2024 17th International Conference on Development in eSystem Engineering (DeSE) | 979-8-3503-6869-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/DeSE63988.2024.10911938

Speech Enhancement Algorithm using Deep Learning and Hahn Polynomials

Ammar S. Al-Zubaidi
Computer Center
University of Baghdad
Baghdad, Iraq

Riyadh Bassil Abduljabbar
Department of Computer Engineering
University of Baghdad
Baghdad, Iraq

Basheera M Mahmmod
Department of Computer Engineering
University of Baghdad
Baghdad, Iraq

Sadiq H. Abdulhussain
Department of Computer Engineering
University of Baghdad
Baghdad, Iraq

Marwah A. Naser
Department of Computer Engineering
University of Baghdad
Baghdad, Iraq

Muntadher Alsabah
Medical Technical College
Al Farahidi University
Baghdad, Iraq

Abir Hussain
Department of Electrical Engineering
University of Sharjah
Sharjah, UAE

Dhiya Al-Jumeily
School of Computer Science and Mathematics
Liverpool John Moores University
Liverpool, UK

***Abstract*—Speech enhancement algorithms and machine learning can play a fundamental role in signal processing to improve speech quality. These techniques can be used to reduce noise and distortions in speech signals, hence ensuring clearer and more intelligible speech. By leveraging advanced machine learning, speech enhancement algorithms not only improve the listener's auditory system, but also increase the efficacy of speech recognition systems. In particular, deep learning is a class of machine learning techniques, which have recently been used in speech enhancement. This paper proposes the use of Discrete Hahn polynomials (DHPs) o extract spectral features from noisy signals using fully connected neural networks and convolutional neural network. Deep learning can efficiently capture the contextual information of speech signals, resulting in superior improvements in speech quality and intelligibility properties. The results are evaluated based on the well-known TIMIT database. The results show that the presented model is able to enhance the speech signal for different conditions.**

***Index Terms*—Deep learning, convolutional neural networks, speech enhancement, Hahn polynomials, Hahn moments.**

I. Introduction

Speech enhancement plays a fundamental role in speech signal processing fields because it improves the overall perceptual quality and intelligibility of the degraded signal. The main objective of developing speech enhancement techniques is to obtain pure speech signals from the attenuated ones [1]. To this end, wide range of techniques have been used over the past years, from conventional to advanced methods. Numerous contributions have been made so far aiming to achieve remarkable improvement in the speech enhancement [2]. At the same time, the recent advancement in deep learning algorithms made it a feasible choice for improving speech enhancement [3]. In particular, the adaptation of deep neural network (DNN) algorithms has an effective role in the process of denoising the speech signals [4]–[7]. DNN algorithms are recently used to improve speech enhancement due to their effectiveness [8], which in turn removes the inherent noise without affecting the original speech signal [9]. In particular, DNN algorithms accurately and reliably transform noisy to clean speech signals by employing non-linear neural network techniques [10].

Background noise could adversely impact the performance of speech-based applications [11]. Several Speech-based applications like automatic speaker identification, mobile communication, speech recognition, and hearing aids, need to improve the intelligibility and quality of speech signals [12]. The measurement of SNR is a widespread technique for detecting noise in speech signals. Numerous research have been made to overcome the diverse effects of background noise [13]–[15]. Recently, DNNs have displaced conventional methods because of their outperforming ability to improve quality of speech signal [16]. In [17], a deep learning-based speech enhancement technique was presented by employing recurrent neural networks (RNNs) alongside DNNs with supervised learning methods. The results depicted an improving performance in both speech intelligibility and quality for different noisy environments. Zhao et al. in [18] proposed an EHNET model, which is a combination of convolutional architectures with RNNs. The performance of the EHNET outperforms other models in terms of five metrics under both visible and unseen noise conditions. Jerjees et al. [19] utilized orthogonal polynomials, specifically Charlier polynomials [20], with FCNN to improve the performance of enhancement process. It is noteworthy that orthogonal polynomials have been utilized in different applications for their robustness in signal representation [21]–[25].

Neural networks are extensively utilized in NN-based speech enhancement algorithms. Among these, fully connected neural networks (FCNN) are a common type that can be utilized here. To this end, a supervised approach is introduced to improve speech quality by learning a mapping function between clean and noisy speech signals using DNNs [26]. The log-power spectral features extracted from pairs of noisy and clean speech data is used to train the regression model. This feature extraction involves applying short-time Fourier analysis to the input speech and then computing the discrete Fourier transform of each overlapping windowed frame [26]. Researchers explored various types of neural networks to leverage the strengths of both convolutional and recurrent neural networks [27]. This combination approach aims to develop a progressive learning model that enhances speech signals in terms of intelligibility and quality.

