Infocommunications Journal

HTE75 1949–2024

A PUBLICATION OF THE SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS (HTE)

March 2024 Volume XVI Number 1 ISSN 2061-2079

MESSAGE FROM THE EDITOR-IN-CHIEF

From PocketQube to Glucose Monitoring – in '24 ..... *Pal Varga* 1

PAPERS FROM OPEN CALL

Deep Learning based DDoS Attack Detection in Internet of Things: An Optimized CNN-BiLSTM Architecture with Transfer Learning and Regularization Techniques ..... *Iqbal Jebril, M. Premkumar, Ghaida Muttashar Abdulsahib, S. R. Ashokkumar, S. Dhanasekaran, Oshamah Ibrahim Khalaf, and Sameer Algburi* 2

Secure post-processing for non-ideal photon arrival time based quantum random number generator ..... *Balázs Solymos, and László Bacsárdi* 12

An Ordered QR Decomposition based Signal Detection Technique for Uplink Massive MIMO System ..... *Jyoti P. Patra, Bibhuti Bhusan Pradhan, and M. Rajendra Prasad* 20

Resonant Radar Reflector On VHF / UHF Band Based on BPSK Modulation at LEO Orbit by MRC-100 Satellite ..... *Yasir Ahmed Idris Humad, and Levente Dudás* 26

Advancements in Expressive Speech Synthesis: a Review ..... *Shaimaa Alwaisi and Géza Németh* 35

Speech synthesis from intracranial stereotactic Electroencephalography using a neural vocoder ..... *Frigyes Viktor Arthur, and Tamás Gábor Csapó* 47

On the Performance of Metamaterial based Printed Circuit Antenna for Blood Glucose Level Sensing Applications: A Case Study ..... *Taha A. Elwi, Hayder H. Al-Khaylani, Wasan S. Rasheed, Sana A. Al-Salim, Mohammed H. Khalil, Lubna Abbas Ali, Omar Almukhtar Tawfeeq, Saba T. Al-Hadeethi, Dhulfiqar Ali, Zainab S. Muqdad, Serkan Özbay, and Marwah.M. Ismael* 56

Enhancing Parkinson's Disease Recognition through Multimodal Analysis of Archimedean Spiral Drawings ..... *Attila Zoltán Jenei, Dávid Sztahó, and István Valálik* 64

Farewell to Tamás Gábor Csapó 73

ADDITIONAL

Guidelines for our Authors 72

Technically Co-Sponsored by

IEEE ComSoc™ IEEE Communications Society

hte

IEEE HUNGARY SECTION

Indexing information

Infocommunications Journal is covered by Inspec, Compendex and Scopus.
Infocommunications Journal is also included in the Thomson Reuters – Web of ScienceTM Core Collection, Emerging Sources Citation Index (ESCI)

Infocommunications Journal

Technically co-sponsored by IEEE Communications Society and IEEE Hungary Section

Supporters

FERENC VÁGUJHELYI – president, Scientific Association for Infocommunications (HTE)

The publication was produced with the support of the Hungarian Academy of Sciences and the NMHH

National Media and Infocommunications Authority • Hungary

Editorial Office (Subscription and Advertisements)**:**
Scientific Association for Infocommunications
H-1051 Budapest, Bajcsy-Zsilinszky str. 12, Room: 502
Phone: +36 1 353 1027 • E-mail: info@hte.hu • Web: www.hte.hu

Articles can be sent also to the following address:
Budapest University of Technology and Economics
Department of Telecommunications and Media Informatics
Phone: +36 1 463 4189 • E-mail: pvarga@tmit.bme.hu

Subscription rates for foreign subscribers: 4 issues 10.000 HUF + postage

Publisher: PÉTER NAGY

HU ISSN 2061-2079 • Layout: PLAZMA DS • Printed by: FOM Media

From PocketQube to Glucose Monitoring – in ’24

Pal Varga

Let's see what the first issue of ICJ in 2024 has brought.

The first paper is by Iqbal Jebril et. al., in which they address the pressing challenge of securing IoT networks against Distributed Denial-of-Service (DDoS) attacks. The authors introduce a robust detection model that integrates three advanced deep learning approaches – CNN, BiLSTM, and transfer learning techniques – enhanced further with regularization to optimize performance. Tthis model not only showcases a high detection accuracy of 99.9%, outperforming previous models, but also demonstrates superior capability in distinguishing between legitimate and malicious traffic across various DDoS attack classes.

Balázs Solymos and László Bacsárdi propose an innovative post-processing framework designed to enhance the reliability and security of optical quantum random number generators (QRNGs) that measure photon arrival times. This framework effectively compensates for potential errors arising from non-ideal system components or external attacks, utilizing minentropy estimation and universal hashing techniques. Their method ensures the generation of a high-quality, uniformly distributed bitstream, even under non-ideal conditions. Results underscore the necessity of minimizing or precisely characterizing error sources to optimize the performance of this QRNG post-processing method in practical applications.

In their paper, Jyoti P. Patra, Bibhuti Bhusan Pradhan, and M. Rajendra Prasad address the computational challenges inherent in massive MIMO (m-MIMO) systems, which are exacerbated by the large number of antennas at the base station. The paper introduces two novel signal detection methods: QR Decompositions (QRD) and Ordered QRD (OQRD). These methods aim to reduce computational complexity while main-taining or improving performance compared to MMSE and other suboptimal methods like Gauss-Seidel and Jacobi. The effectiveness of these proposed techniques is demonstrated through simulations, which show a notable enhancement in symbol error rate (SER) and a reduction in computational complexity. The results suggest that the OQRD method, in particular, offers substantial improvements over traditional approaches, making it a promising candidate for efficient signal detection in uplink massive MIMO systems.

Yasir A. I. Humad and Levente Dudás introduce a new method for tracking and identifying PocketQube satellites using a resonant radar reflector. Their approach utilizes a minimal power VHF/UHF antenna subsystem on the satellite, which does not emit RF signals but reflects a continuous wave RF signal sent from a ground-based illuminator. The onboard microcontroller switches a PIN diode to create BPSK-modulated reflections detectable by ground stations equipped with correlation receivers familiar with the specific BPSK code. The paper emphasizes the efficiency in terms of low power consumption, reduced weight, and compact size, making this method ideal for PocketQube satellites which adhere to global standardization and technology readiness levels.

In their survey on Advancements in Expressive Speech Synthesis, Shaimaa Alwaisi and Géza Németh provide a detailed analysis of the progression and current trends in expressive text-to-speech (TTS) systems. It highlights the significant growth and acceptance of speech synthesis technology, particularly in enhancing the naturalness and expressiveness of synthetic speech. The paper focuses on novel methodologies, such as style transfer and speaker variability enhancement among others, and discusses both subjective and objective metrics used to evaluate the quality of synthesized speech. A unique aspect of this paper is its emphasis on the under-explored area of child speech synthesis, identifying it as a fertile ground for future research.

Frigyes Viktor Arthur and Tamás Gábor Csapó introduce significant advancements in the field of Brain-Computer Interfaces (BCI). They demonstrate the feasibility of synthesizing speech from intracranial stereotactic electroencephalography (sEEG) recordings using advanced deep neural network models and a neural vocoder. Their research present the application of FC-DNN, 2D-CNN, and 3D-CNN architectures for converting sEEG data into Mel spectrograms, a critical step for achieving accurate speech synthesis. The subsequent use of the WaveGlow neural vocoder marks a novel approach, significantly enhancing the naturalness and quality of the synthesized speech compared to traditional methods like the Griffin-Lim algorithm.

In their paper Taha A. Elwi and his co-authors present a novel approach to noninvasive glucose monitoring using a metamaterial (MTM) based antenna sensor. This sensor, integrating a defected patch antenna with an interdigital capacitor, enhances electric field fringing to penetrate the human skin effectively for glucose detection. Operating optimally at 0.6GHz with impressive S11 impedance matching, the sensor demonstrates high efficiency in detecting blood glucose variations through direct skin contact. Experimental validations show the sensor’s ability to measure glucose levels accurately. This technology promises a low-cost, efficient solution for continuous glucose monitoring, highlighting its potential impact on diabetes management.

The study by Attila Zoltán Jenei, Dávid Sztahó, and István Valálik explores the potential of using multimodal datasets to improve Parkinson’s disease (PD) diagnosis. Focusing on both drawing and acceleration data, the research team applied pre-trained models to extract features from transformed spiral drawing images and visual motion data representations. Although motion data initially showed superior predictive performance, statistical analysis via the Mann-Whitney U test indicated no significant difference in the diagnostic efficacy between the two modalities across various classification scenarios. The study’s main discovery is that combining predictions from both drawing and motion data significantly enhances disease recognition.

This again, is a colorful compilation of recent proceedings.

Pal Varga is the Head of Department of Telecommunications and Media Informatics at the Budapest University of Technology and Economics. His main research interests include communication systems, Cyber-Physical Systems and Industrial Internet of Things, network traffic analysis, end-to-end QoS and SLA issues – for which he is keen to apply hardware acceleration and artificial intelligence, machine learning techniques as well. Besides being a member of HTE, he is a senior member of IEEE, where he is active both in the IEEE ComSoc (Communication Society) and IEEE IES (Industrial Electronics Society) communities. He is Editorial Board member in many journals, and the Editor-in-Chief of the Infocommunications Journal.

Deep Learning based DDoS Attack Detection in Internet of Things: An Optimized CNN-BiLSTM Architecture with Transfer Learning and Regularization Techniques

Iqbal Jebril [1], M. Premkumar [2], Ghaida Muttashar Abdulsahib [3], S. R. Ashokkumar [4], S. Dhanasekaran [5], Oshamah Ibrahim Khalaf [6], and Sameer Algburi [7]

***Abstract*—In recent days, with the rapid advancement of technology in informatics systems, the Internet of Things (IoT) becomes crucial in many aspects of daily life. IoT applications have gained popularity due to the availability of various IoT enabler gadgets, such as smartwatches, smartphones, and so on. However, the vulnerability of IoT devices has led to security challenges, including Distributed Denial-of-Service (DDoS) attacks. These limitations result from the dynamic communication between IoT devices due to their limited data storage and processing resources. The primary research challenge is to create a model that can recognize legitimate traffic while effectively protecting the network against various classes of DDoS attacks. This article proposes a CNN-BiLSTM DDoS detection model by combining three deep-learning algorithms. The models are evaluated using the CICIDS2017 dataset against commonly used performance criteria which the models perform well, achieving an accuracy of around 99.76%, except for the CNN model, which achieves an accuracy of 98.82%. The proposed model performs best, achieving an accuracy of 99.9%.**

***Index Terms*—Classification, CNN+BiLSTM, DDOS attacks, deep learning, IoT.**

I. INTRODUCTION

The DDoS attacks are a major threat to wireless sensor networks (WSNs), which are networks of small and low-power devices that collect and transmit data from their surrounding environment. In a WSN, DDoS attacks can be launched to overwhelm the network's resources and disrupt its normal operations, leading to service degradation or complete failure. The WSNs are vulnerable to DDoS attacks due to their limited resources and their distributed nature, which makes it difficult to mitigate attacks. In addition, WSNs may be deployed in harsh and unsecured environments, making them more susceptible to attacks.

The IoT devices are interconnected objects that collect and communicate data over internet, and it often deployed in critical infrastructure such as healthcare, transportation, and industrial control systems.

DDoS attacks in WSNs can take various forms, such as flooding attacks, resource depletion attacks, and sinkhole attacks. Flooding attacks involve creating the traffic, while resource depletion attacks target the network's resources, such as memory or battery, by sending malicious data packets. Sinkhole attacks involve redirecting network traffic to a malicious node, which can intercept or modify the data.

To protect WSNs against DDoS attacks, various defense mechanisms have been proposed, such as intrusion detection systems, data aggregation, and collaborative filtering. These mechanisms aim to detect and mitigate attacks by analyzing network traffic, detecting anomalies, and filtering out malicious packets. The DDoS attacks in WSNs pose a significant threat to security and reliability. This require effective defense mechanisms to ensure their proper functioning.

DDoS attacks in IoT can be launched to overwhelm the devices or network infrastructure with a large volume of traffic, leading to service degradation or complete failure. It can take various forms, such as botnet attacks, amplification attacks, and protocol attacks. Botnet attacks involve compromising a huge figure of IoT devices and using them to launch coordinated DDoS attacks. Protocol attacks involve targeting the vulnerabilities in the communication protocols used by IoT devices, such as the MQTT protocol.

To defend IoT devices against DDoS attacks, various defense techniques have been proposed, such as anomaly detection, traffic filtering, and cloud-based defenses. These mechanisms aim to detect and mitigate attacks by analyzing network traffic, filtering out malicious traffic, and diverting traffic to cloud-based services for further analysis. Overall, to improve the reliability devices and networks, and require effective defense mechanisms to ensure their proper functioning.

[1] Department of Mathematics, Al-Zaytoonah University of Jordan, Amman, Jordan (e-mail: i.jebril@zuj.edu.jo)

[2] Department of Electronics and Communication Engineering, SSM Institute of Engineering and Technology, Dindigul, Tamil Nadu, India (e-mail: prem53kumar@gmail.com)

[3] Department of Computer Engineering, University of Technology, Baghad, Iraq (e-mail: ghaida.m.abdulsaheb@uotechnology.edu.iq)

[4] Centre for Block-Chain and Cybersecurity, Department of Computer and Communication Engineering, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India (e-mail: srashokkumar1987@gmail.com)

[5] Department of Electronics and Communication Engineering, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India (e-mail: dhanselvaraj@gmail.com)

[6] Department of Solar, Al-Nahrain Research Center for Renewable Energy, Al-Nahrain University, Jadriya, Baghdad, Iraq (e-mail: usama81818@nahrainuniv.edu.iq)

[7] Al-Kitab University, College of Engineering Technology, Kirkuk, Iraq (e-mail: sameer.algburi@uoalkitab.edu.iq)

DOI: 10.36244/ICJ.2024.1.1

The authors in [1] proposed IDS for WSNs that uses a rule-based approach to defend the DDoS attacks. The system monitors the traffic at each node and sends alerts to the base station when an attack is detected.

Data aggregation involves collecting and processing data at the nodes near to BS which reduces the amount of traffic. This can help to prevent flooding attacks and reduce the impact of DDoS attacks. The authors in [2] proposed a data aggregation scheme for WSNs that uses a fuzzy logic- to identify and filter out malicious traffic.

Collaborative filtering involves nodes in the network exchanging information to discover the malevolent traffic. Nodes can share information about the types of packets received and the sources of the traffic to defend the attacks. The authors in [3] proposed a collaborative filtering scheme for WSNs that uses a reputation-based approach to defend malevolent traffic.

The ML techniques can be used to train the system to classify patterns in network traffic and detect the DDoS attacks [4]. Dynamic thresholding involves setting thresholds for network traffic based on the network conditions and adjusting them dynamically to accommodate changes in the traffic. The authors in [5] proposed a dynamic thresholding approach in WSNs using the moving average and standard deviation of the network traffic.

II. RELATED WORKS

To defend IoT devices against DDoS attacks, various defense techniques have been proposed. The paper [6] proposes various kind of DDoS attacks and the techniques used to launch them. It also provides an extensive review of different mechanisms used to diminish DDoS attacks. The paper classifies DDoS attacks into various categories, in which the authors discuss the attack characteristics, how they work, and the methods used to mitigate them. It also presents a survey of various tools and technologies used for DDoS attack detection and mitigation.

The several defense mechanisms used to counter the DDoS attacks which include filtering techniques such as packet filtering, source address filtering, and rate limiting. They also discuss other approaches such as anomaly detection, traceback, and redirection. It highlights the limitations of existing defense mechanisms and suggesting areas for future research [6].

Table 1 serves as a comprehensive comparison of literature, key parameters such as accuracy, precision, recall, and F1-score, alongside other essential evaluation metrics for each dataset and corresponding model.

The ML based DDoS detection and mitigation system for SDNs is proposed to categorize the normal or malicious traffic. The system is designed to work in SDNs, which allow for centralized network control and management [7]. From the performance of various ML algorithms, RF algorithm performs the best, with an accuracy of 98.2% and low FPR in SDN environment. It is compared with other IDS in which they find that it outperforms in terms of accuracy, detection rate, and FPR. They suggest that their system can be further improved by incorporating other features, such as flow-based features and temporal features. The paper demonstrates the possible of ML algorithms for DDoS mitigation in SDNs.

Paper [8] proposes an anomaly-based approach to identify DDoS attacks using SVM classifiers. The performance of the SVM classifier is compared with DT and KNN in CAIDA using different evaluation metrics. They find that their SVM classifier can effectively detect attacks with a maximum DR and a minimum FPR. The authors also analyze the SVM classifier under different flooding attack scenarios in which it can detect these attacks with high accuracy and low FPR. This approach can be improved by incorporating additional parameters like packet entropy.

The paper [9] proposes a semi-supervised approach for network traffic classification and fine-grained flow identification using hierarchical deep neural networks. The dataset of network traffic is used to train and test DNN models. Dataset includes both labeled and unlabeled traffic data. The FlowPrint technique is to extract fine-grained flow features from network traffic data. FlowPrint is a representation learning technique that captures the underlying structure of network traffic flows. A hierarchical deep neural network architecture that uses the FlowPrint features for network traffic classification. The hierarchical architecture allows for interpretability and explainability of the classification results.

The performance of the approach is evaluated using different evaluation. The proposed [9] results shows that the approach can accurately classify network traffic with high precision and recall. The authors Zhang et. al [9] conclude that their semi-supervised approach using hierarchical deep neural networks and FlowPrint features is an effective technique for network traffic classification and fine-grained flow identification.

The paper [10] highlights the importance of using big data analytics for DDoS detection, as DDoS attacks generate a large amount of traffic data that needs to be analyzed in real-time. This paper provides an overview about techniques and tools used for big data analytics in DDoS detection, including ML, DL, clustering, and rule-based approaches. It discusses the pros and cons of each technique and tool, and provides examples of recent studies that have used these techniques for DDoS detection. The paper also discusses the challenges and issues involved in DDoS detection, such as the high cost of data storage and processing, and the lack of standardization and interoperability among different tools and techniques. But more efficient and scalable big data analytics techniques for DDoS detection are needed, as well as on improving the accuracy and reliability of these techniques.

The paper [11] provides the details of work carried recently in the field of DDoS attack mitigation techniques. The paper provides an outline of DDoS attacks, characteristics of each type of attack, the vulnerabilities they exploit, and their impacts on the target system. It reviews the different DDoS attack mitigation techniques, including network, host and hybrid level defenses. The challenges and issues of each mitigation method, and provides examples of recent studies are discussed that have used for detection techniques. It also discusses the challenges and issues involved in DDoS attack mitigation, the difficulty of distinguishing between normal and illegitimate activity, and cost of implementing mitigation techniques. The more efficient and effective DDoS attack mitigation techniques, as well as on

improving the collaboration and coordination among different stakeholders in the mitigation process.

The paper [21] provides recent research in the field of DDoS attack mitigation techniques. It highlights the different kind of attacks and the vulnerabilities they exploit, and provide an overview of the different defense techniques that can be used to protect against these attacks. The paper also identifies the challenges and issues involved in DDoS attack mitigation and suggest future research directions to address these challenges.

Analyzing the information presented in Table 1, the research demonstrates that the utilization of machine learning-based methods proves successful in identifying attacks. This effectiveness is notably enhanced when these approaches are combined with supplementary techniques such as feature selection and preprocessing. Moreover, the detection of DDoS attacks in wireless sensor networks introduces unique and specific challenges

TABLE I
COMPARISON OF LITERATURE

Dataset/Model	Author Details	Accuracy	Precision	Recall	F1-Score	Other Evaluation Metrics
NSL-KDD	Mohammed et al [7]	0.999	-	-	-	FPR 0.01%, FNR 0%
NSL-KDD	Garcia et al [12]	0.991	0.972	0.992	0.982	DR 99.2%, FAR 0.8%
CICIDS2017/CNN	Hayyolalam et al [11]	0.99	0.997	0.995	0.996	-
CICIDS2017/SAE	Catak et al [13]	0.99	0.9978	0.999	0.9983	-
CICIDS2017/CNN+LSTM	Nguyen et al [14]	0.985	0.96	0.99	0.97	-
DARPA/MLP	Yin et al [15]	0.996	0.997	0.996	0.996	-
DARPA/DBN	Li et al [16]	0.987	0.991	0.982	0.986	-
DARPA/RF	Farukee et al [17]	0.9795	0.981	0.977	0.979	FPR 1.25%, FNR 2.3%
KDD99/CNN	Ye et al [18]	0.9984	0.998	0.998	0.9982	DR 99.86%, FPR 0.01%
KDD99/GRU	Alghazzawi et al [19]	0.999	-	-	-	FPR 0.07%, FNR 0.02%
NSL-KDD/RNN	Aswad et al [20]	0.9828	0.9738	0.981	0.9762	FPR 2.62%, FNR1.88%
NSL-KDD/CNN	Saini et al [21]	0.9933	0.9929	0.993	0.9927	DR 99.37%, FPR 0.02%
CICIDS2017/CNN	Shang et al [22]	0.9991	-	-	-	FPR 0.03%, FNR 0.05%
UNSW-NB15/LSTM	Yousuf et al [23]	0.9967	0.9968	0.997	0.9967	FPR 0.1%, FNR 0.23%
UNSW-NB15/CNN	Alshehadeh et al [24]	0.9936	0.9944	0.994	0.994	DR 99.4%, FPR 0.04%

III. SYSTEM MODEL

A system model for DDoS attack detection using deep learning is shown in figure 1. It typically involves the first step as building a system is to collect data from the network. This data can include network traffic data, packet header data, and flow data. After data collection, it needs to be preprocessed to prepare it for analysis. The initial stage of the process involves extracting relevant features, normalizing the data, and filtering out unnecessary information. In the following phase, the preprocessed data is used to train DL model. The weights are adjusted to reduce the difference between expected and actual outputs. Once training is complete, a separate dataset is employed to evaluate the model and identify potential issues. Subsequently, the model can be utilized for real-time detection of DDoS attacks in a production setting, following successful training and testing.

At outset of the workflow, the input is obtained, either in its raw form or after preprocessing. Feature extraction is carried out, whereby significant characteristics are identified from the input data, including packet size, packet count, and protocol type. These extracted features are then entered into a deep neural network that may comprise a CNN, RNN, or a hybrid of both. The deep neural network processes the input data and assimilates the patterns and correlations between features that signify DDoS attacks.

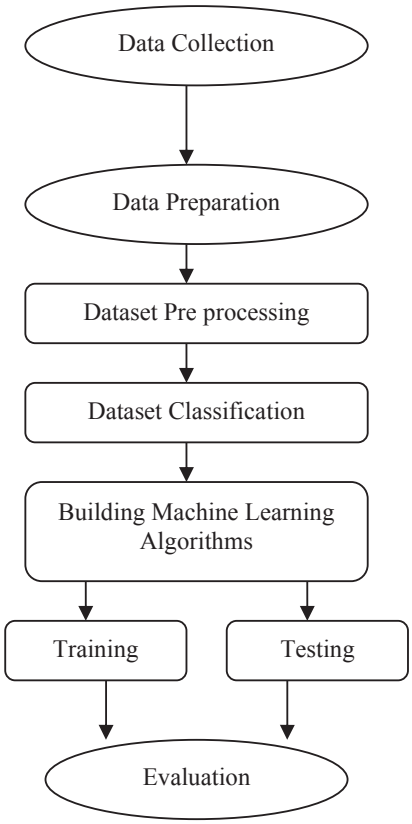


Fig. 1. System model

Finally, the output of DNN is analyzed to decide the data as normal network traffic or a DDoS attack. If a DDoS attack is detected, appropriate mitigation strategies can be employed to prevent it from causing harm to the network.

A. CNN Algorithm

Let X be the input traffic data with shape (batch $_{size,}$ sequence $_{length,}$ input $_{dim}$), where batch $_{size}$ denotes samples count, sequence $_{length}$ is the time sequence length, and input $_{dim}$ denotes number of features in each time step.

The CNN-based deep learning algorithm can be represented as follows:

- Input layer: X with shape (batch $_{size,}$ sequence $_{length,}$ input $_{dim}$)
- Convolutional layer: apply a set of filters with size (filter $_{size,}$ input $_{dim}$) to the input data X, resulting in a set of feature maps.
- Max pooling layer: extract each feature map value to diminish the dimensionality of the feature maps.
- Flatten layer: 2D maps are renewed into a 1D vector.
- Fully connected layer: apply a set of weights to the flattened vector to acquire a hidden value of the input data.
- Output layer: softmax is applied to the hidden representation to obtain predicted class probabilities.

Let W_1, W_2... W_k be the set of convolutional filters, where k is the number of filters. Each filter Wj can be represented as a 2D matrix with size (filter $_{size,}$ input $_{dim}$). The output feature map corresponding to filter Wj can be represented as follows

$$Fj=\max(0,Wj*X+b\ j) \quad (1)$$

where * denotes the convolution operation, bj is the bias term, and max(0, x) is ReLU function. Let V be the weight matrix with shape (num $_{classes,}$ hidden $_{size}$). The output can be represented as follows:

$$H=relu(W*F+b) \quad (2)$$

where W = V^T, b is the bias term, and relu(x) = max(0, x) is ReLU function.

The final predicted class probabilities can be calculated by applying the softmax to the output of the fully connected layer:

$$P=soft\ max(H) \quad (3)$$

where P is a vector of length, representing the predicted class probabilities. The model parameters can be learned by minimizing a suitable loss function using SGD. One approach to train a model for detecting DDoS attacks is to use a labeled dataset of traffic data. In this dataset, each sample is marked as either normal or DDoS traffic.

B. Dataset

The CICIDS2017 [26] is a dataset of network traffic designed for intrusion detection research. It was created by the Canadian Institute for Cybersecurity at the University of New Brunswick in Canada. The dataset includes benign and malicious traffic captured in a real network environment. The malicious traffic includes various kind of attacks such as DoS, DDoS, brute-force attacks, and more. The dataset also includes a variety of network protocols such as HTTP, FTP, TCP, UDP, ICMP, etc.

TABLE II
CICIDS 2017 DATASET DESCRIPTION

Attack Type	Number of Instances	Average Packet Size	Average Flow Duration	Average Fwd Segments	Average Bwd Segments	Max Fwd Packet Length	Max Bwd Packet Length	Fwd Packet Length Std	Bwd Packet Length Std
DoS Hulk	231073	1392.5	46353.5	7.28	3.09	1472	1460	187.6	371.5
DoS GoldenEye	102157	1503.8	36753.1	13.54	3.04	1472	1460	212.1	365.9
DoS slowloris	11586	746.16	40393.5	2.17	2.62	590	589	136.2	142.7
DoS Slowhttptest	549	3104.4	1156624	10.49	7.52	1480	1476	89.15	172.3
Heartbleed	1000	707.6	60483.9	9.62	10.3	177	202	159.7	44.52
Infiltration	36	1352.4	45564.0	22.39	11.3	1472	1472	0.00	0.00
Bot	196907	1468.8	12686.0	10.22	4.90	1472	1472	39.51	237.8
PortScan	158930	882.1	405.44	6.47	4.52	1472	1408	300.6	451.4
Web Attack - XSS	652	1718.2	15838.0	8.28	6.18	1472	1472	206.6	309.0

C. CNN+BiLSTM based deep learning algorithm

CNN+BiLSTM is a DL architecture that combines CNN and BiLSTM in which the CNN is responsible for feature extraction from input data. It consists of multiple filters that slide over the input data and extract local features. Then the output is fed to the BiLSTM layer which is a type of RNN that has the ability to form sequential data. This layer takes the output of the CNN layer and processes it in both forward and backward directions. This allows capturing both past and future contexts of the data. The output of the BiLSTM layer to a fixed number of classes, which in the case of DDoS attack detection corresponds to normal traffic and DDoS traffic. The output is then passed through a softmax function to calculate the final prediction probabilities for each class. The CNN+BiLSTM architecture can be trained using backpropagation and weights are updated iteratively during the training to optimize the model performance in shown in Fig 2.

Let x be a wireless sensor network traffic sequence with m features and n time steps. Let y be the corresponding binary label sequence, where 0 represents non-attack traffic and 1 represents DDoS attack traffic. The mathematical model for identifying DDoS attacks in WSN using CNN+BiLSTM based deep learning algorithm can be represented using the following equations.

Apply a 1D CNN layer with k filters of size f on the input x to extract k feature maps of n-f+1 size. Use ReLU activation function and apply max pooling operation on each feature map to reduce the dimensionality by a factor of p. Let the input features be represented by X ∈ R^(n x m), where n is samples count and m is feature count.

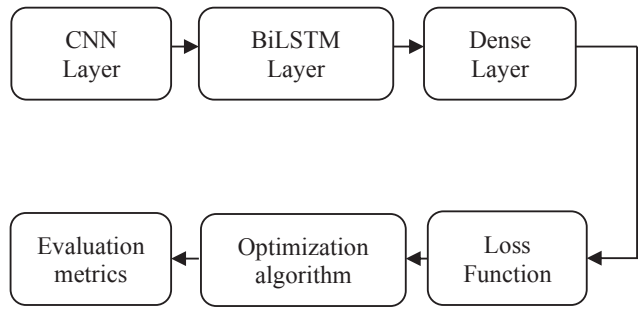


Fig 2. Proposed method workflow

The CNN can be represented using the following equations

$$Yi=\max_{pool}(W*Xi+b) \quad (4)$$

Here, Yi denotes the output feature map of the ith filter, W represents the weight of the ith filter, Xi is the input feature map, b stands for the bias term, and $\max_{pool}$ signifies the max-pooling operation.

The BiLSTM can be represented using the following equations-

$$f_t = \sigma(W_f *[h_{t-1}, x_t]+b_f \quad (5)$$

$$i_t = \sigma(W_i *[ht-1\}, x_t]+b_i \quad (6)$$

$$o_t = \sigma(W_o *[ht-1\}, x_t]+b_o \quad (7)$$

$$g_t = \tanh(W_c *[ht-1\}, x_t]+b_c \quad (8)$$

$$c_t = f_t * c_{t-1} + i_t * g_t \quad (9)$$

$$h_t = o_t * \tanh(c_t) \quad (10)$$

Apply a fully connected dense layer on the BiLSTM layer with o output units and sigmoid to generate o binary predictions.

Let the output layer be represented by $\hat{Y} \in R^n$, where n is the number of samples. The output of the network can be represented using the following equation

$$\hat{Y}=\text{softmax}(W_Y*hn+b_y) \quad (11)$$

where hn is the output of BiLSTM, W_Y is output layer weight, and b_y is the bias term. The DDoS detection can be done by comparing $\hat{Y}$ with Y. If $\hat{Y}$ is significantly different from the actual output Y, then it can be classified as a DDoS attack.

The mathematical model can be trained using backpropagation with cross-entropy loss as the objective function. The training process involves minimizing the objective function with respect to the model parameters. This can be done using gradient descent or any of its variants.

D. Algorithm

1 Load and preprocess the dataset.
2 Convert the text to numerical vectors using word embeddings. Let the input is X = (x1, x2, ..., xn), where each xi is a d-dimensional word embedding vector.
3 Split the data into training and testing sets.
4 Identify the CNN layer with filters of varying sizes. Let the filters have sizes f1, f2, ..., fk, where each fk is a vector of length h. Let the filters count be denoted by m. For each filter fk, convolve it with X to acquire feature maps: fk_i = relu((Wfk * X[i:i+h-1] + bfk)) where Wfk is the weight matrix and bfk is the bias vector associated with filter fk, and relu is the ReLU function.
5 Apply max pooling on the feature maps to get a fixed-length output. For each feature map fk_i, apply max pooling to obtain the maximum value: gk = max(fk_1, fk_2, ..., fk_n-h+1)
6 Concatenate the output from the max pooling layer with BiLSTM layer. Let the concatenated output is Z where each zi is a scalar value obtained by concatenating gk with BiLSTM layer.
7 Define the BiLSTM layer with a certain number of hidden units. Let the hidden size of the BiLSTM layer be denoted by p. Apply a BiLSTM layer to the input sequence X to obtain the output sequence Y.
8 Concatenate the output from the BiLSTM layer with the output from the max pooling layer. Let the concatenated output be denoted by Z.
9 Add a fully connected layer with a softmax activation function for classification. Let the number of classes be denoted by C. Apply a Z to obtain the output vector o, where o = softmax(Wo * Z + bo), and Wo is the weight matrix and bo is the bias vector associated with the fully connected layer.
10 Train the model on the training set using cross-entropy loss. Let the training set be denoted by D and each yi is a one-hot encoded label vector.
11 Calculate the model on the testing set using accuracy or other evaluation metrics.
12 Repeat steps 4-11 with different hyperparameters (e.g., number of filters, filter sizes, number of hidden units) to find the best model.

Some simulation parameters that shown in Table 3 could be used for DDoS detection using a CNN+BiLSTM algorithm.

The first step in setting up a simulation experiment is to choose a dataset. In order to train the CNN+BiLSTM model, a set of input features must be selected. These features could include information such as the IP addresses, the protocol used, and time stamp of each network packet. The hyperparameters of the CNN+BiLSTM model must be defined. These include the number and size of filters in CNN layer, the hidden units in BiLSTM layer, and learning rate used during training. In order to train, a choice of parameters need to be specified, and optimizer to be utilized. Once the model is trained, its performance can be assessed using a separate testing set, with evaluation metrics for classifying a network flow as normal or an attack being among the testing parameters. The Longer simulation duration may be required to achieve higher accuracy levels. Finally, the specifications used to run the simulation should be taken into account. The specifications can include the CPU and GPU, the memory, and the disk space required to store the dataset and model.

IV. RESULTS AND DISCUSSIONS

TABLE III
SIMULATION PARAMETERS

Parameter	Value
Dataset	DARPA 1998
Pre processing	One-hot encoding, normalization
Training-Validation split	70-30
Model Architecture	CNN+BiLSTM
Number of layers	CNN:2; BiLSTM:128
Number of filters	CNN:64, 128; BiLSTM:128
Filter Size	CNN: 3x3, 5x5; BiLSTM: N/A
Dropout rate	0.5
Learning rate	0.001
Batch Size	128
Number of epochs	50
Loss function	Binary cross-entropy
Evaluation metric	Accuracy, precision, recall, F1-score
Hardware	NVIDIA GeForce GTX1080 Ti
Software	Python 3.7, Tensor Flow 2.3.1

The simulation parameters will help in conducting experiments to test the performance of the proposed model for DDoS detection. The aim should be to optimize the hyperparameters and training parameters to achieve the results.
The parameters mentioned above can be customized according to the unique attributes and needs of both the dataset and the model. The confusion matrix presents the counts of TP, FP, FN and TN. Meanwhile, the ROC curve AUC score is utilized to gauge probability thresholds.

Accuracy: The proportion of correctly classified samples out of the total number of samples.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (12)$$

Precision: It measures the ability of the model to correctly identify positive samples.

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (13)$$

Recall: It measures the ability of the model to identify all positive samples.

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (14)$$

F1-score: It provides a balance between precision and recall.

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \quad (15)$$

TABLE IV
PERFORMANCE COMPARISON OF THE PROPOSED METHOD

Model	Accuracy	Error	Precision	Recall	f-1score
RNN [15]	0.996	0.018	0.987	0.983	0.985
BiLSTM [19]	0.998	0.014	0.989	0.992	0.991
LSTM [12]	0.997	0.011	0.911	0.935	0.91
CNN [20]	0.988	0.036	0.977	0.983	0.98
KNN [25]	0.967	0.013	0.976	0.982	0.982
Proposed	0.999	0.012	0.998	0.999	0.997

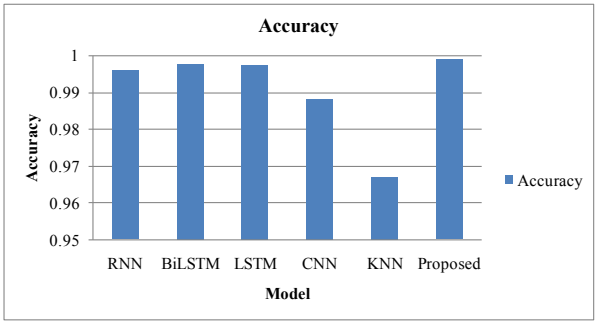


Fig 3. Model Vs Accuracy

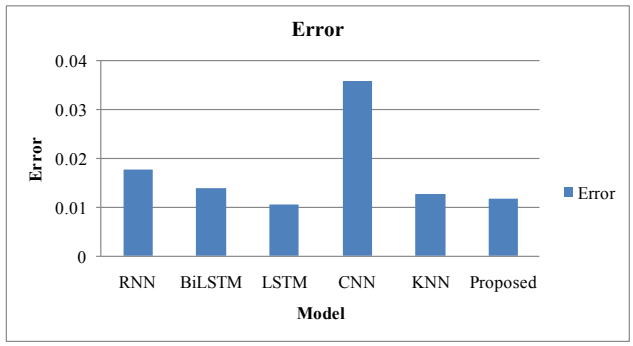


Fig 4. Model Vs Error

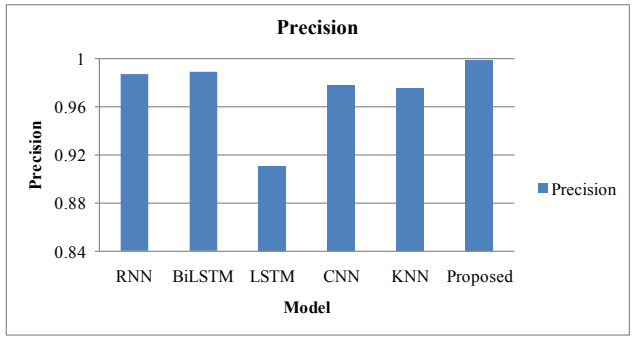


Fig 5. Model Vs Precision

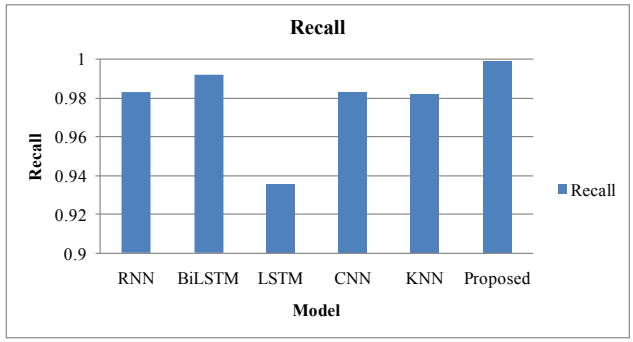


Fig 6. Model Vs Recall

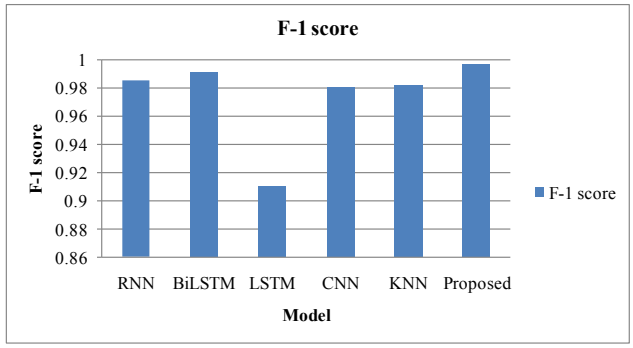


Fig 7. Model Vs F-1 score

Performance of the detection system is evaluated by performance metrics. Figures 3-7 show the plot of detection parameters for various classification models. In figure 3, the

RNN and Proposed models exhibited superior accuracy at 99.9%, closely followed by the BiLSTM model at 99.8%, while the LSTM and CNN models achieved slightly lower accuracies at 99.7% and 98.8% respectively. Concerning the error rate in figure 4, the Proposed, RNN, and BiLSTM models maintained impressively low values at 1.1-1.8%, signifying their robustness in the classification task. In terms of precision the figure 5 shows, the BiLSTM and Proposed models performed exceptionally well at 99%, closely followed by the RNN and CNN models, which achieved high precision values of 98%. In figure 6, the recall values were consistent across most models, with the Proposed, RNN, BiLSTM, and CNN models showcasing strong recall rates at 98-99%. Similarly, the figure 6 shows the F1-score reflected the models' overall performance, with the BiLSTM, Proposed, and RNN models demonstrating the highest scores at 98-99%, followed closely by the CNN and KNN models, which achieved competitive F1-scores at 98%. According to the result in Table 4, the accuracy of the proposed method is 99.9 % which can be improved by increasing the training samples. Error is 1.16%, precision is 99.8%, recall is 99.9% an F-1 score
is 99.7%.

TABLE V
CONFUSION MATRIX FOR THE PROPOSED METHOD

Actual/ Predicted	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Class 1	-180000	9	12	420	55	6
Class 2	30	42000	0	2	0	0
Class 3	3	0	3300	21	2	1
Class 4	80	0	0	76000	0	0
Class 5	16	0	1	1	1800	13
Class 6	9	0	0	0	18	2000

Table 5 displays a confusion matrix, which serves as an assessment tool for a classification model's effectiveness. It operates by contrasting the forecasted labels with the genuine labels of a test dataset. The composition of the confusion matrix for the CNN+BiLSTM approach will be influenced by the specific task at hand and the quantity of classifications present in the data collection.

V. CONCLUSIONS

DDoS attacks are a significant threat and they are difficult to detect because attackers use spoofing technology. Traditional detection systems have been ineffective against historically potent botnets like Mirai and Bashlite. IoT networks, in particular, are at risk of cyberattacks and require strong protective measures. The proposed model achieves an average accuracy of 99.76% in identifying DDoS attacks, surpassing the performance of other tested models. However, the authors caution against overlooking the accuracy of the other three classifiers, which achieve an average accuracy of 99.16%. The research also examines the weaknesses of IoT network construction and identifies potential reasons for its susceptibility to DDoS attacks. Furthermore, the article highlights gaps in prior research on DDoS attacks. The promising results of the proposed model demonstrate its potential to effectively secure IoT network systems in real-world scenarios. Nonetheless, the study's primary limitation is the unavailability of a realistic testing platform, which raises questions about testing reliability. Future research will concentrate on identifying the bottlenecks of IoT network systems concerning their susceptibility to DDoS attacks.

Abbreviations

MQTT : Message Queuing Telemetry Transport
BS : Bootstrap
SDN : Software-Defined Networking
ML : Machine Learning
RF : Random Forest
FPR : False Positive Rate
SVM : Support Vector Machine
CAIDA : Cooperative Association for Internet Data Analysis
KNN : K-Nearest Neighbors
CNN : Convolutional Neural Network
RNN : Recurrent Neural Network
SGD : Stochastic Gradient Descent
DoS : Denial of service attacks
DDoS : Distributed denial of service attacks
IoT : Internet of Things
BiLSTM : Bidirectional long short-term memory

Conflict of interest: The authors have no conflicts of interest to declare.

Data availability statement: The dataset used for this research is available online and has a proper citation within the article's contents.

