19	70.00	70.60	69.16
Entropy (last 20)	60.28	62.24	64.81	60.10	56.84
Zero mean entropy	**90.93**	**91.93**	**92.15**	**92.17**	**91.93**
Zero mean entropy (first 20)	89.14	84.88	886.25	85.38	84.51
Zero mean entropy (last 20)	68.86	71.89	71.61	72.57	70.66

TABLE IV
Fold, test, precision (P), recall (R), F1-score for feature level fusion using CNN classifier on UA-Speech and TORGO.

Data	Features	Fold Acc.	Acc.	P	R	F1
UA-Speech	MFCC+Zero mean	95.65	97.74	97.43	97.61	97.51
	LFCC+Zero mean	94.13	97.79	97.33	97.11	97.20
	LFRCC+Zero mean	94.85	95.66	94.46	94.38	94.41
TORGO	MFCC+Zero mean	91.10	96.97	97.06	96.96	96.98
	LFCC+Zero mean	94.51	94.18	92.43	90.87	91.21
	LFRCC+Zero mean	94.24	95.96	96.14	96.33	96.16

Table 2 and table 3 indicate the accuracies for various forms of entropy and zero mean entropy for UA speech and TORGO datasets, respectively. The first 20 and last 20 features indicate the top and bottom 20 features from the descending arranged sub band entropy values. It can be seen that zero mean entropy outperforms the entropy feature by 9.04% for UA Speech and 17.12% for TORGO datasets. Furthermore, it is evident that the information across all the 40 subbands is essential in order to classify the severity level accurately.

Additionally, feature level fusion of baseline features is performed with best performing entropy feature i.e., zero mean entropy and the fusion based features outperform baseline features by **4.38% (8.35%)** for MFCC, **2.34% (2.36%)** for LFCC, **1.88% (1.76%)** for LFRCC on UA Speech (TORGO), respectively. These experimental results sit right with the spectrographic analysis and emphasize the importance of inclusion of entropy based features for dysarthric severity classification.

These consistent improvements across both datasets highlight the effectiveness of integrating zero mean entropy into feature extraction. This technique not only refines the representation of the audio data but also enhances the overall accuracy of speech recognition models, making them more robust and reliable. The observed gains in accuracy underscore the value of adopting zero mean entropy in the development of advanced speech recognition systems.

VI. Conclusions

This study proposed the use of entropy based features for the application of dysarthric severity classification employing two widely used dysarthric speech corpra: UA-Speech and TORGO. Spectrographic analysis of speech data from individuals with varying degrees of dysarthria reveals distinct entropy patterns associated with each severity level. These patterns help in identifying specific speech alterations that occur as the disorder progresses. For instance, individuals with more severe dysarthria often exhibit higher entropy values due to greater variability in their speech signals. This analysis is supported by several experiments.

Experimental results support the effectiveness of incorporating zero mean entropy into the feature extraction process for classification. These results show that adding entropy information significantly boosts the accuracy of these systems, enabling them to better differentiate between different levels of dysarthria. Furthermore, this work can be extended by inclusion of entropy based information extracted using various filterbanks like linear and mel in the state-of-the-art deep learning models like whisper, wav2vec 2.0 and analyzing the noise robustness ability of entropy which helps us understand the practicality of the proposed methodology.

Acknowledgements

This study has been funded by Ministry of Electronics and Information Technology (MeitY), New Delhi, Govt. of India, under the sponsorship of project 'BHASHINI', (Grant ID: 11(1 2022-HCC.(TDIL)). Authors thank authorities of DA-IICT Gandhinagar for their kind support and cooperation to carry out this research work.

References

[1] P. Lieberman, "Primate vocalizations and human linguistic ability," *The Journal of the Acoustical Society of America*, vol. 44, no. 6, pp. 1574–1584, 1968.

[2] R. Palmer and P. Enderby, "Methods of speech therapy treatment for stable dysarthria: A review," *Advances in Speech Language Pathology*, vol. 9, no. 2, pp. 140–153, 2007.

[3] Y. Kim, R. D. Kent, and G. Weismer, "An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria," 2011.

[4] C. Bhat, B. Vachhani, and S. K. Kopparapu, "Automatic assessment of dysarthria severity level using audio descriptors," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2017, pp. 5070–5074.

[5] A. Farhadipour, H. Veisi, M. Asgari, and M. A. Keyvanrad, "Dysarthric speaker identification with different degrees of dysarthria severity using deep belief networks," *Etri Journal*, vol. 40, no. 5, pp. 643–652, 2018.

[6] A. Kachhi, A. Therattil, A. T. Patil, H. B. Sailor, and H. A. Patil, "Teager energy cepstral coefficients for classification of dysarthric speech severity-level," in *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, IEEE, 2022, pp. 1462–1468.

[7] N. Obin and M. Liuni, "On the generalization of shannon entropy for speech recognition," in *2012 IEEE Spoken Language Technology Workshop (SLT)*, IEEE, 2012, pp. 97–102.

[8] S. P. Mishra, P. Warule, and S. Deb, "Speech emotion recognition using mfcc-based entropy feature," *Signal, image and video processing*, vol. 18, no. 1, pp. 153–161, 2024.

[9] J. Ajmera, I. McCowan, and H. Bourlard, "Speech/music segmentation using entropy and dynamism features in a hmm classification framework," *Speech communication*, vol. 40, no. 3, pp. 351–363, 2003.

[10] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Automatic detection of hypernasal speech signals using nonlinear and entropy measurements.," in *INTERSPEECH*, 2012, pp. 2029–2032.

[11] H. Azami, L. Faes, J. Escudero, A. Humeau-Heurtier, and L. E. Silva, "Entropy analysis of univariate biomedical signals: Review and comparison of methods," *Frontiers in Entropy across the Disciplines: Panorama of Entropy: Theory, Computation, and Applications*, pp. 233–286, 2023.

[12] Z. Tüske, "Discriminative feature modeling for statistical speech recognition," Ph.D. dissertation, Dissertation, RWTH Aachen University, 2020, 2020.

[13] A. Lauraitis, R. Maskeliūnas, R. Damaševičius, and T. Krilavičius, "Detection of speech impairments using cepstrum, auditory spectrogram and wavelet time scattering domain features," *IEEE Access*, vol. 8, pp. 96 162–96 172, 2020.

[14] J. Schröder, S. Goetze, and J. Anemüller, "Spectro-temporal gabor filterbank features for acoustic event detection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2198–2208, 2015.

[15] P. Gupta and H. A. Patil, "Linear frequency residual cepstral features for replay spoof detection on asvspoof 2019," in *2022 30th European Signal Processing Conference (EUSIPCO)*, IEEE, 2022, pp. 349–353.

[16] C. Nadeu, D. Macho, and J. Hernando, "Time and frequency filtering of filter-bank energies for robust hmm speech recognition," *Speech Communication*, vol. 34, no. 1-2, pp. 93–114, 2001.

[17] B. A. Al-Qatab and M. B. Mustafa, "Classification of dysarthric speech according to the severity of impairment: An analysis of acoustic features," *IEEE Access*, vol. 9, pp. 18 183–18 194, 2021.