This paper proposes the use of Discrete Hahn Polynomials (DHPs) to extract spectral features of noisy signals using FCNN and convolutional neural network (CNN). The goal is to optimize speech enhancement algorithm performance by exploring various neural network architectures.

II. Preliminaries

A. FCNN

FCNN is considered a core architecture in deep learning where each neuron in a layer is connected to all neurons in the next layer. Particularly, this type of network is useful for tasks that require learning complex representations of data, such as image recognition, natural language processing, and speech enhancement. The main advantages of FCNN can be summarized as 1) Versatility: this makes it suitable for a wide range of tasks. 2) Simplicity: this makes it conceptually straightforward and easy to implement. 3) Expressive Power: this makes it capable of learning complex relationships. However, despite the advantages of FCNN, there are some disadvantages, such as 1) Computationally Intensive: the requirement for a large number of parameters, leading to high computational cost and memory usage. 2) Overfitting: that makes it prone to overfitting, especially with small datasets, due to its large number of parameters.

Currently, FCNN used in different applications of machine learning, like speech enhancement and object detection [28]–[30]. DNN consists of many layers of interconnected nodes that are similar to the neurons of the brain, each node pass its inputs to the next state, where this state is considered an input to the next layer. An example of DNN with the FCNN as depicted in Fig.1. The implementation of FCNN and DNN is able to model complex and nonlinear relationships when compared to traditional approaches [31]. Modeling of the network is accomplished by supervised and unsupervised techniques [32]. Speech enhancement algorithm based on DNN is proposed in [26], and a powerful modeling is achieved by utilizing a nonlinear regression function in the network architecture.

B. Convolutional neural network (CNN)

CNNs algorithms are a class of deep learning models specifically designed for processing data with grid-like topology, such as images or audio spectrograms. CNNs algorithms are a powerful and versatile tool in deep learning, particularly well-suited for tasks involving spatial data. Their ability to automatically learn and extract hierarchical features, making

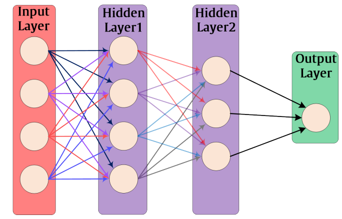

Fig. 1: Schematic representation of FCNN.

them ideal for applications ranging from image recognition to speech enhancement. While they come with computational demands, their performance benefits often justify the costs. There are several advantages of CNNs. These advantages include 1) Translation Invariance: this allows the network to recognize patterns regardless of their position. 2) Parameter Efficiency: sharing weights across spatial dimensions reduces the number of parameters compared to fully connected networks. 3) Automatic Feature Extraction: CNNs automatically learn to obtain relevant features, reducing the need for hand-crafted feature extraction. Besides, there are some disadvantages related to the CNNs. The most common disadvantages include, computational cost and the requirement for large datasets.

C. Computation of Transform Coefficients (Moments)

In this section, we provide the fundamentals mathematics of DHP and its moments.

1) The mathematical definition of DHP: The nth order of the DHPs, $H_n^{\alpha,\beta}(x)$, is defined as follows:

$$H_n^{\alpha,\beta}(x;N) = \frac{(-1)^n(\beta+1)_n(N-n)_n}{n!} \times {}_3F_2\left(\left.\begin{matrix}-n,-x,n+1+\alpha+\beta\\ \beta+1,1-N\end{matrix}\right| 1\right), \tag{1}$$

where ${}_3F_2(\cdot)$ denotes the generalized hypergeometric series [33]. The orthogonality condition is satisfied by the DHPs as

$$\sum_{x=0}^{N-1} H_n^{\alpha,\beta}(x;N)\, H_m^{\alpha,\beta}(x;N)\, \omega_h(x) = \rho_h(n)\delta_{nm} \ , \tag{2}$$

where δ_{nm} denotes the Kronecker delta, ρ_h and ω_h are the norm and weight functions of DHP

$$\omega_h(x) = \frac{\Gamma(\beta+x+1)\Gamma(N+\alpha-x)}{\Gamma(x+1)\Gamma(N-x)} \tag{3}$$

$$\rho_h(n) = \frac{\Gamma(\beta+n+1)\Gamma(\alpha+n+1)(\alpha+\beta+n+1)_N}{(2n+\alpha+\beta+1)\Gamma(N-n)\Gamma(n+1)}. \tag{4}$$

The nth order of the weighted DHP is defined as

$$\hat{H}_n^{\alpha,\beta}(x) = H_n^{\alpha,\beta}(x;N)\sqrt{\frac{\omega_h}{\rho_h}}. \tag{5}$$