REFERENCES

[1] Jianjian, D., Yang, T., & Feiyue, Y. (2018). A novel intrusion detection system based on IABRBFSVM for wireless sensor networks. Procedia computer science, 131, 1113–1121. DOI: 10.1016/j.procs.2018.04.275

[2] Hosseinzadeh, M., Yoo, J., Ali, S., Lansky, J., Mildeova, S., Yousefpoor, M. S., ... & Tightiz, L. (2023). A fuzzy logic-based secure hierarchical routing scheme using firefly algorithm in Internet of Things for healthcare. Scientific Reports, 13(1), 11058. DOI: 10.1038/s41598-023-38203-9

[3] Giotis, K., Apostolaki, M., & Maglaris, V. (2016, April). A reputation-based collaborative schema for the mitigation of distributed attacks in SDN domains. In NOMS 2016-2016 IEEE/IFIP network operations and management symposium (pp. 495–501). IEEE. DOI: 10.1109/NOMS.2016.7502849

[4] Premkumar, M., Ashokkumar, S. R., Jeevanantham, V., Mohanbabu, G., & AnuPallavi, S. (2023). Scalable and energy efficient cluster based anomaly detection against denial of service attacks in wireless sensor networks. Wireless Personal Communications, 129(4), 2669–2691. DOI: 10.1007/s11277-023-10252-3

[5] David, J., & Thomas, C. (2019). Efficient DDoS flood attack detection using dynamic thresholding on flow-based network traffic. Computers & Security, 82, 284–295. DOI: 10.1016/j.cose.2019.01.002

[6] Sahoo, K. S., Tripathy, B. K., Naik, K., Ramasubbareddy, S., Balusamy, B., Khari, M., & Burgos, D. (2020). An evolutionary SVM model for DDOS attack detection in software defined networks. IEEE access, 8, 132 502–132 513. DOI: 10.1109/ACCESS.2020.3009733

[7] Mohammed, S. S., Hussain, R., Senko, O., Bimaganbetov, B., Lee, J., Hussain, F., ... & Bhuiyan, M. Z. A. (2018, October). A new machine learning-based collaborative DDoS mitigation mechanism in software-defined network. In 2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (pp. 1–8). IEEE. DOI: 10.1109/WiMOB.2018.8589104

[8] Rawashdeh, A., Alkasassbeh, M., & Al-Hawawreh, M. (2018). An anomaly-based approach for DDoS attack detection in cloud environment. International Journal of Computer Applications in Technology, 57(4), 312–324. DOI: 10.1504/IJCAT.2018.093533

[9] Zhang, H., Yu, L., Xiao, X., Li, Q., Mercaldo, F., Luo, X., & Liu, Q. (2023). TFE-GNN: A Temporal Fusion Encoder Using Graph Neural Networks for Fine-grained Encrypted Traffic Classification. In Proceedings of the ACM Web Conference 2023 (pp. 2066–2075). DOI: 10.1145/3543507.3583227

[10] Premkumar, M., Ashokkumar, S. R., Mohanbabu, G., Jeevanantham, V., & Jayakumar, S. (2022). Security behavior analysis in web of things smart environments using deep belief networks. International Journal of Intelligent Networks, 3, 181–187. DOI: 10.1016/j.ijin.2022.10.003

[11] Hayyolalam, V., & Kazem, A. A. P. (2018). A systematic literature review on QoS-aware service composition and selection in cloud environment. Journal of Network and Computer Applications, 110, 52–74. DOI: 10.1016/j.jnca.2018.03.003

[12] Garcia, J. F. C., & Blandon, G. E. T. (2022). A deep learning-based intrusion detection and preventation system for detecting and preventing denial-of-service attacks. IEEE Access, 10, 83 043–83 060. DOI: 10.1109/ACCESS.2022.3196642

[13] Catak, F. O., & Mustacoglu, A. F. (2019). Distributed denial of service attack detection using autoencoder and deep neural networks. Journal of Intelligent & Fuzzy Systems, 37(3), 3969–3979. DOI: 10.3233/JIFS-190159

[14] Nguyen, X. H., & Le, K. H. (2023). Robust detection of unknown DoS/DDoS attacks in IoT networks using a hybrid learning model. Internet of Things, 23, 100851. DOI: 10.1016/j.iot.2023.100851

[15] Yin, C., Zhu, Y., Fei, J., & He, X. (2017).A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access, 5, 21 954–21 961. DOI: 10.1109/ACCESS.2017.2762418

[16] Li, Y., Liu, B., Zhai, S., & Chen, M. (2019, April). DDoS attack detection method based on feature extraction of deep belief network. In IOP Conference Series: Earth and Environmental Science (Vol. 252, No. 3, p. 032013). IOP Publishing. DOI: 10.1088/1755-1315/252/3/032013

[17] Farukee, M. B., Shabit, M. Z., Haque, M. R., & Sattar, A. S. (2021). DDos attack detection in iot networks using deep learning models combined with random forest as feature selector. In Advances in Cyber Security: Second International Conference, ACeS 2020, Penang, Malaysia, December 8-9, 2020, Revised Selected Papers 2 (pp. 118–134). Springer Singapore. DOI: 10.1007/978-981-33-6835-4_8

[18] Ye, J., Cheng, X., Zhu, J., Feng, L., & Song, L. (2018). A DDoS attack detection method based on SVM in software defined network. Security and Communication Networks, 2018. DOI: 10.1155/2018/9804061

[19] Alghazzawi, D., Bamasag, O., Ullah, H., & Asghar, M. Z. (2021). Efficient detection of DDoS attacks using a hybrid deep learning model with improved feature selection. Applied Sciences, 11(24), 11634. DOI: 10.3390/app112411634

[20] Aswad, F. M., Ahmed, A. M. S., Alhammadi, N. A. M., Khalaf, B. A., & Mostafa, S. A. (2023). Deep learning in distributed denial-of-service attacks detection method for Internet of Things networks. Journal of Intelligent Systems, 32(1), 20220155. DOI: 10.1515/jisys-2022-0155

[21] Saini, P. S., Behal, S., & Bhatia, S. (2020, March). Detection of DDoS attacks using machine learning algorithms. In 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 16-21). IEEE. DOI: 10.23919/INDIACom49435.2020.9083716

[22] Shang, Y., Yang, S., & Wang, W. (2018, June). Botnet detection with hybrid analysis on flow based and graph based features of network traffic. In International Conference on Cloud Computing and Security (pp. 612–621). DOI: 10.1007/978-3-030-00009-7_55

[23] Yousuf, O., & Mir, R. N. (2022). DDoS attack detection in Internet of Things using recurrent neural network. Computers and Electrical Engineering, 101, 108034. DOI: 10.1016/j.compeleceng.2022.108034

[24] Alshehadeh, A. R., & Al-Khawaja, H. A. (2022). Financial Technology as a Basis for Financial Inclusion and its Impact on Profitability: Evidence from Commercial Banks. Int. J. Advance Soft Compu. Appl, 14(2). DOI: 10.15849/IJASCA.220720.09

[25] Alahmadi, A. A., Aljabri, M., Alhaidari, F., Alharthi, D. J., Rayani, G. E., Marghalani, L. A., ... & Bajandouh, S. A. (2023). DDoS Attack Detection in IoT-Based Networks Using Machine Learning Models: A Survey and Research Directions. Electronics, 12(14), 3103. DOI: 10.3390/electronics12143103

[26] UNB 2017 Intrusion Detection Evaluation Dataset by canadian institute for cybersecurity URL. https://www.unb.ca/cic/datasets/ids-2017.html

Iqbal Jebril comleted his Ph.D. (Mathematical Analysis) Universiti Kebangsaan National University of Malaysia. He is working as professor and Head of the Mathematics Department, Faculty of Science and Information Technology, Al-Zaytoonah University of Jordan, P.O. Box 130 Amman 11733 Jordan

M. Premkumar received the Ph.D. degree in Information and Communication Engineering from Anna University Chennai in 2022. His research interests include wireless Ad hoc networks, security and key management of wireless networks, wireless sensor networks.

Ghaida Muttshar Abdulsahib working as a faculty in computer engineering department, University of Technology, Iraq. She got a lot of awards in computer engineering area. And now Ghaida interested in network and communication area. She also published articles in reputed indexed journals like SCI, WoS and SCOPUS.

S. R. Ashokkumar received the Ph.D. degree in Information and Communication Engineering from Anna University Chennai in 2021. His research interests include Network security and Signal and image processing, wireless sensor networks.

S. Dhanasekaran received his BE degree in Electronics and Communication Engineering in 2008 from Sri Balaji Chockalingam Engineering College, Arani, Tamil Nadu, India. He completed his ME in Communication Systems in 2010 from PSG College of Technology, Coimbatore, Tamil Nadu, and India. He completed his PhD in the year 2022 from Anna University Chennai in the area of Communication systems, MIMO, OFDM, etc.,
He is currently working as Assistant Professor in the department of Electronics and Communication Engineering, Sri Eshwar College of Engineering, Coimbatore. He has around 14 years of teaching experience. He is a life time member of ISTE.

Osamah Ibrahim Khalaf is a Senior Engineering and Telecommunications Lecturer in Al-Nahrain University / College of Information Engineering. He has had many published articles indexed in (ISI/Thomson Reuters/SCI) and has also participated and presented at numerous international conferences. In 2004, he got his B.Sc. in software engineering field from Al-Rafidain University College in Iraq. Then in 2007, he got his M. Sc. in computer engineering field from Belarussian National Technical University. After that, he got his Ph.D. in 2017 in the field of computer networks from faculty of computer systems and software engineering, University Malaysia.

Sameer Saadoon Algburi, got the PhD in 2007 from the University of Technology in Iraq in electrical power systems. A Professor at Al-Kitab University in Kirkuk in teaching undergraduate students and supervised graduate students. interested in power systems, especially renewable energies and climate change and published many papers in local and international conferences and journals. Since 2014, Visiting Researcher at two Lund University centers, the Geographic Information Center GIS and the Center for Middle Eastern Studies CMES. Since 2019, Director and founder of the Swedish-Iraqi Studies Network SISNET and managed it with the help of two boards of management and consulting from Iraq and Sweden. UN/ UNIDO, China, Small Hydropower Systems, Author and UN/ UNIDO/CTCN member. Also the Managing Editor at Al-Kitab Journal for Pure

Secure post-processing for non-ideal photon arrival time based quantum random number generator

Balázs Solymos, and László Bacsárdi, *Member, IEEE*

***Abstract*—Utilizing the inherently unpredictable nature of quantum mechanics, quantum random number generators (QRNGs) can provide randomness for applications where quality entropy (like in the case of cryptography) is essential. We present a post-processing scheme utilizing min-entropy estimation and hashing for optical QRNGs based on measuring individual photon arrival times. Our method allows for the handling of possible errors due to non-ideal components or even a potential attacker, given some basic assumptions to reliably produce a safe, good quality, uniformly distributed bitstream as output. We validate our results with an intentionally non-ideal measurement setup to show robustness, while also statistically testing our final output with four popular statistical test suites.**

***Index Terms*—quantum communication, quantum random number generation, statistical testing, hashing, entropy.**

I. INTRODUCTION

QUALITY randomness is used as a resource in a wide variety of applications, from numerical simulations to classical and even some quantum cryptography protocols [1], [2], that rely on entropy sources as fundamental building blocks. Due to this reliance, using a lower-quality source presents the danger of compromising the correctness of the schemes utilizing its output [3], especially in the field of cryptography. While pseudorandom number generators can provide fast and cheap random-like output, due to their inherently deterministic nature (use of complex but deterministic algorithms) are often considered a liability [4].

Quantum random number generators (QRNGs) [5] aim to harness the unpredictability of quantum mechanical processes. They have the advantage of relying on phenomena proved to be random by the laws of physics, thus giving a solid guarantee of quality in theory. Practical realization of these devices is a formidable engineering challenge, however, as the various imperfections and error sources potentially influencing the measurement have to also be handled. Due to advancements in quantum optics, architectures based on measuring various random properties of light, like path superposition of a photon [6], [7], photon number [8]–[10] or arrival time statistics [11]–[14], amplified spontaneous emission [15], [16], vacuum or phase fluctuations [17], [18], or even Raman scattering [19] have been proposed, while there are already some commercially available products on the market [20] and new chip-based solutions [21]–[23] are also emerging.

We use a simple generator architecture based on photon arrival times, with a continuously running clock, which is different from the ideal case of using a restartable clock but permits simpler and cheaper hardware. Even in the ideal case, the measurement statistics (exponential) differ from the expected uniformly distributed output, so a post-processing step is necessary. For this, various methods can be used, from simply comparing records [11] to utilizing complex algorithms based on entropy estimation and privacy amplification [24], [25]. In this work, we present a post-processing framework that can incorporate possible errors due to non-ideal components or even a potential attacker given some basic assumptions to reliably produce safe, quality output based on universal hashing and entropy estimation. Our framework potentially also enables us to relax minimum hardware requirements at the cost of output speed and the need for robust post-processing. This may prove especially useful for cases, where hardware options are limited either due to physical constraints (e.g., integrated optics), or any other reason (e. g. low budget to spend on quality components.).

II. CONCEPT

A. Generator architecture

Our generator is based on time differences between photon arrival times of an attenuated laser source, counting the number of elapsed clock cycles between detections. Ideally, this statistic follows a geometric distribution, governed by the underlying exponential distribution of the physical process of photon emission, which is different from the expected uniform output, already mandating the need for post-processing. Additionally, effects from the concrete physical realizations and non-idealities further distort the measured statistic, making the generation of guaranteed quality output non-trivial.

In the following sections, we rely heavily on the concept of $H_\infty(D)$ min-entropy to characterize the safely extractable randomness from our measurement results:

$$H_\infty(D) = \min_n(-\log_2 p_n) = -\log_2 \max_n p_n, \qquad (1)$$

where $\max_n p_n = p_{\max}$ is the probability of the most likely measurement result. It is important to note, that attempting to create a uniform output corresponding to more entropy than contained in the measurement results, yields poor quality or even insecure output, while underestimating extractable entropy may only lead to suboptimal output rate, but preserves

The authors are with the Department of Networked Systems and Services, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, Hungary.

(E-mail: solymosb@hit.bme.hu, bacsardi@hit.bme.hu)

B. Solymos thanks the support of the UNKP-22-3-II-BME-238 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development, and Innovation Fund. L. Bacsárdi thanks the support of the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (Grant No. BO/00118/20).

DOI: 10.36244/ICJ.2024.1.2

quality. Our goal is, therefore, to give a safe lower bound for min-entropy (upper bound for $p_{\max}$), which holds even in non-ideal conditions.

B. Hashing for post-processing

Universal hash functions can be used for post-processing [25], since, with them, we can construct a $(k_e, \epsilon, n_e, d_e, m_e)$ extractor, so that

$$\text{Ext} : \{0,1\}^{n_e} \times \{0,1\}^{d_e} \mapsto \{0,1\}^{m_e} \tag{2}$$

for every probability distribution D on $\{0,1\}^{n_e}$ with at least $H_\infty(D) \geq k_e$ min-entropy, the probability distribution $\text{Ext}(D, U_{d_e})$ is ϵ-close statistically to the uniform distribution on $\{0,1\}^{m_e}$. This means, that with the help of a random U_{d_e} seed of d_e bits, we can take a longer, but only partially random stream of n_e bits and create a smaller, but close to uniform output. The reusability of this seed is a crucial requirement for extractors (Since the randomness needed for continual reseeding would exceed the randomness extracted.). Fortunately, universal hash functions are proven to be strong extractors by the Leftover Hash Lemma [26], stating this reusability.

From these, we chose the popular Toeplitz hash to serve as a basis for our randomness extraction method. An $m_e + n_e - 1$ bit long random seed is needed for initialization to construct a random Toeplitz matrix of $n_e \times m_e$. Then during operation, we split our data into n_e long input vectors, and one-by-one multiply them with the initialized random Toeplitz matrix to get m_e long output vectors, which we then assemble into an output bitstream. The k_e extractable entropy contained in the n_e long input defines the possible values for m_e according to

$$m_e = k_e + 2\log\epsilon. \tag{3}$$

Given a target ϵ, from $H_\infty(D)$ and n_e all the other parameters can be derived, so our goal is to present a framework for safely determining these.

C. Error sources

1) Additive noise: Coherent light sources based on stimulated emission like lasers are generally assumed to be Poissonian photon sources [27] (meaning exponentially distributed arrival time differences between photon emissions), due to the underlying physical working principle. In reality, photons from spontaneous emission (e.g. thermal effects) may also have a small superpoissonian contribution to the output distribution of the source, though this effect has been shown to be vanishing with increasing attenuation [28]. Still, we can model this unwanted process by introducing additional photon counts mixed with the ideal exponential statistics. This idea can be extended to include any additive error sources, like afterpulsing effects, or even a potential attacker.

With this in mind, we assume that our count statistic is made up of photons coming from an underlying true exponential source with C_{exp} number of independent counts for a given time period, responsible for the majority of the total counts, and a smaller at most C_{noise} amount of counts coming from noise processes or even potential attackers. This essentially means a limit on noise/attacker intensity, while also assuming an attacker is not capable of influencing photons from the trusted exponential photon source.[1]

The goal is to give a worst-case lower estimate for min-entropy. For this, we propose that there exists an interval series in the joint exponential and noise statistic for which

$$\begin{aligned}\underline{H_\infty(D)} &= -(C_{\text{exp}} - C_{\text{noise}})\log_2 p'_{\max} \\ &= -(C_{\text{exp}} - C_{\text{noise}})\log_2\left(\frac{p_{\max}C_{\text{exp}} + C_{\text{noise}}}{C_{\text{exp}} - C_{\text{noise}}}\right)\end{aligned} \tag{4}$$

is a lower bound in min-entropy[2].

Let $S_0, S_1, ..., S_i, ..., S_{C_{\text{exp}}-1}$ be the arrival times of photons from our ideal source, with $D_0, D_1, ..., D_i, ..., D_{C_{\text{exp}}-1}$ cycle long measured intervals between them and note arrival times of noise/attacker photons with $N_0, N_1, ..., N_i, ..., N_{C_{\text{noise}}-1}$ Since we allow the noise to have any distribution and use any strategy, even allowing dependence on other counts, we do not consider the min-entropy contribution of intervals where noise counts are involved (see Fig. 1), only the entropy contribution of intervals from the sub-series $\{D_j\}_{j\in J}$, where $J = \{j \mid \nexists N_i : S_{j-1} < N_i < S_j\}$ (the intervals not affected by noise counts)

We also have to consider the possible distorting effect the noise can have on the overall distribution and, therefore, the distribution of our remaining considered series. From the point of min-entropy, this means the possible change of the original $p_{\max}$ to a new $p'_{\max}$ (change in the probability of the most frequent result). Assuming a worst-case scenario, additional noise counts can have the following effects:

- Noise is positioned so that all original counts corresponding to $p_{\max}$ (originally most likely outcome of the ideal source) are kept in the considered sub-series.
- Interval statistics from noise counts further increase $p_{\max}$ for at most an additional C_{noise} new counts contributing to the measurement result corresponding to $p_{\max}$.

While the actual physical feasibility of these worst-case effects may at times be questionable, we still consider them, to give a safe lower estimate guaranteed to hold for any possibility. This way, Eq. (4) is a lower bound in min-entropy for the considered sub-series, therefore it is a valid lower bound for the whole series too.

2) Effect of continuous clock and dead time: We can model the effects of using a continuous clock and dead time of the detector on the min-entropy as previously presented in more detail in [29]. Assuming photons arrive according to a Poisson process with rate λ, in the continuous clock case we can divide time into consecutive τ long grids, where τ is the length of a clock cycle, S_i the time of the ith arrival, $T_i = S_i - S_{i-1}$ the ith inter-arrival and γ_i the time between S_i and its preceding τ grid ($0 \leq \gamma_i < \tau$). We measure D_i, the number of τ grids (clock cycles) between S_{i-1} and S_i. An explanatory example case of the model can be seen in Fig. 2. For the distribution

[1] This also means, that, quantum operations, like entangling additional photons with photons from the trusted source, are not allowed either.

[2] For larger values of C_{noise}, where $p_{\max}C_{\text{exp}} + C_{\text{noise}} > C_{\text{exp}} - C_{\text{noise}}$ Eq. (4) can result in negative output. $\underline{H_\infty(D)}$ should be considered 0 in these cases.

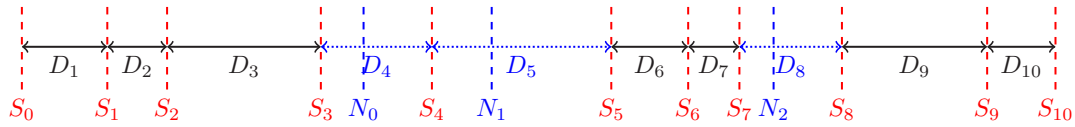


Fig. 1. Example of handling additive noise. Times noted with S_i are counts from the assumed underlying ideal distribution, with D_i intervals between them. After introducing N_i noise counts, we only consider the entropy contribution of intervals not affected by the noise, which are $\{D_1, D_2, D_3, D_6, D_7, D_9, D_{10}\}$ in this pictured example case.

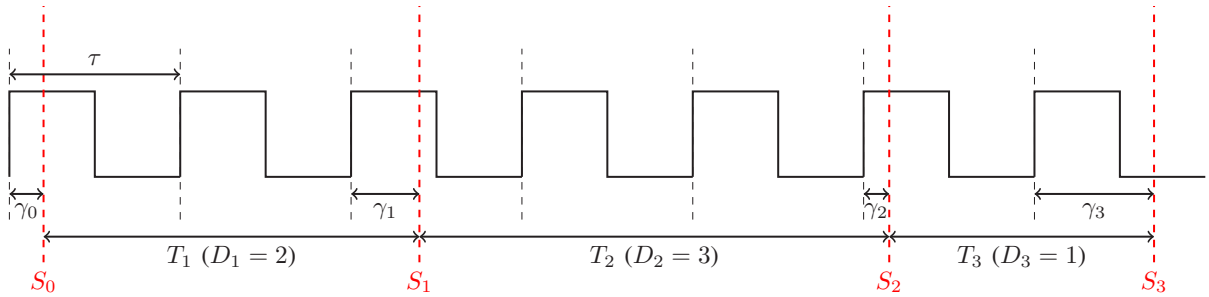


Fig. 2. Continuous clock example from [29]. Photons arrive at times S_i and are counted by a τ resolution running clock. T_i notes the true exponential time between detections and D_i the associated number of counts (our measurement result), while γ_i notes the varying internal starting phases of the counting process.

of D without dead time, we can write

$$\begin{aligned} p_n &= \Pr(D = n \mid \gamma = y) \\ &= \begin{cases} \Pr(y + T < \tau) & \text{if } n = 0, \\ \Pr(n\tau \le y + T < (n+1)\tau) & \text{if } n > 0, \end{cases} \\ &= \begin{cases} 1 - \mathrm{e}^{-\lambda(\tau - y)} & \text{if } n = 0, \\ \left(1 - \mathrm{e}^{-\lambda\tau}\right) \mathrm{e}^{-\lambda(n\tau - y)} & \text{if } n > 0. \end{cases} \end{aligned} \tag{5}$$

To calculate worst-case min-entropy we then maximize p_n:

$$\begin{aligned} &\max_{n,y}(\Pr(D = n \mid \gamma = y)) \\ &\quad = \max_{n,y} \begin{cases} 1 - \mathrm{e}^{-\lambda(\tau - y)} & \text{if } n = 0, \\ \mathrm{e}^{\lambda y}\left(1 - \mathrm{e}^{-\lambda\tau}\right) \mathrm{e}^{-\lambda n\tau} & \text{if } n > 0, \end{cases} \\ &\quad = \max_{n,y} \begin{cases} \left(1 - \mathrm{e}^{-\lambda\tau}\right) & \text{if } n = 0, y \to 0, \\ \mathrm{e}^{\lambda\tau}\left(1 - \mathrm{e}^{-\lambda\tau}\right) \mathrm{e}^{-\lambda n\tau} & \text{if } n > 0, y \to \tau, \end{cases} \\ &\quad = \max_{n,y} \begin{cases} 1 - \mathrm{e}^{-\lambda\tau} & \text{if } n = 0, y \to 0, \\ 1 - \mathrm{e}^{-\lambda\tau} & \text{if } n = 1, y \to \tau, \end{cases} \\ &\quad = 1 - \mathrm{e}^{-\lambda\tau}, \end{aligned} \tag{6}$$

so then the min-entropy is:

$$H_\infty(D) = -\log_2\left(\max_{n,y} p_n\right) = -\log_2\left(1 - \mathrm{e}^{-\lambda\tau}\right). \tag{7}$$

Dead time is a time of detector insensitivity after successful photon detection. Assuming τ_d dead time to be in the form: $\tau_\mathrm{d} = k\tau + \delta$, where k is a non negative integer and $0 \le \delta < \tau$, we can rewrite Eq. (5):

$$\begin{aligned} &\Pr(D = n \mid \gamma = y) \\ &= \begin{cases} 0 \text{ if } n < k, \\ \Pr(y + T + \delta < \tau) \text{ if } n = k, \\ \Pr\left((n-k)\tau \le y + T + \delta < (n-k+1)\tau\right) \text{ if } n = k+1, \\ \Pr\left((n-k)\tau \le y + T + \delta < (n-k+1)\tau\right) \text{ if } n > k+1, \end{cases} \\ &= \begin{cases} 0 \text{ if } n < k, \\ \begin{cases} 1 - \mathrm{e}^{-\lambda(\tau - y - \delta)} & \text{if } y < \tau - \delta, n = k, \\ 0 & \text{if } y \ge \tau - \delta, n = k, \end{cases} \\ \begin{cases} \mathrm{e}^{-\lambda(\tau - y - \delta)}\left(1 - \mathrm{e}^{-\lambda\tau}\right) & \text{if } y < \tau - \delta, n = k+1, \\ 1 - \mathrm{e}^{-\lambda(2\tau - y - \delta)} & \text{if } y \ge \tau - \delta, n = k+1, \end{cases} \\ \left(\mathrm{e}^{-\lambda((n-k)\tau - y - \delta)}\right)\left(1 - \mathrm{e}^{-\lambda\tau}\right) \text{ if } n > k+1. \end{cases} \end{aligned} \tag{8}$$

Maximizing p_n for min-entropy, we then get:

$$\begin{aligned} H_\infty(D) &= -\log_2 \max_{n,y,\tau_\mathrm{d}}(\Pr(D = n \mid \gamma = y)) \\ &= -\log_2\left(1 - \mathrm{e}^{-\lambda\tau}\right), \end{aligned} \tag{9}$$

which is the same result as in the case without dead time.[3] Since this result is also the same as in the "restartable clock without dead time" case [30], we conclude, that using a continuous clock has no adverse effect on extractable min-entropy.

Dead time also has an effect on the detectable photon rate, since during τ_d no detections are possible. Since the bound for min-entropy is calculated using the original λ, and not the λ_d observed rate with dead time, we have to account for this, giving[4]:

$$\lambda = \frac{\lambda_\mathrm{max}}{1 - \lambda_\mathrm{max}\tau_\mathrm{d}}. \tag{10}$$

[3]Note that in this calculation of $H_\infty(D)$ we do not restrict τ_d in any way as in (9) we maximize over all possible τ_d. Due to this, $H_\infty(D) = -\log_2\left(1 - \mathrm{e}^{-\lambda\tau}\right) \le H_\infty(D \mid \tau_\mathrm{d} = Z)$ will hold for any possible Z distribution of τ_d.

[4]Note, that λ_d is maximized in $1/\tau_\mathrm{d}$, so the nominator here always stays positive.

3) Fluctuating λ: The actual value of λ may fluctuate due to various physical imperfections. Since $H_\infty(D)$ is monotonic in λ we can give a lower bound $H_\infty(D)_L$ for min-entropy if we know a $\lambda_{\max}$ upper bound for λ, such that

$$\begin{aligned} H_\infty(D)_L &= -\log_2\left(1-\mathrm{e}^{-\lambda\tau}\right) \leq H_\infty(D) \\ &= -\log_2\left(1-\mathrm{e}^{-\lambda\tau}\right). \end{aligned} \tag{11}$$

$$p_{\max} = 1-\mathrm{e}^{-\lambda\tau} \tag{12}$$

This also means, that by giving an upper bound for τ_d in Eq. (10), we also upper bound λ and lower bound the min-entropy, so this way we can also account for unknown dead time distributions as long as a maximum value for τ_d is known.

4) Attenuation and detector efficiency: The μ quantum efficiency of detectors (the probability of successfully detecting an incoming photon) is analogous to the T_a transmissivity of attenuators. Due to the memoryless property of the exponential distribution, we can account for these effects so that our detected photons arrive according to an $\mathrm{Exp}(\lambda T_\mathrm{a}\mu)$ distribution.

D. Framework for extractor parameter selection

To give a combined min-entropy lower bound for a case containing all the previously investigated noise effects, we can use the following steps:

1) Apply methods from Sections II-C2 and II-C3 to account for the effects of dead time and fluctuations in λ to get a lower bound for min-entropy and, therefore, an upper bound for $p_{\max}$ of the individual intervals corresponding to the ideal operation of the source.[5]
2) Use Eq. (4) with this $p_{\max}$ and appropriately selected $C_{\mathrm{noise}}, C_{\mathrm{exp}}$ values to calculate the final $H_\infty(D)$ for the interval series corresponding to the n_e bit long extractor input.

Note that in the second step, we use the overestimation of $p_{\max}$ of the intervals corresponding to C_{exp}. The reasoning presented in Sec. II-C1 is still valid as potential dependence between intervals due to non-ideal effects during measurement of photons of the ideal source is already accounted for in the overestimated $p_{\max}$ (see Sec. II-C2), therefore, the ability to sum interval min-entropies in Eq. 4 remains.

Also note, that counts from additive noise sources raise the experimentally detected λ, but this overestimation of λ does not lead to any additional security weaknesses, as presented in Sec. II-C3.

Other than min-entropy, we need to choose n_e to fully parameterize the extractor. Generally, choosing n_e to be larger is advantageous, as the scheme becomes more robust against bursty noise, as well as providing a better output ratio for a given ϵ according to Eq. (3). This comes at a cost of increased computational need, however.

[5]Since we usually base estimation on measurement results of the detector, attenuation from attenuators and detector efficiency discussed in II-C4 are already included in these.

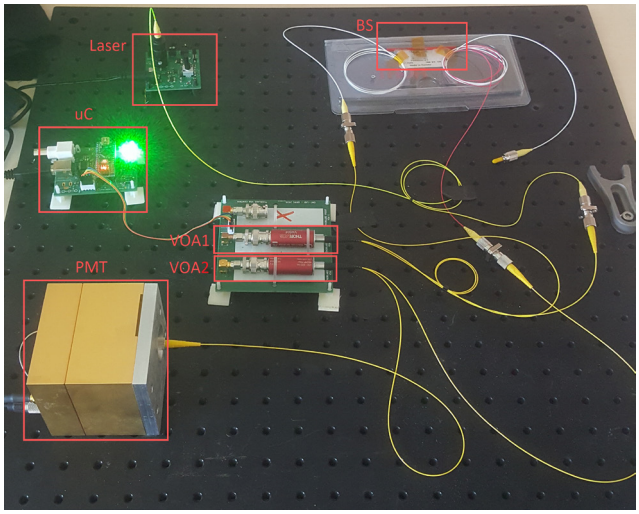


Fig. 3. Photo of the physical setup. uC: microcontroller controlling VOAs, BS: beam splitter, VOA: variable optical attenuator, PMT: photomultiplier tube. Photons travel along the Laser-VOA1-BS-VOA2-PMT optical path.

III. EXPERIMENT

A. Physical setup

Our physical setup presented in Fig. 3 is the same as in [31] and [29]. A Thorlabs LP520-SF15 semiconducting laser (central wavelength 519.9 nm) is attenuated using two successive voltage-controlled variable optical attenuators (Thorlabs V450F) and an optical splitter (Thorlabs TW560R1F1), where the splitter functions as an additional 20 dB attenuator. Photons are then detected by a PicoQuant PMA-175 NANO photomultiplier tube with a $\mu = 21\%$ quantum efficiency. The detector's output voltage pulses are time-tagged by a PicoQuant TimeHarp 260 time-to-digital converter (TDC) card with a base resolution of τ = 250 ps integrated into the PC controlling the measurement and running post-processing. Our detection system (detector and TDC) has a dead time of around 2 ns, very low afterpulsing probability ($\sim 0\%$), and measured dark count rates around 1-10 cps.

B. Parameter selection

We collect and process 2×10^{10} intervals to investigate the validity of our presented framework. During data acquisition, the measured detection rate was around $\lambda_\mathrm{d} = 1.3 \times 10^6$ cps (counts per second) and between $\lambda_{\min} = 1.08 \times 10^6$ cps and $\lambda_{\max} = 1.37 \times 10^6$ cps at all times. We chose not to try mitigating this fluctuation as our goal is to show robustness. We also completely forego using available protective covers made for severely limiting counts from the environment, leading to a $\lambda_\mathrm{n} = 20000$ cps noise rate at our detector. We overestimate our relatively low detector dead time of around 10τ with a conservative $\tau_\mathrm{d} = 50\tau$. In practice, afterpulsing effects are often neutralized by the longer detector dead times compared to them, which is also the case for our hardware, showing negligible afterpulsing probability. To present an example of handling this effect in our framework, we assume a maximum probability of counts caused by afterpulsing of

$P_{\text{after}} = 10^{-4}$ nonetheless. Due to the quality of laser sources, another quantity often considered experimentally negligible is the number of photons created in the source not via stimulated emission. Similarly to the previous case of afterpulsing, this effect could be considered negligible in our setup, but we still assume an exemplary maximum probability for it to be $P_{\text{nonstim}} = 10^{-6}$.

Utilizing our framework presented in Sec. II-D, we can calculate a lower bound for min-entropy using the presented measurement parameters: First, calculate λ from λ_{max} according to Eq. (10), giving $\lambda = 1.3215 \times 10^6$ cps. From this, calculate $p_{\text{max}} = 1 - \mathrm{e}^{-\lambda\tau} = 3.3031 \times 10^{-4}$. Utilizing that count numbers in Eq. (4) can be expressed in terms of detection rate over the investigated timeframe, we can use $\lambda_{\text{id}} = \lambda_{\text{min}}(1 - P_{\text{nonstim}})(1 - P_{\text{after}}) - \lambda_{\text{n}}$ to underestimate the number of counts from the ideal source, and $\lambda_{\text{noise}} = \lambda_{\text{n}} + P_{\text{nonstim}}\lambda_{\text{max}} + P_{\text{after}}\lambda_{\text{max}}$ to overestimate noise.

To be able to determine the C_{noise} counts in a processed data block, we have to first choose n_{e}. As stated before, higher values are beneficial, but since our hashing implementation currently runs on CPU and not on dedicated hardware as its main goal is to serve as proof of concept, we settle for $n_{\text{e}} = 2048$ bits, due to our limited computational resources. With 16-bit long measurement records, 128 records are processed together at once in a block. This means an average C_{noise} of 3 and C_{exp} of 125 (To additionally protect from burst errors C_{noise} can be chosen to be higher if needed.). Using (4) this yields $H_\infty(D) = k_{\text{e}} = 649.682$ bits for an $m_{\text{e}} = 544$ bits with $\epsilon = 2^{\overline{-52.841}} < 2^{-50}$.

For initialization of the Toeplitz hash algorithm (to create the $n \times m$ Toeplitz matrix), we need a $d_{\text{e}} = n_{\text{e}} + m_{\text{e}} - 1 = 2479$ bit long random string, which can come from a different trusted source or can even be a "baked in" string due to its reusability. We used random data collected during a previous different experiment [31] with our setup for initialization.

To further test our framework, we modified our initial measurement record file by artificially inserting counts every 50000 cycles simulating a perfect periodic noise/attacker (and thereby considerably changing the detected distribution too, as there can be no recorded time intervals longer than this inserted periodicity). This accounted for an additional noise source with a 80000 cps rate. Accounting for this (change in λ_{d}, λ_{max}, λ_{min}, λ_{n}), the newly calculated parameters for the hash function, in this case, are: $n_{\text{e}} = 2048$ bits, $k_{\text{e}} = 246.8593$ bits, $m_{\text{e}} = 144$ bits for an $\epsilon < 2^{-51}$ and 1.125 output bits per measurement record accordingly. Note the heavily reduced output efficiency, which is mainly due to the increased unknown noise considered according to Sec. II-C1.

C. Randomness testing

We assess our output files of 8.4 GB and 2.8 GB for the previously mentioned measurement cases with four of the most widely used statistical test suites, namely the NIST STS [32], Dieharder [33], TestU01 [34] and ENT [35] suites. Statistical tests typically try to refute the hypothesis that a source is random, by looking for signs of different kinds of possible non-randomness. Suites are, therefore, composed of batteries of individual tests, each looking for different non-random patterns. Due to the fact that a properly random output contains every possible string, a good generator is also expected to fail some proportion of these tests, so verifying proper operation is tricky and cannot be based on test results alone. To demonstrate this, we also tested unprocessed and not properly parametrized processed versions of our initial measurement data. Still, statistical testing of the output is a handy tool for checking for potential oversights or implementation errors (An uncharacteristically poor performance on tests almost surely indicates some error in operation.).

Results from the NIST STS suite for our first output file are shown in Table I omitting variants of the *NonOverlappingTemplate*, *RandomExcursions* and *RandomExcursionsVariant* tests as these are families of multiple tests producing too many results to be easily presentable in table format. We ran the suite with default settings and 2048 streams to test for both of our files. According to the manual, a case is considered passing if at least 2014 of the streams pass. We found that our data passed all the tests in the assessment.

TABLE I
RESULTS FOR NIST STS TESTS

Test Name	p-value	Proportion	Assessment
Frequency	0.4564	2032/2048	Pass
BlockFrequency	0.7979	2031/2048	Pass
CumulativeSums 1	0.9195	2029/2048	Pass
CumulativeSums 2	0.1850	2025/2048	Pass
Runs	0.5862	2025/2048	Pass
LongestRun	0.6920	2026/2048	Pass
Rank	0.7041	2021/2048	Pass
DFT	0.5450	2022/2048	Pass
OverlappingTemplate	0.1885	2031/2048	Pass
Universal	0.1608	2024/2048	Pass
ApproximateEntropy	0.3294	2025/2048	Pass
Serial 1	0.7997	2025/2048	Pass
Serial 2	0.2548	2028/2048	Pass
LinearComplexity	0.1053	2027/2048	Pass

The Dieharder suite is a collection of many tests expanding upon the original Diehard tests [36]. We present results for our first measurement case from these original (Diehard) tests in Table II. [6] The suit additionally contains other tests, which our data also successfully passed for both of the output files.

We used the *Alphabit* and *Rabbit* batteries recommended for use with hardware RNGs as well as the *SmallCrush* test battery from the TestU01 software library to assess our data. Results from the *SmallCrush* battery for the first output file are presented in Table III. We found that both of our data files passed all these assessments.

The ENT program can test random files in *byte* and *bit* modes, and calculates statistics like symbol occurrences, entropy, approximation of π, and correlation, to assess the randomness of a bitstream. Our files passed these assessments in both modes.

[6]Due to the occasional expected test failures of proper random operation, the dieharder suite also has a *WEAK* assessment result, where the manual advises further investigation. In our case, the tests *diehard_squeeze* and *diehard_sums* originally produced this result, so we ran them with lengthier than standard input data for a stronger examination to make sure of correctness and found them passing.

TABLE II
RESULTS FOR DIEHARD TESTS

Test Name	p-value	Assessment
diehard_birthdays	0.48117807	Pass
diehard_operm5	0.72724586	Pass
diehard_rank_32x32	0.18749969	Pass
diehard_rank_6x8	0.20228745	Pass
diehard_bitstream	0.13044230	Pass
diehard_opso	0.92784321	Pass
diehard_oqso	0.091542557	Pass
diehard_dna	0.60242889	Pass
diehard_count_1s_str	0.30601543	Pass
diehard_count_1s_byt	0.63839715	Pass
diehard_parking_lot	0.91059716	Pass
diehard_2dsphere	0.18188938	Pass
diehard_3dsphere	0.91144842	Pass
diehard_squeeze	0.51632824	Pass
diehard_sums	0.03411320	Pass
diehard_runs 1	0.90873329	Pass
diehard_runs 2	0.86353873	Pass
diehard_craps 1	0.86019516	Pass
diehard_craps 2	0.39312891	Pass

TABLE III
RESULTS FOR TESTU01 TESTS

Test Name	p-value	Assessment
smarsa_BirthdaySpacings	0.86	Pass
sknuth_Multinomial	0.60	Pass
sknuth_Gap	0.76	Pass
sknuth_SimpPoker	0.74	Pass
sknuth_CouponCollector	0.61	Pass
sknuth_MaxOft 1	0.73	Pass
sknuth_MaxOft 2	0.85	Pass
svaria_WeightDistrib	0.82	Pass
smarsa_MatrixRank	0.13	Pass
sstring_HammingIndep	0.97	Pass
swalk_RandomWalk1 H	0.78	Pass
swalk_RandomWalk1 M	0.47	Pass
swalk_RandomWalk1 J	0.20	Pass
swalk_RandomWalk1 R	0.10	Pass
swalk_RandomWalk1 C	0.80	Pass

To demonstrate the nature of statistical testing and the need for proper analysis in addition to passing test results, we also tested unprocessed data (binary datafile only containing the unprocessed 16-bit records) and an additional test case, where we incorrectly parametrized the hash function with $m_e = 2048$. For the first unprocessed case, the ENT test already showed some weaknesses, with an estimated entropy of 7.247 bits per byte (well above the estimated min-entropy in Sec. III-B, below expected 8 of ideal uniform output), and compressibility of 9 percent, while all the other test suites summarily failed the data (which is expected since raw measurement data correspond to a not uniform distribution). Interestingly, the wrongly parametrized processed data also passed our statistical trials, demonstrating, that passing the tests is not a guarantee for secure randomness in itself. This is probably due to the fact that the hashing operation in itself shows behavior similar to pseudo-random number generators, as its main aim is to produce random-like output from any input. While this wrongly parametrized output is clearly not suitable for quality and security-critical use cases, it may still prove useful in cases with less strict output quality criteria, essentially enabling an operation mode realizing a rapidly reseeded pseudo-random number generator, with higher output efficiency. We leave further investigation of such a scheme up for future study.

D. Achievable output rates

The achievable output efficiency and, consequently, the final output rate are heavily influenced by the magnitude of noise effects. Fig. 4. shows that in our test setup, increasing noise can

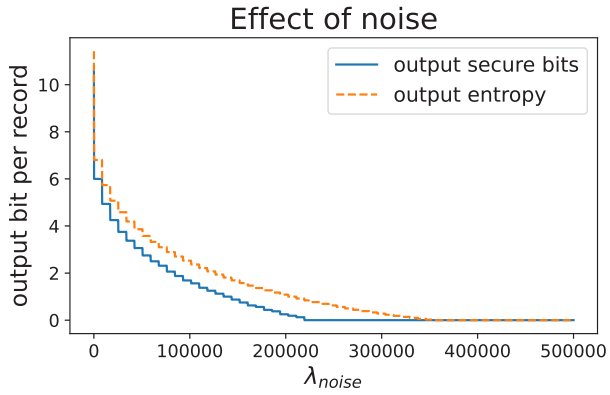


Fig. 4. Effect of different λ_{noise} noise intensities on achievable output bit and entropy rates, with the parameter set presented at the beginning of Sec. III-B, while maintaining $\epsilon < 2^{-50}$.

lead to cases where we can no longer guarantee our goal ϵ for any parameter set (from $\lambda_{\text{noise}} \geq 219352$), or even any secure output at all (from $\lambda_{\text{noise}} \geq 354339$). Furthermore, introducing even small amounts of noise to the system leads to a steep decline in the achievable output rate. For the completely noiseless case, our test setup would have an efficiency of 10.6875 output bits per record, leading to a theoretical max output speed of 13.863 Mbps, while introducing only the example noise from afterpulsing effects and photons not from stimulated emission ($\lambda_{\text{noise}} = 138$) already drops efficiency to 6 output bits per record and output speed to 7.8 Mbps. The two noisy example cases presented before at the start and end of Section III-B have output efficiencies of 4.25 and 1.6875 bits per record and output speeds of 6.598 Mbps and 2.193 Mbps, respectively.

Unfortunately, our current practical implementation presents a computational bottleneck of $\sim 10^5$ records processed per second, limiting our current practically achievable output speeds. This can likely be overcome in the future with a new implementation utilizing either an FPGA or GPU as it has been demonstrated in the literature [37], [38] and therefore, the implementation of a new post-processing program is our next logical practical goal. Better and stricter characterization of possible noise sources is another worthwhile direction to pursue for possible future development, especially for cases with concrete, well-characterized measurement setups, as it may be possible to find tighter lower bounds than Eq. (4) when using less general assumptions.

IV. CONCLUSION

We presented a post-processing framework for optical QRNGs based on the measurement of photon arrival times,

that can be used to safely account for typical distortion effects and hard-to-characterize error sources or attackers given a simple upper limitation on intensity, by strictly underestimating the min-entropy of the measurement results and utilizing this estimate to parameterize a Toeplitz hash-based extractor to provide a guaranteed quality, safe output bitstream. We demonstrated the use of our framework on intentionally non-ideal measurement data, showing its robustness, and assessed the processed outputs with statistical test suites to experimentally verify our proposal's correctness.

We conclude that our method can be used to provide quality output even when paired with noisy and imperfect measurement setups, although at a cost of reduced output efficiency. This drop in efficiency is especially prevalent when adjusting for the effects of error sources considered as unknown, so in practical realizations minimizing or adequately characterizing these should still remain a priority with our framework too.

REFERENCES

[1] L. Gyongyosi, L. Bacsardi, and S. Imre, "A survey on quantum key distribution," *Infocommunications Journal*, no. 2, pp. 14–21, 2019. [Online]. Available: DOI: 10.36244/icj.2019.2.2

[2] D. Chandra, P. Botsinis, D. Alanis, Z. Babar, S.-X. Ng, and L. Hanzo, "On the road to quantum communications," *Infocommunications Journal*, vol. 14, no. 3, pp. 2–8, 2022. [Online]. Available: DOI: 10.36244/icj.2022.3.1

[3] Y. Dodis, S. J. Ong, M. Prabhakaran, and A. Sahai, "On the (im) possibility of cryptography with imperfect randomness," in *45th Annual IEEE Symposium on Foundations of Computer Science. IEEE*, 2004, pp. 196–205. [Online]. Available: DOI: 10.1109/FOCS.2004.44

[4] N. Heninger, Z. Durumeric, E. Wustrow, and J. A. Halderman, "Mining your ps and qs: Detection of widespread weak keys in network devices," in *Presented as part of the 21st {USENIX} Security Symposium ({USENIX} Security 12)*, 2012, pp. 205–220. [Online]. Available: DOI: 10.5555/2362793.2362828

[5] M. Herrero-Collantes and J. C. Garcia-Escartin, "Quantum random number generators," *Reviews of Modern Physics*, vol. 89, no. 1, p. 015004, 2017. [Online]. Available: DOI: 10.1103/RevModPhys.89.015004

[6] T. Jennewein, U. Achleitner, G. Weihs, H. Weinfurter, and A. Zeilinger, "A fast and compact quantum random number generator," *Review of Scientific Instruments*, vol. 71, no. 4, pp. 1675–1680, apr 2000. [Online]. Available: DOI: 10.1063/1.1150518

[7] A. Stefanov, N. Gisin, O. Guinnard, L. Guinnard, and H. Zbinden, "Optical quantum random number generator," *Journal of Modern Optics*, vol. 47, no. 4, pp. 595–598, mar 2000. [Online]. Available: DOI: 10.1080/09500340008233380

[8] M. Fürst, H. Weier, S. Nauerth, D. G. Marangon, C. Kurtsiefer, and H. Weinfurter, "High speed optical quantum random number generation," *Optics Express*, vol. 18, no. 12, p. 13029, jun 2010. [Online]. Available: DOI: 10.1364/oe.18.013029

[9] S. Tisa, F. Villa, A. Giudice, G. Simmerle, and F. Zappa, "High- speed quantum random number generation using CMOS photon counting detectors," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 21, no. 3, pp. 23–29, may 2015. [Online]. Available: DOI: 10.1109/jstqe.2014.2375132

[10] M. Ren, E. Wu, Y. Liang, Y. Jian, G. Wu, and H. Zeng, "Quantum random-number generator based on a photon-number-resolving detector," *Physical Review A*, vol. 83, no. 2, feb 2011. [Online]. Available: DOI: 10.1103%2Fphysreva.83.023820

[11] M. Stipčević and B. M. Rogina, "Quantum random number generator based on photonic emission in semiconductors," *Review of scientific instruments*, vol. 78, no. 4, p. 045104, 2007. [Online]. Available: DOI: 10.1063/1.2720728

[12] H.-Q. Ma, Y. Xie, and L.-A. Wu, "Random number generation based on the time of arrival of single photons," *Applied optics*, vol. 44, no. 36, pp. 7760–7763, 2005. [Online]. Available: DOI: 10.1364/ao.44.007760

[13] M. A. Wayne, E. R. Jeffrey, G. M. Akselrod, and P. G. Kwiat, "Photon arrival time quantum random number generation," *Journal of Modern Optics*, vol. 56, no. 4, pp. 516–522, 2009. [Online]. Available: DOI: 10.1080/09500340802553244

[14] N. Massari, L. Gasparini, M. Perenzoni, G. Pucker, A. Tomasi, Z. Bisadi, A. Meneghetti, and L. Pavesi, "A compact tdc-based quantum random number generator," in *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*. IEEE, 2019, pp. 815–818. [Online]. Available: DOI: /10.1109/icecs46596.2019.8964941

[15] C. R. Williams et al., "Fast physical random number generator using amplified spontaneous emission," *Optics Express*, vol. 18, no. 23, pp. 23 584–23 597, 2010. [Online]. Available: DOI: 10.1364/oe.18.023584

[16] Á. Marosits, Á. Schranz, and E. Udvary, "Amplified spontaneous emission based quantum random number generator," *Infocommunications Journal*, vol. 12, no. 2, pp. 12–17, 2020. [Online]. Available: DOI: 10.36244/icj.2020.2.2

[17] M. Jofre, M. Curty, F. Steinlechner, G. Anzolin, J. P. Torres, M. W. Mitchell, and V. Pruneri, "True random numbers from amplified quantum vacuum," *Optics Express*, vol. 19, no. 21, p. 20665, oct 2011. [Online]. Available: DOI: 10.1364/oe.19.020665

[18] W. Lei, Z. Xie, Y. Li, J. Fang, and W. Shen, "An 8.4 gbps real-time quantum random number generator based on quantum phase fluctuation," *Quantum Information Processing*, vol. 19, no. 11, nov 2020. [Online]. Available: DOI: 10.1007/s11128-020-02896-y

[19] P. J. Bustard, D. Moffatt, R. Lausten, G. Wu, I. A. Walmsley, and B. J. Sussman, "Quantum random bit generation using stimulated raman scattering," *Optics Express*, vol. 19, no. 25, p. 25173, nov 2011. [Online]. Available: DOI: 10.1364/oe.19.025173

[20] "ID Quantique Quantis QRNG chip," https://www.idquantique.com/random-number-generation/products/quantis-qrng-chip/, 2024, (Last accessed 2024/03/05).