2) The definition of DHM: DHMs provided the signals, which are considered here by speech or images that can be projected on the DHP basis. For one-dimensional signals (speech), $f(x)$, the DHMs, η_n, can be computed as

$$\eta_n = \sum_{x=0}^{N_1-1} \hat{H}_n^{\alpha,\beta}(x;N_1)\, f(x), \quad n = 0,1,\ldots,N_1-1 \tag{6}$$

where N_1 is the size of the sample of the signal $f(x)$. The signal can be reconstructed from Hahn moment domain into the time domain as follows

$$\hat{f}(x) = \sum_{n=0}^{N_1-1} \hat{H}_n^{\alpha,\beta}(x; N_1)\,\eta_n, \quad x = 0, 1, \ldots, N_1 - 1 \tag{7}$$

III. Methodology

This section provides an exposition of the specific aspects of the proposed work. The major objective of our work is to enhance the characteristics of the deteriorated signal frame to recover original signal frame from the orthogonal domain. DHT is used to convert the noisy speech signal from time to the moment domain to get the spectral features based on the specified recurrence algorithm. The description of this transform is provided previously, and the implemented recurrence algorithm is presented in the following sections. Furthermore, we conduct the proposed work on the adoption of two types of deep neural networks for the spectral denoising process: FCNN [34] and CNN [35]. While FCNN has fewer parameters than fully connected networks, CNN offers a more lightweight solution that is able to obtain the enhanced speech. A speech enhancement method using robust DHT is proposed, which used a framework based on FCNN and CNN to get a high reduction of noise with minimal speech distortion.

A. The implemented recurrence algorithm

DHP of nth order can be calculated as follows [36]

$$\hat{H}_n^{\alpha,\beta}(x) = \eta_1 \left[\eta_2 \hat{H}_n^{\alpha,\beta}(x-1) + \eta_3 \hat{H}_n^{\alpha,\beta}(x-2)\right] \tag{8}$$

$$n = 0, 1, \ldots, N-1; \quad \text{and} \quad x = 2, 3, \ldots, N-1,$$

with initial values

$$\hat{H}_n^{\alpha,\beta}(0) = (1-N)_n \binom{n+\beta}{n} \sqrt{\frac{\omega_h(0)}{\rho_h(n)}} \tag{9}$$

$$\hat{H}_n^{\alpha,\beta}(1) = \frac{(n+\beta+1)(N-n-1) - n(N+\alpha-1)}{(\beta+1)(N-1)} \sqrt{\frac{\omega_h(1)}{\omega_h(0)}}\, \hat{H}_n^{\alpha,\beta}(0) \tag{10}$$

Symmetry relation is exploited to reduce the required computation time of DHPs coefficients as follows [37]:

$$\hat{H}_n^{\alpha,\beta}(x) = (-1)^n \hat{H}_n^{\alpha,\beta}(N-1-x) \text{ for } \alpha = \beta. \tag{11}$$

B. The implementation of DHP based speech enhancement

DHP is used in the current work as a powerful tool that is used because of its high performances in features extraction [38]. The capability of the current polynomial in information extraction is high. For the additive noise model, let us denote noisy speech signal as the discrete-time speech signal that is degraded by the additive uncorrelated background noise, so the resulting in the observed noisy signal. Then, the noisy signal is transformed into transform domain by applying discrete Hahn transform on the additive model. To get the transform moments of the noisy, noise, and clean signals. Then, a windowing process is applied on the noisy signal to truncated it to an appropriate number of frames. The resulting distorted speech frames are fed to the DNN, preparing them for denoising process. After the enhancement process, the inverse of the DHT is applied to convert the enhanced speech signal back to the time domain.

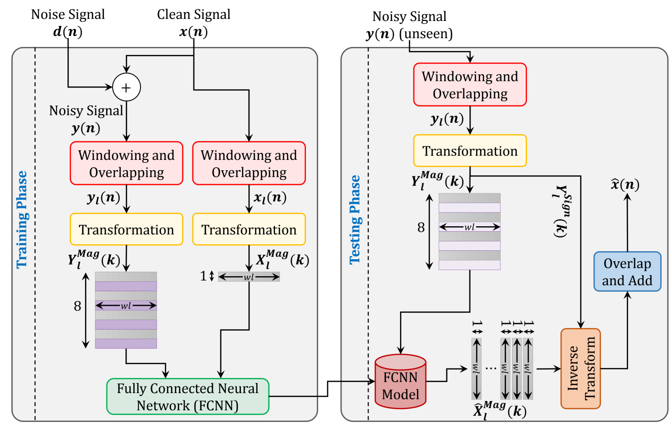

Fig. 2: Flow diagram of SEA using FCNN and CNN.

C. Speech Enhancement Algorithm using DNN

There are two primary categories of DNN-based speech enhancement algorithms: time domain and spectral domain [39]. In the proposed work, the first type of DNN used in the implementation is FCNN. It has been used to reduce the background noise without affecting the quality or intelligibility of the processed signal. During training process, the inputs of the system are noisy speech signal $y(n)$ and clean speech $x(n)$. Noise degrades the clean speech signal, adding noise $d(n)$ to the clean speech $x(n)$. Windowing and overlapping process are used to truncate the input signals. In this paper, the utilized window type is Hamming windows. The signals is partitioned into frames, where the current frame is represented by (l); therefore, the current frame of clean signal is represented by $x_l(n)$ and the noisy signal is represented by $y_l(n)$. The Hahn-polynomials is used to transform the current frame to the DHT domain. The index k denotes the signal index at the DHT domain to acquire $X_l(k)$ and $Y_l(k)$. The magnitude of the clean and noisy signals are denoted by XMag and YMag, respectively. The model is trained to output a single frame of the clean signal, which serves as the target output, and the model uses eight frames of the noisy signal as input features. Finally, the FCNN model is then trained. During the testing process, to obtain the enhanced signal $\hat{x}(n)$, the same procedure of the training is performed using the unseen signal $x(n)$. The output frame from the FCNN is combined with the sign. Then, the inverse is carried out using Hahn polynomials. Finally, the resulting signal is then subjected to an add-and-overlap process, ultimately yielding the enhanced signal $\hat{x}(n)$.

The building blocks of FCNN are comprised of three types of layers. The first layer is the input layer, where we provide the input signal to our model. The number of neurons is determined by the input data (the noisy input signal). The number of input neurons is directly related to the size of the used window, which is the same as the size of the used transform that converts the time domain signal to the transform domain. The second layer is the hidden layers, and in the current model, we have two hidden layers where the output from the input layer is fed into the hidden layers. However, the number of neurons in the input layer is less than the number of neurons in the hidden layers. Each layer of hidden layer produces its output by performing matrix multiplication with adjustable weights and biases, followed by the application of an activation function. The output layer is the final one. In a FCNN, each input has a direct impact on every output in the output vector, as all potential connections between layers are present. The complete topology of the utilized FCNN is adopted from [34]. To optimize the hidden layer output, two units are inserted consecutively after each hidden layer, which are batch normalization and rectified linear units. The output layer consists of M nodes, followed by a regression layer that handles the regression tasks. This study adopted CNN as the second type of deep neural network. The same procedure is implemented, but CNN network has fewer parameters than the FCNN due to its weight-sharing feature. The total number of convolutional layers is 16 [34]. The optimal network architecture achieves the highest performance while utilizing 15 convolutional layers to extract features from the input speech signal. A batch normalization layer and a ReLu layer are applied to optimize each convolutional layer. The final layer implements a convolutional layer with a single filter, followed by a regression layer.

After training the model, the enhancement process extracts the transformed frames and presents them as inputs. Correct training of the model allows to predict the clean signal's moments for each noisy moment under analysis. Then, the inverse DHT with overlap-add is implemented to synthesize the denoised signal in the time domain

IV. Results and Discussion

In this section, we examine the performance of Hahn polynomial-based transform models in reducing noise and preserving speech quality and intelligibility. Specifically, we test these models using a dataset of 32 speech files from the TIMIT corpus, which comprises an equal number of male and female speakers.

In this initial experiment, we assess the performance of the CNN and FCNN models under various parameter settings in a White 0db noise environment. Specifically, we investigate how the Hahn polynomial parameters α and β, overlap percentage ov, and window length win affect the models' performance. To do this, we test different parameter combinations: α and β take on values of 10 and 50, ov is set to 50% and 75%, and win varies across 128, 256, 384, and 512. For each parameter combination, we train and evaluate the models on a dataset, measuring their performance using five evaluation metrics: SegSNR, PESQ, OVL, FWseg, and SNR. The goal of this experiment is to determine how each parameter influences the models' performance and identify the optimal parameter combination for achieving the best results. The obtained results are reported as improvements in TABLES I and II. By improvement, we mean the reported results was the difference between the enhanced speech metrics and noisy speech metrics.

TABLE I: Performance of CNN model for different parameters settings.