[21] P. Keshavarzian, K. Ramu, D. Tang, C. Weill, F. Gramuglia, S. Tan, M. Tng, L. Lim, E. Quek, D. Mandich, M. Stipčević, and E. Charbon, "A 3.3-gb/s spad-based quantum random number generator," *IEEE Journal of Solid-State Circuits*, vol. PP, pp. 1–16, 09 2023. [Online]. Available: DOI: 10.1109/JSSC.2023.3274692

[22] H. Xu, N. Massari, L. Gasparini, A. Meneghetti, and A. Tomasi, "A SPAD-based random number generator pixel based on the arrival time of photons," *Integration*, vol. 64, pp. 22–28, jan 2019. [Online]. Available: DOI: 10.1016/j.vlsi.2018.05.009

[23] F. Regazzoni, E. Amri, S. Burri, D. Rusca, H. Zbinden, and E. Charbon, "A high speed integrated quantum random number generator with on-chip real-time randomness extraction," *arXiv preprint arXiv:2102.06238*, 2021.

[24] B. L. Márton, D. Istenes, and L. Bacsárdi, "Enhancing the operational efficiency of quantum random number generators," *Infocommunications Journal*, vol. 13, no. 2, pp. 10–18, 2021. [Online]. Available:DOI: 10.36244/icj.2021.2.2

[25] X. Ma, F. Xu, H. Xu, X. Tan, B. Qi, and H.-K. Lo, "Postprocessing for quantum random-number generators: Entropy evaluation and randomness extraction," *Physical Review A*, vol. 87, no. 6, jun 2013. [Online]. Available: DOI: 10.1103/physreva.87.062327

[26] R. Impagliazzo, L. A. Levin, and M. Luby, "Pseudo-random generation from one-way functions," in *Proceedings of the twenty-first annual ACM symposium on Theory of computing - STOC '89*. ACM Press, 1989. [Online]. Available: DOI: 10.1145/73007.73009

[27] R. J. Glauber, "Coherent and incoherent states of the radiation field," *Physical Review,* vol. 131, no. 6, pp. 2766–2788, 9 1963. [Online]. Available: DOI: 10.1103/PhysRev.131.2766

[28] M. C. Teich and B. E. A. Saleh, "Effects of random deletion and additive noise on bunched and antibunched photon-counting statistics," *Optics Letters*, vol. 7, no. 8, p. 365, aug 1982. [Online]. Available: DOI: 10.1364/ol.7.000365

[29] B. Solymos and L. Bacsárdi, "Efficiency improvement of photon arrival time based quantum random number generator with hashing," in *IEEE 17th International Symposium on Applied Computational Intelligence and Informatics SACI 2023 : Proceedings*, 2023, pp. 53–58. [Online]. Available: DOI: 10.1109/SACI58269.2023.10158613

[30] A. Schranz and E. Udvary, "Mathematical analysis of a quantum random number generator based on the time difference between photon detections," *Optical Engineering*, vol. 59, no. 4, p. 044104, 2020. [Online]. Available: DOI: 10.1117/1.OE.59.4.044104

[31] Á. Schranz, "Optical solutions for quantum key distribution transmitters," Ph.D. dissertation, Budapest University of Technology and Economics, 2021. [Online]. Available: http://hdl.handle.net/10890/16991

[32] "NIST SP 800-22: Documentation and Software," https://csrc.nist.gov/projects/random-bit-generation/documentation-and-software, 2024, (Last accessed 2024/03/05).

[33] "dieharder by Robert G. Brown, Duke University Physics Department, Durham, NC 27708-0305 Copyright Robert G. Brown, 2019," https://webhome.phy.duke.edu/~rgb/General/dieharder.php, 2024, (Last accessed 2024/03/05).

[34] P. L'Ecuyer and R. Simard, "TestU01: A c library for empirical testing of random number generators," *ACM Transactions on Mathematical Software*, vol. 33, no. 4, pp. 1–40, aug 2007. [Online]. Available: DOI: 10.1145/1268776.1268777

[35] "ENT: A Pseudorandom Number Sequence Test Program," https://www.fourmilab.ch/random/, 2024, (Last accessed 2024/03/05).

[36] G. Marsaglia, ""the marsaglia random number cdrom including the diehard battery of tests of randomness". Florida State University. 1995. archived from the original on 2016-01-25." https://web.archive.org/web/20160125103112/http://stat.fsu.edu/pub/diehard/, (Last accessed 2024/03/05).

[37] X. Zhang, Y.-Q. Nie, H. Liang, and J. Zhang, "FPGA implementation of toeplitz hashing extractor for real time post-processing of raw random numbers," in *2016 IEEE-NPSS Real Time Conference (RT)*. IEEE, jun 2016. [Online]. Available: DOI: 10.1109/rtc.2016.7543094

[38] M. J. Ferreira, N. A. Silva, and N. J. Muga, "Efficient randomness extraction in quantum random number generators," in Anais do II Workshop de Comunicação e Computação Quântica (*WQuantum 2022*). Sociedade Brasileira de Computação, may 2022. [Online]. Available: DOI: 10.5753/wquantum.2022.223591

Balázs Solymos received his B.Sc. degree in 2018, followed by his his M.Sc. degree in early 2020 in Electrical Engineering from the Budapest University of Technology and Economics (BME). He is currently pursuing his PhD at the Department of Networked Systems and Services, BME. He is involved in a research project aiming to establish a quantum random generator service on campus. His current research interests are quantum communications, quantum internet, and quantum computing.

László Bacsárdi (M'07) received his MSc degree in 2006 in Computer Engineering from the Budapest University of Technology and Economics (BME) and his PhD in 2012. He is a member of the International Academy of Astronautics (IAA). Between 2009 and 2020, he worked at the University of Sopron, Hungary in various positions including Head of Institute of Informatics and Economics. Since 2020, he is associate professor at the Department of Networked Systems and Services, BME and head of Mobile Communications and Quantum Technologies Laboratory. His current research interests are quantum computing, quantum communications and ICT solutions developed for Industry 4.0. He is the past chair of the Telecommunications Chapter of the Hungarian Scientific Association for Infocommunications (HTE), Vice President of the Hungarian Astronautical Society (MANT). Furthermore, he is member of IEEE and HTE as well as alumni member of the UN established Space Generation Advisory Council (SGAC). In 2017, he won the IAF Young Space Leadership Award from the International Astronautical Federation.

An Ordered QR Decomposition based Signal Detection Technique for Uplink Massive MIMO System

Jyoti P. Patra, Bibhuti Bhusan Pradhan, and M. Rajendra Prasad

***Abstract*—Signal detection turns out to be a critical challenge in massive MIMO (m-MIMO) system due to the deployment of large number of antennas at the base station. Although, the minimum mean square error (MMSE) is one of the popular signal detection method, but, it requires matrix inversion with cubic complexity. In order to reduce computational complexity, several suboptimal signal detection methods were proposed such as Gauss-Seidel, successive over relaxation, Jacobi, Richardson methods. Although, these methods provide low complexity but their performance are limited to MMSE method. In this paper, we have proposed two signal detection techniques namely QR decompositions (QRD) and ordered QRD (OQRD). Finally, the performances of proposed signal detection methods are compared with various conventional methods in terms of symbol error rate (SER) and computational complexity. The simulation results validate that the proposed methods outperform the MMSE method with substantially lower computational complexity.**

***Index Terms*—Massive MIMO, Signal detction, QRD, OQRD, MMSE, Low complexity.**

I. INTRODUCTION

Massive multiple-input multiple-output (m-MIMO) is the most promising technique in 5G and beyond 5G (B5G) due to its high spectrum and energy efficiency, high spatial resolution, and simple transceiver design. In m-MIMO, a large number of antennas are employed at the base station (BS) [1, 2]. In the uplink transmission, the signals transmitted from mobile terminals are superimposed at the BS which cause interference and reduces the data rate. Due to deployments of large number of antennas, it requires advanced signal processing for data detection. The maximum-likelihood (ML) detection provides optimum bit error rate performance [1, 2]. However, it is not practically possible to employ the maximum likelihood (ML) detector due to its huge computational complexity as it searches all possible combination while performing data detection . The problem is also becoming more complicated when high-order modulation schemes are used and more users are multiplexed. Therefore, many nonlinear signal data detection methods are proposed which includes sphere decoder (SD) [3], tabu search (TS) [4], dirty paper coding [5] etc. Unfortunately, for massive MIMO systems with large number of antennas and higher-order modulation schemes, such methods need still very huge computation complexity. For spatially correlated massive MIMO system, random matrix theory based algorithms such as principal component analysis, eigen analysis, Karhunen-Loeve decomposition, were applied for signal detection [6–8]. However, Most researchers focused on linear signal detection algorithm rather non-linear methods for spatially uncorrelated m-MIMO system. Although, zero forcing and minimum mean square error (MMSE) are the two popular linear signal detection methods, they require matrix inversion with cubic complexity. Even though linear signal detection methods may not offer sufficient performance, still most of the researchers focused on linear methods because of reduced computational complexity. Several linear signal detection methods were proposed by exploiting the Gramian matrix to avoid matrix inversion which include Gauss-Seidel (GS) [9], the Neumann series (NS) [10], the successive overrelaxation (SOR) [11, 12], the Jacobi method [13], the conjugate gradient (CG) [14], the optimized coordinate descent (OCD), and the Richardson (RI) [15]. It has been observed that NS method is lower than the complexity of the detector based on GS, JA, RI and SOR methods, however its performance is the least. A hybrid pseudo-stationary iterative detection algorithm based on Chebyshev polynomial and Weyls inequality was proposed in [16] for uplink massive MIMO systems. This method provides near to ZF method. The authors in [17] proposed a weighted two stage (WTS) method which achieves similar performance to ZF method with lower complexity. Latter, a modified weighted two stage (MWTS) method was proposed in [18] which outperforms the WTS method. However, its performance is lower as compared to MMSE method. In [19], Cayley-Hamilton theorem-based two low complexity signal detection have been proposed to avoid the matrix inversion. This method has lower computational complexity as it does not involve in Gramian matrix. The authors in [20] performed signal detection by QR decomposition of Gram matrix $G = H^H H$. This method has performance limitation to ZF method because the QR decomposition was applied to ZF Gramian matrix. Similarly, in [21], the author applied several matrix decomposition technique such as QR, LDL and Cholesky. These matrix decomposition algorithm were applied to MMSE Grammian matrix, therefore their performances are limited to MMSE method.

In this paper, we proposed two signal detection methods namely QR decomposition (QRD) and order QRD (OQRD) methods for m-MIMO uplink communication system. The QR decomposition is directly applied to original channel matrix H

Jyoti P. Patra and M. Rajendra Prasad are with the Department of Electronics and Communication Engineering, Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India (e-mail: jyotiprasannapatra@gmail.com, rajendraresearch@gmail.com).

Bibhuti Bhusan Pradhan is with the Department of Electronics and Communication Malla Reddy Engineering College, Hyderabad, Telangana, India (e-mail: bibhu.iisc@gmail.com).

DOI: 10.36244/ICJ.2024.1.3

to obtain estimated transmitted signal. Thus, it outperforms other QR decomposition methods [20, 21] where the ZF and MMSE Gramain matrix are decomposed by QR method. Furthermore, the performance of QRD method increases by ordering the column norm of channel matrix in ascending manner. We call this method as OQRD method. The proposed QRD and OQRD methods are compared with various conventional method such as Gauss-Seidel, successive over relaxation, Jacobi, Richardson and MMSE methods in terms of symbol error rate and computational complexity.

We organize our paper as follows. In Section II, we describe the massive MIMO uplink system model. In Section III, we discuss various signal detection methods. In Section IV, we present the proposed signal detection methods. In Section V, we show simulation results of proposed and conventional methods in terms of symbol error rate. Finally, Section VI concludes the paper.

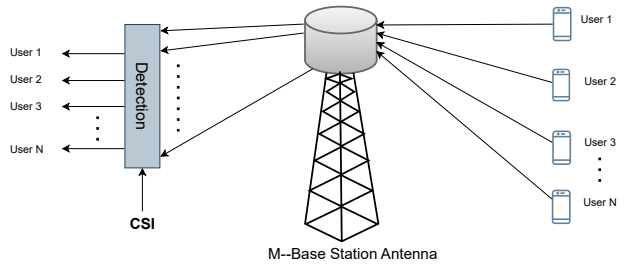


Fig. 1: Block diagram of uplink massive MIMO system with M number of base station antenna and N number of users

II. System Model

The uplink channel is used to transmit data symbols from the user terminal to the base station. In a multiuser uplink massive MIMO system, M number of base station antennas are employed to serve N number of users simultaneously as shown in the Fig 1. Let $\boldsymbol{x}$ denote the complex valued $N \times 1$ simultaneously transmitted signal vector from the N users to the base station. The received signal vector $\boldsymbol{y}$ at the BS can be given by

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

where $\mathbf{H}$ is the channel matrix between the user terminal and the base station with size $M \times N$ and $M > N$. The parameter $\boldsymbol{n}$ is the $M \times 1$ additive white Gaussian noise (AWGN). Although, the maximum likelihood (ML) method is the optimal signal detection method, it is not preferable from the hardware implementation perspective due to its high computational complexity. Therefore, suboptimal linear signal detection techniques such as zero forcing and minimum mean square error (MMSE) methods are widely accepted due to their near-optimal performance with lower computational complexity as compared to ML method. The signal detection based on MMSE method is given by

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{H}^H\mathbf{H} + \frac{N}{SNR}\mathbf{I}_N\right)^{-1}\boldsymbol{H}^H\boldsymbol{y} = \mathbf{A}^{-1}\hat{\boldsymbol{x}}_{MF} \tag{2}$$

where $A = \left(\boldsymbol{H}^H\mathbf{H} + \frac{N}{SNR}\mathbf{I}_N\right)^{-1}$ and $\hat{\boldsymbol{x}}_{MF} = \boldsymbol{H}^H\boldsymbol{y}$. The matrix $\mathbf{I}_N$ is the $N \times N$ identity matrix and SNR is the signal to noise ratio. The MMSE method involves large matrix inversion operations with cubic complexity. To achieve close performance of MMSE with reduce complexity, several signal detection methods have been proposed such as Jacobi, Richardson, Gauss-Seidel, successive over relaxation methods by exploiting the Gram matrix.

III. Conventional Signal Detection Method

In this section, we have discussed various signal detection methods namely Jacobi, Richardson, Gauss-Seidel, successive over relaxation methods for massive MIMO uplink system.

A. Jacobi Method

The Jacobi method was proposed for m-MIMO uplink system in [13]. The Jacobi method approximate the matrix inversion with reduces complexity. The Jacobi method is an iterative approach for finding the solution to a diagonally dominant system. The equation (2) can be rewritten as

$$\mathbf{A}\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}_{MF} \tag{3}$$

Note that when N/M is large, matrix A becomes diagonally dominant. The estimated signal can be obtained as

$$\hat{\boldsymbol{x}}^{(\boldsymbol{n})} = \mathbf{D}^{-1}\left[\hat{\boldsymbol{x}}_{MF} + (\mathbf{D} - \mathbf{A})\hat{\boldsymbol{x}}^{(n-1)}\right] \tag{4}$$

where $\mathbf{D}$ is the digonal matrix of $\mathbf{A}$. The initial values can be selected as

$$\hat{\boldsymbol{x}}^{(0)} = \mathbf{D}^{-1}\hat{\boldsymbol{x}}_{MF}. \tag{5}$$

It can be verified that the first iteration of JA method does not involve matrix multiplication, thus computational complexity decreases.

B. Richardson Method

The Richardson method was proposed in [15]. In this method, the signal detection is performed by iterative process through the exploitation of Gramian matrix $\mathbf{G} = \mathbf{H}^H\mathbf{H}$. Here the convergence rate is very sensitive to a selection of relaxation parameter (ω) where $0 < \omega \leq \frac{2}{\lambda_{max}}$ and the optimum value of ω is defined as $w = \frac{2}{\lambda_{\min}+\lambda_{\max}}$. The parameter λ_{max} and λ_{min} are the maximum and minimum eigenvalues of the symmetric positive definite matrix $\mathbf{A}$ respectively.The estimated signal is obtained as

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} + \omega\left[\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^{(n)}\right] \quad n = 0, 1, 2, \cdots \tag{6}$$

If a prior knowledge of $\mathbf{x}^{(0)}$ is missing, a zero vector can be considered without loss of generality. It can also be selected as $\hat{x}^{(0)} = \mathbf{D}^{-1}\hat{\boldsymbol{x}}_{MF}$ and iteratively refined. The accuracy and the number of computations are highly affected by the value of ω.

C. Gauss-Sidel Method

The Gauss Sidel method computes the solution by an iterative behaviour where the Hermitian positive semi-definite matrix (**A**) is decomposed to a lower triangular matrix (**L**), upper triangular elements (**U**), and the diagonal entries (**D**). In other words, the matrix **A** can be written as

$$\mathbf{A} = \mathbf{D} + \mathbf{L} + \mathbf{U}. \tag{7}$$

This method performs the signal detection in an iterative process as given by

$$\hat{\boldsymbol{x}}^{(n)} = [\mathbf{D} + \mathbf{L}]^{-1} \left[\hat{\boldsymbol{x}}_{MF} - \mathbf{U}\hat{\boldsymbol{x}}^{(n-1)}\right], \quad n = 1, 2, \cdots I_T \tag{8}$$

where I_T is the total number of iterations. Typically, the initial data signal $\hat{\boldsymbol{x}}^{(0)}$ is considered as a zero vector for simplification.

D. Successive Over Relaxation Method

In order to avoid the large dimension inversion matrix, the successive over relaxation(SOR) is a best choice in signal detection. It improves the accuracy of GS method by using a relaxation parameter (ω). The signal is estimated as

$$\hat{\boldsymbol{x}}^{(n)} = \left[\frac{1}{\omega}\mathbf{D} + \mathbf{L}\right]^{-1} \left[\hat{\boldsymbol{x}}_{MF} + \left[\left[\frac{1}{\omega} - 1\right]\mathbf{D} - \mathbf{U}\right]\hat{\boldsymbol{x}}^{[n-1]}\right] \tag{9}$$

Convergence of the SOR method is highly affected by the relaxation parameter (ω). In the MIMO framework, the relaxation parameter (ω) of the SOR technique is typically good when $0 < \omega < 2$. The optimum value is given by $w = \frac{2}{1+\sqrt{1-a^2}}$ where $a = \left(1 + \sqrt{\frac{N}{M}}\right)^2 - 1$. This value of w is fixed throughout the iteration process.

IV. Proposed Signal Detection Methods

In this section, the proposed signal detection methods namely QRD and OQRD are discussed for uplink massive MIMO system.

A. QRD Method

In this paper, we have applied QR decomposition to the channel matrix **H** for performing signal detection. The relation between received and transmitted signal can be written in matrix form as given by

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ h_{M1} & h_{M2} & \cdots & h_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix} \tag{10}$$

where h_{ij} is channel impulse response between jth transmitting antenna to the ith receiving antenna and $j = 1, 2, ..., N-1$ and $i = 1, 2, ..., M-1$. The channel matrix **H** can be decomposed into QR factors as $\mathbf{H} = \mathbf{QR}$ where $Q_{M\times N}$ is an orthonormal matrix and $R_{N\times N}$ is an upper triangular matrix. Substituting $\mathbf{H} = \mathbf{QR}$, the received signal can be written as

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n} = \mathbf{QR}\boldsymbol{x} + \boldsymbol{n} \tag{11}$$

After multiplying $\mathbf{Q}^H$ with the received signal vector $\boldsymbol{y}$, equation (11) can be modified to

$$\tilde{y} = \mathbf{Q}^H\boldsymbol{y} = \mathbf{Q}^H(\mathbf{H}\boldsymbol{x} + \boldsymbol{n}) = \mathbf{R}x + \tilde{n} \tag{12}$$

The equation (12) can be expressed in matrix form as

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_N \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1N} \\ 0 & R_{22} & \cdots & R_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} \tilde{n}_1 \\ \tilde{n}_2 \\ \vdots \\ \tilde{n}_N \end{bmatrix} \tag{13}$$

Since, **R** is a lower triangular matrix, backward substitution method can be applied to obtain the estimated transmitted signal vector and can be written as

$$\hat{x}_N = \Pi[\tilde{y}_N / R_{NN}] \tag{14}$$

$$\hat{x}_k = \Pi\left[\frac{\tilde{y}_k - \sum_{j=k+1}^{N} \tilde{y}_{kj}x_j}{R_{kk}}\right], k = N-1:-1:1 \tag{15}$$

where $\Pi()$ denotes the hard decision function. The detail steps of QRD method is summarized below.

[*Step* 1]: Initialization: $\boldsymbol{y}$, **H** , $\boldsymbol{x}$; $\boldsymbol{y} = \mathbf{Hx} + \boldsymbol{n}$

[*Step* 2]: Decomposition of channel matrix $\mathbf{H} = \mathbf{QR}$,
$\mathbf{y} = \mathbf{Hx} + \mathbf{n} = \mathbf{QR}\boldsymbol{x} + \boldsymbol{n}$

[*Step* 3]: Multiplication of $\mathbf{Q}^H$ with $\boldsymbol{y}$
$\tilde{\boldsymbol{y}} = \mathbf{Q}^H\boldsymbol{y} = \mathbf{R}x + \tilde{n}$

[*Step* 4]: Obtaining the transmitted signal using backward substitution method

$$\hat{x}_N = \Pi[\tilde{y}_N / R_{NN}]$$
$$\hat{x}_k = \Pi\left[\frac{\tilde{y}_k - \sum_{j=k+1}^{N} \tilde{y}_{kj}x_j}{R_{kk}}\right], k = N-1:-1:1$$

B. OQRD Method

The QRD method may suffer from error propagation problem if the initial signal is not detected correctly. Therefore, an order QRD (OQRD) method is proposed which orders the column vector of the channel matrix **H**. The relationship between received and transmitted signal can be written in the column form as

$$\boldsymbol{y} = \mathbf{Hx} + \boldsymbol{w} \ = \boldsymbol{h}_1x_1 + \boldsymbol{h}_2x_2 + \ \cdots + \boldsymbol{h}_Nx_N + \boldsymbol{n} \tag{16}$$

where $\boldsymbol{h}_k$ is the kth column vector of channel matrix **H** and x_k is the kth element of the transmitted signal vector $\boldsymbol{x}$. To perform the ordering of column vector in an ascending manner,

the procedure is as given follows. At first, we calculate the norm of each column vector of the $\mathbf{H}$ matrix as given by

$$norm_k = \|h_k\| \qquad k = 1, 2, ..., N \tag{17}$$

Then, we sort vector $\boldsymbol{norm} = [\ norm_1,\ norm_2,\ \cdots\ , norm_N\ ]$ in an ascending manner and find the ***index*** term

$$\boldsymbol{index} = sort(\boldsymbol{norm}) \tag{18}$$

The column vectors are arranged according to the indices to obtain the ordered banded CFR matrix $\mathbf{H}_o$

$$\boldsymbol{h}_{o,k} = \mathbf{H}_o(:, k) = \mathbf{H}(:, index_k) \qquad k = 1, 2, ..., N \tag{19}$$

where $\boldsymbol{h}_{o,k}$ is the kth column vector of $\mathbf{H}_o$ matrix and $index_k$ denotes the kth element of ***index*** vector. The ordering of the transmitted signal vector can also be represented in terms of the indices as given by

$$x_{o,k} = x(index_k) \qquad k = 1, 2...N \tag{20}$$

Substituting the ordered channel matrix $\mathbf{H}_o$ as defined in (19) and ordered transmitted signal vector $\boldsymbol{x}_o$ (20), the received signal vector $\boldsymbol{y}$ can be expressed as

$$\boldsymbol{y} = \mathbf{H}_o\boldsymbol{x}_o + \ \boldsymbol{n} \tag{21}$$

The order channel matrix $\mathbf{H}_o$ can be decomposed into QR factorization $H_o = Q_o R_0$. After multiplication $\mathbf{Q}_o^H$ on both sides of (21), it yields

$$\begin{aligned} \tilde{\mathbf{y}} &= \mathbf{Q}_o^H \boldsymbol{y} = \mathbf{Q}_o^H(\mathbf{H}_o\boldsymbol{x}_o + \boldsymbol{n}) \\ &= \mathbf{Q}_o^H(\mathbf{Q}_o\mathbf{R}_o\boldsymbol{x}_o + \boldsymbol{n}) = \mathbf{R}_o\boldsymbol{x}_o + \tilde{\mathbf{n}} \end{aligned} \tag{22}$$

The transmitted signal is obtained after performing the backward substitution method. The detail steps of QRD method is summarized below.

[*Step* 1]: Initialization: $\boldsymbol{y}$, $\mathbf{H}$, $\boldsymbol{x}$

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{w} \ = \boldsymbol{h}_1 x_1 + \boldsymbol{h}_2 x_2 + \ \cdots + \boldsymbol{h}_N x_N + \boldsymbol{n}$$

[*Step* 2]: Calculating norm of Channel matrix $\mathbf{H}$ and index the norm in ascending order

$$norm_k = \|h_k\| \ , \qquad k = 1, 2, ..., N$$
$$\boldsymbol{index} = sort(\boldsymbol{norm})$$

[*Step* 3]: Modifying the channel matrix $\mathbf{H}$ and signal vector $\boldsymbol{x}$ in terms of ascending order

$$\mathbf{H}_o(:, k) = \mathbf{H}(:, index_k) \quad k = 1, 2, ..., N$$
$$x_{o,k} = x(index_k) \qquad k = 1, 2..., N$$

[*Step* 4]: Decomposition of channel matrix $\mathbf{H}_o = \mathbf{Q}_o\mathbf{R}_0$,

$$\mathbf{y} = \mathbf{H}_o\mathbf{x}_o + \mathbf{n} = \mathbf{Q}_o\mathbf{R}_o\boldsymbol{x}_o + \boldsymbol{n}$$

[*Step* 5]: Multiplication of $\mathbf{Q}_o^H$ with $\boldsymbol{y}$

$$\tilde{\mathbf{y}} = \mathbf{Q}_o^H\boldsymbol{y} = \mathbf{R}_o\boldsymbol{x}_o + \tilde{\mathbf{n}}$$

[*Step* 6]: Obtaining the transmitted signal using backward substitution method

$$\hat{x}_{oN} = \Pi[\tilde{y}_{oN}/R_{oNN}]$$

$$\hat{x}_{ok} = \Pi\left[\frac{\tilde{y}_{ok} - \sum_{j=k+1}^{N} \tilde{y}_{okj} x_o j}{R_{okk}}\right], k = N-1 : -1 : 1$$

This method eliminates the error propagation problem of QRD by detecting the stronger signal at first and then cancels its effects before detection of weaker signal.

C. Computational Complexity

In this section, the computational complexity of the proposed QRD, OQRD methods are analysed in terms of multiplications. Then, the complexity of the proposed metod is compared with various signal detection methods which includes MMSE, Jacobi, Richardson, Gauss Seidel and SOR methods. The QR decomposition of channel matrix H requires $N^{2.529}$ [22]. Multiplying Q^H with Y requires NM^2 complexity. To obtain the estimated transmitted data signal requires backward substitution algorithm as given in [step 4] of the proposed QR method requires $2N(N-1)$ complexity. Thus, the QRD method requires a total of $N^{2.529} + 4NM^2 + 2N(N-1)$ complexity. The ordered QRD (OQRD) require same complexity as QRD method. In addition to that, OQRD method requires to find the norm of the column vector of matrix $\mathbf{H}$ which needs $4MN$ operations. Thus, total complexity involves in OQRD method is $N^{2.529} + 4NM^2 + 2N(N-1) + 4MN$. The computational complexity of the proposed methods are compared with conventional methods and is given in Table 1.

TABLE I
COMPUTATIONAL COMPLEXITY

Method	Multiplications
MMSE	$2MN^2 + (10/3)N^3 + 4MN + 4N^2$
Jacobi [13]	$(4M + 4I_T + 1)N^2 + 2NM$
Richardson [15]	$(4M + 4I_T)N^2 + 2NM$
GS [9]	$(4M + 4I_T - 2)N^2 + 2(N - 2I_T + 1)N$
SOR [12]	$(4M + 4I_T - 2)N^2 + 2(M - I_T + 1)N$
QRD	$N^{2.529} + 4NM^2 + 2N(N-1)$
OQRD	$N^{2.529} + 4NM^2 + 2N(N-1) + 4MN$

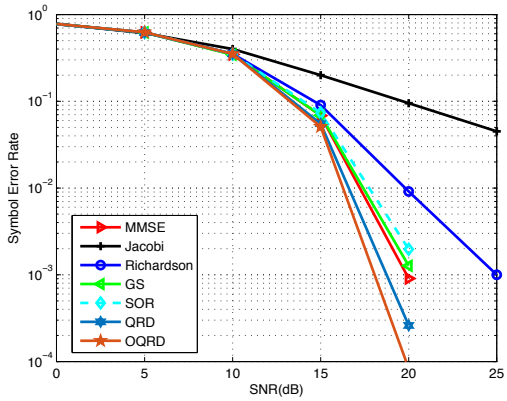


Fig. 2: SER performance comparison of various signal detection method forr NR = 24, NT = 64

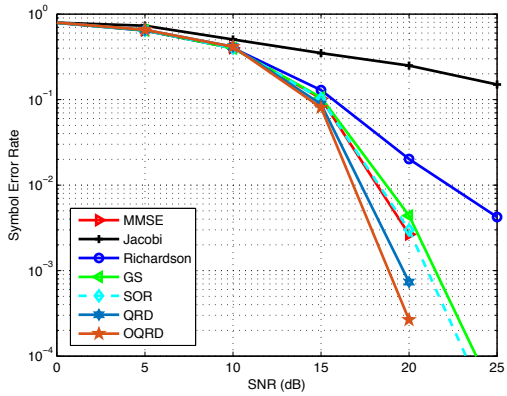


Fig. 3: SER performance comparison of various signal detection method for NR = 36, NT = 64

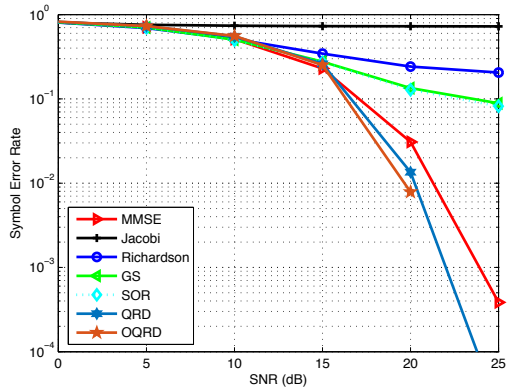


Fig. 4: SER performance comparison of various signal detection method for NR = 48, NT = 64

V. RESULTS

In this section, the performance of proposed QRD and OQRD methods are compared with various conventional signal detection methods for uplink massive MIMO system in terms of symbol error rate (SER) and computational complexity. The SER performance of various signal detection methods are carried out based on Monte Carlo simulation using MATLAB. We have considered $M = 64$ number of receiving antennas at the base station and N number of users with each user equipped with single transmitting antenna. For simulation, the antenna configuration $(N \times M)$ are as follows: 24×64, 36×64 and 48×64. The baseband signal modulation technique uses 16QAM, and for each SNR value, we simulate at least 48000 symbols. The transmission channel is considered as non-correlated Rayleigh fading channel. The perfect channel state information (CSI) is assumed to be known at the receiver terminal.

Fig 2, Fig.3 and Fig 4 show the SER performance comparison of proposed QRD and OQRD methods with various conventional signal detection methods for number of users $N = 24$, $N = 36$ and $N = 48$ respectively. From the simulation results it is seen that the Jacobi method has significantly lower performance. It is observed that the performance of Richardson method is much better Jacobi methods. The simulation results shows that the performance of SOR significantly improves and outperforms all the conventional methods when the number of users increases. The GS method is much better than Jacobi and Richardson methods. The performance of SOR provides slightly better when the ratio between user to BS i.e. N/M increases. Since, all the signal detection methods are derived from MMSE method based on several approximate matrix inversion methods, therefore, their performance are always lower than the MMSE methods. The performance of proposed QRD significantly outperforms the MMSE method. It is observed that the performance of OQRD method gives better performance than QRD method as it performs addition ordering of the channel matrix.

The SER vs number of users (N) performance comparison for various signal detection methods at $20dB$ SNR is shown in the Fig 5. The simulation results shows Jacobi and Richardson achieves considerable performance for lower number of users. But, as the number of users increases their performance significantly decreases. It is observed that the performance of GS and SOR methods performs close to MMSE method for lower number of users and the performance gap increase with a large number of users. It can be seen that the performance of proposed QRD method outperforms the MMSE method for lower to medium number of users. From the result, it is also observed that the performance of proposed OQRD method significantly outperform the MMSE method. Although the gap between OQRD and MMSE method decreases with very high number of users but still the OQRD method is significantly outperforms the MMSE method.

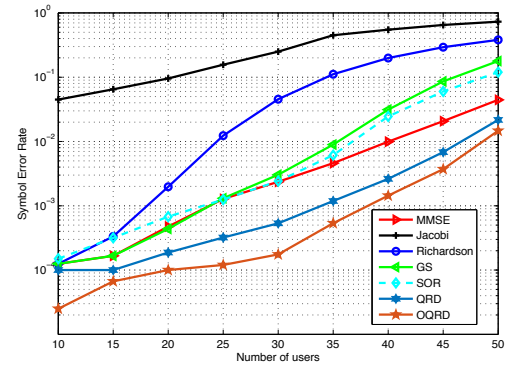


Fig. 5: SER Vs number of user (N)performance comparison of various signal detection method at 20dB SNR

VI. CONCLUSION

In this paper, we have proposed QRD and OQRD based signal detection methods for massive MIMO uplink system. The QRD method is based on the QR factorization of the original channel matrix to obtain estimated transmitted signal. Furthermore, the OQRD method is proposed which enhances the performance of QRD method. The OQRD method is based on the QR decomposition of the column norm ordering of the

channel matrix. These proposed methods are compared with various conventional signal detection methods which include Jacobi, Richardson, Gauss-Sidel, SOR and MMSE in terms of SER and computational complexity. The simulation results show that the proposed methods significantly outperforms conventional signal detection method with complexity lower than MMSE method. Therefore, the proposed OQRD method can be considered as a suitable signal detection technique for uplink massive MIMO system.

REFERENCES

[1] M. A. Albreem, W. Salah, A. Kumar, M. H. Alsharif, A. H. Rambe, M. Jusoh, and A. N. Uwaechia, "Low complexity linear detectors for massive mimo: A comparative study," *IEEE Access*, vol. 9, pp. 45 740 –45 753, 2021, DOI: 10.1109/ACCESS.2021.3065923.

[2] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive mimo detection techniques: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3109–3132, 2019, DOI: 10.1109/COMST.2019.2935810.

[3] T. Datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Random-restart reactive tabu search algorithm for detection in large-mimo systems," *IEEE Communications Letters*, vol. 14, no. 12, pp. 1107–1109, 2010, DOI: 10.1109/LCOMM.2010.101210.101587.

[4] L. G. Barbero and J. S. Thompson, "Fixing the complexity of the sphere decoder for mimo detection," *IEEE Transactions on Wireless communications*, vol. 7, no. 6, pp. 2131–2142, 2008, DOI: 10.1109/TWC.2008.060378.

[5] C. KurisummoottiL Thomas and D. Slock, "Reduced-order zero-forcing beamforming vs optimal beamforming and dirty paper coding and massive mimo analysis," in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2018, pp. 351–355, DOI: 10.1109/SAM.2018.8448769.

[6] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive mimo 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 232–257, 2020, DOI: 10.1109/TCOMM.2019.2945792.

[7] "Eigen analysis of flipped toeplitz covariance matrix for very low snr sinusoidal signals detection and estimation," *Digital Signal Processing*, vol. 129, p. 103 677, 2022, DOI: 10.1016/j.dsp.2022.103677.

[8] O. Sharifi-Tehrani, M. F. Sabahi, and M. Raees Danaee, "Efficient gnss jamming mitigation using the marcenko pastur law and karhunenloeve decomposition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 3, pp. 2291–2303, 2022, DOI: 10.1109/TAES.2021.3131400.

[9] L. Dai, X. Gao, X. Su, S. Han, I. Chih-Lin, and Z. Wang, "Low-complexity soft-output signal detection based on gauss–seidel method for uplink multiuser large-scale mimo systems," *IEEE Transactions on Vehicular Technology,* vol. 64, no. 10, pp. 4839–4845, 2014, DOI: 10.1109/TVT.2014.2370106.

[10] L. Shao and Y. Zu, "Joint newton iteration and neumann series method of convergence-accelerating matrix inversion approximation in linear precoding for massive mimo systems," *Mathematical Problems in Engineering*, vol. 2016, 2016, DOI: 10.1155/2016/1745808.

[11] C. N. Cuong, T. T. Hong, and L. D. Khai, "Hard-ware implementation of the efficient sor-based massive mimo detection for uplink," in *2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF)*. IEEE, 2019, pp. 1–6, DOI: 10.1109/RIVF.2019.8713667.

[12] X. Gao, L. Dai, Y. Hu, Z. Wang, and Z. Wang, "Matrix inversion-less signal detection using sor method for uplink large-scale mimo systems," in *2014 IEEE Global Communications Conference*. IEEE, 2014, pp. 3291–3295, DOI: 10.1109/GLOCOM.2014.7037314.

[13] J. Zhou, Y. Ye, and J. Hu, "Biased mmse soft-output detection based on jacobi method in massive mimo," in *2014 IEEE International Conference on Communiction Problem-solving*. IEEE, 2014, pp. 442–445, DOI: 10.1109/ICCPS.2014.7062317.

[14] Y. Hu, Z. Wang, X. Gaol, and J. Ning, "Low-complexity signal detection using cg method for uplink large-scale mimo systems," in *2014 IEEE international conference on communication systems*. IEEE, 2014, pp. 477–481, DOI: 10.1109/ICCS.2014.7024849.

[15] B. Kang, J.-H. Yoon, and J. Park, "Low-complexity massive mimo detectors based on richardson method," *ETRI Journal*, vol. 39, no. 3, pp. 326–335, 2017, DOI: 10.4218/etrij.17.0116.0732.

[16] S. Shafivulla and A. Patel, "Linear complexity zf-based linear precoder for massive-mimo systems," in *2021 IEEE 26th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. IEEE, 2021, pp. 1–6, DOI: 10.1109/CAMAD52502.2021.9617813.

[17] Y. Liu, J. Liu, Q. Wu, Y. Zhang, and M. Jin, "A near-optimal iterative linear precoding with low complexity for massive mimo systems," *IEEE Communications Letters*, vol. 23, no. 6, pp. 1105–1108, 2019, DOI: 10.1109/lcomm.2019.2911472.

[18] M. Chinnusami, C. Ravikumar, S. Priya, A. Arumainayagam, G. Pau, R. Anbazhagan, P. S. Varma, and K. Sathish, "Low complexity signal detection for massive mimo in b5g uplink system," *IEEE Access*, 2023, DOI: 10.1109/ACCESS.2023.3266476.

[19] N. Jain, V. A. Bohara, and A. Gupta, "Pci-mf: Partial canonical identity and matrix factorization framework for channel estimation in mmwave massive mimo systems," *IEEE Open Journal of Signal Processing*, vol. 1, pp. 135–145, 2020, DOI: 10.1109/OJSP.2020.3020002.

[20] A. Boukharouba, M. Dehemchi, and A. Bouhafer, "Low-complexity signal detection and precoding algorithms for multiuser massive mimo systems," *SN Applied Sciences*, vol. 3, no. 2, p. 169, 2021, DOI: 10.1007/s42452-020-04085-z.

[21] S. Shahabuddin, M. H. Islam, M. S. Shahabuddin, M. A. Albreem, and M. Juntti, "Matrix decomposition for massive mimo detection," in *2020 IEEE Nordic Circuits and Systems Conference (NorCAS)*. IEEE, 2020, pp. 1–6, DOI: 10.1109/NorCAS51424.2020.9264998.

[22] C. Camarero, "Simple, fast and practicable algorithms for cholesky, lu and qr decomposition using fast rectangular matrix multiplication," *arXiv preprint*, 2018, DOI: 10.48550/arXiv.1812.02056.

Jyoti Prasanna Patra received his Ph.D. and M.Tech. degrees from the National Institute of Technology Rourkela, India in 2018 and 2012, respectively. He has completed his B.Tech. degree from Biju Patnaik University of Technology, Odisha, India in 2008. Currently, he is working as an Associate Professor in the Department of Electronics and Communication Engineering at Vidya Jyothi Institute of Technology, Hyderabad, Telangana India. His research interests include signal processing for wireless communication.

Bibhuti Bhusan Pradhan received his Ph.D. and M.Tech. degrees from the National Institute of Technology Rourkela, India, in 2018 and 2012, respectively. He has completed his B.Tech. degree from BPUT University, Odisha, India, in 2007. He is currently working as a Senior Assistant Professor in the Department of Electronics and Communication Engineering at Malla Reddy Engineering College, Hyderabad, Telangana India. His research interests concentrate on performance analysis of wireless communication systems over fading channels.

M. Rajendra Prasad obtained his B. Tech from SK University, M.E from Osmania University and was awarded Ph.D. in Internet of Things from Osmania University. He is currently serving as Professor and Head of the Department of Electronics and Communication Engineering at Vidya Jyothi Institute of Technology, Hyderabad, Telangana India. His main research interests are Security algorithms for IoT Networks and Embedded System Development for mobile applications.

Resonant Radar Reflector On VHF / UHF Band Based on BPSK Modulation at LEO Orbit by MRC-100 Satellite

Yasir Ahmed Idris Humad, and Levente Dudás

Abstract—**This paper presents a novel method for identifying and tracking PocketQube satellites: the MRC-100 satellite is a model, and this method is based on a resonant radar reflector. The resonant reflector's basic concept is that the resonant reflector uses a VHF/UHF communication subsystem antenna; there is no radiated RF signal, which means the power consumption is only some Milliampere (mA). The continuous wave (CW) illuminator RF source is on the ground, and the onboard antenna receives the CW RF signal from the Earth. The microcontroller (uC) periodically switches PIN diode forming BPSK modulated signal reflection so that another Earth station can receive the backscattered Binary Phase Shift Keying (BPSK) modulated signal. Also, it can detect the satellite if the ground station receiver can use a matched filter like a correlation receiver. If the ground station receiver knows the BPSK code of the satellite, it can detect it. If not, there is no way to detect the satellite. This method is similar to Radio Frequency Identification (RFID) applications, but the reader is the ground station, and the tag is the satellite.**

Index Terms—**PocketQube, Student Satellite, Resonant Radar Reflector, Ground Station**

I. INTRODUCTION

MOST launches in the past involved a single large satellite being launched on a specialized launch vehicle. Small satellites were sometimes ' dropped off ' on the route to the primary payload's orbit or rode along with the primary payload to the final orbit. In either case, identifying primary and secondary payloads based on size and operational parameters was usually clear. By launching CubeSats close together in space, they are difficult to differentiate from one another; by launching them close together in time, Sorting out which object is which can take weeks or months at times, and some objects may never be individually identified at all. After launch, it is difficult to identify the satellite if there is no radio connection between the satellite and the ground station [1], [3], [4], [14].

Yasir Ahmed Idris Humad is a Ph.D student at the dept. of Broadband Infocommunications and Electromagnetic Theory, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics (BME), Budapest, Hungary. (e-mail: yasirahmedidris.humad@edu.bme.hu)

As a supervisor, Levente Dudás Ph.D. is a communicational and system engineer for the MRC-100 student satellite project at the dept. of Broadband Infocommunications and Electromagnetic Theory, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics (BME), Budapest, Hungary. (e-mail: dudas.levente@vik.bme.hu)

This paper's primary goal is to present a novel method for tracking and identifying PocketQuebe satellites without needing a costly tracking system. This method considers the satellites' weight, size, and power consumption while maintaining compatibility with the technology readiness level (TRL) and the global standardization of the PocketQube satellite's standard. The most recent satellite mission from BME University is the MRC-100 satellite, which is shown in the following sections to demonstrate the ability to identify and track the PocketQube satellites based on the resonant radar reflector described.

Little research has been done on the topic; the most closely related work in this field finds the French radar surveillance system's reflected signal (GRAVES). The radar-based space surveillance system Graves emits continuous waves at 143,05 MHz on the VHF band [15]. In this article, the authors could detect many hundred-kilogram satellites in low Earth orbit. PocketQube satellite has a much smaller radar-cross section (RCS) compared with hundred-kilogram satellites, so although it seems possible, research has to be done to determine the feasibility of small targets.

The PocketQube Satellite is the most recent type of nano-satellite to be proposed. It limits developers to a volume of roughly ($5 \times 5 \times 5$) cm for one unit and a mass range of 0.1 to 1 kilogram. The Microwave Remote Sensing Laboratory at BME University, in the Department of Broadband Info-communications and Electromagnetic Theory, developed various PocketQube Satellites experiments. MRC-100 is the new PocketQube Satellite and was developed over three years through the collaboration of lecturers, researchers, and students. It was given that name in honor of the Muegyetemi Radio Club (MRC), which will celebrate its 100^{th} anniversary in 2024 [16]. All small satellites were developed with significant help from the club [2], [6], [11].

MRC-100 is a 3-PQ (PocketQube) satellite with (50 × 50 × 178) mm dimension and 587 grams total mass. The main subsystems of MRC-100 are COM - Communication System, OBC - On-Board Computer, and EPS - Electrical Power System. MRC-100 contains several scientific payloads: **Resonant Radar Reflector**, spectrum analyzer 30 - 2600 MHz, 1 Mbit/s S-band down-link, automatic identification system receiver for vessel traffic services, memory-based single

DOI: 10.36244/ICJ.2024.1.4

event detector, special thermal insulator test, UHF-band LoRa-GPS Tracking, total ionizing dose measurement system, active magnetic attitude control, horizon + Sun camera, and satellite GPS + LoRa downlink (satellite identification) [1], [3], [4]. The three dimension model and the cross-sectional view of MRC-100 subsystems can be seen in Fig. 1. and Fig. 2.

The article is structured as follows: Section I overviews the introduction of the PocketQube satellite and the base sub-systems of the MRC-100 satellite. Section II discusses the power budget estimation and the orbital motion estimation of MRC-100 satellite. Section III discusses the MRC-100 communication system and ground stations. Section IV discusses the concept of the proposed resonant radar reflector based on BPSK modulation. Section V discusses the link budget estimation of the proposed System. Section VI shows the preliminary idea of the reflector and laboratory measurements results. Finally, the article's conclusion and prospects for the MRC-100 satellite are presented in Section VII.

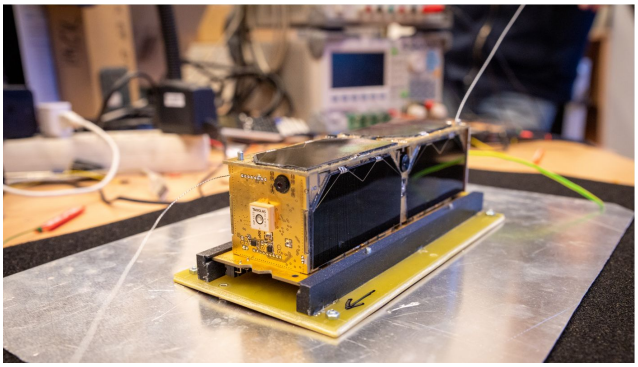


Fig. 1: MRC-100 flight module

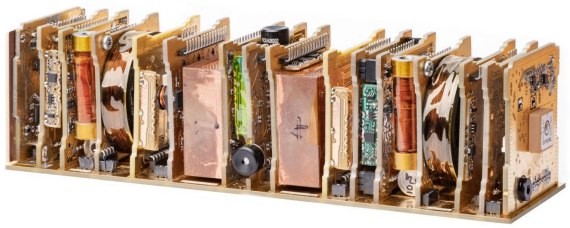


Fig. 2: Cross-sectional view of MRC-100 subsystems.

II. Power Budget and Orbital Motion Estimation of MRC-100

The MRC-100 satellite trajectory is scheduled to follow a polar, circular, and sun-synchronous Low Earth Orbit (LEO) with 600-kilometer apogee and perigee distances.

A. Power budget estimation

About 150 million kilometers is the distance between the Sun and the Earth. $1360\frac{W}{m^2}$ is the average power density around the Earth. As shown in Fig. 3. MRC-100 is covered with eight (80×40) mm three-layer solar panels from AzurSpace [7].

Due to the atmosphere, the solar power density on Earth's surface is only $1000\ \frac{W}{m^2}$ (mainly by the ozone layer). The MRC-100's three-layer solar cells have a $40\ mm \times 80\ mm$ size, a 28% efficiency, and 1.1 W of DC (Direct Current) output. The LEO of MRC-100 lasts 90 minutes, spending 60% of that time in light and 40% in darkness. As a result, the DC input averages 0.68 W with a peak of 1.7 W (on LEO, the DC input will be 36% higher) [3], [5], [12], [13]. The onboard systems of the MRC-100 is a single-point-failure tolerant and cold-redundant.

The three-layer solar cell dimension (80×40) mm and the cut-off edge of the solar cell (13.5×13.5) mm for a 1U cube (100×100) mm are both important factors in estimating the peak power of $1.7\ W$.

$$\frac{[(80 \times 40) - (13.5 \times 13.5)] \cdot 2}{(100 \times 100)} = 60\% \tag{1}$$

The solar power density around the Earth's surface equals $1000\frac{W}{m^2}$ (for $10\ cm^2$ cube equal to $10\frac{W}{cm^2}$). In equations (2) - (5) the estimation of overall DC power, the maximum DC input, the mean DC power, and the mean DC input in one orbital period (90 minutes) [3].

$$\text{Overall DC power } = 10\ W \cdot\ 60\% = 6\ \text{W} \tag{2}$$

$$\text{Maximum DC input } = 6\ W \cdot\ 28.5\% = 1.71\ \text{W} \tag{3}$$

$$\text{Mean DC power } = 1.71\ W \left(\frac{4\ sides}{6\ sides}\right) = 1.14\ W \tag{4}$$

$$\text{Mean DC input } = 1.14\ W \cdot\ 60\% = 0.684\ W \tag{5}$$

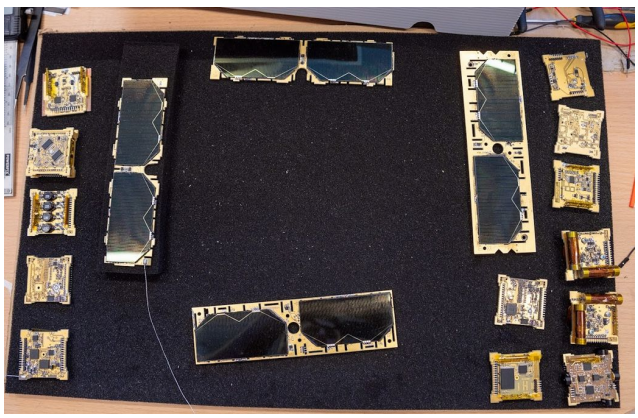


Fig. 3: MRC-100's solar cells

B. The estimation of orbital motion

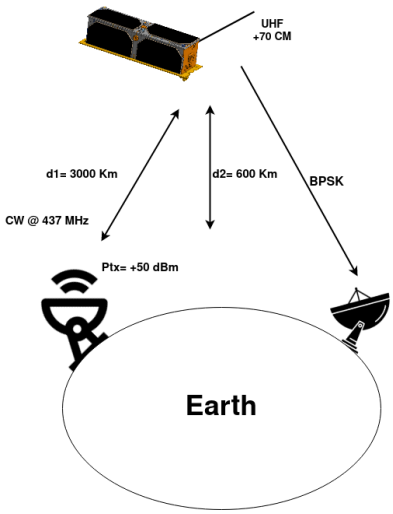


Fig. 4: Graphical depiction of the satellite horizon.

MRC-100 maximum distance from the ground station (communication in zero degrees elevation angle - horizon) is calculated by equation (6) at 600 km apogee/perigee of the orbit. Fig. 4. explains the graphical depiction of the satellite horizon and the theoretical estimation of the maximal distance between the MRC-100 satellite and the ground station, as well as the speed of the MRC-100 satellite in a circular orbit.

$$d = \sqrt{(R+h)^2 - R^2} \tag{6}$$

Where d is the maximal distance between the satellite and the ground station (where h = 600 km, R = 6,371 km, and d = 2830 km).

In a circular orbit, the speed of a satellite is calculated by equation (7).

$$v = \sqrt{\frac{g \cdot R}{1 + H/R}} = 7.55 \ \frac{km}{s} \tag{7}$$

Where g is the gravitational acceleration on the Earth's surface.

III. MRC-100 Communication System and Ground Stations

The communication system of the MRC-100 is based on an external microcontroller and Acsip LoRa and FSK radio module type S68F. As shown in Fig. 7 and Fig. 8, the communication antenna is a V-shaped dipole type 2 x quarter wavelength radiator made by space-qualified bicycle brake Bowden that emits a "quasi"-omnidirectional 3D radiation pattern. This eliminates the fading effect brought on by the 3-PQ's uncontrolled movement on the LEO. The 3D radiation pattern of the antenna on the cube-skeleton and the communication subsystem antenna can be seen in Fig. 6. As shown in Fig. 5. two independent telecommand receivers and two independent telemetry transmitters are linked to the antenna realizing cold-redundant, half-duplex functioning on the UHF radio amateur band due to the subsystem-level redundancy [9], [10], [12].

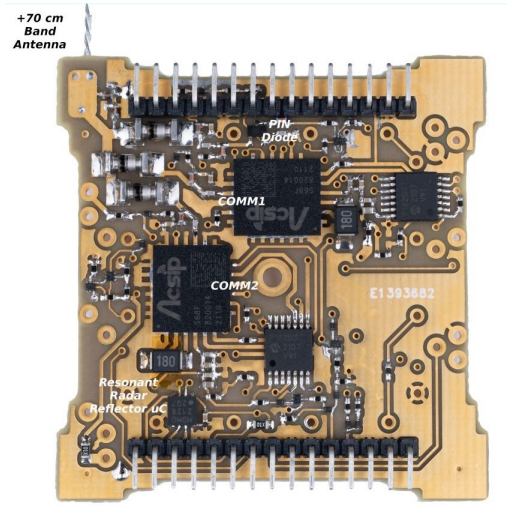


Fig. 5: The printed circuit board of the flight module.

The MRC-100 telecommand and telemetry system uses OOK (On-Off Keying), 2-GMSK (Gaussian Minimal Shift Keying as Frequency Shift Keying), and LoRa-type linear FM (Frequency Modulation) chirp. The 5000 $\frac{bit}{s}$ 2-GMSK modulation and OOK (Morse code) is used to establish a slow basic telemetry data connection. The bandwidth of the communication system of MRC-100 is licensed for operation at 12.5 kHz for uplink and 20 kHz for downlink in accordance with the international amateur radio union (IARU), and the international telecommunication union (ITU). [12], [13].

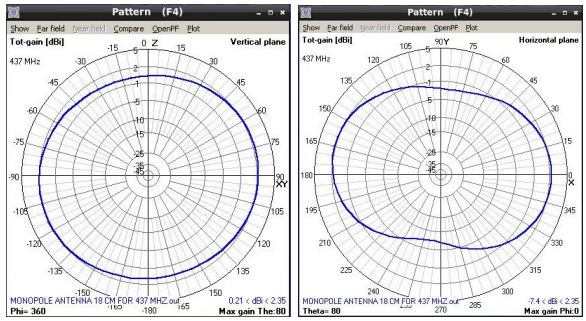


Fig. 6: The radiation pattern of the communication subsystem on Horizontal and Vertical plane.

The primary ground station (GND) is in Budapest (BME) University, as shown in Fig. 9. It has a 4.5-meter parabolic reflector-type aperture antenna with a circular back-fire helix primary radiator operating within the UHF 437 MHz band. This antenna has a notably focused main lobe, with angular dimensions of 8 degrees (-3 dB), 18 degrees (-10 dB), and 22 degrees (between null points). The antenna

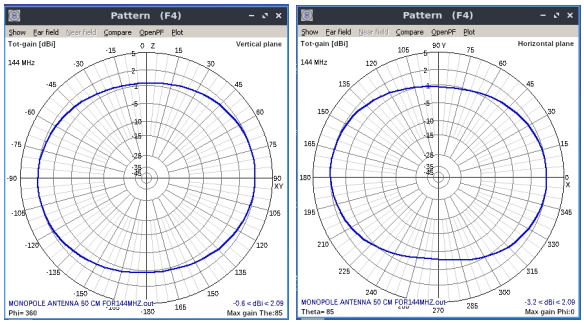


Fig. 7: The radiation pattern of the 2 meters Band on Horizontal and Vertical plane.

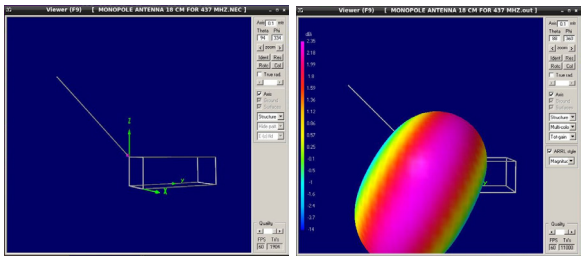


Fig. 8: The antenna on cube-skeleton and the 3D radiation pattern.

Fig. 9: Automated satellite tracking and remote-control. [6].

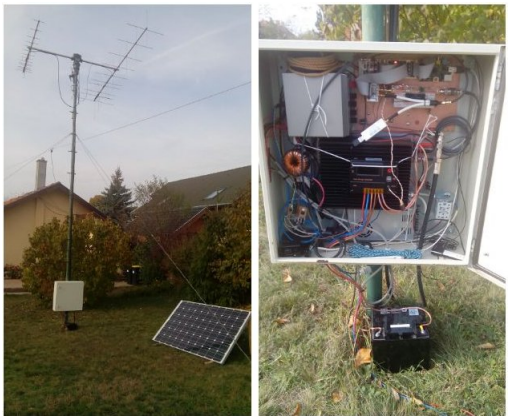


Fig. 10: The secondary GND satellite tracking [6].

gains 21 dBi for linearly polarized RF signals (or 24 dBi for circular polarization) within the region between the null points. In addition to the BME main ground station, Fig. 10. The secondary GND is located in Erd, approximately 20 kilometers from Budapest. [8], [9].

The main ground stations (GNDs) are fully automated and remote-controlled through the Internet. The core of their control system is a Raspberry PI single-board computer (SBC), capable of performing a wide range of functions. This includes the accurate azimuth-elevation antenna rotation, the detailed tracking of the satellite's path, the calculation of the Doppler shift in the RF signal, and the control of the complete radio transceiver suite including a low noise amplifier (LNA), power amplifier (PA), and coaxial relay. The primary GND's output RF power is 400 W RF + 21 dBi antenna gain, while the secondary GND's output RF power is 120 W RF + 16 dBi antenna gain. These GNDs are known about the actual operational digital data link of the MRC-100 3-PQ (5^{th}) Hungarian satellite. [8], [9].

IV. Resonant Radar Reflector Based On BPSK

The main idea of the resonant radar reflector: the antenna of the communicational system can be used as a resonant reflector if the loading RF PIN diode can form short-circuit reflection (-1) and open-loop reflection (+1) as BPSK modulation of its Radar Cross Section (similar to the conventional RFID, but the distance between the reader and the tag can reach 3000 km). From the Earth Station, it is necessary to have an illuminator RF CW signal. This signal will be reflected by the onboard antenna with binary code modulation, and other ground stations with SDRs can sense this signal. If the code is known by the ground station (GND), it can detect and identify the satellite. The satellite has no RF radiation, and the consumed DC power is a few mW. We devised this resonant radar reflector and have designed a demonstrational system using a transmitter & receiver of software-defined radio (TX-RX SDR).

A. The proposed system concept

The idea behind the proposed resonant reflector is to use an antenna for a VHF/UHF communication subsystem. No radio frequency signal is radiated onboard the satellite, so power consumption is reduced to a few milliamperes (mA). Only +3 mA is used if the loading RF PIN diode can form the shape of a short-circuit, and 0 mA if the PIN diode is open-loop. With a regulated standard bus voltage of +3.3V, the power usage is under ten mW or an average of +3.5 mA. The continuous wave (CW) illuminator RF source used by the system is based on the ground, and the onboard antenna receives the CW RF signal from the Earth. Then, the microcontroller (uC) onboard the satellite alternates the PIN diode regularly to produce binary phase shift keying (BPSK) modulated signal reflection, which enables the backscattered signal to be received by additional ground stations.
Additionally, the system can detect the satellite if the ground station receiver uses a matching filter to determine the satellite's BPSK code. On the other hand, only if the receiver

knows the satellite's BPSK code can the satellite be identified. Fig. 11. and Fig. 12. explains the proposed system's block diagrams and the ground illuminator RF source.

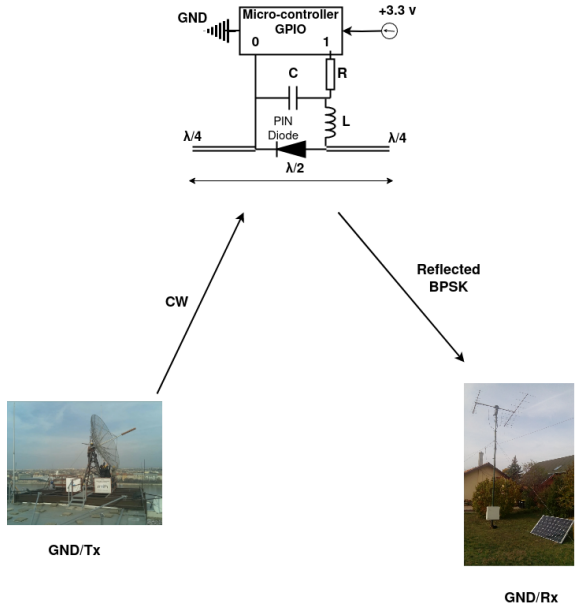


Fig. 11: Block diagram of the proposed system.

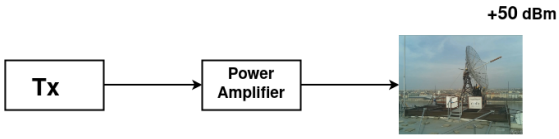


Fig. 12: Block diagram of the illuminator RF source.

The receiver part (ground segment) is realized within the proposed system as a coherent receiver, with a matched filter to precisely identify the satellite's Binary Phase Shift Keying (BPSK) code. This single-sideband (SSB) receiver receives and processes the backscattered modulated signals. It operates with an audio generator with an input signal that has a bandwidth of 2700 Hz. This signal is interfaced with an analog-to-digital converter (ADC), with seamless integration facilitated through the computer's sound card. The resulting outcome is the construction of the 'Received Vector Range and Velocity (R.V.) matrix.' In this matrix, the 'range' part is the time delay, while the 'velocity' part is the Doppler shift. This matrix contains the received BPSK code after performing a thorough and precise analysis of the received backscattered signal. Fig. 13. Shows the block diagram of the receiver.

V. Link Budget Estimation of the Proposed System

As mentioned before, the proposed trajectory of MRC-100 is polar and circular, with 600 kilometers of distance (apogee and perigee). The maximum distance of the satellite from the ground station is approximately 3000 kilometers (on the horizon), and 600 kilometers where the satellite is

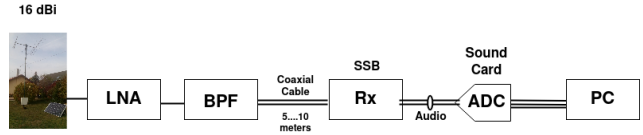


Fig. 13: Block diagram of the receiver.

perpendicular to the ground station (on the zenith). So, estimating the link budget on the UHF band (437 MHz) of the reflected BPSK modulated signal from the satellite is necessary.

The transmitted power from the ground station will be 100 W (+50 dBm). The gain of the ground station (G_{tx}) is 21 dBi, the antenna's gain onboard the satellite is 0 dBi, and (G_{rx}) is 16 dBi. The modulation loss is 10 dB, and equation (8) can estimate the free space loss on the horizon:

$$a_0 = 20 \lg \frac{4\pi d}{\lambda} = 155 \ dB \tag{8}$$

The received power (P_{r_x}) by the onboard satellite's antenna (0 dBi), where the satellite is at the horizon (3000 km) with free space loss 155 dB, and the transmitted power from the ground station is 100 W (+50 dBm) with antenna gain (21 dBi), can be estimated by equation(9).

$$P_{r_x} = P_{t_x} + G_{t_x} - a_0 = -84 \ dBm \tag{9}$$

According to the estimated received power (-84 dBm) by the onboard satellite's antenna, we can estimate the reflected power from the satellite to the ground station when the modulation loss is 10 dB and the satellite at the horizon.

$$P_{reflected} = P_{r_x} - Modulation \ Loss = -94 \ dBm \tag{10}$$

The reflected power from the satellite to the ground station when the modulation loss of 10 dB is (-94 dBm), So the received reflected power by the ground station can be estimated by (11), the antenna gain of the ground station 16 dBi.

$$P_{r_x} = P_{reflected} + G_{gnd} + G_{sat} - a_0 = -233 \ dBm \tag{11}$$

The ground station can receive thermal noise power (P_n) radiated by the environment. (12):

$$P_n = 10 \lg(kTB) + 30 = -144 \ dBm \tag{12}$$

Where k is the Boltzmann-constant, B is the bandwidth (1000 Hz), and T is the 300 K noise power level of the Earth.

The signal-to-noise ratio (SNR) can be estimated by (13) :

$$SNR = P_{rx} - P_n = -89 \ dB \tag{13}$$

The gold code generator inside the satellite's microcontroller is a linear feedback shift register (LSFR) with 10 bits length and the code length will be estimated by equation(14) .

$$Code\ Length = 2^{10\ bits} - 1 = 1023 \tag{14}$$

The processing gain can be estimated by (15) :

$$Processing\ Gain = 10\lg(Code\ Length) = 30\ dB \tag{15}$$

The received reflected power by the ground station is (-233 dBm), and the processing gain is 30 dB; the received power at the matched filter can be calculated by (16):

$$P_{r_x} = P_{rx}gnd + Processing\ Gain = -202\ dB_m \tag{16}$$

After the matched filter, we can estimate the bandwidth of the received signal, the thermal noise power level, and the signal-to-noise ratio (17) - (19) .

$$BW_{MF} = \frac{1000\ Hz}{\text{Code Length}} \simeq 1\ Hz \tag{17}$$

$$P_n = 10\lg(kTB) + 30 = -174\ dBm \tag{18}$$

Where B is the Bandwidth of the received signal after the matched filter (1 Hz), k is the Boltzmann-constant, and T is the 300 K noise power level of the Earth.

The signal-to-noise ratio after the matched filter can be calculated by (19):

$$SNR = P_{rx} - P_n = -29\ dB \tag{19}$$

The code length time can be estimated by (20) :

$$Code_T = \frac{Code\ Length}{\text{BW}} = 1.024\ s \tag{20}$$

TABLE I
HORIZON / ZENITH LINK BUDGET ESTIMATION ON UHF BAND.

Parameters	Horizon 3000 Km	Zenith 600 Km
Frequency	437 MHz	437 MHz
Wavelength	0.69 m	0.69 m
P_{t_x}	+50 dBm	+50 dBm
G_{gnd1}	21 dBi	21 dBi
G_{sat}	0 dBi	0 dBi
G_{gnd2}	16 dBi	16 dBi
a0 to Satellite	155 dB	141 dB
P_{r_x}@ sat	-84 dBm	-70 dBm
modulation loss	10 dB	10 dB
$P_{ref}from$ sat	-94 dBm	-80 dBm
P_{r_x}@ gnd	-233 dBm	-205 dBm
Bandwidth	1000 Hz	1000 Hz
Noise Temp	300 K	300 K
Thermal Noise Power	-144 dBm	-144 dBm
SNR	-89 dBm	-61 dBm
Gold Code Generator Length	10 bits	10 bits
Code Length	1024 chips	1024 chips
Processing Gain	30 dB	30 dB
Code Time	1.024 s	1.024 s
P_{r_x}@ MFout	-202 dBm	-175 dBm
BW@ MFout	1 Hz	1 Hz
P_n@ MFout	-174 dBm	-174 dBm
SNR@ MFout	-29 dB	-1 dB

The proposed reflector's system can be evaluated within the Very High Frequency (VHF) band when the satellite is at a distance of 3,000 kilometers. The transmitted power from the ground station amounts to 100 watts, equivalent to +50 decibels-milliwatts (dBm). The ground station exhibits a transmit gain (G_{tx}) of 16 decibels isotropic (dBi) and a receive gain (G_{rx}) of 13 dBi, and the satellite's antenna gain is measured at 0 dBi. The modulation loss introduces an attenuation of 10 decibels (dB), while the Gold code length is 10 bits. Consequently, the estimation of the link budget is summarized in Table II.

TABLE II
HORIZON / ZENITH LINK BUDGET ESTIMATION ON VHF BAND.

Parameters	Horizon 3000 Km	Zenith 600 Km
Frequency	144 MHz	144 MHz
Wavelength	2.08 m	2.08 m
P_{t_x}	+50 dBm	+50 dBm
G_{gnd1}	16 dBi	16 dBi
G_{sat}	0 dBi	0 dBi
G_{gnd2}	13 dBi	13 dBi
a0 to Satellite	145 dB	131 dB
P_{r_x}@ sat	-79 dBm	-65 dBm
modulation loss	10 dB	10 dB
$P_{ref}from$ sat	-89 dBm	-75 dBm
P_{r_x}@ gnd	-221 dBm	-193 dBm
Bandwidth	1000 Hz	1000 Hz
Noise Temp	300 K	300 K
Thermal Noise Power	-144 dBm	-114 dBm
SNR	-77 dBm	-80 dBm
Gold Code Generator Length	10 bits	10 bits
Code Length	1024 chips	1024 chips
Processing Gain	30 dB	30 dB
Code Time	1.024 s	1.024 s
P_{r_x}@ MFout	-191 dBm	-163 dBm
BW@ MFout	1 Hz	1 Hz
P_n@ MFout	-174 dBm	-174 dBm
SNR@ MFout	-17 dB	+11 dB

Considering the orbital movement and Two-Line Element (TLE) data, MRC-100 passes over Hungary three to six times daily, lasting approximately 10 minutes. Consequently, it becomes imperative to gauge the integration time required for the back-scattered Binary Phase Shift Keying (BPSK) modulated signal. Our experiment operates for 10 seconds every minute, equating to 100 seconds during one pass. Thus, the cumulative experimentation duration totals 300 seconds for three passes and 600 seconds for six passes. The overall integration time (T_i) on the horizon equals 10000 seconds. Referencing equation (21)., we determine the optimal Integration Gain (IG) for the received back-scattered BPSK modulated signal after the matched filter according to our experiment time.

$$IG = 10\log_{10}\cdot(\frac{T_i}{CodeTime}) = 40\ dB \tag{21}$$

Based on Equation (21), the integration gain (IG) is based on the measurement time, considering that the 1024 BPSK samples correspond to a bit time of one millisecond, resulting in a total code time of 1024 milliseconds. Consequently, IG is equivalent to +40 dB, signifying that the receiver can effectively discern the backscattered BPSK code.

VI. PRELIMINARY CONCEPT OF THE REFLECTOR AND LABORATORY MEASUREMENTS RESULTS

The resonant radar reflector system was functionally validated using a proof-of-concept technique based on laboratory measurements. The proposed scenario refers explicitly to creating the demonstrational system using a transmitter and receiver of software-defined radio (B200 Tx/Rx SDR) connected to two log periodic antennas. The log periodic antenna works in the range of the DVB-T band with linear gain (6 dBi). The reflector's system consists of a Raspberry Pi loaded with an RF PIN diode and a quarter-wave antenna. The RF PIN diode form short-circuits (-1 reflection) and open-loop (+1 reflection) as BPSK modulation. The maximum distance between the transmitter and the reflector's system is 22 meters, and the transmitted power (P_{tx}) is -43 dBm. Fig. 14. Explains the block diagram of the experimental model, Fig. 15 and Fig. 16. Shows the realized experimental model of the reflector's system with the antenna.

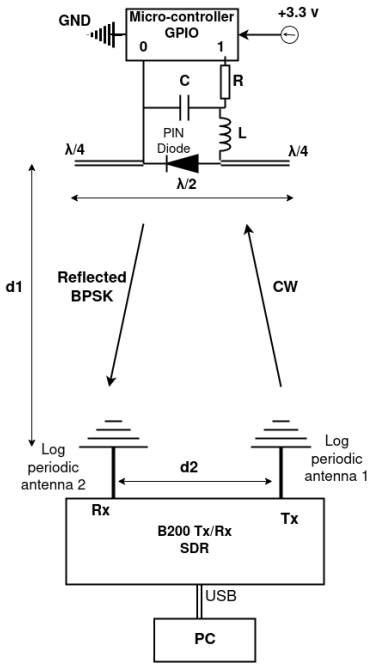


Fig. 14: Block diagram of the experimental model.

The realized measurement - SDR screen, as shown in Fig. 17. the upper part represents a waterfall diagram of the received reflected signal from the reflector, the middle part represents the real and imaginary part of the received reflected signal, and the bottom one represents the spectrum of the received reflected signal.

As shown in Fig. 18., the upper part is a waterfall diagram of the received reflected signal from the reflector when the transmitter is sending a continuous wave (CW) at 437 MHz (signal variation in time). The middle is the real and imaginary part of the received reflected signal. The fundamental part of the signal is higher than the imaginary part because the received backscattered signal is BPSK modulated signal. The bottom one is the spectrum of the received reflected signal, which seems like $\frac{sinx}{x}$.

Fig. 15: The realized experimental model of the reflector's system.

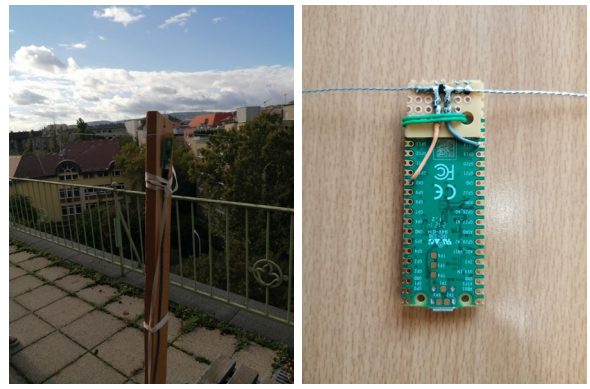


Fig. 16: The realized reflector model with the antenna.

Fig. 19. Shows the peak sidelobe level (PSL) of the received BPSK code from the reflector after the matched filter at the receiver part, and also it can be estimated theoretically by (22) , when the code length is equal to 1024.

$$PSL = 20 \log Code\ Length = 60.21\ dB \qquad (22)$$

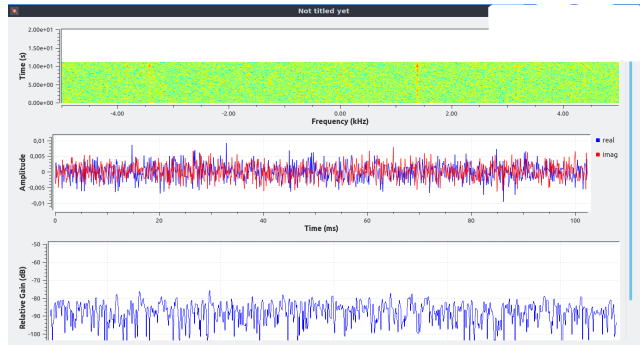


Fig. 17: The realized measurement - SDR screen.

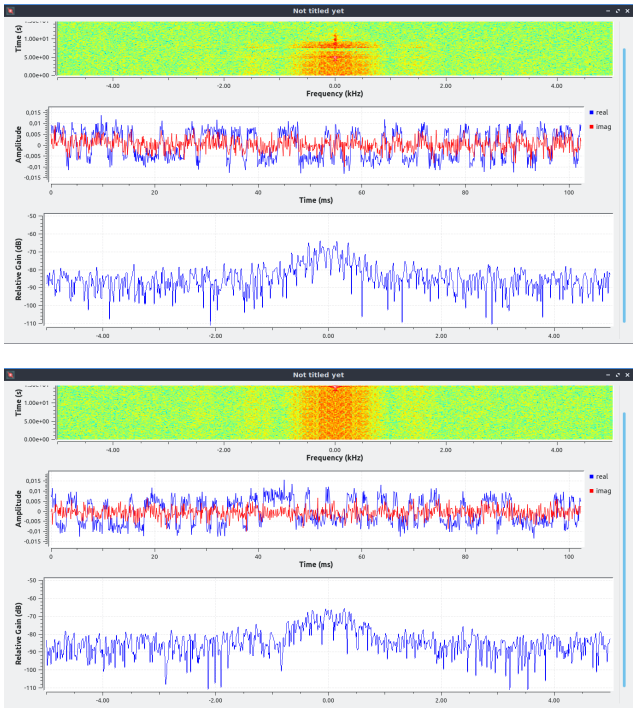


Fig. 18: 2 Figures of received reflected signal from the reflector's.

Fig. 19: 2 Figures of the received code as backscattered BPSK modulated signal.

VII. CONCLUSION

The work presented in this paper described how to identify and track PocketQube satellites based on a resonant radar reflector, and the MRC-100 satellite is a model. The goal of this paper was to establish a new method to track and identify PocketQube satellites without an expensive tracking system from the perspective of power consumption, weight, and size compatible with the global standardization of the PocketQube satellite's standard and the technology readiness level (TRL). We estimated the link budget of the system on the VHF/UHF band when the satellite on the horizon region like communication in zero degrees elevation angle and the zenith region when the satellite is perpendicular to the ground station, and we estimated the optimal integration gain that the receiver could effectively discern the backscattered BPSK code. Furthermore, the Resonant Radar Reflector was functionally validated at the laboratory. It can receive a continuous wave (CW) from the ground stations and backscatter the received signal as a Binary Phase Shift Keying (BPSK) modulated signal with a maximum distance between the transmitter and the reflector's system of 22 meters. The transmitted power is -43 dBm, and the Doppler shift is 0 Hz. We started to record the received backscattered BPSK modulated signal from the MRC-100 satellite, and the extended version will contain the Range and Velocity matrix (time delay and Doppler shift). The MRC-100 was successfully installed on the satellite platform in February 2023 and launched into outer space via a Falcon-9 rocket from the USA on 12 June 2023. The first signal from the MRC-100 was received on 22 June 2023.

REFERENCES

[1] Humad, Y. A. I., Dudás, L. "Extended Wide-band Spectrum Monitoring System from 2.2 GHz to 2.6 GHz by MRC-100 3-PocketQube Class Student Satellite." In 2022 24th International Microwave and Radar Conference (MIKON), pp. 1–5, IEEE, 2022. DOI: 10.23919/MIKON54314.2022.9924834

[2] Humad, Y. A. I., Dudás, L. "GPS Type Tracker Based On LoRa Transmission for MRC-100 3-PocketQube Student Satellite." In 2022 13th International Symposium on Communication System, Network and Digital Signal Processing (CSNDSP), pp. 98–102. IEEE, 2022. DOI: 10.1109/CSNDSP54353.2022.9907953

[3] Humad,Y. A. I., Dudás, L."Wide-band Spectrum Monitoring System from 30MHz to 1800MHz with limited Size, Weight and Power Consumption by MRC-100 Satellite," Infocommunication Journal, vol. 14, no. 2, pp. 56–63, 2022. DOI: 10.36244/icj.2022.2.6.

[4] Humad, Y. A. I., Dudás, L. "A Wide-band Spectrum Monitoring System as a scientific payload for MRC-100 3-PQ (Pocket Qube) student satellite." In 7th International Conference on Research, Technology and Education of Space April., pp. 26–30. 2022. https://space.bme.hu/wp-content/uploads/2022/09/Proceedings_Papers_HSPACE-2022.pdf.

[5] Dudás, L., Varga, L., Seller, R. "The communication subsystem of Masat-1, the first Hungarian satellite." In Photonics Application in Astronomy, Communication, Industry, and High-Energy Physics Experiments 2009, vol. 7502, pp. 184–193. SPIE, 2009. DOI: 10.1117/12.837484

[6] "Home Page - SMOG - 1," Home Page - SMOG - 1. [Online]. Available: http://152.66.80.46/smog1/satellites.pdf. [Accessed: April 3, 2023].

[7] "– Home - AZUR SPACE Solar Power GmbH," – Home - AZUR SPACE Solar Power GmbH. [Online]. Available: http://www.azurspace.com/images/0003429-01-01_DB_3G30C-Advanced.pdf. [Accessed: April 1, 2023].

[8] Dudás, L., Pápay, L., Seller, R. "Automated and remote controlled ground station of Masat-1, the first Hungarian satellite." In 2014 24th International Conference Radioelekteronika, pp. 1–4. IEEE, 2014. DOI: 10.1109/Radioelek.2014.6828410

[9] "Home Page - SMOG - 1," Home Page - SMOG - 1. [Online]. Available: https://gnd.bme.hu/smog. [Accessed: April 5, 2023]

[10] Takács, D., Markotics, B., Dudás, L. "Processing and Visualizing the Low Earth Orbit Radio Frequency Spectrum Measurement Results From the SMOG Satellite Project," Infocommunication Journal, vol. 13, no. 1, pp. 18–25, 2021. DOI: 10.36244/icj.2021.1.3.

[11] Dudás, L., Gschwindt, A. "Filling the Gap in the ESA Space Technology Education." In 4th International Conference on Research, Technology and Education of Space Feb., pp. 15–16. 2018. https://space.bme.hu/wp-content/uploads/2019/01/Proceedings_abstracts_HSPACE2018.pdf

[12] Dudás, L., Gschwindt, A. "The communication and spectrum monitoring system of Smog-1 PocketQube class satellite." In 2016 21st International Conference on Microwave, Radar and Wireless Communication (MIKON), pp. 1–4. IEEE, 2016. DOI: 10.1109/MIKON.2016.7491999

[13] Dudás, L., Szucs, L., Gschwindt, A. "The Spectrum Monitoring System of Smog-1 Satellite". 14th Conference on Microwave Techniques, pp. 143–146, ISBN: 978-1-4799-8121-2.CO-MITE 2015. DOI: 10.1109/COMITE.2015.7120316

[14] Humad, Y. A. I., TagElsir, A., Daffalla, M. M. "Design and implementation of communication subsystem for ISRASAT1 Cube Satellite." In 2017 International Conference on Communication, Control, Computer and Electronic Engineering (ICCCCEE), pp. 1–4. IEEE, 2017. DOI: 10.1109/ICCCCEE.2017.7867651

[15] Mieczkowska, D., Wójcicki, J., Szewczak, P., Kubel-Grabau, M., Zaborowska, M., Zielińska, U., ... & Szabó, V. (2017, September). Detection of objects on LEO using signals of opportunity. In 2017 SIGNAL Processing Symposium (spsympo) (pp. 1–6). IEEE. 2017. DOI: 10.1109/SPS.2017.8053660

[16] Herman, T., & Dudás, L. (2023). Picosatellite identification and Doppler estimation using passive radar techniques. Infocommunications Journal, 15(3), 11–17. 2023. DOI: 10.36244/ICJ.2023.3.2

Yasir Ahmed Idris Humad Ph.D. candidate at the Faculty of Electrical Engineering and Informatics, Department of Broadband Infocommunications and Electromagnetic Theory, Budapest University of Technology and Economics (BME). He received his Bachelor's Degree in Electronic Engineering (Telecommunication) from AL-Neelain University, Faculty Of Engineering, in 2012. He continued his graduate studies at AL-Neelain University Faculty of Engineering and received his MSc in Electronic Engineering (Data and Communication Networks) in 2015. He enrolled as a research assistant at the National Center for Research, Institute of Space Research and Aerospace (ISRA) Sudan in 2014. His current research areas include CubeSat and PocketQube-type satellite development; analog RF hardware and antenna design; automated and remote-controlled satellite control station development; Software Defined Radio-based signal processing for satellites. Yasir is a member of the MRC-100 satellite development team, and he is responsible to developed Three scientific payloads: spectrum analyzer (30–2600) MHz, UHF-band LoRa-GPS Tracking, and Resonant Radar Reflector.
https://gnd.bme.hu/smog.

Levente Dudás At the Budapest University of Technology and Economics, he got his MSc in electrical engineering in 2007 and his Ph.D in 2018 in Radar and Satellite Applications of Radio and Antenna Systems (BME). He is currently a senior lecturer at BME's Department of Broadband Infocommunications and Electromagnetic Theory, an electrical engineer in the Microwave Remote Sensing Laboratory, and the president of the BME Radio Club. Microwave Remote Sensing (RADAR) lab. is working on: active and passive radars; CubeSat and PocketQube satellite development; analog RF hardware and antenna design; automated and remote-controlled satellite control station development; Software Defined Radio based signal processing for satellite and radar applications are just a few of his research interests.
http://radarlab.hvt.bme.hu/, https://gnd.bme.hu/.

Advancements in Expressive Speech Synthesis: a Review

Shaimaa Alwaisi and Géza Németh

***Abstract*—In recent years, we have witnessed a fast and widespread acceptance of speech synthesis technology in, leading to the transition toward a society characterized by a strong desire to incorporate these applications in their daily lives. We provide a comprehensive survey on the recent advancements in the field of expressive Text-To- Speech systems. Among different methods to represent expressivity, this paper focuses the development of expressive TTS systems, emphasizing the methodologies employed to enhance the quality and expressiveness of synthetic speech, such as style transfer and improving speaker variability. After that, we point out some of the subjective and objective metrics that are used to evaluate the quality of synthesized speech. Finally, we point out the realm of child speech synthesis, a domain that has been neglected for some time. This underscores that the field of research in children's speech synthesis is still wide open for exploration and development. Overall, this paper presents a comprehensive overview of historical and contemporary trends and future directions in speech synthesis research.**

***Index Terms*—Speech style, Expressivity, Emotional speech, Expressive TTS, Prosody modification, Multi- lingual and multi-speaker TTS**

I. INTRODUCTION

THE objective of this study is to explore the latest advancements in speech synthesis research. It is primarily intended for researchers involved in the development and enhancement of Text-to-Speech (TTS) systems, as well as professionals in various fields that utilize TTS applications, including such as customer service, navigation systems, and language education [1].

TTS is a process that converts written text into speech like that of humans [2]. Speech serves is a crucial element in human interaction and verbal communication. Throughout history, people have relied on speech as an effective means of conveying information, expressing themselves, and revealing their emotional state [3]. We communicate using various speech styles, which can differ based on factors such as the subject, environment, and culture [2]. In other words, speech styles depend on the content, context, and audience. They can range from formal to casual.

In recent years, advancements in speech technology have led to the development of artificial speech that closely resembles human speech in terms of naturalness and intelligibility. This technology, also known as speech synthesis, takes text as input and generates speech as output. Modern TTS systems have evolved from a long history of efforts to create synthesized human language from written text.

Numerous TTS applications have achieved impressive levels of naturalness and intelligibility. Key factors contributing to naturalness include expressiveness, emotion, and speech style. Modern TTS systems need to deliver synthesized speech in the desired style for users. Expressivity pertains to the manner in which thoughts, emotions, and information are conveyed through a specific expressive style [1] [4].

Speech style in speech synthesis is influenced by various factors, such as the topic, language, speech rate and intensity, and regional culture of the spoken language. Linguistically, expressivity refers to communicating positive or negative ideas or emotions in a style that is relevant to the listener. Emotional expression serves as a vocal indicator of emotions, which is evident in the speech waveform [5]. In addition to speech styles, emotions are also considered expressions. Different expressive styles can be generated based on two approaches: corpus-driven and prosodic-phonology approaches. The corpus-driven approach involves analyzing large datasets of speech to extract patterns of prosody associated with different emotions. This data is then used to train a machine learning model to predict the appropriate prosody for a given text.

The prosodic-phonology approach, on the other hand, involves modeling the underlying linguistic and phonological features of speech that give rise to different emotional expressions. This approach involves analyzing the sound units such as fundamental frequency F0 and duration [6].

Department of Telecommunications and Media Informatics Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, Hungary
(E-mail: shaima.alwaisi@edu.bme.hu, nemeth@tmit.bme.hu)

DOI: 10.36244/ICJ.2024.1.5

Neutral Speech Synthesis Systems (NSS) generate speech from written text in a single style, often referred to as neutral or flat sound [7]. Fig 1. presents a framework for a Neutral Speech Synthesis System (NSS). The text analysis and processing stage known as front-end involves the analysis of the written text to extracting linguistic features. This stage provides linguistic and acoustic features to the back-end stage, where the acoustic features of the speech signal are generated.

The back-end stage is where the speech signal is synthesized from the linguistic features extracted in the text analysis and processing stage. This stage involves converting the linguistic features into acoustic features, such as pitch, duration, intensity, and spectral characteristics, to generate a natural-sounding speech signal [8]. Linguistic features are derived through syntactic, semantic, and lexical analysis steps, which guide the synthesis process to produce neutral speech [9]. To generate speech with a specific expressive style, the desired expressive style is incorporated as an additional input to the TTS model, as depicted in Fig. 1 [10] [11].

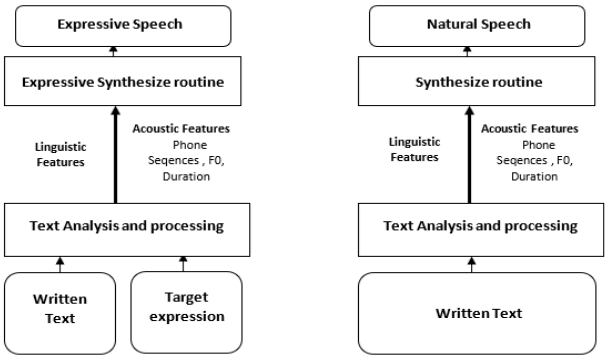


Fig 1: TTS and Expressive TTS System Architecture. On the left side, a schematic diagram illustrates the expressive TTS system. This system processes input text along with the desired expressive elements. On the right side, natural speech is generated by a natural TTS system [12].

The benchmark for evaluating speech technologies is human ratings. Traditionally, listeners are tasked with listening to speech samples and providing ratings, either in isolation or within a context. However, researchers face challenges in evaluation of the new system, since Ratings are subjective, varying from person to person. This subjectivity becomes more pronounced when listeners have limited context and training [13].

The most common method of evaluation is the MOS (Mean Opinion Score) test [14]. This test involves collecting MOS scores from listeners who evaluate each utterance in isolation. In this method, listeners assign scores to individual utterances, typically on a five-point scale, with 5 score representing highly natural speech and 1 score representing highly unnatural speech. Unlike MOS tests, where ratings are provided in isolation, the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) test involve listeners in a multiple comparison test. MUSHRA test offers enhanced the sensitivity to subtle differences between stimuli compared to MOS tests [13].

This paper follows the following structure: Section II, since current TTS systems based on the advance Deep learning algorithms, we first introduce several Deep learning models widely used in TTS systems. In Section III and IV, we review some of subjective and objective metrics that are used for evaluating TTS models. In Section V, we summarize style representation and transfer methods. Section VI. discusses prosody modelling in speech synthesis. In Section VII. we take attempt to point out some challenges in child speech synthesis. Finally, in Section VIII, the paper concludes with a summary of the findings and possible future directions.