Parameters				SegSNR	PESQ	OVL	FWseg	SNR
a	b	ov	win					
10	10	50%	128	6.243	0.207	0.519	2.256	8.306
			256	5.509	0.144	0.440	1.775	7.361
			384	4.984	0.096	0.376	1.046	6.525
			512	4.637	0.073	0.280	1.041	6.044
10	50	50%	128	5.971	0.154	0.491	1.643	8.039
			256	5.218	0.121	0.420	1.201	6.974
			384	4.706	0.089	0.341	1.095	6.232
			512	4.381	0.072	0.257	1.022	5.715
10	10	75%	128	6.181	0.197	0.516	2.307	8.234
			256	4.634	0.073	0.286	1.044	5.970
			384	5.017	0.105	0.369	1.290	6.637
			512	4.752	0.072	0.267	1.317	6.228
10	50	75%	128	5.916	0.147	0.477	1.649	7.958
			256	5.225	0.120	0.402	1.364	6.966
			384	5.042	0.096	0.328	1.704	6.346
			512	4.383	0.059	0.239	1.033	5.693

From TABLE I, it is cleat that the CNN model achieves its best performance when $\alpha = 10$, $\beta = 10$, $ov = 75\%$, and $win = 128$. This combination yields the highest SegSNR (6.181), PESQ (0.197), OVL (0.516), FWseg (2.307), and SNR (8.234) values. These results demonstrate that the CNN model is most effective when the overlap percentage is high and the window size is small.

TABLE II: Performance of FCNN model for different parameters settings.

Parameters				SegSNR	PESQ	OVL	FWseg	SNR
a	b	ov	win					
10	10	50%	128	4.453	0.110	0.330	2.045	4.298
			256	3.833	0.047	0.164	1.248	3.315
			384	3.431	0.024	0.108	0.674	2.692
			512	3.219	0.017	0.087	0.389	2.392
10	50	50%	128	4.281	0.079	0.297	1.333	4.098
			256	3.734	0.038	0.161	0.679	3.064
			384	3.564	0.024	0.124	0.576	2.794
			512	4.367	0.067	0.243	1.204	4.594
10	10	75%	128	6.308	0.231	0.592	2.818	8.555
			256	5.795	0.162	0.443	2.493	7.644
			384	5.152	0.112	0.344	1.966	6.601
			512	4.626	0.079	0.271	1.546	5.777
10	50	75%	128	6.054	0.177	0.552	2.211	8.272
			256	5.508	0.133	0.394	1.968	7.311
			384	5.042	0.096	0.328	1.704	6.346
			512	4.693	0.072	0.281	1.428	5.698

TABLE II shows that the FCNN model also achieves its best performance when $\alpha = 10$, $\beta = 10$, $ov = 75\%$, and $win = 128$. This combination yields the highest SegSNR (6.308), PESQ (0.231), OVL (0.592), FWseg (2.818), and SNR (8.555) values. The FCNN model tends to outperform the CNN model in terms of SegSNR, PESQ, and OVL values. These results show that the FCNN model is a the best for this task. However, in the following experiments, we have trained and tested both the FCNN and CNN models for different noise types using the optimal parameters achieved from the aforementioned experiments.

TABLE III reported the improvement results of the CNN model for different types of noise (White, Pink, and F16) and

TABLE III: The improvement results of the CNN model for different types of noise.

Type of Nosie	Noise Level	Seg SNR	PESQ	Csig	Cbak	OVL	FWSeg	SNR
White	0	6.181	0.197	0.917	0.469	0.516	2.307	8.234
	5	4.679	0.339	1.092	0.434	0.702	2.869	5.673
	10	3.589	0.539	1.012	0.465	0.769	3.315	3.859
Pink	0	4.972	0.168	0.654	0.431	0.399	1.898	7.148
	5	3.808	0.338	0.630	0.423	0.490	2.260	5.091
	10	2.210	0.392	0.373	0.331	0.385	1.451	2.658
F16	0	4.252	0.142	0.371	0.406	0.271	1.572	6.176
	5	3.248	0.257	0.353	0.382	0.324	1.808	4.316
	10	2.027	0.335	0.251	0.323	0.306	1.336	2.550

noise levels (0, 5, and 10)dB. Overall, the improvement is satisfactory for the FCNN. The model has achieved decent results, with SegSNR values ranging from 2.027 to 6.181, and PESQ values ranging from 0.142 to 0.539. Notably, the White noise type has the highest Seg SNR and SNR values, while the Pink noise type has the lowest PESQ values. These results demonstrate that the FCNN model is effective in enhancing speech, although its performance varies depending on the noise type and level.

TABLE IV: The improvement results of the FCNN model for different types of noise.