II. DEEP LEARNING BASED SPEECH SYNTHESIS

Deep Neural Networks (DNNs) [15] play a crucial role in modern speech synthesis approaches, such as WaveNet [16] and Tacotron [17]. The shift from decision trees to deep learning methods has led to significant improvements in the quality of synthesized speech. This shift also involves a move from (HMM) to frame prediction using deep learning models, contributing to the notable improvements in speech synthesis quality. However, a closer examination of the literature reveals several challenges, including the need for substantial computational resources and large speech datasets to train TTS models. Moreover, recording speech datasets with professional speakers can be costly. These challenges have been addressed through various knowledge transfer approaches, such as fine-tuning, transfer learning, and multi-task learning [18].

A. Back-End Synthesizer

WaveNet, proposed by Google DeepMind in 2016, is a deep learning-based autoregressive approach. This fully probabilistic and autoregressive model generates synthesized speech that closely resembles natural audio waveforms. WaveNet's architecture is based on a Convolutional Neural Network (CNN) trained with speech samples to predict natural speech, with each sample depending on the previously generated ones. WaveNet serves as a vocoder for TTS models, with inputs consisting of linguistic features, predicted log fundamental frequency (F0), and phoneme durations [18].

For expressiveness prediction, non-autoregressive WaveNet blocks outperform the original WaveNet [19]. Multi-speaker WaveNet vocoders have demonstrated higher performance compared to traditional methods [20]. Parallel WaveNet combines WaveNet and the Inverse Autoregressive Flow (IAF) method. Inverse-autoregressive flows (IAFs) are generative models used for high dimensional observable samples. High-dimensional observable samples typically refer to a large set of acoustic features that are extracted from speech signals, such as the fundamental frequency, spectral envelope, and time-varying spectral parameters [21]. Capable of generating speech in a wide range of styles (emotional, neutral, conversational, long-form reading, news briefing, and singing), Parallel WaveNet [22] uses a vocoder trained on a multi-speaker emotional dataset to convert Mel spectrograms from a neutral style to various emotional styles [23].

A novel speech synthesis system called Autovocoder has been proposed by [24] to generates high-quality audio and outperforms other waveform generation systems in terms of computational cost. Autovocoder is trained as a denoising autoencoder and generates a waveform at a speed 5 times greater than Griffin-Lim algorithm [25] and 14 times faster than the neural vocoder HiFi-GAN [26].
Autovocoder utilized parallel computing and data parallelism techniques by leveraging fast, Differentiable Digital Signal Processing DSP operations, a purely convolutional residual network, and a learned representation to achieve efficient and fast waveform generation.
Generative Adversarial Networks (GANs) have emerged as a powerful tool for generating high-quality audio, including speech synthesis. GAN-based vocoders are a type of vocoder that uses GANs to generate raw waveform audio from acoustic features and linguistic information. This approach offers several advantages over traditional vocoders, such as improved audio quality, expressiveness, and robustness to noise. Among the various GAN-based vocoders that have been developed, prominent instances of well-known models include HiFi-GAN [26], SnakeGAN [27],Parallel WaveGAN [28], and BigVGAN [29].

B. Linguistic Analysis and Prosody Front-End

Front-end models play a crucial role in processing input text into intermediate representations, often involving linguistic features or phonetic information. Tacotron is an end-to-end Text-to-Speech (TTS) system that uses deep neural networks to generate natural-sounding speech from text input. It operates by predicting mel-spectrograms from text characters, which are then converted into time-domain waveforms using a vocoder [30]. Tacotron2 is a generative model that combines an encoder-decoder architecture with a soft attention mechanism to generate spectrograms from a given text [31]. The primary concept of the attention mechanism is to identify the most relevant characters for each Mel spectrogram frame and determine weights for each character embedding [31].

Tacotron2 has been employed to enhance the expressivity of multi-speaker end-to-end TTS models. The expressivity of latent representation is used for predictions made by the encoder to derive emotion [32]. Text-Predicting Global Style Token (TP-GST) is combined with Tacotron to generate speech in a specific style. Style attention, prosody encoder, and style embedding are added to Tacotron. During the training phase, a combination of trainable embeddings is extracted to be shared across the entire text, driving the Global Style Tokens [33].

FastSpeech [34], another notable front-end model, introduces a novel feed-forward network that generates mel-spectrograms in parallel, utilizing feed-forward Transformer blocks, a length regulator, and a duration predictor. FastSpeech incorporates a phoneme duration predictor to ensure hard alignments between phonemes and mel-spectrograms, reducing the ratio of skipped words and repeated words and contributing to high audio quality. FastSpeech aims to address several challenges present in traditional autoregressive text-to-speech (TTS) models. These challenges include slow inference speed, lack of robustness leading to word skipping and repeating, and limited controllability over voice speed and prosody.

TABLE 1
WAVENET-BASED EXPRESSIVE VOCODERS

Reference	Expressions	Evaluation method	Parameters	Findings
[23]	Happy, angry, sad	Mean Opinion Score (MOS)	linear-scale log magnitude spectrograms and mel spectrograms Using dynamic time warping (DTW) to align them	Implementing WaveNet vocoder to generate speech from melspectrograms led to overall improvement regarding the quality of synthesized speech of each emotion.
[98]	Normal, happy, angry	Mean Opinion Score (MOS)	mel-spectrum parameters and Emotion ID (EID)	Proposed model successfully gener emotional speech t into account mel-spectrum parameters
[22]	Emotional, Neutral, Conversational Long-form reading, News briefing and Singing	MUSHRA	mel-spectrograms	The proposed method synthesized speech with various styles and languages in real-time

FastPitch [35], a fully parallel text-to-speech model, draws its foundation from Fast Speech. FastPitch conditioned on fundamental frequency contours. It predicts pitch contours during inference, allowing for more expressive and engaging speech. The model retains the favorable, fully parallel Transformer architecture.

On the other hand, FastSpeech2 [36] represents a paradigm shift in text-to-speech modeling as a non-autoregressive system. FastSpeech2 simplifies the training pipeline, improves voice quality, and introduces more variation information of speech, such as pitch and energy, as conditional inputs. It provides variance information such as pitch, energy, and more accurate duration. The model architecture includes a pitch predictor, pitch contour, and pitch spectrogram, allowing for manual manipulation of pitch, duration, and energy in synthesized speech.

III. SUBJECTIVE METRICS

1- Mean Opinion Score (MOS): This is a widely used subjective measurement for evaluating the quality of synthesized speech. Listeners rate the speech using a numerical scale ranging from 1 to 5, where 5 signifies excellent speech quality and 1 represents the lowest quality. MOS is a subjective method recommended by standardization bodies such as IEEE Subcommittee. During the listening test, listeners complete a questionnaire that may include sections on overall impression, listening effort, pronunciation, speaking rate, articulation, and voice pleasantness [37].

2- AB Preference Test: In AB Preference Test, participants are presented with audio samples from two distinct speech synthesis models, denoted as model A and model B. Participants listen to samples from both systems and express their preference [38].

3- ABX Preference Test: Participants listen to three speech versions—A, B, and X—with X being the target speech and A and B being two synthesized speech sentences generated by different models. Test subjects are asked to choose which synthesized version is closer to the target speech X[38].

4- MUSHRA (Multiple Stimuli with Hidden Reference and Anchor): In a MUSHRA test, participants evaluate systems on a scale ranging from 1 to 100. They accomplish this by listening to stimuli for the same text presented side-by-side, in comparison to a high-quality reference. This method facilitates a comprehensive assessment of multiple systems, allowing for a nuanced ranking based on perceived quality [39].

IV. OBJECTIVE METRICS

Objective measurements involve the quantitative evaluation of speech synthesis systems, providing a mathematical assessment of the quality of synthesized speech.

A. Itakura-Saito measure

This method is like most objective methods for the evaluation of TTS Models divides the speech signal into frames. Let $s(i)$ and $s'(i)$ be two sampled speech signals, and $x_n(i)$ and $x'_n(i)$ are two windowed frames generated from implementing a window equation $w(i)$, where n is the frame index designating the window location.

$$x_n(i) = w(i)s(i + n) \quad (1)$$

$$c = w(i)s'(i + n) \quad (2)$$

We indicate the z-transform of $x_n(i)\ and\ x_n(i)$ by $X_n(z)$ and $X'_n(z)$. The Fourier transform is derived by assessing the z-transform on the unit circle, i.e., $z = e^{j\omega}$. The $X_n(e^{j\omega})$ and $X'_n(e^{j\omega})$ are utilized to represent the Fourier transforms of two signals that have been windowed, respectively. Then for each pair of $X_n(e^{j\omega})$ and $X'_n(e^{j\omega})$, spectral distortion $p[Xn, X']$ is defiend as the dissimilarity among $X_n(e^{j\omega})$ and $X'_n(e^{j\omega})$, the Itakura-Saito formula for speech analysis is defined as below [40].

$$p_{is}[X_n, X'_n] \triangleq \int_{-\pi}^{\pi}\left[\left(\frac{|X_n(e^{j\omega})|^2}{|X'_n(e^{j\omega})|^2}\right) - \Lambda(\omega) - 1\right]\frac{d\omega}{2\pi} \quad (3)$$

where

$$\Lambda(\omega) = log\ |X_n(e^{j\omega})|^2 - log\ |X'_n(e^{j\omega})|^2 \quad (4)$$

B. *Root mean square (RMSE)*

RMSE is a mathematical measure used to evaluate log f_0 trajectories produced by TTS models, it is stated as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log(F0_i) - \log(F0'_i))^2} \quad (5)$$

Where $F0_i$ and $F0'_i$ stand for the original and predicted F0 features, respectively. and N is the length of the F0 sequence [41].

C. Gross pitch error (GPE)

GPE is the proportion of segments that are measured as voiced for natural and generated speech having relative pitch error higher than a certain threshold (usually taken as 20% in speech analysis) [42].

V. STYLE REPRESENTATION AND TRANSFER

A. Global Style Token

In text-to-speech, Global Style Tokens (GSTs) are a recently proposed method for extracting style embedding features that reflect specific speech styles. GSTs introduce an auxiliary input vector to the speech synthesis model to control the global style of the synthesized speech. Style tokens are global features of speech style that can be adjusted to synthesize speech in a target style. Modern GST architectures have been developed to learn latent representations of high-dimensional speech data [43] . An attention mechanism calculates attention weights for style tokens, and the sum of style tokens is used for style embeddings. During the training phase, style tokens are initially created randomly, and they learn speech styles in an unsupervised manner.

In [44], a Global Style Token (GST) network is combined with an augmented version of Tacotron to capture expressive variations in speech style. The GST network processes GST combination style embeddings as expressive style labels that are jointly predicted within Tacotron. The TP-GST network extracts weights or style embedding space from text alone, without explicit labels during training phases. Two text-prediction pathways, Predicting Combination Weights (TPCW) and Predicting Style Embeddings (TPSE), are used to extract style tokens during inference time. TP-GST methods successfully generate expressive speech without background noise. Other studies [45] [46] [47] [48] have also utilized GSTs in various ways to synthesize expressive speech, control speaking styles, and explore fine-grained control of speech generation.

Inspired by the GST module, [49] proposes using global speaker embeddings (GSEs) to control the style of synthesized speech. GSE has a unique purpose and functionality that differs from GST such as focusing on capturing the speaker-specific characteristics within a given text, enabling the identification of speakers from their speech patterns. In contrast, GSTs are designed to capture the stylistic elements of a text, such as reading or formal styles. They enable the modification of text style while preserving its content.

In general, GSTs are an effective method for controlling global stylistic features of synthesized speech. However, they have limitations and challenges, such as requiring a sufficient amount of speech samples during the training step to effectively synthesize speech in the desired style. As GSTs are designed to capture global style features, they may not be an effective tool for controlling the nuances of the desired style, such as intonation or rhythm.

B. Style disentanglement

Speech style disentanglement refers to the process of extracting various style factors, such as prosody, speaker, and linguistic-related factors, which enables fine-grained control of multi-reference speech style on separate speech datasets. Disentangling different informative factors in speech synthesis is essential for highly controllable speech style transfer. One of the significant challenges in speech technology is separating intertwined informative factors. Therefore, separating representations of these factors can enhance the robustness of expressive speech synthesis systems [50]. Traditional latent space representation learning algorithms predict general style embeddings with limited fine-grained control.

In [51], disentangled latent space representations based on adversarial learning are adopted to improve the robustness of highly controllable style transfer in voice conversion (VC). An Adversarial Mask-And-Predict (MAP) network is designed to explicitly disentangle the extracted speech representations, which include content, timbre, and two additional factors related to prosody, rhythm, and pitch. (MAP) network consists of a gradient reverse layer (GRL) and a stack of prediction head layers. During training, one of the four speech representations is randomly masked, and the adversarial network attempts to infer the masked representation from the other three representations. The prediction head layers in the MAP module are composed of a fully connected layer, GeLU activation, layer normalization, and another fully connected layer. The MAP network is trained to predict the masked representation as accurately as possible by minimizing the adversarial loss. However, during backward propagation, the gradient is reversed, which encourages the representations learned by the encoder to contain as little mutual information as possible.

The adversarial MAP network aims to increase the correlation between the masked and other speech representations, while the speech representation encoders try to disentangle the representations to decrease the correlation using the inversed gradient of the adversarial MAP network. The proposed method enhances the quality of synthesized speech in voice conversion across multiple factors [52]. A single model is trained for multiple speakers using the adversarial learning framework, instead of building a separate model for each target speaker. The proposed method has two training phases, resulting in significant improvements in the quality of synthesized voice. In [53], a zero-shot style transfer approach using disentangled speech representation learning is adopted to transfer speech styles with non-parallel datasets. The disentanglement process improves style transfer accuracy.

In general, disentanglement speech representation learning is a promising approach for highly controllable speech style transfer. However, this method comes with computational complexity that requires substantial computing power. This issue needs to be carefully considered.

C. Cross-speaker style transfer

Cross-speaker style transfer (CSST) is a cutting-edge technique for synthesizing expressive speech. It aims to transfer multiple speaking styles from various supporting speakers to a target speaker while maintaining the target speaker's identity and timbre [54]. Unlike traditional speaking style transfer methods that collect style embeddings from reference speech and use them as auxiliary inputs to synthesize stylized speech [55], [56], modern cross-speaker style transfer conveys different speaking styles between speakers without requiring text-paired reference speech [57]. Numerous studies have adopted CSST to transfer speech styles between multiple speakers.

In [58], a chunk-wise multi-scale cross-speaker style and adversarial classifiers are proposed for style transfer. Multi-scale cross-speaker style is trained in two phases to predict both global style embeddings (GSE) and local prosody embeddings using an adversarial training approach. An adequate amount of speech style data from non-target speakers is needed during the training process. In [59], a multi-speaker acoustic system called Daft-Expert is employed to transfer highly expressive prosodic styles from both seen and unseen speakers. FiLM conditioning layers are used to embed prosody information in the TTS system. FiLM conditioning layers is a general-purpose conditioning technique for neural networks known as FiLM (Feature-wise Linear Modulation) proposed by [60]. FiLM layers influence neural network computations through a straightforward feature-wise affine transformation, utilizing conditioning information. The proposed model is combined with both FiLM layers and adversarial learning for highly accurate cross-speaker transfer.

Cross-speaker transfer with data augmentation techniques has been successfully used in low-resource expressive TTS systems. A recent study [61] applied data voice conversion VC-based augmentation for cross-speaker style transfer, where expressive speech datasets are not available for the target speaker. The adopted method uses two models: Pitch-Shift PS-based data augmentation and voice conversion VC-based data augmentation. Pitch-shift PS-based augmentation involves altering the fundamental frequency of the speech signal, providing a technique to modify the perceived pitch without changing the speaker identity .PS-based augmentation is used for source and target speaker samples to enhance the stability of the training stage, while short-time Fourier transform (STFT)-based optimization is adopted for the voice conversion training stage.

FastSpeech Multi-language TTS system [62] applied cross-language style transfer to synthesize speech in any speaker style in the target language, overcoming the challenge of non-authentic accent issues in cross-speaker style transfer. Conditional variational encoder and adversarial learning are used in the training process. Cross-speaker style transfer still faces challenges since multiple speakers have varying styles and timbres. Several studies have applied different techniques, such as speaker normalization [63] [64] [65] to model speaker attributes, data augmentation [66] [67] [68] [69] , and multi-task learning [70] [71], to generalize TTS systems to new speakers.

D. Speaker adaptation

TTS systems that employ speaker adaptation techniques aim to adjust a pre-trained model with a large-scale corpus to accommodate unseen speakers during the training process, even when there is a limited amount of speech data. Speaker adaptation is an effective technique when only a few minutes of target style data are available, as its primary role is to transfer speaking styles from a source speaker to a new speaker with limited adaptation data [72]. Adaptation strategies can be divided into two main categories. The first category of TTS systems uses pre-trained additional encoding networks to predict speaker attributes, which are then combined with linguistic characteristics as inputs to the synthesizer model [73] [74] [75] [76]. On the other hand, the second category fine-tunes the weights of the pre-trained multi-speaker TTS system to mimic a new speaker [77] [78]. Bayesian optimization (BO) has achieved high performance in fine-tuning TTS models.

A novel method called BOFFIN TTS (Bayesian Optimization for Fine-tuning Neural TTS) has been able to transfer styles for voice cloning in TTS systems under data-scarcity constraints [79]. This proposed method finds the optimal weights for hyperparameters for any target speaker in a functional and automatic manner. One of the critical aspects of this approach is its ability to intelligently search the hyperparameter space while minimizing the required computational resources. This is achieved through the use of Gaussian processes, which model the target function and provide a measure of uncertainty to guide the search for optimal hyperparameters. By exploiting this uncertainty, the algorithm can effectively balance exploration and exploitation during optimization. Another advantage of the Bayesian optimization approach is its flexibility in incorporating various constraints and domain knowledge into the optimization process. For example, one can introduce regularization terms or prior information on the hyperparameters to improve the adaptation performance. This can be particularly useful when dealing with challenging scenarios, such as limited data or highly diverse speaker characteristics.

Some recent works have also explored the combination of Bayesian optimization with other machine learning techniques, such as transfer learning and multi-task learning, to further improve the adaptation process [80]. By leveraging the shared information between different speakers or tasks, these approaches can achieve better performance, even with limited adaptation data. Despite the promising results, there are still challenges in applying Bayesian optimization for speaker adaptation in TTS systems. One of the main issues is the scalability of the optimization process, as the complexity of Gaussian process regression grows with the number of observations. This can limit the applicability of the method to large-scale problems or high-dimensional hyperparameter spaces. Moreover, the choice of the surrogate model and acquisition function, as well as the initialization of the optimization process, can significantly impact the overall performance.

VI. PROSODY MODELLING IN SPEECH SYNTHESIS

Prosody is a crucial aspect of speech synthesis that focuses on the rhythmic, melodic, and expressive features of speech [76]. The primary components of prosody include pitch, duration, intensity, and pauses, which collectively contribute to the overall expressiveness and naturalness of synthetic speech. Prosody helps convey emotions, emphasis, and linguistic structure in spoken language, thus playing a significant role in making synthetic speech sound more natural [77]. Hidden Markov Models (HMMs) have been used for capturing prosodic and linguistic features of speech, where decision trees are used to tie contextual features to individual nodes of the decision tree [81]. This approach enables more accurate modeling of prosody, allowing for generating natural and expressive synthetic speech. A new approach has been applied for prosody modeling [82]. This approach enhances prosody by integrating pre-trained cross- utterance (CU) representations from Wav2Vec2.0 and BERT into Fastspeech2. It improves speech naturalness and expressiveness in Mandarin and English but heavily relies on pre-trained models and lacks evaluation on other languages. Further investigation into model layers is needed for better prosody modeling.

A. Pitch Contour Modeling

Pitch contour modeling is the process of estimating and generating the fundamental frequency (F0) of speech, which corresponds to the perceived pitch. Accurate pitch contour modeling is essential for achieving natural-sounding prosody in speech synthesis. Many studies have been conducted to enhance the robustness of pitch. Among them, FastPitch [35] has gained popularity for its ability to control pitch and duration at the phoneme level during the synthesis of speech by conditioning these values. VocGAN-PS [83] and the FastPitch training algorithms have been proposed to improving pitch controllability. VocGAN-PS is a timbre-preserving pitch shift method that expands the pitch range without altering vocal characteristics. It avoids the need for additional algorithms like pitch tracking, however, may struggle with precise pitch estimation during transitions. The FastPitch training algorithm utilizes pitch-augmented speech data generated by VocGAN-PS to enhance FastPitch's pitch control and robustness, but its effectiveness relies on the quality and diversity of the augmented datasets.

There are different techniques for pitch contour modeling, including rule-based methods, statistical parametric methods [84], and deep learning approaches [81]. Rule-based methods use linguistic and phonetic rules to generate pitch contours, while statistical parametric methods (e.g., hidden Markov models or Gaussian mixture models) learn the relationship between linguistic features and pitch contours from data. Recently, deep learning methods like recurrent neural networks (RNNs) and convolutional neural networks (CNNs) have also been employed for pitch contour modeling, leveraging their ability to learn complex patterns and capture long-range dependencies in the data [67].

B. Duration Modeling

Duration modeling deals with predicting the duration of phonemes, syllables, or words in synthetic speech. Accurate duration modeling is vital for natural-sounding speech, as it contributes to the overall rhythm and pace of the spoken language [85].

Reference [86] propose an unsupervised text-to-speech (UTTS) system. In this system, a Speaker-Aware Duration Prediction module takes the phoneme sequence and speaker embedding as input to predict the speaker-aware duration for each phoneme. The phoneme sequence is first passed into a trainable look-up table to obtain the phoneme embeddings. Then, a multi-layer attention module is used to extract the latent phoneme representation, followed by a conv-1D module to combine the latent phoneme representation with the speaker embedding. A linear layer is then applied to generate the predicted duration in the logarithmic domain. During training, the Mean Squared Error (MSE) is utilized to calculate the difference between the predicted duration and the target duration obtained from forced alignment extracted by Montreal Forced Alignment (MFA).

During inference, the duration predictor rounds up the predicted duration and expands the phoneme sequence to form an estimated forced alignment. This estimated forced alignment is then used in the UTTS system for speech synthesis.

In [87] zero-shot TTS model utilized duration modeling as part of the conditioning process, enabling rhythm transfer and extracts disentangled embeddings between rhythm-based

speaker characteristics and acoustic-feature-based ones.The proposed method captures rhythm-based speaker characteristics , leading to higher perceived speaker similarity.

Another study [88] proposed two approaches to improve duration modeling in TTS systems. The first approach is a duration model conditioned on phrasing, which enhances predicted durations and provides better modeling of pauses. The second approach is a multi-speaker duration model called Cauliflow, which utilizes normalizing flows to predict durations that better match the target duration distribution. The proposed models improved naturalness of speech and variable durations for the same prompt, as well as variable levels of expressiveness.

C. Intensity Modeling

Intensity modeling is concerned with estimating and generating the energy or intensity of speech signals. Intensity contributes to the perceived loudness and stress patterns of synthetic speech and is an essential factor for natural-sounding prosody.

Different approaches have been proposed for intensity modeling, ranging from rule-based approaches to statistical methods and deep learning techniques. Rule-based methods rely on linguistic and phonetic rules to generate intensity patterns, while statistical methods, such as Gaussian mixture models or hidden Markov models, learn the relationship between linguistic features and intensity from data. Recently, deep learning methods, including recurrent neural networks (RNNs) and convolutional neural networks (CNNs), have been employed for intensity modeling, leveraging their capacity to learn complex patterns in the data [80].

D. Pause Modeling

Pauses play a crucial role in speech synthesis, as they help convey the structure of spoken language, provide time for the listener to process information, and contribute to the naturalness of synthetic speech. Pause modeling involves predicting the timing and duration of pauses in speech synthesis.

Many techniques have been proposed for pause modeling, including rule-based approaches, statistical methods, and deep learning techniques. Rule-based methods rely on linguistic and syntactic rules to predict pause locations and durations, while statistical methods learn these relationships from data [89]. Deep learning techniques, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), can also be employed for pause modeling, as they are capable of learning complex patterns and capturing long-range dependencies in the data.

VII. CHILDREN SPEECH SYNTHESIS

Children's speech synthesis is the process of creating artificial voices that sound like children, which is useful in developing interactive systems and robots for children's education and entertainment [81], [90]. However, this area of research poses several challenges. First, obtaining high-quality and phonetically balanced speech data from children is difficult. Additionally, children's voices have distinct characteristics that set it apart from adult speech. Mispronounced words, disfluencies, and ungrammatical utterances often characterize child speech. Furthermore, children exhibit linguistic differences compared to adult speech across different levels, such as prosody, vocabulary, grammar, and sizeable acoustic variability of child speech [91]. Moreover, synthesizing expressive conversational speech is a further challenge, as it requires the inclusion of paralinguistics and emotions in the synthesized speech [92]. Evaluating the quality of children's speech synthesis is also not straightforward, as it involves prolonged exposure to the synthetic voice.

Despite these challenges, researchers are exploring various approaches, such as speaker adaptive. a study conducted by [93] explored the acoustic characteristics of children's speech, encompassing aspects such as duration and pitch the results indicated that certain vowel sounds have longer durations in children compared to adults. Moreover, synthesizing expressive conversational speech is a further challenge, as it requires the inclusion of paralinguistics and emotions in the synthesized speech [92].

Evaluating the quality of children's speech synthesis is also not straightforward, as it involves prolonged exposure to the synthetic voice. Despite these challenges, researchers are exploring various approaches, such as speaker-adaptive HMM-based speech synthesis and deep learning techniques, to develop efficient and accurate methods for children's speech synthesis.

The goal is to make dialogue systems more inclusive and accessible for younger users. Hidden Markov Models (HMMs) have been used in child speech synthesis to find suitable initial models and speaker adaptation methods [94], [95]. Nevertheless, HMM-based systems for synthesizing child speech often face difficulties in achieving high naturalness and accurately replicating the subtleties of children's speech. In this study [91],

The researchers introduced deep neural vocoders within a TTS framework to achieve child speech synthesis. Their method involves fine-tuning both the acoustic model Tacotron2 and a pre-trained WaveRNN vocoder. Moreover, they performed additional fine-tuning of the WaveRNN vocoder on a dedicated child speech dataset, improving the quality of child speech

synthesis [96]. In [97], a hybrid system that combines DNN with HMM was utilized for automatic speech recognition, using approximately 10 hours of Italian child speech data. This hybrid DNN-HMM approach proved effective in enhancing speech recognition accuracy specifically for Italian child speech.

VIII. Discussion and Conclusion

Speech synthesis has come a long way since the early days of simple rule-based systems. Today, there are a variety of approaches and techniques that can be used to generate natural-sounding synthetic speech. This survey offers an overview of the development of expressive Text-to-Speech (TTS) systems and the diverse methodologies employed to synthesize expressive speech from written text. The selected articles presented a range of TTS and speech synthesis models that aim to enhance the quality and expressiveness of synthetic speech. This survey encapsulates the contemporary as well as conventional methods that are utilized in TTS systems. We discussed deep learning-based speech synthesis, emotional speech synthesis and style transfer in speech synthesis. Additionally, we have reviewed several objective metrics such as Itakura-Saito measure, Root mean square (RMSE), Gross pitch error (GPE) and subjective metrics such as MOS and MUSHRA utilized to access the quality of the synthesized speech are examined. In addition, our focus was on the representation and transfer approaches for style to comprehensively illustrate the significance of style representation in enhancing the expressiveness of synthesized speech in Text-to-Speech (TTS) systems. Further, we reviewed both deep learning-based autoregressive model such as Parallel WaveNet and non-autoregressive model such as FastSpeech that are used in the front-end and back-end of TTS system.

Finally, we point out the challenges in child speech synthesis, which involves the difficulty of obtaining high-quality and phonetically balanced speech data from children. Additionally, we address the unique characteristics of children's speech, differentiating it from adult speech, including linguistic variations and expressive conversational patterns.

We hope this paper will offer a clear overview for readers to understand the current status of expressive speech synthesis models, inspiring continuous research efforts on expressive TTS systems. This, in turn, aims to promote future modern in the field of study expressive TTS systems, especially in the field of child speech synthesis.

IX. Acknowledgements

This paper is supported by the European Union's HORIZON Research and Innovation Programme under grant agreement No 101120657, project ENFIELD (European Lighthouse to Manifest Trustworthy and Green AI) and by the Ministry of Innovation and Culture and the National Research, Development and Innovation Office of Hungary within the framework of the National Laboratory of Artificial Intelligence. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union and the granting authorities. Neither the European Union nor the granting authorities can be held responsible for them.

References

[1] X. Tan, T. Qin, F. Soong, and T.-Y. Liu, "A Survey on Neural Speech Synthesis," Jun. 2021, [Online]. arXiv preprint *arXiv:2106.15561*, 2021. Available: http://arxiv.org/abs/2106.15561.

[2] N. Tits, "Controlling the emotional expressiveness of synthetic speech: a deep learning approach," *4OR*, vol. 20, no. 1, pp. 165–166, Mar. 2022, DOI: 10.1007/s10288-021-00473-2.

[3] P. Alexander. Taylor, Text-to-speech synthesis. Cambridge University Press, 2009. DOI: 10.1017/CBO9780511816338

[4] Y. Ning, S. He, Z. Wu, C. Xing, and L. J. Zhang, "Review of deep learning-based speech synthesis," *Applied Sciences (Switzerland)*, vol. 9, no. 19. MDPI AG, Oct. 01, 2019. DOI: 10.3390/app9194050.

[5] K. R. Scherer, "Vocal affect expression: a review and a model for future research.," *Psychol Bull*, vol. 99, no. 2, p. 143, 1986. DOI: 10.1037/0033-2909.99.2.143

[6] J. F. Pitrelli, R. Bakis, E. M. Eide, R. Fernandez, W. Hamza, and M. A. Picheny, "The IBM expressive text-to-speech synthesis system for american english," *IEEE Trans Audio Speech Lang Process*, vol. 14, no. 4, pp. 1099–1108, Jul. 2006, DOI: 10.1109/TASL.2006.876123.

[7] M. Mahrishi, K. K. Hiran, G. Meena, and P. Sharma, Machine learning and deep learning in real-time applications. IGI global, 2020. DOI: 10.4018/978-1-7998-3095-5.ch009

[8] D. H. Klatt, "Review of text-to-speech conversion for English," *J Acoust Soc Am*, vol. 82, no. 3, pp. 737–793, 1987. DOI: 10.1121/1.395275

[9] D. H. Klatt, "Software for a cascade/parallel formant synthesizer." *Acoust Soc Am*, vol. 67, no. 3, pp. 971–995, 1980, DOI: 10.1121/1.383940

[10] J. Tao, Y. Kang, and A. Li, "Prosody conversion from neutral speech to emotional speech," *IEEE Trans Audio Speech Lang Process*, vol. 14, no. 4, pp. 1145–1153, Jul. 2006, DOI: 10.1109/tasl.2006.876113

[11] N. Campbell, W. Hamza, H. Hoge, J. Tao, and G. Bailly, "Editorial Special Section on Expressive Speech Synthesis," *IEEE Trans Audio Speech Lang Process*, vol. 14, no. 4, pp. 1097–1098, Jun. 2006, DOI: 10.1109/tasl.2006.878306.

[12] D. Govind and S. R. M. Prasanna, "Expressive speech synthesis: a review," *Int J Speech Technol*, vol. 16, pp. 237–260, 2013. DOI: 10.1007/s10772-012-9180-2

[13] C. Valentini-Botinhao, M. S. Ribeiro, O. Watts, K. Richmond, and G. E. Henter, "Predicting pairwise preferences between TTS audio stimuli using parallel ratings data and anti-symmetric twin neural networks," *International Speech Communication Association, INTERSPEECH*, 2022. DOI: 10.21437/interspeech.2022-10132

[14] International Telecommunication Union (ITU), "Methods for subjective determination of transmission quality," 1996. DOI: 10.18356/16e04175-en.

[15] H. Zen, A. Senior, and M. S. Google, "Statistical Parametric Speech Synthesis Using Deep Neural Networks." *International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2013. DOI: 10.1109/icassp.2013.6639215

[16] A. van den Oord et al., "WaveNet: A Generative Model for Raw Audio," Sep. 2016, arXiv preprint *arXiv:1609.03499*, DOI: 10.48550/arXiv.1609.03499

[17] Y. Wang et al., "Tacotron: Towards End-to-End Speech Synthesis," *International Speech Communication Association, INTERSPEECH*, 2017, DOI: 10.21437/interspeech.2017-1452

[18] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10. pp. 1345–1359, 2010. DOI: 10.1109/TKDE.2009.191.

[19] X. Zhuang, T. Jiang, S. Y. Chou, B. Wu, P. Hu, and S. Lui, "Litesing: Towards Fast, Lightweight And Expressive Singing Voice Synthesis," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 7078–7082. DOI: 10.1109/icassp39728.2021.9414043.

[20] T. Okamoto, K. Tachibana, T. Toda, Y. Shiga, and H. Kawai, "An Investigation Of Subband Wavenet Vocoder Covering Entire Audible Frequency Range With Limited Acoustic Features," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2018, pp. 5654–5658. DOI: 10.1109/icassp.2018.8462237

[21] A. Van Den Oord et al., "Parallel WaveNet: Fast High-Fidelity Speech Synthesis," In *International conference on machine learning* (pp. 3918–3926). PMLR, 2018. arXiv preprint *arXiv:1711.10433* DOI: 10.48550/arXiv.1711.10433

[22] Y. Jiao, A. Gabryś, G. Tinchev, B. Putrycz, D. Korzekwa, and V. Klimkov, "Universal Neural Vocoding with Parallel Wavenet," In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 6044–6048. 6048. DOI: 10.1109/icassp39728.2021.9414444

[23] H. Choi, S. Park, J. Park, and M. Hahn, "Emotional Speech Synthesis For Multi-Speaker Emotional Dataset Using Wavenet Vocoder," In *2019 IEEE International Conference on Consumer Electronics (ICCE)*, IEEE, 2019, pp. 1–2. DOI: 10.1109/icce.2019.8661919

[24] J. J. Webber, C. Valentini-Botinhao, E. Williams, G. E. Henter, and S. King, "Autovocoder: Fast Waveform Generation from a Learned Speech Representation Using Differentiable Digital Signal Processing," In *ICASSP – 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Jun. 2023, pp. 1–5. DOI: 10.1109/ICASSP49357.2023.10095729.

[25] D. Griffin and J. Lim, "Signal Estimation from Modified Short-Time Fourier Transform," *IEEE Trans Acoust*, vol. 32, no. 2, pp. 236–243, 1984. DOI: 10.1109/tassp.1984.1164317

[26] J. Kong, J. Kim, and J. Bae, "Hifi-Gan: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis," *Adv Neural Inf Process Syst*, vol. 33, pp. 17022–17033, 2020. https://proceedings.neurips.cc/paper/2020/hash/c5d736809766d46260d816d8dbc9eb44-Abstract.html

[27] S. Li et al., "SnakeGAN: A Universal Vocoder Leveraging DDSP Prior Knowledge and Periodic Inductive Bias," in *2023 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2023, pp. 1703–1708. DOI: 10.1109/icme55011.2023.00293

[28] R. Yamamoto, E. Song, and J.-M. Kim, "Parallel Wavegan: A Fast Waveform Generation Model Based on Generative Adversarial Networks with Multi-Resolution Spectrogram," in *ICASSP 2020, - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May 2020, pp. 6199–6203. DOI: 10.1109/ICASSP40776.2020.9053795.

[29] S. Lee, W. Ping, B. Ginsburg, B. Catanzaro, and S. Yoon, "Bigvgan: A Universal Neural V ocoder With Large-Scale Training,", 2022. DOI: 10.48550/arXiv.2206.04658

[30] Y. Zheng, J. Tao, Z. Wen, and J. Yi, "Forward-backward decoding sequence for regularizing end-to-end TTS," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 27, no. 12, pp. 2067–2079, Dec. 2019, DOI: 10.1109/taslp.2019.2935807.

[31] J. Shen et al., "Natural TTS Synthesis By Conditioning Wavenet On Mel Spectrogram Predictions," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, 2018, pp. 4779–4783. https://doi.org/10.1109/icassp.2018.8461368

[32] A. Kulkarni, V. Colotte, and D. Jouvet, "Improving Transfer of Expressivity For End-To-End Multispeaker Text-To-Speech Synthesis." In *2021 29th European Signal Processing Conference (EUSIPCO)*, pp. 31–35. IEEE, 2021. DOI: 10.23919/eusipco54536.2021.9616249

[33] R. J. Skerry-Ryan et al., "Towards End-to-End Prosody Transfer for Expressive Speech Synthesis with Tacotron," in *ICML* 2018. http://proceedings.mlr.press/v80/skerry-ryan18a.html

[34] Y. Ren et al., "Fastspeech: Fast, Robust And Controllable Text To Speech," *Advances in neural information processing systems Adv Neural Inf Process Syst*, vol. 32, 2019. https://proceedings.neurips.cc/paper_files/paper/2019/hash/f63f65b503e22cb970527f23c9ad7db1-Abstract.html

[35] A. Łańcucki, "Fastpitch: Parallel Text-To-Speech With Pitch Prediction," in I*CASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 6588–6592. DOI: 10.1109/icassp39728.2021.9413889

[36] Y. Ren et al., "Fastspeech 2: Fast And High-Quality End-To-End Text To Speech," 2020, arXiv preprint *arXiv:2006.04558*, DOI: 10.48550/arXiv.2006.04558.

[37] P . C. Loizou, "Speech quality assessment." *Multimedia analysis, processing and communications 2011*, pp. 623–654. DOI: 10.1007/978-3-642-19551-8_23

[38] B. Sabine, J. Latorre, and K. Yanagisawa. "Crowdsourced assessment of speech synthesis." Crowdsourcing for Speech Processing: Applications to Data Collection, Transcription and Assessment (2013): 173–216, DOI: 10.1002/9781118541241.ch7

[39] I. Recommendation, "1534-1,'Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA),'" *International Telecommunications Union*, Geneva, Switzerland, vol. 2, 2001. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1534-3-201510-I!!PDF-E.pdf

[40] B.-H Juang, "On Using the Itakura-Saito Measures for Speech Coder Performance Evaluation," *AT&T Bell Laboratories Technical Journal*, vol. 63, no. 8, pp. 1477–1498, 1984, DOI: 10.1002/j.1538-7305.1984.tb00047.x

[41] C.-C. Wang, Z.-H. Ling, B.-F. Zhang, and L.-R. Dai, "Multi-Layer F0 Modeling For HMM-Based Speech Synthesis," in *2008 6th International symposium on Chinese spoken language processing*, IEEE, 2008, pp. 1–4. DOI: 10.1109/chinsl.2008.ecp.44

[42] O. Babacan, T. Drugman, N. d'Alessandro, N. Henrich, and T. Dutoit, "A Comparative Study of Pitch Extraction Algorithms on a Large Variety of Singing Sounds," Dec. 2019, DOI: 10.1109/icassp.2013.6639185

[43] Y. Wang et al., "Style Tokens: Unsupervised Style Modeling, Control and Transfer in End-to-End Speech Synthesis," in *ICML* 2018. https://proceedings.mlr.press/v80/wang18h.html?ref=https://githubhelp.com

[44] D. Stanton, Y. Wang, and R. Skerry-Ryan, "Predicting Expressive Speaking Style From Text In End-To-End Speech Synthesis," In *IEEE Spoken Language Technology Workshop (SLT)*, 2018. DOI: 10.1109/slt.2018.8639682

[45] Y. Wang et al., "Style Tokens: Unsupervised Style Modeling, Control and Transfer in End-to-End Speech Synthesis," 2018. https://proceedings.mlr.press/v80/wang18h.html

[46] S. Liu, S. Yang, D. Su, and D. Yu, "Referee: Towards Reference-Free Cross-Speaker Style Transfer With Low-Quality Data For Expressive Speech Synthesis," In *ICASSP ,IEEE International Conference on Acoustics Speech and Signal , Processing (ICASSP)*, pp. 6307–6311. IEEE, 2022. DOI: 10.1109/icassp43922.2022.9746858

[47] C. Yu et al., "DurIAN: Duration Informed Attention Network for Multimodal Synthesis," *International Speech Communication Association, INTERSPEECH*, 2019, pp. 2027–2031, DOI: 10.21437/interspeech.2020-2968

[48] Y. Lee and T. Kim, "Robust And Fine-Grained Prosody Control Of End-To-End Speech Synthesis," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5911–5915. IEEE, 2019. DOI: 10.1109/icassp.2019.8683501

[49] W. Lu et al., "One-Shot Emotional Voice Conversion Based On Feature Separation," *Speech Commun*, vol. 143, pp. 1–9, Sep. 2022, DOI: 10.1016/j.specom.2022.07.001.

[50] D. Wang, L. Li, Y. Shi, Y. Chen, and Z. Tang, "Deep Factorization for Speech Signal," " *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5094–5098. IEEE, 2018. DOI: 10.1109/icassp.2018.8462169

[51] J. Wang, J. Li, X. Zhao, Z. Wu, S. Kang, and H. Meng, "Adversarially Learning Disentangled Speech Representations For Robust Multi-Factor Voice Conversion," *International Speech Communication Association, INTERSPEECH*, 2021, pp. 846–850, DOI: 10.21437/interspeech.2021-1990

[52] J. Chou, C. Yeh, H. Lee, and L. Lee, "Multi-target Voice Conversion without Parallel Data by Adversarially Learning Disentangled Audio Representations," *International Speech Communication Association, INTERSPEECH*, 2018, pp. 501–505, DOI: 10.21437/interspeech.2018-1830

[53] S. Yuan, P. Cheng, R. Zhang, W. Hao, Z. Gan, and L. Carin, "Improving Zero-shot V oice Style Transfer via Disentangled Representation Learning," *International Conference on learning representation* 2021, https://openreview.net/forum?id=TgSVWXw22FQ

[54] Y. Shin, Y. Lee, S. Jo, Y. Hwang, and T. Kim, "Text-driven Emotional Style Control and Cross-speaker Style Transfer in Neural TTS," *International Speech Communication Association, INTERSPEECH*, 2022, pp. 2313–2317, DOI: 10.21437/interspeech.2022-10131

[55] Y. Bian, C. Chen, Y. Kang, and Z. Pan, "Multi-reference Tacotron by Intercross Training for Style Disentangling,Transfer and Control in Speech Synthesis," Apr. 2019, arXiv preprint *arXiv:1904.02373* DOI: 10.48550/arXiv.1904.02373

[56] M. Whitehill, S. Ma, D. McDuff, and Y. Song, "Multi-Reference Neural TTS Stylization with Adversarial Cycle Consistency," *International Speech Communication Association, INTERSPEECH*, 2020, pp. 4442–4446, DOI: 10.21437/interspeech.2020-2985

[57] S. Pan, "Cross-speaker Style Transfer with Prosody Bottleneck in Neural Speech Synthesis." *International Speech Communication Association, INTERSPEECH*, 2021, pp. 4678–4682 DOI: 10.21437/interspeech.2021-979

[58] X. Li, C. Song, X. Wei, Z. Wu, J. Jia, and H. Meng, "Towards Cross-speaker Reading Style Transfer on Audiobook Dataset," *International Speech Communication Association, INTERSPEECH*, 2022, pp. 5528–5532, DOI: 10.21437/interspeech.2022-11223

[59] J. Zaïdi, H. Seuté, B. van Niekerk, and M.-A. Carbonneau, "Daft-Exprt: Cross-Speaker Prosody Transfer on Any Text for Expressive Speech Synthesis," *International Speech Communication Association, INTERSPEECH*, 2022, pp. 4591–4595, DOI: 10.21437/interspeech.2022-10761

[60] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "Film: Visual Reasoning with A General Conditioning Layer," in *Proceedings of the AAAI conference on artificial intelligence*, 2018. DOI: 10.1609/aaai.v32i1.11671

[61] R. Terashima et al., "Cross-Speaker Emotion Transfer for Low-Resource Text-to-Speech Using Non-Parallel Voice Conversion with Pitch-Shift Data Augmentation," *International Speech Communication Association, INTERSPEECH*, 2022, pp. 3018–3022, DOI: 10.21437/interspeech.2022-11278

[62] Z. Shang, Z. Huang, H. Zhang, P. Zhang, and Y. Yan, "Incorporating Cross-Speaker Style Transfer For Multi-Language Text-To-Speech," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2021, pp. 3406–3410. DOI: 10.21437/Interspeech.2021-1265.