Type of Nosie	Noise Level	Seg SNR	PESQ	Csig	Cbak	OVL	FWSeg	SNR
White	0	6.349	0.235	1.014	0.523	0.593	2.835	8.601
	5	5.050	0.387	1.151	0.509	0.766	3.446	6.188
	10	3.294	0.481	0.976	0.433	0.727	3.238	3.350
Pink	0	5.173	0.186	0.645	0.462	0.407	2.358	7.439
	5	3.963	0.300	0.551	0.434	0.439	2.546	5.278
	10	2.325	0.362	0.372	0.345	0.377	1.791	2.714
F16	0	4.829	0.164	0.429	0.489	0.323	2.502	6.813
	5	3.632	0.238	0.305	0.429	0.302	2.313	4.760
	10	2.081	0.238	0.091	0.302	0.186	1.268	2.427

The performance of the FCNN model is reported in TABLE IV, which highlights its ability to handle various noise types and levels. The model improvement is clear because it is able to achieve improvement between the enhanced and noisy speech signals. Notably, the results reveal a significant range of SegSNR values (2.081-6.349) and PESQ values (0.164-0.481).

To evaluate the performance of the presented models based on Hahn polynomials, we have performed a comparison of the OVL and FWseg evaluation metrics in Figures 3 and 4. For each model, the results represent the improvement between enhanced and noisy speech for the presented models (FCNN-Hahn and CNN-Hahn) and existing models based on Charlier polynomials (FCNN-Charlier) [19].

From Fig. 3, it is clear that for a noise level of 0, FCNN-Hahn and CNN-Hahn have relatively close OVL performance across all noise types, with FCNN-Hahn having a slight edge, while FCNN-Charlier has lower OVL performance. For a noise level of 5, FCNN-Hahn has the highest OVL performance across all noise types, followed closely by CNN-Hahn, with FCNN-Charlier still having lower OVL performance.

To further assess the performance, another comparison is reported for the evaluation metric FWseg.

Fig. 4 shows the performance of three models, FCNN-Charlier, FCNN-Hahn, and CNN-Hahn, under different noise

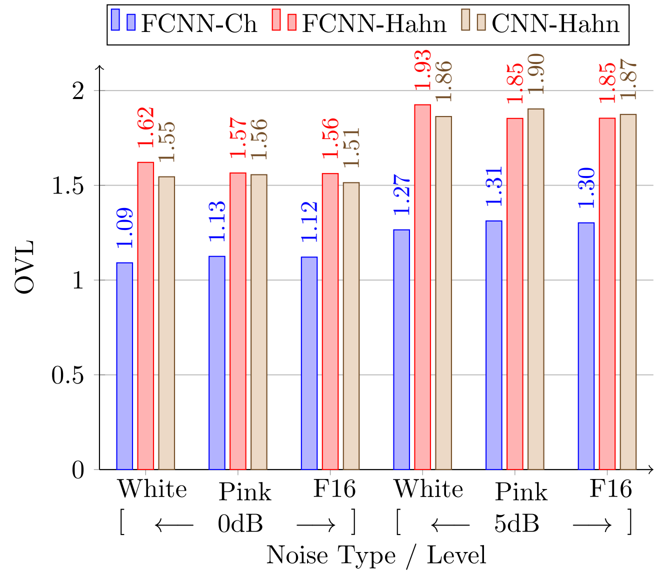

Fig. 3: A performance comparison between different types of noise and SNR levels using OVL.

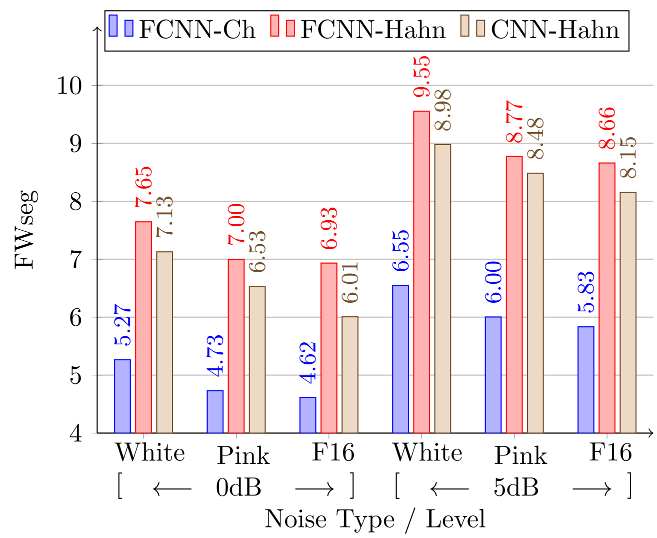

Fig. 4: A performance comparison between different types of noise and SNR levels using FWseg.

levels and types. For both of the noise levels 0dB and 5dB, FCNN-Hahn and CNN-Hahn have similar performance across all noise types, with FCNN-Hahn having a slight edge, i.e., both models are robust to different types of noise. FCNN-Charlier has lower FWseg performance compared to the other two models. To sum up, the presented models based on Hahn polynomials outperforms the existing works.