[63] P. Wu et al., "Cross-speaker Emotion Transfer Based on Speaker Condition Layer Normalization and Semi-Supervised Training in Text-To-Speech," Oct. 2021, arXiv preprint *arXiv:2110.04153*, DOI: 10.48550/arXiv.2110.04153

[64] C. Qiang, P. Yang, H. Che, X. Wang, and Z. Wang, "Style-Label-Free: Cross-Speaker Style Transfer by Quantized VAE and Speaker-wise Normalization in Speech Synthesis," Dec. 2022, in *13th International Symposium on Chinese Spoken Language Processing (ISCSLP)* Dec. 2022, DOI: 10.1109/iscslp57327.2022.10038135

[65] S. Aryal and R. Gutierrez-Osuna, "Accent Conversion Through Cross-Speaker Articulatory Synthesis," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2014, pp. 7694–7698. DOI: 10.1109/icassp.2014.6855097

[66] G. Huybrechts, T. Merritt, G. Comini, B. Perz, R. Shah, and J. Lorenzo-Trueba, "Low-Resource Expressive Text-To-Speech Using Data Augmentation," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6593–6597. IEEE, 2021. DOI: 10.1109/icassp39728.2021.9413466

[67] J. Wu, A. Polyak, Y. Taigman, J. Fong, P. Agrawal, and Q. He, "Multilingual Text-To-Speech Training Using Cross Language Voice Conversion And Self-Supervised Learning Of Speech Representations," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, pp. 8017–8021. DOI: 10.1109/icassp43922.2022.9746282

[68] Z. Byambadorj, R. Nishimura, A. Ayush, K. Ohta, and N. Kitaoka, "Multi-Speaker TTS System For Low-Resource Language Using Cross- Lingual Transfer Learning And Data Augmentation," in *2021 Asia- Pacific Signal and Information Processing Association Annual Summit and Conference (APSIP A ASC)*, IEEE, 2021, pp. 849–853. https://ieeexplore.ieee.org/abstract/document/9689505

[69] Z. Zhang, Y. Zheng, X. Li, and L. Lu, "WeSinger: Data-augmented Singing Voice Synthesis with Auxiliary Losses," *International Speech Communication Association, INTERSPEECH*, 2022, pp. 4252–4256, DOI: 10.21437/interspeech.2022-454

[70] Y. Nakai, Y. Saito, K. Udagawa, and H. Saruwatari, "Multi-Task Adversarial Training Algorithm for Multi-Speaker Neural Text-to- Speech," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, IEEE, 2022, pp. 743–748 DOI: 10.23919/apsipaasc55919.2022.9980331

[71] X. Zhang, J. Wang, N. Cheng, and J. Xiao, "TDASS: Target Domain Adaptation Speech Synthesis Framework for Multi-speaker Low-Resource TTS," International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE, 2022, DOI: 10.1109/ijcnn55064.2022.9892596

[72] K. Inoue, S. Hara, and M. Abe, "Module Comparison of Transformer-TTS For Speaker Adaptation Based On Fine-Tuning," in *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIP A ASC)*, IEEE, 2020, pp. 826–830. https://ieeexplore.ieee.org/abstract/document/9306250

[73] C. Du, Y. Guo, X. Chen, and K. Yu, "Speaker Adaptive Text-to-Speech with Timbre-Normalized Vector-Quantized Feature," *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2023). DOI: 10.1109/taslp.2023.3308374

[74] A. R. Mandeel, M. S. Al-Radhi, and T. G. Csapó, "Speaker Adaptation Experiments with Limited Data for End-to-End Text-To-Speech Synthesis using Tacotron2," *Infocommunications Journal*, vol. 14, no. 3, pp. 55–62, 2022. DOI: 10.36244/icj.2022.3.7

[75] C.-P. Hsieh, S. Ghosh, and B. Ginsburg, "Adapter-Based Extension of Multi-Speaker Text-to-Speech Model for New Speakers," *International Speech Communication Association, INTERSPEECH*, 2023, pp. 3028–3032 DOI: 10.21437/interspeech.2023-2313

[76] Y. Jia et al., "Transfer Learning from Speaker V erification to Multi-speaker Text-To-Speech Synthesis." *Advances in neural information processing systems*, 2018. https://proceedings.neurips.cc/paper_files/paper/2018/hash/6832a7b24bc06775d02b7406880b93fc-Abstract.html

[77] K. Inoue, S. Hara, M. Abe, T. Hayashi, R. Yamamoto, and S. Watanabe, "Semi-Supervised Speaker Adaptation For End-To-End Speech Synthesis With Pretrained Models," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7634–7638. IEEE, 2020. DOI: 10.1109/icassp40776.2020.9053371

[78] M. Zhang, X. Zhou, Z. Wu, and H. Li, "Towards Zero-Shot Multi-Speaker Multi-Accent Text-to-Speech Synthesis," *IEEE Signal Processing Letters* 2023. DOI: 10.1109/lsp.2023.3292740

[79] H. B. Moss, V. Aggarwal, N. Prateek, J. González, and R. Barra-Chicote, "BOFFIN TTS: Few-Shot Speaker Adaptation by Bayesian Optimization," In *ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7639–7643. IEEE, 2020. DOI: 10.1109/icassp40776.2020.9054301

[80] J. P. H. Van Santen, R. Sproat, J. Olive, and J. Hirschberg, Progress in speech synthesis. Springer Science & Business Media, 2013. DOI: 10.1007/978-1-4612-1894-4_15

[81] N. Kaur and P. Singh, "Conventional and contemporary approaches used in text to speech synthesis: a review," *Artif Intell Rev*, vol. 56, no. 7, pp. 5837–5880, Jul. 2023, DOI: 10.1007/s10462-022-10315-0.

[82] Y. J. Zhang, C. Zhang, W. Song, Z. Zhang, Y. Wu, and X. He, "Prosody Modelling with Pre-Trained Cross-Utterance Representations for Improved Speech Synthesis," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 31, pp. 2812–2823, 2023, DOI: 10.1109/TASLP.2023.3278184.

[83] H. Bae and Y.-S. Joo, "Enhancement of Pitch Controllability using Timbre-Preserving Pitch Augmentation in FastPitch," *International Speech Communication Association, INTERSPEECH*, 2022, pp. 6–10 DOI: 10.21437/interspeech.2022-55

[84] N. Adiga and S. R. M. Prasanna, "Acoustic Features Modelling for Statistical Parametric Speech Synthesis: A Review," *IETE Technical Review*, vol. 36, no. 2, pp. 130–149, 2019. DOI: 10.1080/02564602.2018.1432422

[85] J. Ni, Y. Shiga, and H. Kawai, "Duration Modeling with Global Phoneme- Duration Vectors.," I*nternational Speech Communication Association, INTERSPEECH*, 2019, pp. 4465–4469. DOI: 10.21437/interspeech.2019-2126

[86] J. Lian, C. Zhang, G. K. Anumanchipalli, and D. Yu, "Unsupervised TTS Acoustic Modeling for TTS with Conditional Disentangled Sequential V AE," *IEEE/ACM Trans Audio Speech Lang Process*, 2023. DOI: 10.1109/taslp.2023.3290423

[87] K. Fujita, T. Ashihara, H. Kanagawa, T. Moriya, and Y. Ijima, "Zero-Shot Text-To-Speech Synthesis Conditioned Using Self-Supervised Speech Representation Model," in *IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)* 2023. DOI: 10.1109/icasspw59220.2023.10193459

[88] A. Abbas et al., "Expressive, variable, and controllable duration modelling in TTS," *International Speech Communication Association, INTERSPEECH*, 2022. DOI: 10.21437/interspeech.2022-384

[89] Y. Stylianou, "Applying the Harmonic Plus Noise Model in *Concatenative Speech Synthesis*," 2001. DOI: 10.1109/89.890068

[90] J. Y. Zhang, A. W. Black, and R. Sproat, "Identifying Speakers in Children's Stories for Speech Synthesis." In *Eighth European Conference on Speech Communication and Technology*. 2003. DOI: 10.21437/eurospeech.2003-586

[91] R. Jain, M. Y. Yiwere, D. Bigioi, P. Corcoran, and H. Cucu, "A Text-to-Speech Pipeline, Evaluation Methodology, and Initial Fine-Tuning Results for Child Speech Synthesis," *IEEE Access*, vol. 10, pp. 47 628–47 642, 2022, DOI: 10.1109/access.2022.3170836

[92] A. Borgh, K. ; Dickson, W. Patrick, K. Borgh, and W. P. Dickson, "DOCUMENT RESUME ED 277 007 CS 210 188 The Effects on Children's Writing of Adding Speech Synthesis "permission to reproduce this material has been granted by," 1986. DOI: 10.1080/08886504.1992.10782629

[93] C. Terblanche, M. Harty, M. Pascoe, and B. V Tucker, "A Situational Analysis of Current Speech-Synthesis Systems for Child Voices: A Scoping Review of Qualitative and Quantitative Evidence," *Applied Sciences*, vol. 12, no. 11, p. 5623, 2022.

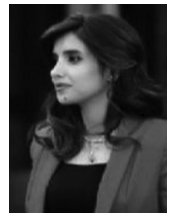
[94] A. Govender, F. de Wet, and J.-R. Tapamo, "HMM Adaptation For Child Speech Synthesis," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015. DOI: 10.21437/interspeech.2015-379

[95] A. Govender and F. De Wet, "Objective Measures To Improve The Selection Of Training Speakers In HMM-Based Child Speech Synthesis," in *Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*, IEEE, 2016, pp. 1–6. DOI: 10.1109/robomech.2016.7813193

[96] D. Giuliani and B. BabaAli, "Large Vocabulary Children's Speech Recognition with DNN-HMM and SGMM Acoustic Modeling," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015. DOI: 10.21437/interspeech.2015-378

[97] P. Cosi, "A Kaldi-Dnn-Based Asr System For Italian," in *International Joint Conference On Neural Networks (IJCNN)*, IEEE, 2015, pp. 1–5. DOI: 10.1109/ijcnn.2015.7280336

[98] Matsumoto, Kento, Sunao Hara, and Masanobu Abe. "Speech-Like Emotional Sound Generation Using WaveNet." *IEICE TRANSACTIONS on Information and Systems* 105, no. 9 (2022): 1581–1589. DOI: 10.1587/transinf.2021edp7236



Shaimaa Alwaisi was born in, Iraq. She got a BSc degree in Computer Engineering at Diyala University, higher Diploma from Iraqi commission for computer and informatics ICCI and a MSc degree in Computer Engineering at Selcuk University, Turkey. She currently PhD student at the Speech Technology and Smart Interactions Laboratory in the Budapest University of Technology and Economics. She is working on neural vocoders and acoustic models for speech synthesis. her current interests are signal processing, expressive speech synthesis, Child speech synthesis, Deep learning, acoustic models, and voice conversion.

Géza Németh was born in 1959. He obtained his MSc in electrical engineering, major in Telecommunications at the Faculty of Electrical Engineering of BME in 1983. Also, at BME: dr. univ., 1987, PhD 1997. He is an associate professor at BME. He is the author or co-author of more than 170 scientific publications and 4 patents. His research fields include speech technology, service automation, multilingual speech and multimodal information systems, mobile user interfaces and applications. He is the Head of the Speech Technology and Smart Interactions Laboratory of BME TMIT.

Speech synthesis from intracranial stereotactic Electroencephalography using a neural vocoder

Frigyes Viktor Arthur and Tamás Gábor Csapó

***Abstract*—Speech is one of the most important human biosignals. However, only some speech production characteristics are fully understood, which are required for a successful speech-based Brain-Computer Interface (BCI). A proper brain-to-speech system that can generate the speech of full sentences intelligibly and naturally poses a great challenge. In our study, we used the SingleWordProduction-Dutch-iBIDS dataset, in which speech and intracranial stereotactic electroencephalography (sEEG) signals of the brain were recorded simultaneously during a single word production task. We apply deep neural networks (FC-DNN, 2D-CNN, and 3D-CNN) on the ten speakers' data for sEEG-to-Mel spectrogram prediction. Next, we synthesize speech using the WaveGlow neural vocoder. Our objective and subjective evaluations have shown that the DNN-based approaches with neural vocoder outperform the baseline linear regression model using Griffin-Lim. The synthesized samples resemble the original speech but are still not intelligible, and the results are clearly speaker dependent. In the long term, speech-based BCI applications might be useful for the speaking impaired or those having neurological disorders.**

***Index Terms*—human-computer interaction, sEEG, BCI, brain-computer interface**

I. Introduction

It is expected that 0.4% of the European population suffers from a speech impairment [1], [2], [3]. Digital applications using speech technology could significantly help their everyday communication. Loss of speech can cause social isolation, and feelings of loss of identity and can lead to clinical depression [4]. Augmentative and alternative communication (AAC) technologies, such as brain-computer interfaces (BCIs) might directly read brain signals to restore lost speech capabilities [5]. In the future, the application of speech neuroprostheses have the potential to help patients with neurological disorders or speech impairment.

Brain-computer interfaces enable direct control of computers without physical activity, with potential applications as rehabilitation devices for motor-impaired persons (e.g., input system for writing, prosthetic control). Ideally, BCI applications operate in naturalistic scenarios, requiring a neural input with good temporal resolution, minimal preprocessing needs and relative ease of measurement. There are several available modalities for neuroimaging, including electroencephalography (EEG) [6], stereotactic depth electrodes [7], intracranial electrocorticography (ECoG) [8], Magnetoencephalography (MEG) [9], Local Field Potential (LFP) [8].

Budapest University of Technology and Economics Department of Telecommunications and Media Informatics (BME), Hungary
corresponding author (e-mail: arthur@tmit.bme.hu) (e-mail: csapot@tmit.bme.hu)

DOI: 10.36244/ICJ.2024.1.6

From the above, EEG has been the most widely studied one for BCI [6], [10]. EEG is a non-invasive method for measuring small electrical currents on the scalp, which reflect brain activity. It allows one to assess cortical excitability and effective connectivity in clinical and basic research without extensive invasive surgical installation. However, obtaining clean and usable EEG recordings (e.g., signals, data) is challenging due to the various bio-physiology-related artifacts that contaminate the electroencephalographic signal. In biomedical applications, such as monitoring brain activity during surgery or in sleep studies, EEG measurements typically utilize multiple electrodes, ranging from 32 to 256, with sampling rates around 256–2048 Hz. Relative to other methods recording electric potentials from the brain (ECoG, MEG, LFP), at the cost of poorer SNR and lower spatial resolution [8], EEG is non-invasive, cheap, and can be obtained even with wearable devices that allow for measurements outside the lab [11].

Csapó et al. [12] present a novel multimodal analysis method that combines EEG, articulatory movements, and speech signals for multimodal analysis, combining brain signal analysis during speech with ultrasound-based articulatory data. This study developed a fully connected deep neural network (FC-DNN) to predict articulatory movements using EEG signals. The study has demonstrated a clear relationship between EEG and articulatory movements and therefore provides valuable insights for future research in speech BCI.

Arthur and Csapó [13] discuss using deep learning to process EEG brain signals and synthesize speech. EEG signals were processed and used in this study to estimate the mel-spectral parameters of speech using deep learning models. Although not intelligible, the synthesized speech resembled the original speech signal, presenting a promising avenue for further investigation.

While initial results are encouraging, it is important to recognize the current limitations and challenges facing EEG-based BCI systems in the context of speech synthesis. Although these systems show potential, especially for aiding individuals with speech impairments, the extent of their effectiveness and practical applicability remains an area of ongoing research. The journey towards refining these technologies to reliably and effectively synthesize speech involves overcoming significant technical and scientific hurdles. Continued research and development are crucial to enhance our understanding and to push the boundaries of what is achievable with EEG-based BCIs. Ultimately, the goal is to leverage these advancements to improve the quality of life for those facing communication challenges, but it is essential to maintain a realistic perspective on the current state of the technology and the work that still

lies ahead.

More invasive methods offer increased insights into brain activity compared to EEG. Still, invasive EEG-based speech-BCIs (e.g., brain-to-speech and brain-to-text) are not yet successful due to the fact that the input brain signal and the target speech signal or text are spatially, acoustically, and temporally too distant from each other. All studies related to this topic [14], [15], [16], [17], [18], except for a feasibility experiment [19], use estimated or "indirect" articulatory information, meaning that they consider the articulatory data derived from the speech signal or the textual content. Recently, a novel database featuring parallel speech and intracranial stereotactic Electroencephalography recordings has been introduced (SingleWordProduction-Dutch-iBIDS dataset, [7]). This dataset employs a baseline linear regression method for sEEG-to-speech mapping, utilizing the Griffin-Lim algorithm for speech generation. As highlighted by Verwoert et al. in [7], the application of neural vocoders in conjunction with deep neural networks for sEEG-to-speech prediction has not been previously explored.

A. Goal of the current study

Speech is one of the most important human biosignals, but not all the characteristics of speech production are fully understood, which are required for a successful speech-based BCI [20]. A proper brain-to-speech system capable of generating full sentences in an intelligible and natural manner presents significant challenges and necessitates multidisciplinary approaches. In this paper, we apply deep neural networks for sEEG-to-speech synthesis, using neural vocoders.

In our study, we employed the Griffin-Lim algorithm as a baseline method for speech generation and used linear regression for mapping brain signals to speech features, following the methodology of Verwoert et al. [7]. This choice maintains consistency with existing literature and enables direct comparison of our results. The simplicity and ease of implementation of both techniques provide easily replicable and interpretable baselines, highlighting the improvements offered by advanced methods, such as deep learning-based solutions compared to traditional techniques.

II. Related work

A. Brain-to-speech synthesis

There has been some research on non-invasive EEG-to-speech synthesis [21], [22]. As EEG provides information only from the surface of the scalp, this process is extremely difficult, and until now there has been no successful approach to predict fully intelligible synthesized speech. On the other hand, typically more invasive methods have been tested for speech BCI [20]. With participants implanted using sEEG, audible speech could be reliably generated in real-time [23].

With intracranial electrocorticography (ECoG), another highly invasive procedure, continuous speech decoding could be solved [15]. Verwoert et al. [7] applied the Griffin-Lim algorithm in combination with linear regression to show that sEEG-to-speech mapping is feasible. According to the correlations that they received during cross-validation and comparison of 10 speakers, the results are highly dependent on the speaker, most probably because of the location of the sEEG electrodes in the individual subjects.

Another recent article, Lesaja et al. [24] presents brain2vec, a self-supervised model for learning speech-related hidden unit representations from unlabeled intracranial EEG data. Brain2vec's performance rivals that of competitive supervised learning methods on speech activity detection and word classification tasks, indicating potential practical applications in speech decoding using intracranial EEG data.

The BrainBERT model, introduced as a transformer-based model, marks a significant advancement in analyzing neural signals recorded from the human brain for natural language decoding [25]. This model, an adaptation of the well-established BERT (Bidirectional Encoder Representations from Transformers) in Natural Language Processing, is specifically designed to translate brain signals into natural language. Unlike traditional methods that predominantly rely on labeled data, BrainBERT employs self-supervised learning from extensive unlabeled data, potentially enhancing its performance. As per the original BERT model, BrainBERT records context from both directions of the input data (in this case, brain signals), which allows it to understand the temporal dependency between signals [26]. Recent studies have examined BrainBERT using sEEG data, with promising results [25].

B. Neural vocoders in speech synthesis

Since the introduction of WaveNet in 2016 [27], neural vocoders have been instrumental in generating highly natural raw samples of speech. These vocoders, including recent variants like WaveGlow [28], synthesize high-quality speech by transforming mel-spectrograms or other spectral feature inputs into audio waveforms. WaveGlow, in particular, stands out as a flow-based network capable of real-time, high-quality speech synthesis from mel-spectrograms. Its simplicity and efficiency in speech generation offer considerable advantages. This approach has been effectively utilized in various applications, such as in the work of Csapó et al. [29], who integrated WaveGlow into an ultrasound-based articulatory-to-acoustic conversion process. Similarly, Cao and colleagues demonstrated the successful use of WaveGlow for synthesizing speech from Electromagnetic Articulography (EMA) data of tongue movements [30].

C. Speaker adaptation in Text-To-Speech synthesis

A significant area of research in this field has focused on the development of natural-sounding speech synthesis. Csapó et al. have extensively explored the role of prosodic variability methods in a corpus-based unit selection text-to-speech system [31], and have worked on enhancing the naturalness of synthesized speech [32]. More recently, Mandeel et al. [33] demonstrate successful speaker adaptation experiments using Tacotron2, a state-of-the-art text-to-speech synthesis system.

These advances together show rapid progress in brain-to-speech synthesis, neural vocoders, and text-to-speech synthesis. It is anticipated that the integration of cutting-edge methods and innovative approaches will provide significant

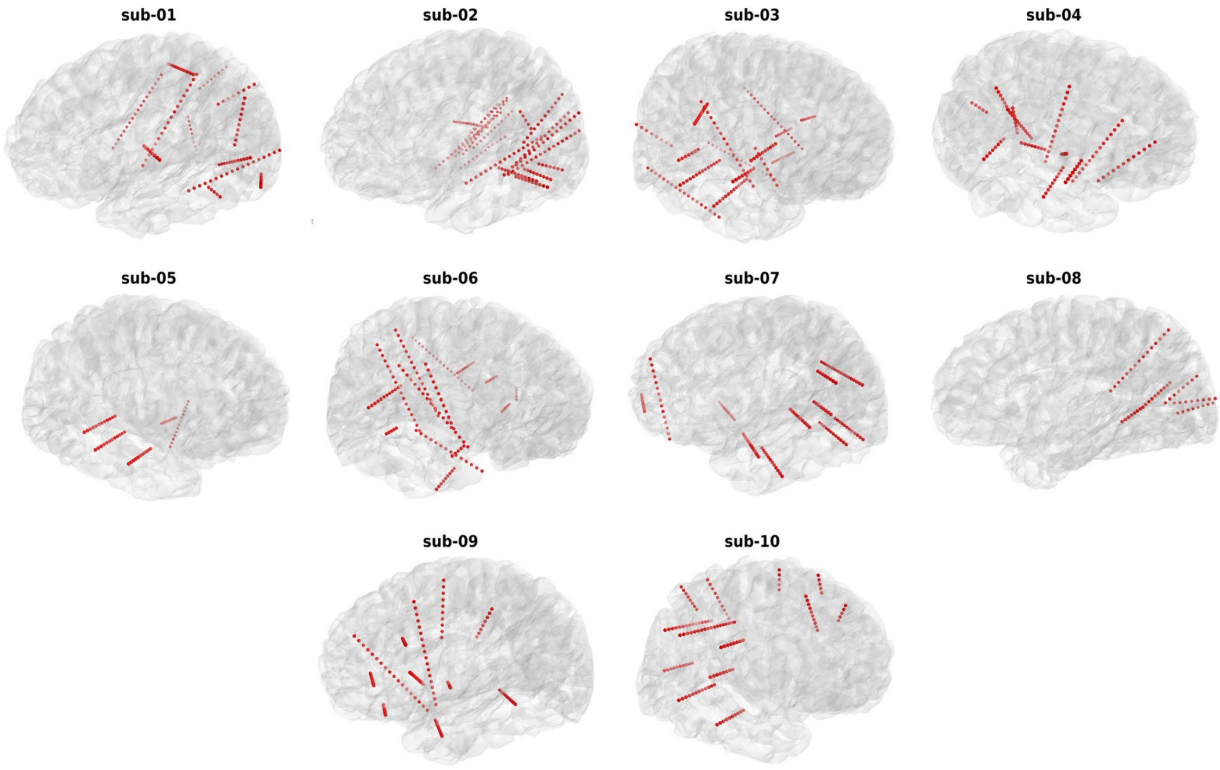


Fig. 1. The electrode locations of each participant were visualized on the surface reconstruction of their native anatomical MRI, as sourced from the SingleWordProduction-Dutch-iBIDS [7] dataset. Each red sphere in the figure represents an implanted electrode channel. This visualization is pivotal to our study as it illustrates the diverse and individualized placement of sEEG electrodes across participants, all of whom are part of the dataset used in this research. The variation in electrode placement is dictated by clinical requirements for treating epilepsy.

advancements in communications technology in the future, particularly for individuals with speech impairments.

III. METHODS

A. Data

We used the SingleWordProduction-Dutch-iBIDS dataset ([7], https://osf.io/nrgx6/) that contains in total 10 speakers with drug-resistant epilepsy (mean age 32.4 +/- 12.6 years; 5 male, 5 female). sEEG electrodes (Fig. 1.) were implanted as part of the clinical management of their epilepsy. The location of the electrodes was determined solely on the basis of clinical need. All participants were native Dutch speakers. Participants' voices were pitch-shifted to ensure anonymity. A total of 100 words were recorded for each participant, resulting in a total recording time of 300 seconds. Participants were implanted with platinum-iridium sEEG electrode arrays. Neural data were recorded using one or two Micromed SD LTM amplifier(s) with 128 channels each. Electrode connections were mapped to a common white matter contact. Data were recorded at 1024 Hz or 2048 Hz and downsampled to 1024 Hz. The audio was recorded at 48 kHz.

Recording of brain and speech signals using separate but time-aligned devices was already provided with the dataset. Synchronization is essential to ensure that each segment of EEG data corresponds to the specific speech output. This is achieved through a precise time-stamping process during recording, which aligns the EEG signals with the respective speech segments.

B. Preprocessing the brain and speech signals

On the sEEG brain signal, we followed a detailed preprocessing protocol as described in the publication we acquired the data set from [7], using the tools at https://github.com/neuralinterfacinglab/SingleWordProductionDutch/.
Specifically, we executed several steps to refine the EEG data:

Extraction of the Hilbert Envelope: We targeted the high-frequency activity (70–170Hz) for each electrode contact using a bandpass filter (4th order IIR). This step was crucial for isolating significant neural activity relevant to speech processes. Hilbert transform provides several advantages for sEEG signal analysis, including the construction of analytic signals, extraction of instantaneous amplitude and phase information, improved time-frequency analysis, envelope detection, cross-frequency coupling analysis, and applicability to non-linear and non-stationary signals. These advantages can help better understand the underlying brain activity.

Attenuation of Line Noise: To minimize electrical interference, particularly the harmonics of 50Hz line noise, we employed two bandstop filters (4th order IIR).

Temporal Windowing and Stacking: We averaged the filtered signal over 50ms windows with a 10ms frameshift. To incorporate temporal context, which is vital for understanding the

dynamics of brain activity, we stack features from multiple time windows. Specifically, for each time window of interest, we include features from the 4 preceding and 4 succeeding windows alongside the current window, totaling nine windows per feature set.

Normalization: For each feature, we normalized the data to zero mean and unit variance using the statistics from the training data. This normalization was then consistently applied to the evaluation data to maintain data integrity across different sets.

After preprocessing the sEEG signal, we calculate 80-dimensional mel-spectrogram of the speech using the 'librosa' library. During synthesis, we obtain the estimated speech using the WaveGlow model with inverse STFT transform [28], using a pre-trained model provided by NVIDIA, https://drive.google.com/file/d/1cjKPHbtAMh_4HTHmuIGNkbOkPBD9qwhj/view?usp=sharing.

Regarding the database split, we used a standard approach where the dataset was divided into training and testing subsets. Specifically, 80% of the data was used for training, and the remaining 20% for testing. This split was performed on a per-speaker basis, ensuring that the model's performance could be evaluated on unseen data from each subject.

C. Linear regression (baseline)

The baseline study [7] reconstructed the log-mel spectrogram from the high-gamma features using linear regression models. In these models, the high-gamma feature vector is multiplied with a weight matrix to reconstruct the log-mel spectrogram. The weights are determined using a least-squares approach. For the waveform reconstruction, they utilized the Griffin-Lim method.

D. Deep learning architectures

Next, we train the deep learning algorithms, which receive windowed sEEG Hilbert transformed components as input and produce 80-dimensional mel-spectral coefficients as output.

As for the hyperparameters, the learning rate, number of epochs, and other training parameters were selected through a series of preliminary experiments aimed at optimizing model performance. The number of epochs was set to 100, with early stopping using a patience of three, to prevent overfitting. The learning rate was initially set to a standard value of 0.001 and was adjusted based on the model's performance during the validation phase. Regarding learning rate scheduling, we used a dynamic approach where the learning rate was halved if there was no improvement in model performance on the validation set for two epochs.

Our method is illustrated in Figure 2, which shows the general flow from the raw sEEG input to the final synthetic speech. In order to obtain an analytical signal from the sEEG data, the Hilbert transform is used to acquire both amplitude and phase information (as detailed in Sec. III-B). We then apply the transformed signal as the input of our neural network models, including FC-DNN, 2D-CNN, and 3D-CNN. Based on the sEEG input, these models are trained to predict the mel-spectrograms of speech, thereby creating a mapping between brain activity and acoustic representations of speech. WaveGlow neural vocoder is used to convert the predicted mel-spectrogram into audible speech.

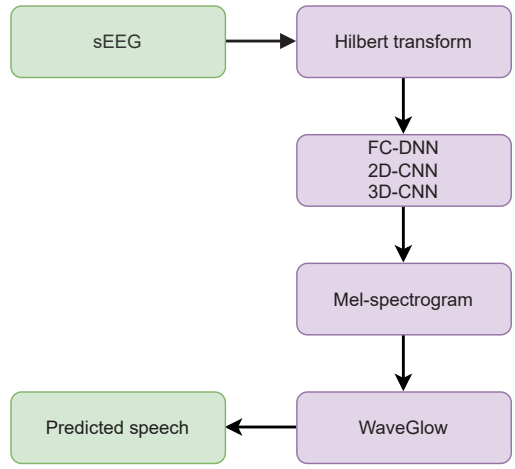


Fig. 2. General block diagram of our methods: from sEEG input, we predict mel-spectrogram of speech, which is synthesized to audio using a neural vocoder.

1) FC-DNN Architecture: We utilized a Fully Connected, Feed-Forward Deep Neural Network (FC-DNN) as our foundational model. This architecture incorporates five hidden layers, each consisting of 1000 neurons. We employed a Rectified Linear Unit (ReLU) as the activation function. The network's input layer has a dimensionality of 1143, which represents features calculated from a combination of 127 EEG channels and 9 temporal windows, as detailed in Section III-B. The output layer features 80 neurons, corresponding to the number of mel-spectral coefficients.

2) 2D-CNN: Our 2D convolutional network starts with two convolutional layers, each equipped with a 5x5 kernel size, having swish activation. The input data is formatted as 9x127 dimensions (9 temporal windows with 127 features in each). After a maxpooling layer, there is a third convolutional layer. The filter sizes are 30, 60 and 70. Dropout layers with a rate of 0.2 are used. Subsequent to the convolutional layers, the network architecture includes two fully connected layers. The first fully connected layer contains 1000 neurons. The final layer in our 2D-CNN model is the output layer, having linear activation, and designed with 80 neurons to match the number of mel-spectral bands for the waveform reconstruction.

3) 3D-CNN: Standard CNN considers 2D images to extract features by convolving 2D filters over images. Therefore, to model temporal information, a third dimension has to be considered [34], [35]. Here we use a 3D-CNN variation by adding a third dimension as (2+1)D CNN which shows good performance in video action recognition task [36]. It also shows good results when used with ultrasound images and it could be considered as a substitute of CNN+LSTM [37]. This network processed 5 frames of input that were 6 frames apart (6 is the stride parameter of the convolution along the time axis) [37]. Following the concept of (2+1)D convolution, the 5 frames were first processed only spatially, and then got combined along the time axis just below the uppermost dense

TABLE I
MCD SCORES ON THE TEST SET.

Speaker	Mel-Cepstral Distortion (dB)			
	Linear Regression with Griffin-Lim	FC-DNN with WaveGlow	3D-CNN with WaveGlow	2D-CNN with WaveGlow
sub-01	6.25	4.63	4.86	4.64
sub-02	6.41	4.95	5.19	4.98
sub-03	5.52	4.39	4.50	4.51
sub-04	5.28	4.16	4.86	4.50
sub-05	6.20	6.12	6.08	6.39
sub-06	4.36	3.63	4.16	4.10
sub-07	5.50	4.32	5.39	4.31
sub-08	5.03	5.00	5.50	5.13
sub-09	5.12	4.29	5.56	5.15
sub-10	4.26	4.01	4.34	4.13
Mean	5.39	4.55	5.04	4.78

layer.

Our 3D model begins with an input layer that handles the reshaped sEEG data, formatted into a 9x127 dimension. To accommodate the 3D processing, the data is expanded into a five-dimensional structure, ensuring compatibility with the subsequent 3D convolutional layers. The core of our 3D-CNN comprises three convolutional layers, each utilizing a kernel size of (5, 13, 13), strides set to (6, 2, 2), and having swish activation. These layers are designed to extract and analyze both spatial and temporal features from the sEEG data. There is a maxpooling layer after the second convolution. The filter sizes are 30, 60 and 70. Dropout layers with a rate of 0.2 are used. Subsequent to the convolutional layers, the network architecture includes two fully connected layers, similarly to the 2D-CNN, finally predicting the 80-dimensional mel-spectrogram.

After the trainings with the above deep neural networks, the predicted spectrograms of the test data are used to synthesize speech using the WaveGlow vocoder (Sec. III-B).

IV. RESULTS

A. Demonstration sample

Fig. 3 a) shows the spectrogram of a natural utterance and b–e) those of synthesized speech from sEEG input with linear regression (baseline from [7]) and various DNNs. The synthesized speech has a similar envelope as the natural speech, but few of the spectral details are included. Although the speech reconstructed from the mel-spectral parameters estimated on the test pile resembles the original speech, it is noisy and difficult to understand. However, in some parts, sections of synthesized speech (e.g. vowels) are similar to the original audio. Synthesized samples are available at http://smartlab.tmit.bme.hu/icj2023_sEEG.

B. Objective evaluation

To check whether the proposed DNNs can reproduce the features of the original speech, we evaluated the spectral differences between natural speech and synthesized speech using Mel-Cepstral Distortion (MCD) [38], which is a standard metric for text-to-speech synthesis evaluation. As MCD is an error measure, lower values indicate higher similarity between the original and synthesized speech. Table I displays the MCD values calculated on the test data for each speaker. In

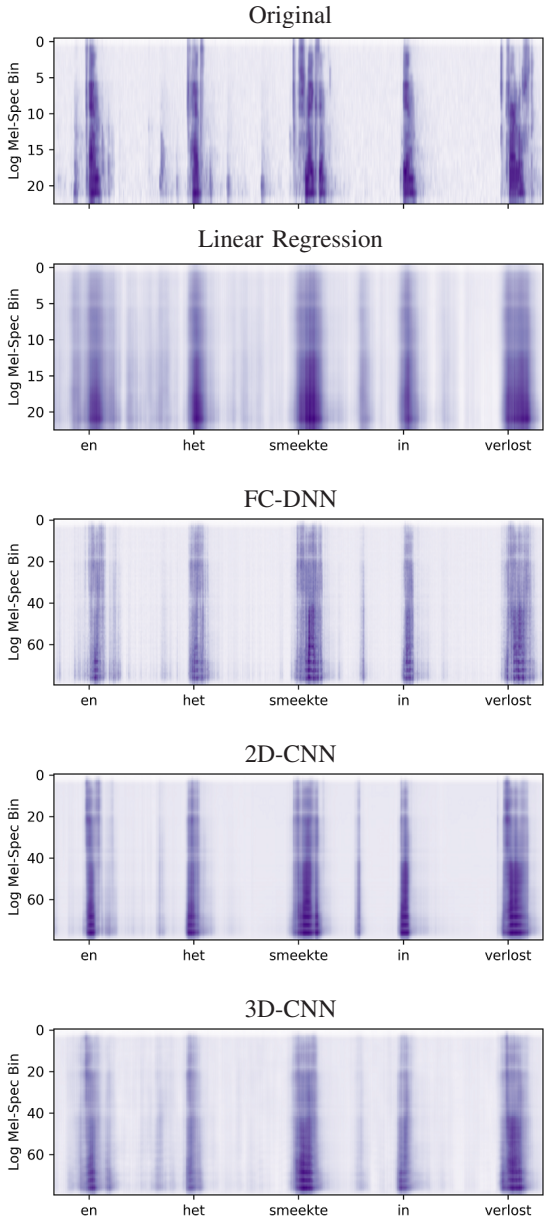


Fig. 3. Speech samples from speaker sub-06: a) original, b) synthesized using LR (baseline) c) FC-DNN, d) 2D-CNN, e) 3D-CNN.

combination with a WaveGlow vocoder, the Fully Connected Deep Neural Network (FC-DNN) model consistently produced the lowest MCD values across all speakers tested. Therefore, this combination of models and vocoders is the most effective means of reproducing speech that is resembling the original.

It is interesting to note the variation in MCD values among different speakers. For instance, speaker sub-06 consistently showed lower MCD values across all models, indicating that the acoustic features of this speaker might be easier for the models to learn and reproduce. This observation suggests that individual characteristics of each speaker's data, and most probably, the electrode positions can significantly influence the performance of speech synthesis models. The comparison with the correlations in [7] provided intriguing insights. Speakers with higher brain-speech signal correlation generally had lower MCD values, reinforcing the potential link between these two metrics. Prior studies [15] have also suggested a possible connection between neural correlates and the quality of speech synthesis.

C. Subjective evaluation

In order to determine which proposed version is closer to natural speech, we conducted an online MUSHRA-like test [39].

Our aim was to compare the natural words with the synthesized words of the baseline and the proposed approaches. In the test, the listeners had to rate the naturalness of each stimulus in a randomized order relative to the reference (which was the natural utterance), from 0 (very unnatural) to 100 (very natural). Out of the 10 speakers used in the earlier analysis, we selected four speakers for the listening test, based on the correlation analysis between brain and speech signals (Fig. 4 of [7]): 'sub-04/F', 'sub-06/M' (high correlation), and 'sub-01/F', 'sub-02/M' (low correlation). We selected four words from the test set of each speaker (altogether 16 words, each being 2 seconds long). The variants appeared in randomized order (different for each listener).

Each word was rated by non-Dutch speakers: altogether 9 listeners participated in the test; 7 males, 2 females; ages: 23-39 (avg: 32). The test took 5–28 minutes (avg: 11 minutes) to complete. Fig. 4 top shows the average naturalness scores for the tested approaches. The benchmark (Linear Regression) version achieved the lowest scores, while the natural words were rated the highest, as expected. The proposed DNN and neural vocoder based versions were performed over the baseline system for all speakers. In the overall figure, we can see a slight preference towards the FC-DNN, compared to the convolutional neural networks. To check the statistical significances, we conducted Mann-Whitney-Wilcoxon ranksum tests with a 95% confidence level. Based on this, the differences between FC-DNN, 2D-CNN, and 3D-CNN are not statistically significant.

When vizualizing the results speaker by speaker (Fig. 4 bottom), we can see the following trends: for the female speakers (sub-01 and sub-04), the 2D-CNN was preferred most, whereas this is not the case for the male speakers (sub-02 and sub-06). Based on the earlier correlation analysis on

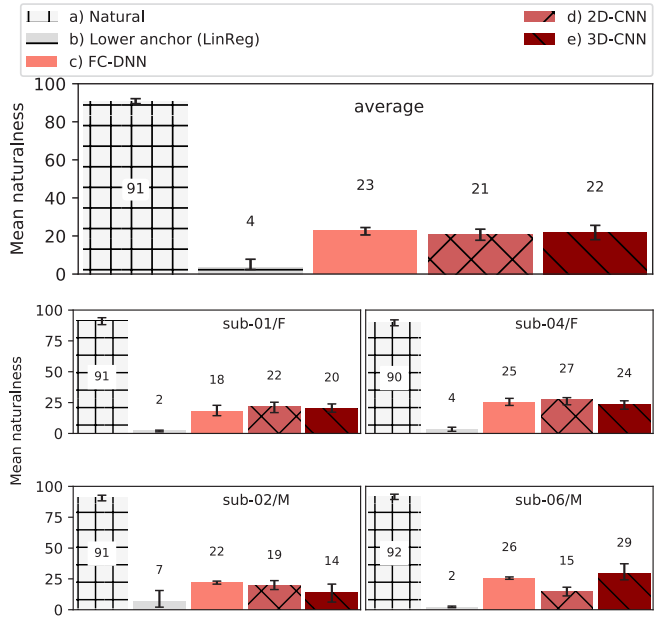


Fig. 4. *Results of the subjective evaluation with respect to naturalness, speaker by speaker (top) and average (bottom). The errorbars show the 95% confidence intervals.*

the speakers in [7] we have seen that sub-04 and sub-06 had a higher overall correlation between brain and speech signals, and this is clearly reflected in the speaker-by-speaker results of the listening test: both of them achieved reasonably higher naturalness scores compared to sub-01 and sub-02. Regarding MWM ranksum tests, the only case when the results are statistically significant, is sub-06: here, the 2D-CNN was ranked significantly lower than FC-DNN and 3D-CNN, while the difference between the latter two is not significant, but 3D-CNN is slightly preferred.

As a summary of the evaluation, the objective MCD score was not always found to be helpful in our case (i.e, it does not highly correspond to the correlations of [7]), but clearly, the subjective listening test could show the differences between the speakers of low and high correlation. The relatively low naturalness scores (18–29) indicate that sEEG-based synthesized speech is far from being intelligible, but at least, has properties similar to the natural speech signal.

V. Discussion and Conclusions

In this paper, we applied deep neural networks (FC-DNN, 2D-CNN, and 3D-CNN) for sEEG-to-melspectrogram prediction. Next, we synthesized speech using the WaveGlow neural vocoder. Our objective evaluation (Mel-Cepstral Distortion) has shown that the DNN-based approaches with neural vocoder outperform the baseline linear regression model using Griffin-Lim for speech generation [7].

Various studies have demonstrated the feasibility of ECoG-to-text [40] and ECoG-to-speech [15] conversion using different methodological approaches, such as linear regression and deep neural networks. However, their applicability in sEEG-to-speech conversion remained largely unexplored. Our work, therefore, complements these efforts and provides an

alternative approach to sEEG-to-speech synthesis. Compared to traditional methods such as Griffin-Lim, neural vocoders represent an advance in generating more natural-sounding speech than traditional methods. While the complexity of sEEG data presented significant challenges, our approach utilizing deep neural networks and a neural vocoder showed promising results in comparison to the baseline linear regression model.

However, we acknowledge that the quality of synthesized speech remains an area for improvement. Our models produced speech that has distinct speech-like characteristics but was not yet fully understandable. This is a common issue encountered in the field of brain-to-speech synthesis, including studies utilizing EEG and ECoG data.

The reason why the 2D-CNNs and 3D-CNNs produced samples with larger errors in the current study might be that the amount of training data is extremely limited (i.e., only 100 words / 300 seconds), and more complex networks cannot learn the necessary mapping. Another explanation for the low 2D-CNN and 3D-CNN results might be that as our sEEG input data is put together in a specific way (i.e, brain signal is windowed, and Hilbert-transformed values are stacked together), this type of image is difficult to process for a convolutional neural network. On the other hand, the differences are highly dependent on the speaker (and thus, most probably on the electrode positioning) : with sub-06, who had the highest correlations in [7], the 3D-CNN performed best, indicating that there is potential in applying convolutional neural networks for this task.

Both the subjective listening tests and objective evaluations show that the neural network-based approaches outperformed the linear regression baseline. The relatively low naturalness scores (18–29) indicate that sEEG-based synthesized speech is far from being intelligible, but clearly, has properties similar to the natural speech signal, both visually on the spectrograms, and when listening to the samples. Therefore, we expect that our results might help future speech-based Brain-Computer Interfaces.

VI. Future work

Deep learning is vast and ever-evolving, providing ample opportunity to refine our sEEG-to-speech prediction models. One approach to enhance the current results could involve experimenting with different architectures and types of deep learning models. For instance, Transformer models [41], known for their effectiveness in various natural language processing tasks, could be explored for sEEG-to-speech synthesis. We may be able to gain valuable insights into how different brain regions contribute to speech production through the attention mechanism in Transformers, potentially enabling us to improve our predictive abilities [41]. We acknowledge that the efficacy of complex models like Transformers is contingent on the availability of substantial training data. However, we expect that as more and more research groups are dealing with speech and brain signal recording and processing, such larger datasets might be available in the future.

Our feature extraction process currently involves windowing the raw sEEG data and applying the Hilbert transform. However, future work could involve more sophisticated feature extraction techniques like Wavelet Transform [42] or Fourier Transform [43]. These techniques could capture different aspects of the sEEG signals, leading to improved performance of the models [44].

In terms of data, our current study is based on the SingleSpeechProductionDutch dataset [7]. While this dataset has provided valuable insights, we recognize the potential benefits of using a more extensive and diverse dataset. Consequently, we intend to record our database, expanding the pool of speakers and potentially improving the generalizability and robustness of the model. Nevertheless, it is important to note that we will use EEG signals rather than sEEG for our planned dataset, which may present new challenges and opportunities.

Furthermore, it may be beneficial to explore applying more advanced post-processing techniques. The WaveGlow neural vocoder is currently employed for speech synthesis, but future work could investigate the use of more recent vocoding techniques, like AutoVocoder [45], to enhance the quality of the speech synthesised.

The positions of sEEG electrodes in the dataset were determined by clinical needs in the treatment of epilepsy, which can influence the quality of synthesized speech [7]. This is supported by existing literature, which shows that electrodes placed closer to key speech areas, particularly in the left frontal lobe, are more likely to capture neural signals that are crucial for accurate speech synthesis. This theoretical understanding, underpinned by neurophysiological insights into speech production processes, suggests that variations in electrode arrangements could result in differences in the quality of synthesized speech. However, a detailed correlation analysis between electrode positions and synthesized speech quality was beyond the scope of our current study, presenting a valuable direction for future research.

Finally, we see many potential applications for sEEG-to-speech synthesis in the future. Due to rapid advances in deep learning, we anticipate improving our models and contributing to the development of speech-based Brain-Computer Interfaces in the future, as well as improving their performance.

VII. Acknowledgements

The research was partially funded by the National Research, Development and Innovation Office of Hungary (FK 142163 grant). T.G.Cs. was supported by the Bolyai János Research Fellowship of the Hungarian Academy of Sciences and by the ÚNKP-22-5-BME-316 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.

References

[1] D. Dupré and A. Karjalainen, “Employment of disabled people in Europe in 2002,” *Statistics in focus*, pp. 3–26, 2003.

[2] Hungarian Central Statistical Office, “2011. évi népszámlálás, 11. fogyatékossággal élők,” *Tech. Rep.*, 2011. [Online]. Available: http://www.ksh.hu/docs/hun/xftp/idoszaki/nepsz2011/nepsz_11_2011.pdf

[3] M. Lecerf, “Employment and disability in the European Union,” *European Parliamentary Research Service (EPRS)*, no. May, pp. 1–7, 2020.

[4] "White Rose restoring the larynx has https://www.whiterose.ac.uk/collaborationfunds/silent-speech-restoring-the-power-of-speech-to-people-whose-larynx-has-been-removed/

[5] E. F. Chang and G. K. Anumanchipalli, "Toward a speech neuroprosthesis," *JAMA*, vol. 323, no. 5, pp. 413–414, feb 2020, DOI: 10.1001/JAMA.2019.19813.

[6] D. J. McFarland and J. R. Wolpaw, "EEG-based brain–computer inter-faces," *Current Opinion in Biomedical Engineering*, vol. 4, pp. 194–200, dec 2017, DOI: 10.1016/J.COBME.2017.11.004.

[7] M. Verwoert, M. C. Ottenhoff, S. Goulis, A. J. Colon, L. Wagner, S. Tousseyn, J. P. van Dijk, P. L. Kubben, and C. Herff, "Dataset of speech production in intracranial electroencephalography," *Scientific Data 2022 9:1*, vol. 9, no. 1, pp. 1–9, jul 2022. Available: https://www.nature.com/articles/s41597-022-01542-9 [Online]. DOI: 10.1038/s41597-022-01542-9.

[8] G. Buzsáki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes," *Nature Reviews Neuroscience*, vol. 13, no. 6, pp. 407–420, may 2012. Available: https://www.nature.com/articles/nrn3241 [Online]. DOI: 10.1038/nrn3241.

[9] D. Dash, P. Ferrari, A. Babajani-Feremi, A. Borna, P. D. Schwindt, and J. Wang, "Magnetometers vs Gradiometers for Neural Speech Decoding," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2021, pp. 6543– 6546, nov 2021, DOI: 10.1109/EMBC46164.2021.9630489.

[10] F. Lotte, L. Bougrain, and M. Clerc, "Electroencephalography (EEG)-Based Brain–Computer Interfaces," *Wiley Encyclopedia of Elec- trical and Electronics Engineering*, pp. 1–20, sep 2015, DOI: 10.1002/047134608X.W8278.

[11] A. J. Casson, "Wearable EEG and beyond," *Biomedical engineering letters*, vol. 9, no. 1, pp. 53–71, feb 2019. Available: https://pubmed.ncbi.nlm.nih.gov/30956880/ [Online]. DOI: 10.1007/S13534-018-00093-6.