V. Conclusion

This study leveraged a Hahn polynomials-based transform to extract enhanced signals from noisy inputs using both FCNN and CNN. By employing deep learning techniques, the approach effectively captured context information from speech signals, resulting in enhanced speech with improved intelligibility and quality. The investigation explored various noise levels and discrete transform parameters to optimize speech enhancement algorithm through deep learning. The results showed that the Hahn polynomials-based transform, with parameters set to $\alpha = 10$ and $\beta = 10$, yielded high-quality and intelligible speech across different noise types. For future trends, the success of this approach paves the way for exploring other discrete transforms in deep learning frameworks, potentially advancing speech enhancement algorithms.

REFERENCES

[1] S. Leglaive, X. Alameda-Pineda, L. Girin, and R. Horaud, "A recurrent variational autoencoder for speech enhancement," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 371–375.

[2] S. A. Nossier, J. Wall, M. Moniri, C. Glackin, and N. Cannings, "An experimental analysis of deep learning architectures for supervised speech enhancement," *Electronics*, vol. 10, no. 1, p. 17, 2020.

[3] A. Pandey and D. Wang, "A new framework for CNN-based speech enhancement in the time domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 7, pp. 1179–1188, 2019.

[4] Y. Zhao, B. Xu, R. Giri, and T. Zhang, "Perceptually guided speech enhancement using deep neural networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 5074–5078.

[5] M. Gogate, K. Dashtipour, and A. Hussain, "Robust real-time audio-visual speech enhancement based on dnn and gan," *IEEE Transactions on Artificial Intelligence*, 2024.

[6] K. Mraihi and M. A. Ben Messaoud, "Deep learning-based empirical and sub-space decomposition for speech enhancement," *Circuits, Systems, and Signal Processing*, pp. 1–31, 2024.

[7] N. K. Shukla, F. H. Shajin, and R. Rajendran, "Speech enhancement system using deep neural network optimized with battle royale optimization," *Biomedical Signal Processing and Control*, vol. 92, p. 105991, 2024.

[8] K. Tan and D. Wang, "Towards model compression for deep learning based speech enhancement," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 29, pp. 1785–1794, 2021.

[9] S. K. Roy, A. Nicolson, and K. K. Paliwal, "DeepLPC: A deep learning approach to augmented Kalman filter-based single-channel speech enhancement," *IEEE Access*, vol. 9, pp. 64 524–64 538, 2021.

[10] A. Nicolson and K. K. Paliwal, "Deep learning for minimum mean-square error approaches to speech enhancement," *Speech Communication*, vol. 111, pp. 44–55, 2019.

[11] A. Pandey and D. Wang, "On cross-corpus generalization of deep learning based speech enhancement," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 28, pp. 2489–2499, 2020.

[12] A. Kumar and D. Florencio, "Speech enhancement in multiple-noise conditions using deep neural networks," *arXiv preprint arXiv:1605.02427*, 2016.

[13] L. Wang and A. Cavallaro, "Deep Learning Assisted Time-Frequency Processing for Speech Enhancement on Drones," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 6, pp. 871–881, dec 2021.

[14] G. Park, W. Cho, K.-S. Kim, and S. Lee, "Speech Enhancement for Hearing Aids with Deep Learning on Environmental Noises," *Applied Sciences*, vol. 10, no. 17, p. 6077, sep 2020.

[15] A. Gabbay, A. Shamir, and S. Peleg, "Visual speech enhancement," *arXiv preprint arXiv:1711.08789*, 2017.

[16] Y. Xia, S. Braun, C. K. A. Reddy, H. Dubey, R. Cutler, and I. Tashev, "Weighted speech distortion losses for neural-network-based real-time speech enhancement," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 871–875.

[17] N. Saleem, M. I. Khattak, M. Al-Hasan, and A. B. Qazi, "On learning spectral masking for single channel speech enhancement using feedforward and recurrent neural networks," *IEEE Access*, vol. 8, pp. 160 581–160 595, 2020.

[18] H. Zhao, S. Zarar, I. Tashev, and C.-H. Lee, "Convolutional-recurrent neural networks for speech enhancement," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2401–2405.

[19] S. A. Jerjees, H. J. Mohammed, H. S. Radeaf, B. M. Mahmmod, and S. H. Abdulhussain, "Deep Learning-Based Speech Enhancement Algorithm Using Charlier Transform," in *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*, 2023, pp. 100–105.

[20] B. M. Mahmmod, W. N. Flayyih, S. H. Abdulhussain, F. A. Sabir, B. Khan, M. Alsabah, and A. Hussain, "Performance enhancement of high degree Charlier polynomials using multithreaded algorithm," *Ain Shams Engineering Journal*, vol. 15, no. 5, p. 102657, may 2024.

[21] H. S. Radeaf, B. M. Mahmmod, S. H. Abdulhussain, and D. Al-Jumaeily, "A steganography based on orthogonal moments," in *Proceedings of the International Conference on Information and Communication Technology - ICICT '19*, ser. ICICT '19. New York, USA: ACM Press, 2019, pp. 147–153.

[22] Z. N. Idan, S. H. Abdulhussain, B. M. Mahmmod, K. A. Al-Utaibi, S. A. R. Al-Hadad, and S. M. Sait, "Fast Shot Boundary Detection Based on Separable Moments and Support Vector Machine," *IEEE Access*, vol. 9, pp. 106 412–106 427, 2021.

[23] S. H. Abdulhussain, B. M. Mahmmod, J. Flusser, K. A. AL-Utaibi, and S. M. Sait, "Fast Overlapping Block Processing Algorithm for Feature Extraction," *Symmetry*, vol. 14, no. 4, p. 715, apr 2022.

[24] S. H. Abdulhussain, B. M. Mahmmod, A. AlGhadhban, and J. Flusser, "Face Recognition Algorithm Based on Fast Computation of Orthogonal Moments," *Mathematics*, vol. 10, no. 15, p. 2721, aug 2022.

[25] D. A. Abdulqader, M. S. Hathal, B. M. Mahmmod, S. H. Abdulhussain, and D. Al-Jumeily, "Plain, Edge, and Texture Detection Based on Orthogonal Moment," *IEEE Access*, vol. 10, pp. 114 455–114 468, 2022.

[26] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "A Regression Approach to Speech Enhancement Based on Deep Neural Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 1, pp. 7–19, jan 2015.

[27] A. Li, M. Yuan, C. Zheng, and X. Li, "Speech enhancement using progressive learning-based convolutional recurrent neural network," *Applied Acoustics*, vol. 166, p. 107347, sep 2020.

[28] J. Du, Q. Wang, T. Gao, Y. Xu, L.-R. Dai, and C.-H. Lee, "Robust speech recognition with speech enhanced deep neural networks," in *Fifteenth annual conference of the international speech communication association*, 2014.

[29] M. M. Faisal, M. S. Mohammed, A. M. Abduljabar, S. H. Abdulhussain, B. M. Mahmmod, W. Khan, and A. Hussain, "Object Detection and Distance Measurement Using AI," in *2021 14th International Conference on Developments in eSystems Engineering (DeSE)*. IEEE, dec 2021, pp. 559–565.

[30] M. S. Mohammed, A. M. Abduljabar, M. M. Faisal, B. M. Mahmmod, S. H. Abdulhussain, W. Khan, P. Liatsis, and A. Hussain, "Low-cost autonomous car level 2: Design and implementation for conventional vehicles," *Results in Engineering*, vol. 17, p. 100969, mar 2023.

[31] J. Zou, Y. Han, and S.-S. So, "Overview of Artificial Neural Networks," 2008, pp. 14–22.

[32] H. Kim, J. Kim, S. Won, and C. Lee, "Unsupervised deep learning for super-resolution reconstruction of turbulence," *Journal of Fluid Mechanics*, vol. 910, 2021.

[33] B. M. Mahmmod, W. N. Flayyih, Z. H. Fakhri, S. H. Abdulhussain, W. Khan, and A. Hussain, "Performance enhancement of high order Hahn polynomials using multithreading," *PLOS ONE*, vol. 18, no. 10, p. e0286878, oct 2023.

[34] S. R. Park and J. Lee, "A fully convolutional neural network for speech enhancement," *arXiv preprint arXiv:1609.07132*, 2016.

[35] D. Liu, P. Smaragdis, and M. Kim, "Experiments on deep learning for speech denoising," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[36] H. Zhu, M. Liu, H. Shu, H. Zhang, and L. Luo, "General form for obtaining discrete orthogonal moments," *IET Image Processing*, vol. 4, no. 5, pp. 335–352, 2010.

[37] A. Daoui, M. Yamni, O. E. Ogri, H. Karmouni, M. Sayyouri, and H. Qjidaa, "New algorithm for large-sized 2D and 3D image reconstruction using higher-order Hahn moments," *Circuits, Systems, and Signal Processing*, 2020.

[38] B. M. Mahmmod, S. H. Abdulhussain, T. Suk, and A. Hussain, "Fast Computation of Hahn Polynomials for High Order Moments," *IEEE Access*, vol. 10, pp. 48 719–48 732, 2022.

[39] S. Drgas, "A survey on low-latency dnn-based speech enhancement," *Sensors*, vol. 23, no. 3, p. 1380, 2023.