[12] T. G. Csapó, F. V. Arthur, P. Nagy, and A. Boncz, "A beszéd artikulációs mozgásának predikciója agyi jel alapján – kezdeti eredmények," in *XIX. Magyar Számítógépes Nyelvészeti Konferencia, MSZNY 2023*, 2023, pp. 357–368. [Online]. Available: https://m2.mtmt.hu/api/publication/33599995

[13] F. V. Arthur and T. G. Csapó, "Deep learning alapú agyi jel feldolgozás és beszédszintézis előkészítő munkálatai," in *XVIII. Magyar Számítógépes Nyelvészeti Konferencia: MSZNY 2022,* 2022, pp. 185–198. [Online]. Available: https://m2.mtmt.hu/api/publication/32636136

[14] J. Chartier, G. K. Anumanchipalli, K. Johnson, and E. F. Chang, "Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex," *Neuron,* vol. 98, pp. 1042–1054.e4, 6 2018, DOI: 10.1016/j.neuron.2018.04.031.

[15] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, pp. 493– 498, 4 2019, DOI: 10.1038/s41586-019-1119-1.

[16] J. Brumberg, E. Wright, D. Andreasen, F. Guenther, and P. Kennedy, "Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech motor cortex," *Frontiers in Neuroscience*, vol. 5, 2011, DOI: 10.3389/fnins.2011.00065.

[17] G. Le Godais, "Decoding speech from brain activity using linear methods," Theses, Université Grenoble Alpes [2020-....], Jun. 2022. [Online]. Available: https://theses.hal.science/tel-03852448

[18] F. H. Guenther, J. S. Brumberg, E. J. Wright, A. Nieto-Castanon, J. A. Tourville, M. Panko, R. Law, S. A. Siebert, J. L. Bartels, D. S. Andreasen, P. Ehirim, H. Mao, and P. R. Kennedy, "A wireless brain-machine interface for real-time speech synthesis," *PLOS ONE*, vol. 4, no. 12, pp. 1–11, 12 2009, DOI: 10.1371/journal.pone.0008218.

[19] S. Lesaja, C. Herff, G. D. Johnson, J. J. Shih, T. Schultz, and D. J. Krusienski, "Decoding lip movements during continuous speech using electrocorticography," in *2019 9th International IEEE/EMBS University power Consortium – Silent of speech to people Speech: whose Available: been removed." [Online]. Conference on Neural Engineering (NER)*, 2019, pp. 522–525, DOI: 10.1109/NER.2019.8716914.

[20] S. Luo, Q. Rabbani, and N. E. Crone, "Brain-computer interface: Applications to speech decoding and synthesis to augment communication," *Neurotherapeutics 2022*, vol. 1, pp. 1–11, jan 2022, DOI: 10.1007/S13311-022-01190-2.

[21] G. Krishna, C. Tran, Y. Han, M. Carnahan, and A. H. Tewfik, "Speech synthesis using EEG," in *Proc. ICASSP*, online, 2020, pp. 1235–1238, DOI: 10.1109/ICASSP40776.2020.9053340.

[22] G. Krishna, C. Tran, M. Carnahan, and A. H. Tewfik, "Advancing speech synthesis using EEG," *International IEEE/EMBS Conference on Neural Engineering, NER*, vol. 2021-May, pp. 199–204, may 2021, DOI: 10.1109/NER49283.2021.9441306.

[23] M. Angrick, M. C. Ottenhoff, L. Diener, D. Ivucic, G. Ivucic, S. Goulis, J. Saal, A. J. Colon, L. Wagner, D. J. Krusienski, P. L. Kubben, T. Schultz, and C. Herff, "Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity," *Communications Biology*, vol. 4, no. 1, p. 1055, 2021, DOI: 10.1038/s42003-021-02578-0.

[24] S. Lesaja, M. Stuart, J. J. Shih, P. Soroush, T. Schultz, M. Manic, and D. J. Krusienski, "Self-supervised learning of neural speech representations from unlabeled intracranial signals," *IEEE Access*, 2022, DOI: 10.1109/ACCESS.2022.3230688.

[25] C. Wang, V. Subramaniam, A. U. Yaari, G. Kreiman, B. Katz, I. Cases, and A. Barbu, "BrainBERT: Self-supervised representation learning for intracranial recordings," 2023, DOI: 10.48550/arxiv.2302.14367.

[26] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2019, DOI: 10.48550/arxiv.1810.04805.

[27] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu, "WaveNet: A Generative Model for Raw Audio," *CoRR*, vol. abs/1609.0, 2016, DOI: 10.48550/arXiv.1609.03499.

[28] R. Prenger, R. Valle, and B. Catanzaro, "Waveglow: A flow-based generative network for speech synthesis," in *Proc. ICASSP*, Brighton, UK, 2019, pp. 3617–3621, DOI: 10.1109/ICASSP.2019.8683143.

[29] T. G. Csapó, Zainkó, Cs., L. Tóth, G. Gosztolya, and A. Markó, "Ultrasound-based articulatory-to-acoustic mapping with WaveGlow speech synthesis," in *Proc. Interspeech, 2020*, pp. 2727–2731, DOI: 10.21437/Interspeech.2020-1031.

[30] B. Cao, A. Wisler, and J. Wang, "Speaker adaptation on articulation and acoustics for articulation-to-speech synthesis," *Sensors*, vol. 22, no. 16, p. 6056, 2022, DOI: 10.3390/S22166056.

[31] T. G. Csapó, Cs. Zainkó, and G. Németh, "A study of prosodic variability methods in a corpus-based unit selection text-to-speech system," *Infocommunications Journal*, vol. LXV, p. 2010, 01 2010.

[32] T. G. Csapó, G. Németh, and M. Fék, "Szövegfelolvasó természetességének növelése," *Híradástechnika*, vol. LXIII, p. 2008, 05 2008.

[33] A. R. Mandeel, M. S. Al-Radhi, and T. G. Csapó, "Speaker adaptation experiments with limited data for end-to-end text-to-speech synthesis using tacotron2," *Infocommunications Journal*, vol. 14, pp. 55–62, 2022, DOI: 10.36244/ICJ.2022.3.7.

[34] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221–231, 2012, DOI: 10.1109/TPAMI.2012.59.

[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, nov 1997, DOI: 10.1162/neco.1997.9.8.1735.

[36] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, "A closer look at spatiotemporal convolutions for action recognition," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, pp. 6450–6459, DOI: 10.48550/arXiv.1711.11248.

[37] L. Tóth and A. H. Shandiz, "3D Convolutional Neural Networks for Ultrasound-Based Silent Speech Interfaces," in *Proc. ICAISC*, Zakopane, Poland, 2020, DOI: 10.48550/arXiv.2104.11532.

[38] R. F. Kubichek, "Mel-cepstral distance measure for objective speech quality assessment," in *Proceedings of IEEE Pacific Rim Conference on Communications Computers and Signal Processing*, Victoria, Canada, 1993, pp. 125–128, DOI: 10.1109/pacrim.1993.407206.

[39] "ITU-R Recommendation BS.1534: Method for the subjective assessment of intermediate audio quality," 2001.

[40] C. Herff, D. Heger, A. de Pesters, D. Telaar, P. Brunner, G. Schalk, and T. Schultz, "Brain-to-text: decoding spoken phrases from phone representations in the brain," *Frontiers in Neuroscience*, vol. 8, 2015, DOI: 10.3389/fnins.2015.00217.

[41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 2017-Decem. Neural information processing systems foundation, jun 2017, pp. 5999–6009, DOI: 10.48550/arxiv.1706.03762.

[42] A. Graps, "An introduction to wavelets," *IEEE Computational Science and Engineering*, vol. 2, no. 2, pp. 50–61, Summer 1995, DOI: 10.1109/99.388960.

[43] R. Bracewell, *The Fourier Transform and Its Applications*, ser. Circuits and systems. McGraw-Hill, 1978.

[44] A. K. Singh and S. Krishnan, "Trends in EEG signal feature extraction applications," *Frontiers in Artificial Intelligence*, vol. 5, p. 1072801, jan 2023, DOI: 10.3389/FRAI.2022.1072801/BIBTEX.

[45] J. J. Webber, C. Valentini-Botinhao, E. Williams, G. E. Henter, and S. King, "Autovocoder: Fast Waveform Generation from a Learned Speech Representation using Differentiable Digital Signal Processing," in *Proc. ICASSP*, Rhodes, Greece, 2023, DOI: 10.48550/arxiv.2211.06989.

Frigyes Viktor Arthur is a PhD student at BME in the field of computer science. He has a background in biomedical engineering, with a master's degree, and has significant programming experience from various side-projects mainly related to medical image processing, such as Intracranial Hemorrhage Detection, and native and cross-platform Android and iOS mobile application development. His research interests include deep learning-based brain signal and speech processing, Brain-Computer Interfaces, EEG, ECG, and multimodal biological signals.

Tamás Gábor Csapó, PhD is currently a senior research fellow at the Speech Technology and Smart Interactions Laboratory (BME-SmartLab), Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics. He received his PhD at BME in 2014 about text-to-speech synthesis, with Géza Németh as the supervisor. Meanwhile, in 2014, he was a Fulbright scholar at Indiana University in Dr. Steven Lulich's lab, where he started his research on ultrasound tongue imaging and automatic contour tracking. Between 2016–2021, he was an active member of MTA–ELTE Lingual Articulation Research Group, where he started Hungarian articulatory research using ultrasound). Between 2017–2022, he had two national OTKA postdoc projects (FK-17 and PD-18) on articulatory-to-acoustic mapping using ultrasound/lip/vocal tract MRI. Currently, he is the PI of an FK-22 OTKA project on the analysis of articulation and brain signals for speech-based brain-computer interfaces, using EEG and UTI. He has 150+ publications, most of them in high-ranked conferences (e.g., Interspeech or ICASSP) and in top-ranked journals (e.g., IEEE Selected Topics in Signal Processing, Journal of the Acoustical Society of America – Express Letters).

On the Performance of Metamaterial based Printed Circuit Antenna for Blood Glucose Level Sensing Applications: A Case Study

Taha A. Elwi[1,2]*, Hayder H. Al-Khaylani[3,4], Wasan S. Rasheed[5], Sana A. Al-Salim[6], Mohammed H. Khalil[7], Lubna Abbas Ali[8], Omar Almukhtar Tawfeeq[6], Saba T. Al-Hadeethi[7], Dhulfiqar Ali[1], Zainab S. Muqdad[8], Serkan Özbay[9], and Marwah. M. Ismael[10]

***Abstract*—Due to the urgent need to develop technologies for continuous glucose monitoring in diabetes individuals, potential research has been applied by invoking the microwave techniques. Therefore, this work presents a novel technique based on a single port microwave circuit, antenna structure, based on Metamaterial (MTM) transmission line defected patch for sensing the blood glucose level in noninvasive process. For that, the proposed antenna is invoked to measure the blood glucose through the field leakages penetrated to the human blood through the skin. The proposed sensor is constructed from a closed loop connected to an interdigital capacitor to magnify the electric field fringing at the patch center. The proposed antenna sensor is found to operate excellently at the first mode, 0.6GHz, with S11 impedance matching less than -10dB. The proposed sensor performance is tested experimentally with 15 cases, different patients, through measuring the change in the S11 spectra after direct touching to the sensor a finger print of a patient. The proposed sensor is found to be effectively very efficient detector for blood glucose variation with a low manufacturing cost when printed on an FR-4 substrate. The experimental measurements are analyzed mathematically to obtain the calibration equation of the sensor from the curve fitting. Finally, the theoretical and the experimental results are found to be agreed very well with a percentage of error less than 10%.**

***Index Terms*—Glucose, sensor, MTM, noninvasive.**

[1] International Applied and Theoretical Research Center (IATRC), Baghdad Quarter, Iraq, (e-mail: taelwi82@gmail.com)

[2] Islamic University Centre for Scientific Research, The Islamic University, Najaf, Iraq

[3] Laser and Optoelectronics Engineering Department, University of Technology, Baghdad, Iraq

[4] Computer Techniques Engineering, Al Hikma University College, Baghdad, Iraq

[5] Department of Information and Communications Engineering, Al-Khwarizmi College, University of Baghdad, Iraq

[6] Department of Mechatronics, Al-Khwarizmi College of Engineering, Baghdad University, Iraq

[7] Media Technology and Communication Engineering Department/College of Engineering, University of Information Technology and Communications, Baghdad, Iraq

[8] Electrical Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq

[9] Electrical Electronics Engineering Department, Gaziantep University, Turkey

[10] Department of Information and Communication Engineering, College of Information Engineering, Al-Nahrain University, Baghdad, Iraq.

I. INTRODUCTION

Viktor Veselago first presented metamaterials (MTM) in 1967 [1]. Such structures may be called artificial materials with nontraditional properties [2]. These materials possess negative permittivity (ε) and negative permeability (μ) that in turn support the backward wave propagation of electromagnetic waves [3]. In 1999 an interesting subwavelength element realized as split ring resonator (SRR) to achieve negative permeability was proposed by Pendry [4]. Basically, SRR can be represented as tank of LC circuit possessing equivalent inductance (L) and the capacitance (C) between two concentric rings resonating at specific frequency [4]. The SRR has normally a size much less than, around wavelength (λ/10), the guided wavelength. Therefore, many researchers applied SRR as technique to retrieve the characterizations of materials [4].

In such process, the changes in the scattering parameters (S-parameters) with respect to a sample under test (SUT) introduction can be transferred through certain algorithm to retrieve the materials characterizations.

Microwave sensing is a reliable method for liquid characterization that has been employed in the past decade [5]. SRRs [6], complementary SRRs (CSRRs) [3], open CSRRs (OCSRRs) [4], closed ring resonator [5] and other miniaturized microwave resonators [6] based on transmission lines attracted a considerable attention of researchers for biomedical applications [6]. The electromagnetic properties of these structures depend on their operation frequency and quality factor changes with respect to different liquid introductions [7]. Through measuring S-parameters coefficients, complex dielectric parameters of SUT, the liquid characterizations can be addressed [8]. This has provided a new sensing platform for the biological, pharmaceutical, and fuel industries [9]. Most microwave biological sensors are mounted under ultra-thin cylindrical pipes [10] or slotted cylindrical tubes [11]. Based on measuring S-parameters magnitudes at certain frequency, materials under test losses can be extracted for quality detection [12]. Such technology, however, the resolution of high-losses was found a significant concern due to the issue of skin depth penetration [13]. On the other hand, most suggested methods require micro fluidic channel tubes to contain SUT; that add

DOI: 10.36244/ICJ.2024.1.7

extra losses and difficulty of penetrations [14]. Nevertheless, many sensors detection process is based on certain volume of SUT that limits their use for real-time monitoring applications [15]. Moreover, an additional size and fabrication cost could be ensure field retardation and phase change to analyze the dielectric characterization [17] in specific for biological solvent detection.

Based on the current development of microwave sensors technologies, fractal based MTM structures attracted researchers to realize effective and efficient sensors to achieve a high selectivity [18]. On top of that, MTM realizes excellent performance in the microwave ranges due to their nontraditional properties [19]. Moreover, MTM based fractal geometries use remain very excellent candidates in the biomedical aspects due to their size reduction in comparison to traditional microwave structures. For example, a MTM defected patch-based monopole antenna was presented in [20] for pollution detections. Authors in [21] proposed a study of using a traditional microstrip transmission line for liquid properties detection. Then, an extended study based on a fractal MTM structure was presented for blood glucose sensing using a microstrip transmission line loaded with carbon nanotube patch in [22]. The use of MTM patch-based nanoscale structures was introduced for gas detection in [23] and [24]. MTM based fractal structures were applied to realize as MTM defect on the ground plane for cancer cell detection based on their electrical properties' changes [25].

In this work, the proposed work is a design of a sensor antenna structure based on MTM inclusion for sensing applications. This paper, a new approach based on a single port antenna element via MTM transmission line structure is proposed for blood glucose level detection. The MTM is constructed from an interdigital capacitor structure with a closed loop ring. The proposed structure is designed for glucose detection. The paper is organized as follows: The geometrical details are discussed in section II. The analysis process and the parametric study are presented in section III. The experimental measurements are explained in section IV after conducting 15 cases for the proposed study. The paper is concluded in section V.

II. Sensor Design and Details

The proposed antenna is mounted on an FR-4 substrate of a dielectric constant ε_r=4.3 and loss tangent tanδ=0.025 with thickness h=1.6 mm. The copper metal is 35µm. The proposed antenna is compacted on 30×30 mm2 size. The antenna is fed with a 50Ω microstrip transmission line of 7.25mm width and extended to touch the patch at length of 1.5mm to be connected directly to the radiating structure. The proposed closed loop design details are shown in Fig1(a). The back panel is covered completely copper as appeared in Fig1(b).

In addition to the proposed closed loop in this design, a copper interdigital capacitor (C_{int}), see Fig1(c), is conducted to maximize the electrical field intensity [12] at the center of the patch where SUT would be positioned. All geometrical details of Fig1 are listed in Table 1.

III. MTM Patch Theoretical Analysis

As mentioned later, the proposed radiating patch consists of a closed loop coupled with C_{int}. Therefore, the proposed patch is structured in such way to provide maximum fringing from the C_{int} edges. The C_{int} unit cell is constructed as interfaced strip lines with an effective length (L_n) to provide the desirable resonant frequency (f_r). The resulted capacitance value of the used C_{int} can be calculated analytically from the following equation [26]:

$$C = \frac{\varepsilon_{re}\times 10^{-3}}{18\pi}\frac{K(k)}{K'(k)}(n-1)\frac{L}{10^{-6}} \quad (1)$$

where C is the capacitance in pF, ε_{re} is the effective relative permittivity, n is the capacitor figure number, L is the finger length, K and K' are elliptical integral coefficients and they are given as [27]:

$$K(k) = \int_0^{\frac{\pi}{2}}[1 - ksin(t)^2]^{-0.5}dt \quad (2)$$

$$K'(k) = \int_0^{\frac{\pi}{2}}\left[1 - \left(\sqrt{1-k^2}\right)sin(t)^2\right]^{-0.5}dt \quad (3)$$

where, k is the argument and can be calculated as [26]:

$$k = \left(tan\left(\frac{a\pi}{4b}\right)\right)^2 \quad (4)$$

where:

$$a = \frac{W}{2} \text{ and } b = \frac{W+Z1}{2} \quad (5)$$

where, W is the finger width and S is the separation distance between fingers.

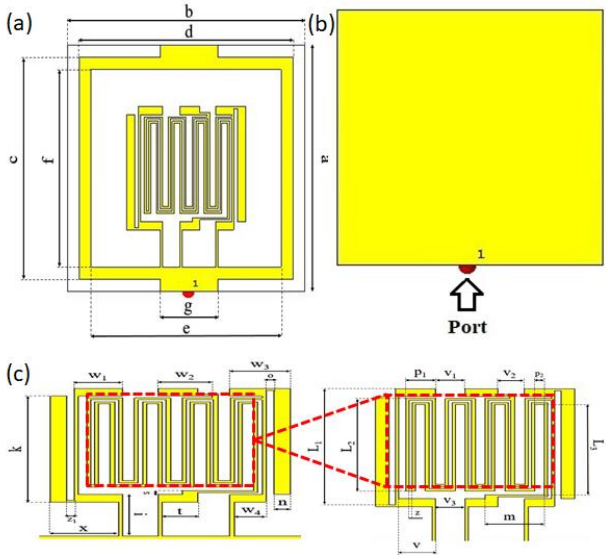


Fig1; Antenna design: (a) Front view, (b) Back view, and (c) Interdigital capacitor design.

TABLE I
GEOMETRICAL DETAILS OF THE PROPOSED ANTENNA SENSOR.

Parameter	Value (mm)	Parameter	Value (mm)
A	30	K	14
B	30	Z	0.25
C	27	Z_1	0.5
D	27	I	5.5
E	24	S	0.5
F	24	N	1
G	7.25	V_1	2.25
W_1	3	V_2	2
W_2	2.5	V_3	2.25
W_3	3.75	V	2.75
W_4	2	P_1	2.25
O	0.5	P_2	0.75
L_1	14.75	T	2.25
L_2	11.75	M	4.5
L_3	11.5	X	4.25

Now, the proposed sensor is analytically decomposed from the equivalent circuit diagram that is shown in Fig2. From Fig2, the proposed C_{int} is connected to two strip lines. The strip lines provide L-C branch (L_{strip} and C_{strip}) connection in parallel with C_{int}. OCSSR structure is connected to C_{int} through three shoring posts to be presented by (L_{short}) in the equivalent circuit diagram. It is good to mention that the effects of the inherent stray inductors (L_s) and stray capacitors (C_s) are proposed in the equivalent circuit diagram. Nevertheless, the feed line equivalently is considered as an inductor (L_{feed}). The equivalent representation for the closed loop (CL) is considered as an inductor (L_{CL}) and can be calculated according to the following equation [23]:

$$L = \mu_o a \left\{ ln\left(\frac{8a}{Z1}\right) - 2 \right\} \quad (6)$$

Based on the circuit model in Fig2, the authors calculated the relative lumped elements with an initial gauss from equations (1 and 6) for the proposed design at the desired frequency band using ADS software package parametrically. The S-parameters are calculated from the circuit model to be shown in Fig. 2. It is found that the proposed sensor shows a resonance mode at 0.63GHz from the lumped elements that are listed in Table 2.

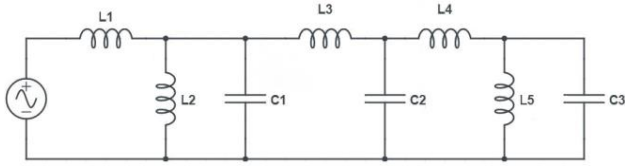


Fig2; Equivalent circuit for the proposed resonator.

TABLE II
CALCULATED LUMPED ELEMENTS

Lumped element	Value
L1	1.01nH
L2	1.89nH
L3	11.6mH
L4	23.04nH
L5	17.45nH
C1	11.12pF
C2	2.06pF
C3	1.91pF

From the proposed circuit model, the equivalent impedance is calculated based on the second branch to be described by L_2 and C_2 be noted as $Z_{T,2}$ following:

$$Z_{T,2} = \frac{j\omega L_2}{1-\omega^2 L_2 C_2} \quad (7)$$

Therefore, the resonant frequency, $f_{r,2}$ is expressed as

$$f_{r,2} = \frac{1}{2\pi\sqrt{L_2 C_2}} \quad (8)$$

In Fig. 2, the equivalent circuit of the proposed patch can be represented as two series of L_1C_c circuits and one parallel L_2C_2 circuit and the total impedance, $Z_{T,1}$ in addition to the band pass resonant frequency, $f_{r,1}$ which can be given as following:

$$Z_{T,1} = \frac{2(1-\omega^2 L_1 C_c)}{j\omega C_c} + \frac{j\omega L_2}{1-\omega^2 L_2 C_2} \quad (9)$$

$$f_{r,1} = \frac{1}{2\pi\sqrt{L_1 C_c}} \quad (10)$$

By employing (2) and (4), relation (3) can be rewritten as following:

$$Z_{T,1} = \frac{2(1-(\frac{\omega}{\omega_{r,1}})^2)}{j\omega C_c} + \frac{j\omega L_2}{1-(\frac{\omega}{\omega_{r,2}})^2} \quad (11)$$

$Z_{T,1}$ in relation (5) has two resonant frequencies, lower and upper frequencies which are represented by $\omega_{r,1}$ and $\omega_{r,2}$ respectively. When $\omega = \omega_{r,1}$ a maximum transmission could occur, where minimum $Z_{T,1}$ is ($Z_{T,1} \geq j\omega L_2$), while zero transmission occurs at $\omega = \omega_{r,2}$ where $Z_{T,1}$ maximum is achieved. In order to confirm prior discussion, an electromagnetic 3D simulation based on CST software packages in invoked to validate the obtained results from the circuit model. In Fig3, the patch is simulated when mounted in closed to a transmission line to evaluate the S_{11} spectrum. It is found that there are three resonant frequencies, $f_{r,1}$, $f_{r,2}$ and $f_{r,3}$, all these frequencies are the same resonant frequencies of the proposed patch. Therefore, these frequencies can be used as indicators for characterizing SUT after applying sensitivity analysis at the resonant frequencies.

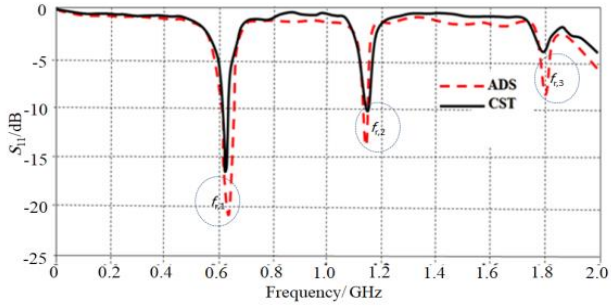


Fig 3; S_{11} spectra comparison between ADS and CST software packages.

IV. SENSITIVITY ANALYSIS

The proposed sensor is invented based on a quasi-static small antenna [12] in which designed with an interdigital capacitor surrounded with closed loop. Due to such combination a current circulation occurs at three legs of connection between the capacitor and the closed loop. The variation in the capacitance of the structure generally concerns on the variation in the permittivity of SUT. Therefore, the performance variation in the proposed structure with

introducing different permittivity values is discussed as following:

A. Resonant frequencies analysis

After introducing different SUT in the CST MWS environment, the frequency shifts are recorded to be tested later experimentally. Therefore, SUT must cover the whole area for the efficient perturbation of the *E*-field. The resonant frequencies of S_{11} in Fig.4 are considered as the reference of unloaded filter ($f_{r,1}$ and $f_{r,2}$). The proposed sensor is then loaded with the SUT, where the dielectric constant of the sample is changed randomly in a broad range from 76 to 90. The resonant frequencies ($f_{r,1}$ and$f_{r,2}$) corresponding to each sample are extracted, which are also plotted with variation of the dielectric constant as shown in Fig. 4 (a). For preferable conception the change in the resonant frequencies (Δf_r= unloaded ($f_{r,1}$ and $f_{r,2}$) – corresponding loaded ($f_{r,1}$and $f_{r,2}$)) is plotted in Fig. 4 (b).

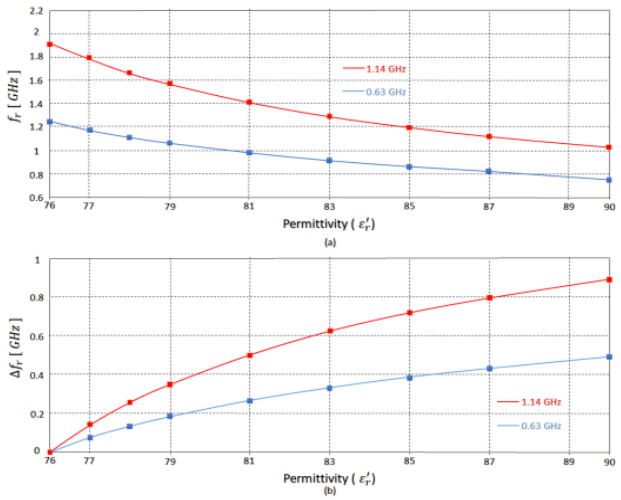


Fig. 4 Performance variation in terms of frequency resonance of the proposed sensor with changing ε_r: (a) f_r change and (b) Δf_r^{-1}.

From Fig. 4 (a) it can be noted that the resonant frequency, $f_{r,1}$is about 170 MHz greater than resonant frequency, $f_{r,2}$. Moreover, the relative change $f_{r,1}$ with dielectric constant is about 150 MHz greater than $f_{r,2}$ as depicted in Fig. 4 (b). In another meaning, for these resonant frequencies, $f_{r,1}$ shows to be a good preference to obtain high sensitivity for supposed dielectric constant as compared to that of resonant frequency, $f_{r,1}$.

B. Quality factor analysis

For general resonators the quality factor, Q may be presented as [34]:

$$Q = \omega_o \frac{W}{P_L} \tag{12}$$

where, ω_0 is the angular resonant frequency, W is the electric and magnetic stored energy and P_L represents the average power dissipated per cycle. The previous equation also can be rewritten as:

$$Q = \frac{f_r}{\Delta f} \tag{13}$$

where f_r is the resonant frequency and Δf represents the relative 3dB bandwidth of the resonator frequency response. The proposed sensor performance is simulated with different values of loss tangent from 0.01 to 0.15 and corresponding ε'_r variation from 76 to 90. The relative quality factor is computed and depicted in Fig. 5.

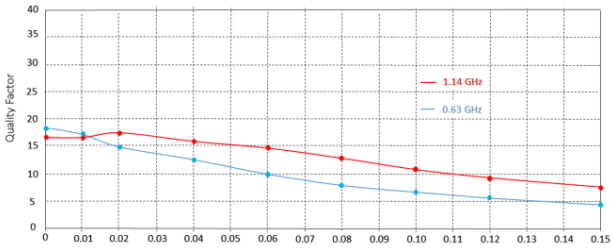


Fig. 5; Quality factor variation of the proposed sensor calculated for $f_{r,1}$ and$f_{r,2}$.

From Fig. 5, it is fully interesting to observe that the slope of quality factor is identical to resonant $f_{r,1}$ in comparison to $f_{r,2}$ slope with loss tangent change. While the slope of quality factor relative to $f_{r,1}$ is greater than in $f_{r,2}$ for the same condition. Hence the resonant $f_{r,1}$ is utilized for characterizing the SUT.

V. Data Analysis

Now, the obtained results from the previous section are analyzed to evaluate the calibration sensitivity to be compared to the experimental results. In order to describe the tested samples, a numerical type is desired which generally plots the measured parameters (e.g., the resonant frequency and the quality factor) to the relative permittivity of the SUT.

A. Real permittivity calibration effects

The proposed sensor frequency resonance is found to be changed due to samples loading as can be observed in Fig.6; in which the inverse square of the resonant frequency (f_r) is extracted from the S_{11} spectra. The results with the corresponding real permittivity (ε'_r) of SUT are depicted in Fig. 6.

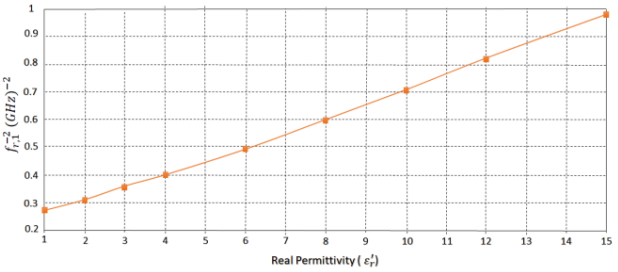


Fig. 6. $f^{-2}_{r,1}$variation with respect of changing ε'_r.

The obtained results in Fig. 6 confirm the results in equation (10) that is achieved from the equivalent circuit model. From equation (10), the values of L_2 and C_2 are supposed to be constant due to the solid values of the overall length of OCSRRs and ε_r of the substrate. It is interesting to note that the inverse square of the resonant frequency is directly proportional to the real permittivity of the SUT. Thus, in order to combine all the above effects, the dielectric constant of the SUT is mathematically expressed in terms of the resonant frequency (f_r) as following:

$$\varepsilon'_r = -3.519(f_r^{-2})^2 + 23.84(f_r^{-2}) - 5.007 \tag{14}$$

The above relation is obtained from employing the curve fitting tools, which supplies a numerical model of the proposed

sensor to determine the real permittivity of SUT in terms of the measured resonant frequency ($f_{r,1}$). It should be noted that all SUT has with a fixed 3mm thickness.

B. *Imaginary permittivity calibration effects*

After founding the numerical relations to determine the dielectric constant of SUT, an identical analysis is completed to find a numerical relation for computing the loss tangent (tanδ) of SUT. As explained earlier that the resonant $f_{r,1}$ provides a quality factor greater than those obtained at $f_{r,2}$. Hence, the resonant $f_{r,1}$ is employed for calculating tanδ of SUT. Therefore, at first, the dielectric constant in the range of 76 to 90 are possessed and the tanδ values combatable for each dielectric SUT is changed from 0 to 0.15, and the relative simulated results of $f_{r,1}$change is depicted in Fig. 7.

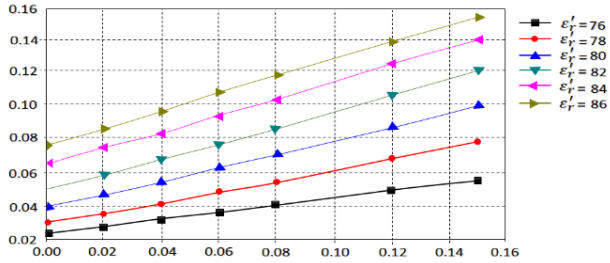


Fig. 7. Inverse of Q-factor in terms of tanδ for various values of ε'_r (Linear relevance between Q^{-1}_{SUT} and tanδ for all values of ε'_r).

The quality factor (Q_{SUT}) for each case is determined from the simulated response of S_{11}spectra, after that the inverse of Q_{SUT} values and the corresponding tanδ are depicted in Fig. 7. The relation between the tanδ and the Q_{SUT} can be specified as following [12]:

$$Q_{MUT} = \frac{1}{\tan\delta} = \frac{\varepsilon_r^{//}}{\varepsilon_r^{/}} \qquad (15)$$

where ε_r' and ε_r" are the real and imaginary parts of the relative permittivity in equation (15). From Fig. 7, it is noted that the alteration of Q^{-1}_{SUT} with tanδ is linear compound with a rising values depend on ε'_r of SUT. Thus, to deduce the tanδ of SUT, which relies on the loaded quality factor as well as the ε'_r of SUT, a curve fitting tool is utilized to conclude the numerical model as presented below:

$$\tan\delta = \exp\left(\frac{Q_{MUT}^{-1} + 0.02393}{0.2183 + 0.03131 \times \varepsilon_r}\right) - 1.16477 \quad (16)$$

After deciding the ε'_r from equation (14) and tanδ from (16), the imaginary part of the complex permittivity can be determined using (15).

VI. Sensor Fabrication

The proposed sensor is fabricated using printed circuit board technology as shown in Fig. 9. The sensor is fabricated from using chemical wit etching process in the laboratory. The FR4 substrate is considered as the plate form layer for the proposed sensor.

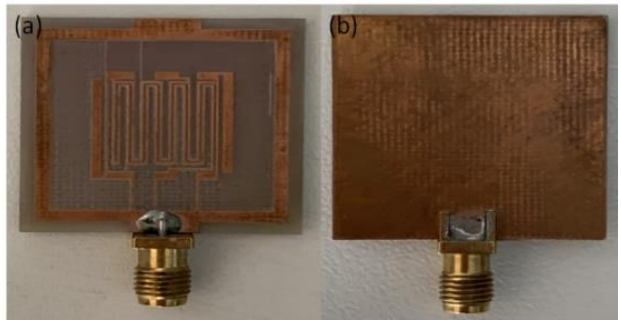


Fig. 8; Fabricated sensor structure: (a) front view and (b) back view.

Now, the proposed sensor performance is measured in terms of S_{11} spectrum as seen in Fig. 10 without introducing any SUT. The obtained results from measurements are compared to those obtained from simulation results to show excellent agreement as seen in Fig. 9.

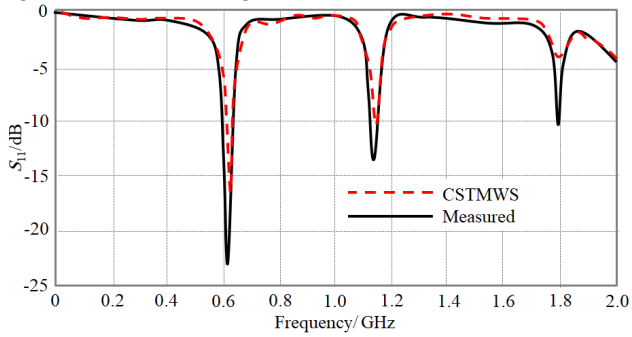


Fig 9; Experimental validation.

The measurement results are conducted to PNA8720 network analyzer after applying a single port calibration process. From the measured S_{11} spectra, the proposed sensor shows a frequency resonance at 0.63 GHz with $|S_{11}|$=16dB, and a bandwidth from 0.6GHz to 0.65GHz. This frequency is considered to ensure excellent penetration through the human tissues with minimum skin depth loss [6].

VII. Measurements and Validations

In this section, the proposed sensor measurement operation is based on placing a finger on the interdigital capacitor part to monitor the variation in the S_{11} magnitude and frequency resonance shift. The field penetration through the finger skin is affected by the blood glucose variation [12]. Such variation is attributed to the blood glucose change that could be reflected on the effective permittivity of the blood as discussed [23]. Therefore, the effects of touching the proposed sensor by 15 patients at three different times to realize 45 recorders are listed in Table 3 to analyze the sensor performance. The recorded data in Table 3 are collected based on S_{11} spectra change in terms of S_{11} magnitude, frequency resonance, phase change, bandwidth, and quality factor.

A. *Sensing Process*

The S_{11} spectra of the proposed sensor are obtained according to the samples listed in Table 3. The sensor is designed to ensure that the first resonant position is located around 0.63 GHz. Therefore, the fabricated sensor S_{11} spectra

changes are evaluated after introducing the patient finger touch. Thus, the prepared design is experimentally tested using the PNA8720 network analyzer. The obtained changes in the S_{11} spectra are monitored in terms of $|S_{11}|$, frequency shift, phase change, quality factor, and bandwidth.

TABLE III
CLARIFICATION OF THE RESULTS OF MEASURING DIFFERENT BLOOD SAMPLES.

Case	BMI	Age	Sex	Glucose level	$f_{r,1}$ shift/ GHz	Q_{SUT}/ %	ε_r'	ε_r"
				112	0.1091	12	25.91	0.311
1	18.9	8	M	210	0.112	11	25.76	0.283
				150	0.115	11	25.59	0.281
				119	0.106	12	26.08	0.313
2	23.1	49	M	103	0.115	15	25.59	0.3833
				210	0.111	22	25.81	0.567
				250	0.123	9	25.16	0.226
3	22.9	65	F	225	0.193	8	21.59	0.172
				245	0.191	21	21.69	0.455
				131	0.103	31	26.25	0.813
4	27.8	50	F	159	0.109	9	25.92	0.233
				177	0.108	12	25.97	0.311
				331	0.121	11	25.27	0.278
5	26.5	59	F	193	0.111	13	25.81	0.335
				168	0.156	17	23.43	0.398
				102	0.116	15	25.54	0.383
6	29.9	38	F	131	0.182	10	22.13	0.221
				109	0.174	9	22.52	0.202
				158	0.091	16	26.9	0.43
7	30.2	44	M	146	0.098	13	26.52	0.344
				126	0.092	10	26.85	0.268
				110	0.109	11	25.92	0.282
8	28.5	43	M	191	0.106	12	26.08	0.313
				198	0.110	11	25.87	0.284
				104	0.122	7	25.22	0.176
9	24.1	48	F	107	0.133	7	24.63	0.172
				115	0.124	9	25.11	0.226
				119	0.111	11	25.81	0.283
10	32.6	41	F	131	0.110	13	25.87	0.336
				114	0.113	12	25.7	0.308
				121	0.091	10	26.9	0.269
11	36.1	53	M	180	0.094	7	26.74	0.187
				140	0.092	11	26.85	0.295
				390	0.109	9	25.92	0.233
12	32.6	67	F	331	0.101	7	26.36	0.184
				378	0.105	12	26.14	0.313
				190	0.189	11	21.79	0.239
13	22.4	61	F	143	0.188	10	21.83	0.218
				126	0.177	9	22.37	0.201
				129	0.109	12	25.92	0.311
14	23.5	62	M	130	0.110	11	25.87	0.284
				120	0.112	9	25.76	0.231
				121	0.195	4	21.5	0.086
15	24.9	54	M	189	0.196	5	21.45	0.107
				134	0.179	8	22.28	0.178

B. Sensing Validation

The variation in the S_{11} spectra of the proposed sensor is measured after placing a finger on it as a non-invasive technique. Therefore, the glucose level is monitored through a normal device glucose meter, PRODIGY Autocode, and the results are recorded in Table 3. Then, the patient finger is placed on the proposed sensor and the frequency resonance shift and S_{11} magnitude change are listed in Table 3. Next, the measured glucose level is compared with respect to the relative values of ε_r' and ε_r" that are listed in Table 3. Therefore, the measured $f_{r,1}$ and Q_{SUT} values are applied in equation (14) to (15) to calculate the relative values of ε_r' and ε_r" from the measured data. The calculated values of ε_r' and ε_r" are compared to their relatives from measurements in Table 3. Thus, in Table 4, the relative errors between the measured and calculated ε_r' and ε_r" values are calculated. It is found a good agreement between the measured and calculated values. Therefore, from this comparison between the relative values of ε_r' and ε_r", the glucose level can be detected according to Table 4.

TABLE IV
COMPARISON RELATIVE ERRORS BETWEEN THE MEASURED AND CALCULATED ε_R' AND ε_R" VALUES.

Case number	Measured values		Calculated values		Error rate for	Error rate for	Total error rate
	ε_r'	ε_r"	ε_r'	ε_r"	ε_r'	ε_r"	
	25.91	0.311	26.84	0.322	3.46%	3.41%	6.87%
1	25.76	0.283	25.87	0.284	0.42%	0.35%	0.77%
	25.59	0.281	26.52	0.291	3.5%	3.43%	6.93%
	26.08	0.313	27.01	0.324	3.44%	3.39%	6.83%
2	25.59	0.383	26.52	0.397	3.5%	3.52%	7.02%
	25.81	0.567	26.73	0.588	3.44%	3.57%	7.02%
	25.16	0.226	26.09	0.234	3.56%	3.41%	6.97%
3	21.59	0.172	22.51	0.180	4.08%	4.44%	8.52%
	21.69	0.455	22.61	0.474	4.06%	4%	8.06%
	26.25	0.813	27.17	0.842	3.38%	3.44%	6.82%
4	25.92	0.233	26.84	0.241	3.42%	3.31%	6.73%
	25.97	0.311	26.90	0.322	3.45%	3.41%	6.86%
	25.27	0.278	26.19	0.288	3.51%	3.47%	6.98%
5	25.81	0.335	26.73	0.347	3.44%	3.45%	6.89%
	23.43	0.398	24.36	0.414	3.81%	3.86%	7.67%
	25.54	0.383	26.46	0.396	3.47%	3.28%	6.75%
6	22.13	0.221	23.05	0.230	3.99%	3.91%	7.90%
	22.52	0.202	23.45	0.211	3.96%	4.26%	8.22%
	26.9	0.43	27.83	0.445	3.34%	3.37%	6.71%
7	26.52	0.344	27.44	0.356	3.35%	3.37%	6.72%
	26.85	0.268	27.77	0.277	3.31%	3.24%	6.55%
	25.92	0.285	26.84	0.295	3.46%	3.38%	6.84%
8	26.08	0.313	27.01	0.324	3.44%	3.39%	6.83%
	25.87	0.284	26.79	0.182	3.43%	56.04%	59.47%
	25.22	0.176	26.14	0.230	3.51%	23.47%	26.98%
9	24.63	0.172	25.56	0.234	3.63%	26.49%	30.12%
	25.11	0.226	26.03	0.294	3.53%	23.12%	26.65%
	25.81	0.283	26.73	0.348	3.44%	18.67%	22.11%
10	25.87	0.336	26.79	0.319	3.43%	5.32%	8.75%
	25.7	0.308	26.63	0.319	3.49%	3.44%	6.93%
	26.9	0.269	27.83	0.278	3.34%	3.23%	6.57%
11	26.74	0.187	27.66	0.193	3.32%	3.26%	6.58%
	26.85	0.295	27.77	0.305	3.31%	3.27%	6.58%
	25.92	0.233	26.84	0.241	3.42%	3.31%	6.73%
12	26.36	0.184	27.28	0.190	3.37%	3.15%	6.52%
	26.14	0.313	27.06	0.324	3.39%	3.39%	6.78%
	21.79	0.239	22.71	0.249	4.05%	4.01%	8.06%
13	21.83	0.218	22.76	0.227	4.08%	3.96%	8.04%
	22.37	0.201	23.30	0.209	3.99%	3.77%	7.76%
	25.92	0.311	26.84	0.322	3.42%	3.41%	6.83%
14	25.87	0.284	26.79	0.294	3.25%	3.40%	6.65%
	25.76	0.231	25.82	0.232	0.23%	0.43%	0.66%
	21.5	0.086	22.42	0.089	4.23%	3.37%	7.60%
15	21.45	0.107	22.37	0.111	4.11%	3.60%	7.71%
	22.28	0.178	23.20	0.185	3.96%	3.78%	7.74%
					Total error ratio for all measurements		9.76%

The total calculated error is evaluated from Table 4 according to the following equation:

$$*\text{error} = \frac{\|measured\ values - calculated\ values\|}{calculated\ values} \times 100 \qquad (17)$$

It is found the maximum error from the total values is less than 10%.

VIII. CONCLUSION

The proposed sensor is presented to characterize blood glucose level through measuring the relative values of ε_r' and ε_r" for different blood samples. The proposed sensor is constructed as a single-port network; therefore, it is designed based on an interdigital capacitor patch to sense the blood glucose level non-invasively. The reason of that, the field fringing from the proposed sensor is found to be magnified and easy to penetrate through the human skin to the blood vassals. It is found that the proposed sensor shows different frequency resonances within the band of interest. However, it is decided to consider only the first frequency resonance ($f_{r,1}$) for sensing

where the maximum sensitivity can be achieved. In this case, $f_{r,1}$ and Q_{SUT} measurement are gathered from different patients. Therefore, from the measured values, an analytical model is synthesized based on curve fitting analysis. In such process, fifteen patients are submitted to the proposed sensor for estimating the level of glucose in the blood, ending with results very similar to the results measured by traditional commercial methods. The measured values of ε_r' and ε_r" are found to be agree very well with those obtained from the calculated results based on curve fitting analysis with less than 15% errors. It is found that the proposed sensor is a suitable choice for biomedical applications including blood glucose measurements. The proposed measurements point out the total error is about 10%. Finally, a future work on metamaterial-based printed circuit antennas for blood glucose level sensing applications includes optimizing antenna design, integrating with biosensors, miniaturization for wearable devices, ensuring biocompatibility, employing advanced signal processing techniques, conducting clinical validation, ensuring long-term stability, improving cost-effectiveness, and exploring multiparameter sensing capabilities. These efforts aim to enhance accuracy, reliability, and practicality for diabetes management and healthcare monitoring.

REFERENCES

[1] D. Whiting et al., "IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030". Diabetes Res. Clin. Pract. 94(3), 311–321 (2011). DOI: 10.1016/j.diabres.2011.10.029.

[2] "American Diabetes Association. Diagnosis and classification of diabetes mellitus". Diabetes Care 37, S81–S90 (2014). DOI: 10.2337/dc14-S081.

[3] "Pre diabetes diagnosis and treatment: a review". World J. Diabetes 6(2), 296-303 (2015). DOI: 10.4239/wjd.v6.i2.296.

[4] M. G. Burt et al., "Brief report: comparison of continuous glucose monitoring and finger-prick blood glucose levels in hospitalized patients administered basal-bolus insulin". Diabetes Technol. Ter. 15(3), 241–245 (2013). DOI: 10.1089/dia.2012.0282.

[5] W. V. Gonzales et al., "A progress of glucose monitoring - a review of invasive to minimally and non-invasive techniques, devices and sensors". Sensors 19(4), 800 (2019). DOI: 10.3390/s19040800.

[6] T. Lin et al., "Non-invasive glucose monitoring: a review of challenges and recent advances". Curr. Trends Biomed. Eng. Biosci. 6(5), 1–8 (2017). DOI: 10.19080/CTBEB.2017.06.555696.

[7] Spegazzini, N. et al. "Spectroscopic approach for dynamic bioanalyte tracking with minimal concentration information". Sci. Rep. 4, 7013 (2015). DOI: 10.1038/srep07013.

[8] Kuroda, M. et al. "Effects of daily glucose fluctuations on the healing response to everolimus-eluting stent implantation as assessed using continuous glucose monitoring and optical coherence tomography". Cardiovasc. Diabetol. 15(1), 79 (2016). DOI: 10.1186/s12933-016-0395-4.

[9] J. Y. Sim et al., "In vivo microscopic photoacoustic spectroscopy for non-invasive glucose monitoring invulnerable to skin secretion products". Sci. Rep. 8(1), 1059 (2018). DOI: 10.1038/s41598-018-19340-y.

[10] M. Goodarzi et al., "Selection of the most informative near infrared spectroscopy wavebands for continuous glucose monitoring in human serum". Talanta 146, 155–165 (2016). DOI: 10.1016/j.talanta.2015.08.033.

[11] J. Yadav et al., "Prospects and limitations of non-invasive blood glucose monitoring using near infrared spectroscopy". Biomed. Signal Process. Control 18, 214–227 (2015). DOI: 10.1016/j.bspc.2015.01.005.

[12] Y. Alnaiemy et al., "Electromagnetic Characterizations of Cement Using Free Space Technique for The Application of Buried Object Detection". Academic Science Journal, 2015, Volume 11, Issue 4, Pages 1 -10. DOI: 10.13140/RG.2.2.19212.56965.

[13] R. K. Abdulsattar et al., "A Theoretical Study to Design a Microwave Resonator for Sensing Applications". Journal of Engineering and Sustainable Development (JEASD), 2021, Volume, Issue Conference proceedings 2021, Pages 1-42-1-48. DOI: 10.31272/jeasd.conf.2.1.6.

[14] R. K. Abdulsattar et al., "Artificial Neural Network Approach for Estimation of Moisture Content in Crude Oil by Using a Microwave Sensor". International journal of microwave and optical technology, Vol.18, No.5, September 2023. DOI: 10.1016/j.diabres.2011.10.029.

[15] A. I. Anwer et al., "Minkowski Based Microwave Resonator for Material Detection over Sub-6 GHz 5G Spectrum," 2023 2nd International Conference on 6G Networking (6GNet), Paris, France, 2023, pp. 1-4. DOI: 10.1109/6GNet58894.2023.10317726.

[16] T. A. Elwi "Metamaterial based a printed monopole antenna for sensing applications". Int J RF Microw Comput Aided Eng. 2018; 28:e21470. DOI: 10.1002/mmce.21470.

[17] R. K. Abdulsattar et al., "A New Microwave Sensor Based on the Moore Fractal Structure to Detect Water Content in Crude Oil". Sensors 2021, 21, 7143. DOI: 10.3390/s21217143.

[18] A. A. Al-Behadili et al., "Differential Microstrip Sensor for Complex Permittivity Characterization of Organic Fluid Mixtures". Sensors 2021, 21, 7865. DOI: 10.3390/s21237865.

[19] A. I. Anwer et al., "A Fractal Minkowski Design for Microwave Sensing Applications". Journal of Engineering and Sustainable Development, 2022, 26(5), 78–83. DOI: 10.31272/jeasd.26.5.7.

[20] M. Alibakhshikenari et al., "Design of a Planar Sensor Based on Split-Ring Resonators for Non-Invasive Permittivity Measurement". Sensors. 2023; 23(11):5306. DOI: 10.3390/s23115306.

[21] R. K. Abdulsattar et al., "Optical-microwave sensor for real-time measurement of water contamination in oil derivatives", AEU - International Journal of Electronics and Communications, Volume 170, 2023. DOI: 10.1016/j.aeue.2023.154798.

[22] A. I. Anwer et al., "A theoretical study to design a microwave sensor for biomedical detections". AIP Conference Proceedings. Vol. 2787. No. 1. AIP Publishing, 2023. DOI: 10.1063/5.0148154.

[23] R. K. AbdulSattar et al., "Metamaterial Based Sensor Using Fractal Hilbert Structure for Liquid Characterization," 2023 International Conference on Electromagnetics in Advanced Applications (ICEAA), Venice, Italy, 2023, pp. 480-483. DOI: 10.1016/j.sbsr.2020.100395.

[24] T. A. Elwi et al., "A Passive Wireless Gas Sensor Based on Microstrip Antenna with Copper Nanorods," Progress In Electromagnetics Research B, Vol. 55, 347-364, 2013. DOI: 10.2528/PIERB13082002.

[25] Ali˙, D., et al., (2021). Metamaterial-based Printed Circuit Antenna for Biomedical Applications. Avrupa Bilim Ve Teknoloji Dergisi(26), 12-15.

[26] Saha, S. et al. "A glucose sensing system based on transmission measurements at millimetre waves using micro strip patch antennas". Sci. Rep. 7, 6855 (2017). DOI: 10.1038/s41598-017-06926-1.

[27] S. Mohan et al., "Wireless integration, design, modeling, and analysis of nanosensors, networks, and systems: a system engineering approach". Proceedings Volume 7291, Nanosensors, Biosensors, and Info-Tech Sensors and Systems; 72910G (2009). DOI: 10.1117/12.821554.

Taha A. Elwi received his B.Sc. in Electrical Engineering Department (2003) (Highest Graduation Award), and Postgraduate M.Sc. in Laser and Optoelectronics Engineering Department (2005) (Highest Graduation Award) from Al-Nahrain University Baghdad, Iraq. From April 2005 to August 2007, he worked with Huawei Technologies Company, in Baghdad, Iraq. On January 2008, he joined the University of Arkansas at Little Rock and he obtained his Ph.D. in December 2011 in system engineering and Science. He is considered of Stanford University's top 2% scientists in 2022. His research areas include wearable and implantable antennas for biomedical wireless systems, smart antennas, WiFi deployment, electromagnetic wave scattering by complex objects, design, modeling, and testing of metamaterial structures for microwave applications, design and analysis of microstrip antennas for mobile radio systems, precipitation effects on terrestrial and satellite frequency re-use communication systems, effects of the complex media on electromagnetic propagation and GPS. His research is conducted to consider wireless sensor networks based on microwave terminals and laser optoelectronic devices. The nano-scale structures in the entire electromagnetic spectrum are a part of his research interest. Also, his work is extended to realize advancements in reconfigurable intelligent surfaces and control the channel performance. Nevertheless, the evaluation of modern physics phenomena in wireless communication networks including cognitive radio networks and squint effects is currently part of his research. His research interests include pattern recognition, signal and image processing, machine learning, deep learning, game theory, and medical image analysis-based artificial intelligence algorithms and classifications. He serves as an editor in many international journals and publishers like, MDPI, IEEE, Springer, and Elsevier. He is currently the head of the International Applied and Theoretical Research Center (IATRC), Baghdad Quarter, Iraq. Also, he has been a member of the Iraqi scientific research consultant since 2016. He is leading three collaborations around the world regarding biomedical applications using microwave technology. He is the supervisor of many funded projects and Ph.D. theses with corresponding of more than 150 published papers and holding 10 patents.

Hayder H. Al-khaylani received B.Sc. degree in Electrical Engineering Department, Faculty of Engineering, from University of Baghdad, Iraq, 2007. M.Tec. Electronics and Communication Engineering Department from Sam Higginbottom University, India, 2012 and PhD. in Electrical and Computer Engineering Department, Faculty of Engineering, Altinbas University, Istanbul, Turkey. His research interests are in the areas of smart antennas and wearable systems.

Saba T. Al-Hadeethi was born in 1984 in Baghdad, Iraq. She received a BSc degree in electronics and communication Engineering in 2005 from AL-Nahrain University Engineering college, Baghdad. She obtained his MSc degree in Electronic and Communication Engineering in 2009 from the Department of Electronics and Communication Engineering from AL-Nahrain University in Baghdad and PhD. in Electrical and Computer Engineering Department, Faculty of Engineering, Altinbas University, Istanbul, Turkey. Her research interests are in the areas of antennas design for 5G applications.

Sana A. Nasser was born in 1988 in Baghdad, Iraq. She received a BSc degree in communication Engineering in 2015 from Al-Mamuan college, Baghdad. She obtained his MSc degree in Electronic and Communication Engineering from the Department of Electrical Engineering, University of Technology, in 2021. Currently, he is a lecturer in the department of Mechatronics Engineering, Alkawarizmi College, at University of Baghdad. Her fields of research are Artificial Neural Networks.

Omar Almukhtar T. Najm was born in 1990 in Baghdad, Iraq. He received a BSc degree in communication Engineering in 2011 from University of Baghdad, Baghdad, Iraq. He obtained his MSc degree in Electronic and Communication Engineering from the Department of Electronic and Communication Engineering, from University of Baghdad, Baghdad, Iraq, in 2015. His fields of research are Digital communication, STBC-MIMO and robotic control systems.

Zainab S. Muqdad received her B.Sc. degree in Electrical Engineering (2017) and her M. Sc. degree in Electronics and Communication Engineering (2022), both from the Mustansiriyah University, Baghdad, Iraq. Her research interests include antenna, microwave applications, microwave radiology imaging, neural network, metamaterials, biomedical wireless systems, and cancer detection.

Mohammed H. Khaleel received the B.Sc. degree in Electrical Engineering Department, Faculty of Engineering, from University of Baghdad, Iraq, 2007 and M.Tec. Electronics and Communication Engineering from Sam Higginbottom University, India, 2012 research interests are in the areas of wireless communication.

Wasan S. Rasheed was born in 1989 in Baghdad, Iraq. She received the B.Sc. degree in Communication Engineering in 2011 from University of Technology, Iraq. She obtains his MSc in Communication Engineering, from Electrical Engineering department, University of Technology, Iraq, 2013. Her current research interests are in the field of antennas design.

Enhancing Parkinson's Disease Recognition through Multimodal Analysis of Archimedean Spiral Drawings

Attila Zoltán Jenei[1], Dávid Sztahó[1], and István Valálik[2]

Abstract—**Parkinson's disease is one of the most common neurodegenerative diseases, which is incurable according to recent clinical knowledge. Evaluating motor symptoms across diverse modalities such as speech, handwriting, and movement composes a conventional diagnostic approach. However, concurrently utilizing multimodal datasets encompassing drawing and acceleration data remains an underexplored field. Our investigation involved examining drawing and movement data of 45 Parkinson's disease (PD) patients and 47 healthy individuals (HC). The PD group presented mild symptoms in the right hand. We transformed drawing data into spiral images and used visual representations of motion data, employing pre-trained models for feature extraction and classifiers. While motion representations exhibited superior performance compared to drawing images, a comprehensive evaluation with the Mann-Whitney U test at a significance level of 0.05 revealed no statistically significant difference between the efficacy of movement and drawing data in all classification scenarios. Significant improvements were made by combining the drawing data predictions with the motion data predictions. The key finding of the research is that the recognition of the disease can be improved by connecting (post-model) the two modalities. Furthermore, it can be concluded that with the present approach, neither the drawing nor the movement data produced lower results on average.**

Index Terms—**Acceleration Data, Classification, Parkinson's disease, Pre-trained Models, Mann-Whitney U Test**

I. INTRODUCTION

Parkinson's disease (PD) is one of the most common neurological disorders, which affects mainly the aging population. According to current clinical knowledge, the disease is incurable, promoting this area as an extensive research field. The goal is typically to recognize the disease early enough to alleviate symptoms, slow disease progression, and maintain quality of life.

Its prevalence is 1% in people over 60 and 3% in people over 80 [1]. These values tend to increase due to aging societies, environmental factors, and accessibility of health care (more people get recognized). The tendency to the disease is increased by the male sex, certain chemicals, and genetic factors [2].

[1] Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, Hungary (e-mail: jenei.attila.zoltan@vik.bme.hu, sztaho.david@vik.bme.hu; orcid id: 0000-0003-1007-9907, 0000-0002-7361-4260).

[2] Department of Neurosurgery, St. John's Hospital, Budapest, Hungary (e-mail: valalik@parkinson.hu).

DOI: 10.36244/ICJ.2024.1.8

The destruction of dopaminergic neurons with subsequent depigmentation of the susbtantia nigra pars compacta (SNpc) and the appearance of Lewy bodies can be observed in the development of PD [1]. The importance of early detection of PD is shown by the fact that with the current diagnostic procedure, 60% of the dopamine-producing cells are already dead and cause problematic symptoms.

Non-motor symptoms appear earlier on the onset of PD. These include, for example, loss of smell, memory loss, digestive problems, and difficulty with sleeping [3]. These symptoms can appear even years earlier than motor symptoms. However, using these symptoms is difficult because they can indicate other illnesses, and not everyone develops the same symptoms.

In addition, the motor symptoms appear later in time, of which the three main are slowed movement (bradykinesia), muscle stiffness, and limb tremors at rest (resting tremor) [4]. These are vital symptoms taken into account by the neurologist to a large extent when establishing the diagnosis. It is important to emphasize that the diagnostic procedure relies heavily on the visual assessment of symptoms, imaging procedures, and drug tests. Still, there is currently no objective test for PD. Furthermore, the assessment can be influenced by the physician's subjectivity [5].

Because of the former, many researchers use AI and different modalities of data to recognize the disease using motor symptoms that help to increase the diagnosis accuracy and objectivity. Moreover, it allows the possibility of personalized care. Speech can be such a modality since 70% of PD patients develop dysphonic speech [6]. In addition, drawings/handwriting [7][8] and different forms of movement [9][10] are often used to analyze limb symptoms and help the diagnosis process.

In the present research, we investigate whether drawing (more precisely, drawing a spiral pattern) as pictorial information or the acceleration data measured during drawing provides a better recognition of the disease. Furthermore, we attempt to use the two modalities jointly to see whether recognition performance improves.

The paper's main contributions are 1) *comparing the spiral drawing and the simultaneously recorded movement data as image representation with the same processing approach to detect PD* and 2) *attempting to improve PD detection with the joint usage of the drawing and the simultaneously recorded motion data*.

The following section presents the **literature** on the problem statement, highlighting the key points. Then, in the **methods** section, we discuss the applied procedure. In the **results** and **discussion** section, we present and consider the possible outcomes. Finally, in the **summary and conclusion** section, we recap our work and highlight the essential findings and concerns.

II. LITERATURE STUDY

A requirement for technology to support diagnosis is that it does not overtax the patient. Ideally, this means short-in-time, simple, non-invasive tests. The machine learning algorithm makes these automatic evaluations possible, which can provide an objective output. This can significantly contribute to the physician's opinion and provide a universal measurement procedure.

One such test could be the recording and analysis of drawing or handwriting, which tries to capture the motor symptoms in PD patients' hands. Currently, this is not part of the criteria for diagnosis. However, McLennan et al. pointed out that 5% of patients have handwriting difficulties before the onset of motor symptoms, and 30% of those patients report deteriorating handwriting later [11].

Changes in fine motor skills, such as writing/drawing speed, continuity, and text or shape size, can be seen in PD patients. The reduction of the writing size is called micrography, of which two categories are consistent and progressive micrography [12]. Presumably, one and the other develop depending on the involvement of different brain areas and respond differently to dopaminergic drug treatment [13]. Similar changes appear in the drawings as well.

Many patterns are common, such as spirals, waves, or lines. They are mostly made on a tablet device, while it was more common to use traditional pen and paper in the past. The drawing should be simple so that it does not require special drawing skills but complex so that fine motor changes can be detected.

Kotsavasiloglou and his colleagues [14] conducted experiments involving 20 healthy (Healthy Control - HC) and 24 PD individuals by drawing a horizontal line. They extracted a number of features related to the speed of the pen tip and vertical deviations (deviations from a straight line). With multiple classifiers (Naïve Bayes, AdaBoost, Log. Regression, Support Vector Machine [SVM], Random Forest [RF], J48), 79.4-88.5% accuracy was achieved.

Sharma et al. [15] investigated the recognizability of PD by involving several databases using the Modified Gray Wolf Optimization algorithm on classifiers. Their drawing database was the Hand PD database created by Botucatu Medical School, São Paulo State University, which included 105 PD patients and 53 HC individuals (mean age 44.2 and 58.8, respectively). The classification algorithms were k-Nearest Neighbors (k-NN), RF, and Decision Tree (DT). With the help of predetermined features, 73.4-92.4% accuracy was achieved on the spiral drawings and 72.8-93.0% accuracy on the meander drawings.

In addition, further research deals with monitoring the drawing task with acceleration sensors and examining the usability of the data generated in this way for recognition.

Ali et al. [16] investigated Essential Tremor (ET) with acceleration data acquired while drawing spiral patterns. 17 ET patients and 18 HC individuals were included in the databases. Three sensors were placed at three points: on the dorsum of the hand, on the posterior forearm, and the posterior upper arm. SVM was used to classify the power spectral density (calculated from the acceleration data after creating a single vector magnitude). 74.3-85.7% accuracy was achieved.

Pereira et al. [17] used time series data from sensors mounted in a pen to detect PD. Their research used the HandPD database with 14 PD patients and 21 HC. Images were created from the sensor data (sound, finger grip, axial pressure of ink, acceleration and tilt data in X, Y, and Z directions). These were classified using image processing algorithms (ImageNet, CIRAF-10, LeNet) and a baseline model (Optimum-Path Forest [OPF]) on the raw data. 85.0-87.1% accuracy was achieved with ImageNet on meander data, and 77.9-83.8% accuracy with OPF on spiral drawings as the best performance.

Savalia and his colleagues [18] similarly used digital pen-provided time signals as images with image processing algorithms. They named their database newHandPD, which included 35 HC and 31 PD patients. With the help of Convolutional Neural Network (CNN) and EffNet-based classifiers, they achieved 84.8-88.8% accuracy.

Taleb et al. [19] experimented with several approaches: classification of raw signals from a pen and classification with a spectrogram (a 2D representation of the raw signals). They pointed out that combining several writing tasks improves the classifier's performance. Their best result with data augmentation was 97.6% accuracy.

Cascarano and his colleagues [20] examined the drawing patterns of 21 PD and 11 HC individuals. Descriptive characteristics were calculated from geometric, dynamic, and muscle activity data. 90.8% accuracy was achieved with spiral drawing without feature selection, while 93.8% accuracy was achieved with selection (with Multi-Objective Genetic Algorithm).

The literature study shows that the performance of the classifiers varies widely and may result from the database, feature extraction, and classification algorithms. Paper-based drawing also created a need to record dynamic data, which nowadays is easier to do with the help of tablets and digital pens. With this, the drawn pattern can be used as an image (the drawing itself) and as a set of time series. We also saw an approach where the researchers created images (2D representations) from time series, taking advantage of the performance of image processing algorithms. We could also find examples where combining several time signals/drawing tasks improves recognition. We wish to contribute to this with our present research, in which the participants draw a spiral pattern. During the process, a sensor attached to the wrist records the acceleration data. We examine these two modalities separately and together.

III. METHODOLOGY

A. Database

The database contains drawing and drawing-related sensor data of 45 PD patients and 47 HC individuals. All participants were informed beforehand about the research details and gave their consent by signing a consent form. Participants volunteered to participate in the research and were informed that their participation could be withdrawn without explanation.

There are 37 men and eight women in the PD class, whose average age is 66.0 years with a standard deviation of 14.2 years. Twenty-one people were recorded with drug onset, 18 people with Deep Brain Stimulation (DBS) switched on (10 people belonged to both categories). The severity score of PD patients was defined based on the Unified Parkinson's Disease Rating Scale (UPDRS) scale. The resting tremor (3.17), postural tremor (3.15), rigidity (3.3), and finger tapping (3.4) tasks were evaluated by the neurologist. The average severity was 1.10, where 0 means a healthy (normal) stage, and 4 means the severe stage. Severity scores for the right hand are presented since data from the right hand were examined.

There were 20 men and 27 women in the HC class, with an average age of 56.7 years and a standard deviation of 15.2. According to their admission, the members of the HC class did not have Parkinson's disease or any other disease that affects their movement.

The drawing and motion recordings were made using an application developed for Android tablets. In the case of the drawing, the participants followed the pattern of an Archimedean spiral by moving between the lines from the inside to the outside of the spiral template. The drawing data was sampled at a maximum of 110 Hz. A sensor attached to the wrist provided acceleration data in three dimensions (X, Y, and Z) with a sampling frequency of 50 Hz to record the movement.

Recordings were anonymized before use. Their metadata, such as gender and age, were used only to describe the classes.

B. Preprocessing

The drawings were plotted along X and Y coordinates and resized to 224x224 pixels with 24-bit depth. The average along the corresponding coordinate was subtracted from the acceleration data as standardization (remove the bias of the measurement device). Furthermore, a 4-order bandpass Butterworth filter was applied to the signals between 2 and 15 Hz. Finally, one data vector was created from the three coordinates, which contained the length of the vector pointing to the point described by the three coordinates at each instant of time. Two-dimensional representations were made, as shown in Fig. 1, such as *MarkovTransitionField*, *RecurrencePlot*, and *GramianAngularField*. For this data representation, the *pyts* (v0.12.0) Python package was used with default parameters [21]. These images were also resized to 224x224 pixels with 24-bit depth.

The feature vectors from the resulted input images were derived from the feature extraction part of pre-trained deep learning algorithms. For this, the Keras API was used with the *Tensorflow* (v2.0.0) machine learning platform. Among the available models, *Xception* [22], *ResNet50* [23], and *MobileNet* [24] were used. We removed the classification (*Dense*) layer from the end of each model and assigned a *GlobalAveragePooling2D* layer to get one-dimensional feature vectors. Finally, the feature vectors were standardized using the *StandardScaler* function of sklearn (v0.0.post7).

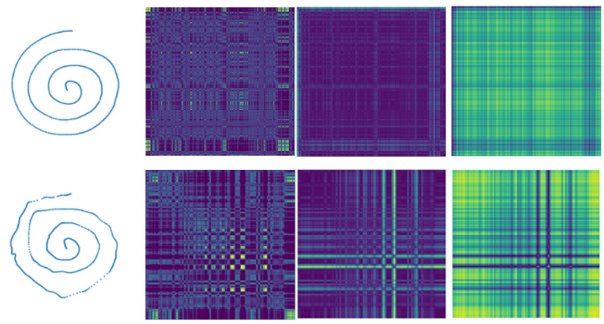


Fig. 1. The spiral drawing and the 2D movement representations for an HC person (upper images) and a PD patient (lower images).

C. Classification and Evaluation

To classify the feature vectors, SVM, RF, and k-NN classifiers were used with nested cross-validation at 10-fold numbers.

The test set (10% of the database) was separated in the outer cycle, independent of the model learning. On the remaining data, the best 100 features were selected based on ANOVA F-value [25]. The same features were selected in the test set accordingly.

The validation set (10% of the 90% data) was separated from the training dataset in the internal cycle. This training set was used to train the classifiers, and the optimization was carried out with the validation set. The parameters below were probed available for *sklearn* [26] models:

- *kernel* (linear, rbf), *C* value (0.001, 0.01, 0.1, 1, 10, 100), and *gamma* (10, 1, 0.1, 0.01, 0.001, 0.0001) for **SVM**,
- the *max*_depth (10, 30, 50, 70, 90, None), max*_features* (auto, sqrt), and *min_samples_leaf* (1, 2, 4) for **RF**,
- the *number of neighbors* (2, 3, 5, 10, 20) for **k-NN**.

The test sets were finally estimated with the models trained with the best parameters (on 90% of the data). The sensitivity, specificity, macro f1 score, and area under the receiver operating characteristic (ROC) curve (AUC) values were derived from the estimates.

The predictions of the two different modalities (drawing and movement) were aggregated according to Eq. 1, where $y_s \in \{0,1\}$ is the prediction from spiral drawing and $y_m \in \{0,1\}$ is the prediction from the movement data. $Q \in \{0,1\}$ is the weight factor between the two predictions, $y_{final} \in \{0,1\}$ is the final prediction.

If the value of Q is 0, then the final prediction is equal to the prediction from the movement. In the case of $Q = 1$, the final prediction is equal to the prediction from the drawing. The final prediction for Q between 0 and 1 is a ratio between the two modalities.

$$y_{final} = y_s Q + y_m(1 - Q) \quad (1)$$

The samples are assigned to classes based on the models' estimates with a decision limit 0.5. If the estimate is above this, the sample is classified as positive (PD) and below it as negative (HC).

Mann-Whitney U non-parametric test [27] was used to compare the performance of modalities and aggregations. The comparison was made with the macro f1 scores. The significance level was chosen as 0.05 [28].

IV. RESULTS

A. Separate Classification of Drawings and Movements

TABLE I shows the results achieved on the spiral drawings and the representation of the movement. The marking of the spiral drawing is *Spiral*, while the representations are named according to the names described in the III. Methodology section. The second column contains the names of the feature extractors, and the third column contains the classification algorithms. The last four columns contain the sensitivity (sens), specificity (spec), macro f1 (f1), and AUC (auc) values, respectively. The best results per representation are marked with **bold** style.

In the case of the *Spiral*, the MobileNet feature extractor resulted in the highest macro f1 score on average (67.9%). ResNet50 and Xception achieved an average of 59.7% and 57.6% macro f1, respectively. Regarding the classification algorithms, RB achieved the best macro f1 score on average (66.7%), while k-NN and SVM achieved macro f1 values of 59.5% and 59.0%, respectively. The best result among all cases was achieved with the MobileNet feature extractor and RF classification model for the spiral drawing (macro f1 score 70.6%).

In the case of *RecurrencePlot*, the feature extractors' results no longer deviate spectacularly from each other. MobileNet, ResNet50, and Xception achieved 67.8%, 63.9%, and 66.6% macro f1 scores, respectively. The result is similar according to the classifiers: 64.4% (k-NN), 65.2% (RF), and 68.8% (SVM). Compared with the data of the Spiral drawing, it can be seen that the values have improved on average, whether examining the feature extractors or the classifiers. The average difference between the two modalities is 4.4% in the macro f1 score. The *RecurrencePlot* approach obtained the best result with the MobileNet feature extractor and SVM classification model (74.9% macro f1 score). Compared to *Spiral's* best result, it provided a 4.3% better result. Regarding the other metrics, the specificity improved by an average of 0.7%, the sensitivity by 8.1%, and the auc value by 0.088 compared to the *Spiral* metrics.

In the *GramianAngularField* cases, the averages according to the feature extractors are 67.7% (MobileNet), 59.7% (ResNet50), and 61.2% (Xception) in macro f1 score. According to the classifiers, the average values are 61.1% (k-NN), 65.9% (RF) and 60.9% (SVM). In this case, the average difference compared to the *Spiral* is 0.9% in the macro f1 score. The best result was achieved with MobileNet feature extraction and RF classifier with a macro f1 score of 72.7%. This is 2.1% better than the *Spiral's* best result. Regarding the other metrics, the sensitivity improved by an average of 7.7% and the auc value by 0.039 compared to the *Spiral* metrics, but the specificity decreased by an average of 5.9%.

TABLE I
RESULTS OBTAINED WITH DIFFERENT MODALITIES, FEATURE EXTRACTORS, AND CLASSIFICATION ALGORITHMS.

	Extractor	Classifier	sens	spec	f1	auc
Spiral	**MobileNet**	k-NN	55.6%	78.7%	66.8%	0.686
		RF	**71.1%**	**70.2%**	**70.6%**	**0.757**
		SVM	64.4%	68.1%	66.3%	0.720
	ResNet50	k-NN	46.7%	78.7%	61.9%	0.574
		RF	60.0%	59.6%	59.8%	0.650
		SVM	53.3%	61.7%	57.5%	0.619
	Xception	k-NN	46.7%	53.2%	49.9%	0.533
		RF	68.9%	70.2%	69.6%	0.699
		SVM	53.3%	53.2%	53.3%	0.576
RecurrencePlot	**MobileNet**	k-NN	51.1%	80.9%	65.4%	0.752
		RF	64.4%	61.7%	63.0%	0.750
		SVM	**71.1%**	**78.7%**	**74.9%**	**0.834**
	ResNet50	k-NN	71.1%	46.8%	58.2%	0.689
		RF	66.7%	70.2%	68.4%	0.749
		SVM	64.4%	66.0%	65.2%	0.710
	Xception	k-NN	66.7%	72.3%	69.5%	0.721
		RF	71.1%	57.4%	64.0%	0.709
		SVM	66.7%	66.0%	66.3%	0.689
GramianAngularField	**MobileNet**	k-NN	75.6%	59.6%	67.3%	0.774
		RF	**80.0%**	**66.0%**	**72.7%**	**0.768**
		SVM	64.4%	61.7%	63.0%	0.720
	ResNet50	k-NN	62.2%	46.8%	54.2%	0.606
		RF	64.4%	66.0%	65.2%	0.685
		SVM	57.8%	57.4%	57.6%	0.635
	Xception	k-NN	60.0%	63.8%	61.9%	0.653
		RF	64.4%	55.3%	59.7%	0.664
		SVM	60.0%	63.8%	61.9%	0.659
MarkovTransitionField	MobileNet	k-NN	71.1%	51.1%	60.6%	0.635
		RF	75.6%	61.7%	68.4%	0.749
		SVM	51.1%	63.8%	57.4%	0.661
	ResNet50	k-NN	68.9%	70.2%	69.6%	0.757
		RF	**77.8%**	**74.5%**	**76.1%**	**0.802**
		SVM	75.6%	72.3%	73.9%	0.822
	Xception	k-NN	71.1%	55.3%	62.9%	0.634
		RF	64.4%	74.5%	69.4%	0.736
		SVM	66.7%	76.6%	71.6%	0.727

In the case of *MarkovTransitionField*, the average macro f1 scores are 62.1% for MobileNet, 73.2% for ResNet50 and 68.0% for Xception. Regarding to the classifiers, k-NN achieved 64.3%, RF 71.3%, and SVM 67.6% macro f1 score. Compared to the *Spiral*, ResNet50 and Xception performed better with 13.5% and 10.4% macro f1 scores, respectively. When comparing classification algorithms, on average, all three performed better (4.8%, 4.6%, and 8.6% better results for k-NN, RF, and SVM) in *MarkovTransitionField* representation than in the *Spiral*. Overall, this approach outperformed the *Spiral* by 6.0% in macro f1 score. The best result was obtained with the ResNet50 feature extractor with the RF classifier (76.1% macro f1 value). This is better with a 5.4% macro f1 score than the *Spiral's* best result. Regarding the other metrics, the specificity improved by an average of 0.7%, the sensitivity by 11.4%, and the auc value by 0.079 compared to the *Spiral* metrics.

Overall, it can be seen that the image representation of the movement provides a better result in average value: 4.4% (*RecurrencePlot*), 0.9% (*GramianAngularField*), 6.0% (*MarkovTransitionField*) in macro f1 score. According to the best results, better results were also achieved with movement representations by 4.3% (*RecurrencePlot*), 2.1% (*GramianAngularField*), and 5.4% (*MarkovTransitionField*) macro f1 score. However, ***no significant difference between the two modalities can be established*** with the Mann-Whitney U statistical test. The *p-values* of the tests are 0.331 (*Spiral vs RecurrencePlot*), 0.791 (*Spiral vs GramianAngularField*), and 0.102 (*Spiral vs MarkovTransitionField*). This probably stems from the observation that the two modalities have no clear trend according to the performance.

B. Joint examination of the modalities

Aggregating the predictions achieved on the modalities were based on Eq. 1. Fig. 2 shows the progression of macro f1 scores by connecting *Spiral* and *RecurrencePlot.* The weight factor or voting factor (Q) is shown on the horizontal axis. If Q is zero, then only the prediction of the movement. If $Q = 1$, then the prediction of the drawing applies as the two extreme points of the axis. The macro f1 scores between 0.5 and 0.9 are shown on the vertical axis.

The average macro f1 score achieved with *RecurrencePlot* is 66.1%, while with *Spiral* it is 61.7%. The average of the maximum points shown in the figure was 71.4% macro f1 score. The average of the maximum deviation (maximum point – modality with lower performance) was 11.7% macro f1 score, while the average of the minimum deviation (maximum point – modality with higher performance) was 3.2%. In two cases, no improvement was observed by combining the two modalities: Xception with k-NN classifier and Xception with SVM classifier. The possible reason is that the difference between the two modalities is high (19.6% and 13.0%). In the other cases, the difference between the modalities was minor (6.2% on average). The most considerable improvement was achieved with the MobileNet and SVM classifier. The macro f1 score on the *RecurrencePlot* is 74.9%. On the *Spiral* it is 66.3%, while the maximum value is 82.5%. By comparing the modalities with the maximum points pairwise using the Mann-Whitney U test, we experienced a ***significant improvement***. The p-value is 0.010 between the maximum and the *Spiral*, and 0.047 between the maximum and the *RecurrencePlot*.

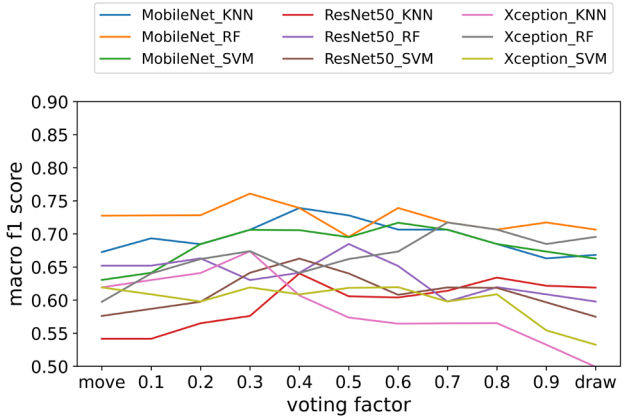


Fig. 3. The aggregated macro f1 scores of *Spiral* and *GramianAngularField* in proportion to the voting factor.

Fig. 3 shows the result of aggregating *Spiral* and *GramianAngularField*. The markings in the figure are the same as those in Fig. 2. In this case, the average macro f1 score of the *GramianAngularField* is 62.6%. This macro f1 score for the *Spiral* is 61.7%. The average of the maximum scores is 69.1%. In all cases, improvement was observed with the prediction aggregation. The maximum improvement is an average of 9.6% macro f1 score, and the minimum improvement is an average of 4.1%. With the present approach, the highest macro f1 score was provided by MobileNet and RF, with 76.1%. The same case achieved 72.7% on *GramianAngularField* and 70.6% on *Spiral*. ResNet50 achieved the most significant improvement (from both modalities) with the SVM classifier. It achieved 57.6% macro f1 score on movement data, 57.5% on drawing data, and 66.3% at the maximum point. We found a ***significant improvement*** by combining the modalities using the Mann-Whitney U test. The p-values are 0.030 between the maximum and the *Spiral* and 0.017 between the maximum and the *GramianAngularField*.

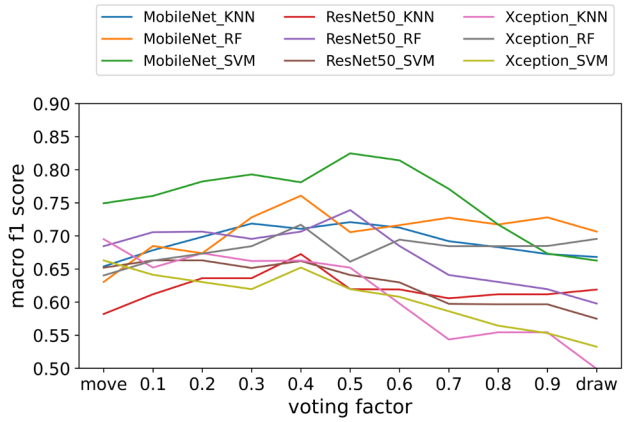


Fig. 2. The aggregated macro f1 scores of *Spiral* and *RecurrencePlot* in proportion to the voting factor.

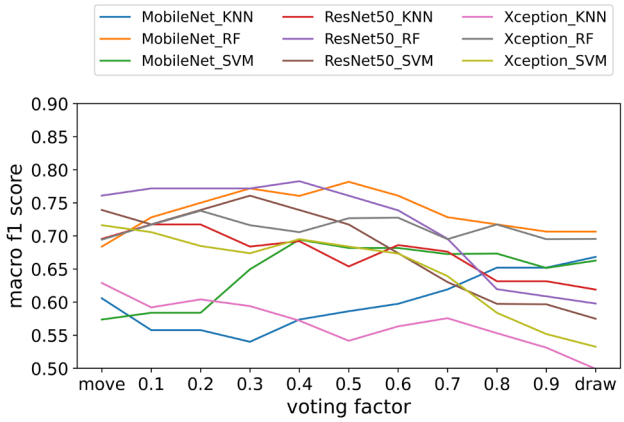


Fig. 4. The aggregated macro f1 scores of *Spiral* and *MarkovTransitionField* in proportion to the voting factor.

Fig. 4 shows the *Spiral* and the *MarkovTransitionField* representation. The markings of the figure are the same as those of the previous two figures. The average macro f1 score of the movement is 67.8%, that of the drawing is 61.7%, while that of the maximum points is 71.5%. The biggest change is 11.7% in macro f1 score on average, while the minimum is 1.8%. In this case too, we experienced cases where aggregation did not improve the results: Xception with k-NN classifier, Xception with SVM classifier and MobileNet with k-NN classifier. The first two showed no improvement using the *RecurrencePlot* either. The results of the two modalities show a similarly large difference compared to each other as with the *RecurrencePlot*. In the case of MobileNet k-NN, the difference between the two modalities is not that large, but it does not show improvement. The highest result was obtained with the ResNet50 feature extraction and the RF classifier in the macro f1 score of 78.3%. It achieved 76.1% on motion and 59.8% macro f1 score on drawing. This was approached by the MobileNet and RB with a macro f1 score of 78.2%.

The same case achieved 68.4% on movement and 70.6% on drawing. With the Mann-Whitney U test, we found a ***significant improvement*** between the maximum and *Spiral*, ***but not*** between the maximums and the *GramianAngularField*. The p-value is 0.009 between maximums and drawing results and 0.102 between maximums and movement results.

TABLE II summarises the best results of the single and joint modalities. The single modalities are marked with the names from TABLE I. The joint results are marked with the name of the representation approaches. The descriptor metrics are analogous to TABLE I. The movement representations appeared superior to the spiral drawing from the single use of the modalities. Utilizing joint predictions, improvements were achieved. The top performance was received with *RecurrencePlot* (sens: 75.6%, spec: 89.4%, macro f1 score: 82.5%).

TABLE II
SUMMARY TABLE OF THE BEST RESULTS OBTAINED FROM THE SINGLE AND JOINT MODALITIES.

	case	sens	spec	bacc	f1	auc
Single	Spiral	71.1%	70.2%	70.7%	70.6%	0.757
	RecurrencePlot	71.1%	78.7%	75.0%	74.9%	0.834
	GramianAngularField	80.0%	66.0%	72.8%	72.7%	0.768
	MarkovTransitionField	77.8%	74.5%	76.1%	76.1%	0.802
Joint	**RecurrencePlot**	**75.6%**	**89.4%**	**82.6%**	**82.5%**	**0.825**
	GramianAngularField	77.8%	74.5%	76.1%	76.1%	0.761
	MarkovTransitionField	80.0%	76.6%	78.3%	78.3%	0.783

V. DISCUSSION

Examining the modalities separately shows that the representations created from movement performed better than the drawing when examining the best results per modality (TABLE II). However, looking at the single results as a whole, we did not find any significant differences using the Mann-Whitney U test. This can be influenced by the methodology for generating the representations and the nature of the feature extraction models. It shows that the different data types generally have the same detection performance with these vision-based examinations on the applied classifiers. This may be a consideration when the physician wants to use a minimal tool in the shortest possible time. Nevertheless, it can be seen from TABLE II that a few percent better results can be achieved by selecting the best-performing models with motion data.

By using the modalities together, we experienced an improvement in all paired cases. This improvement proved significant in the *RecurrencePlot* and the *GramianAngularField*, whether we compared the best values to the drawing or the movement. In the case of the *MarkovTransitionField*, the improvement only reached a significant result compared to the drawing. This suggests that the decision is more confident and accurate when multiple sensors are used for the same task (even with the same processing scheme). However, by observing the voting factor, different optimal points may result. This implies that the modalities may play a different role in the final decision.

The present results provide insight into how the same task can provide better recognition by examining several sensors. There are also studies in the literature where, for example, video cameras and motion sensors help to recognize PD-related episodes. There is a multi-sensory examination of the same task similarly.

The limitation of the research is the heterogeneous database. A database with more elements may be necessary for filtering according to various factors (medication, brain stimulation). Another direction of development is sex equality. Although research [29] shows that the drawing of the Archimedean spiral does not differ significantly between men and women, this requires further support. Studies [30] are underway to investigate the separability of the sexes, where the authors have already shown that there is a significant difference when copying shapes. However, most studies have looked at participants between 18 and 30 years old and have used manual features that have not been linked to Parkinson's disease. However, this raises the need for further investigation.

Finally, it should be mentioned that the present results are based on image data processing with out-of-domain feature extraction algorithms. These play a role in avoiding overlearning by extracting a more general set of features. Presumably, fine-tuning the extractor models on the database specific to the task can increase the performance of the classifiers. However, a database with few elements can also cause overlearning.

VI. SUMMARY AND CONCLUSION

PD is becoming one of the most common neurological diseases of our time. The importance of research related to it is that there is no cure according to current clinical knowledge.

Current clinical knowledge is firmly based on motor symptoms, which are typically limb tremors at rest, bradykinesia, and muscle stiffness (rigidity). Recognizing them in the early stages is also not clear since the appearance of the symptoms and the lateral involvement may differ from person to person.

The support of artificial intelligence (including machine learning algorithms) can be desirable in the diagnostic procedure since it can effectively recognize even minor deviations and can create a more objective evaluation.

Previously, manually extracted features were used to analyze movements and drawings, which proved effective for learning the nature of the disease. On the other hand, deep learning algorithms allow the model to learn the characteristics that support automatic feature extraction. From this point of view, MobileNet, Xception, and ResNet50 were used, and they were previously trained on an extensive database that enabled cross-sectional recognition. We used networks trained in this way to recognize the disease, keeping the original weights.

In the literature research, it can be seen that several modalities are available to recognize PD, such as speech, drawing, or movement. These can achieve significant recognition performance by themselves. Comparing them is difficult because each modality goes through a different processing process, and there are also differences in the databases. In addition, the joint application of several modalities seems to improve PD recognition.

In our work, we examined the spiral drawings in connection with diagnosing the disease through two types of modalities. In the spiral drawing task, the X and Y coordinates of the actual drawing were recorded for each person, as well as the acceleration data along the X, Y, and Z axes with a wrist-mounted sensor.

Image representations were created from the data: 1) production of the actual drawing from X, and Y drawing data, 2) image representations from the resulting vector of movement data (*MarkovTransitionField*, *RecurrencePlot*, *GramianAngularField*). From these input images, we determined features with the pre-trained models, and then we trained classification algorithms on the features using nested cross-validation.

Based on the single modality results, image representations of the ***movement data reached at least a similar performance as the drawing itself*** (no significant difference). This result suggests that using only the spiral drawing alone or with the same processing of only the motion data from the drawing, PD. can be detected with the same efficiency.

Furthermore, the joint approach improved the recognition performance of the PD (significance difference)**, *highlighting the possibility of measuring the same task with various sensors.*** The result shows that even though the task the participant performs is the same, improving detection with different sensors is still possible.

Further investigation is required regarding the database composition.

REFERENCES

[1] R. Balestrino and A. H. V. Schapira, 'Parkinson disease', *Euro J of Neurology*, vol. 27, no. 1, pp. 27–42, Jan. 2020, DOI: 10.1111/ene.14108.

[2] T. Pringsheim, N. Jette, A. Frolkis, and T. D. L. Steeves, 'The prevalence of Parkinson's disease: A systematic review and meta-analysis', *Movement Disorders*, vol. 29, no. 13, pp. 1583–1590, Nov. 2014, DOI: 10.1002/mds.25945.

[3] R. F. Pfeiffer, 'Non-motor symptoms in Parkinson's disease', *Parkinsonism & Related Disorders*, vol. 22, pp. S119–S122, Jan. 2016, DOI: 10.1016/j.parkreldis.2015.09.004.

[4] S. Sveinbjornsdottir, 'The clinical symptoms of Parkinson's disease', *Journal of Neurochemistry*, vol. 139, no. S1, pp. 318– 324, Oct. 2016, DOI: 10.1111/jnc.13691.

[5] M. Gil-Martín, J. M. Montero, and R. San-Segundo, 'Parkinson's Disease Detection from Drawing Movements Using Convolutional Neural Networks', *Electronics*, vol. 8, no. 8, p. 907, Aug. 2019, DOI: 10.3390/electronics8080907.

[6] L. Moro-Velazquez, J. A. Gomez-Garcia, J. D. Arias-Londoño, N. Dehak, and J. I. Godino-Llorente, 'Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects', *Biomedical Signal Processing and Control*, vol. 66, p. 102418, Apr. 2021, DOI: 10.1016/j.bspc.2021.102418.

[7] Z. Li, J. Yang, Y. Wang, M. Cai, X. Liu, and K. Lu, 'Early diagnosis of Parkinson's disease using Continuous Convolution Network: Handwriting recognition based on off-line hand drawing without template', *Journal of Biomedical Informatics*, vol. 130, p. 104 085, Jun. 2022, DOI: 10.1016/j.jbi.2022.104085.

[8] N. Basnin, T. A. Sumi, M. S. Hossain, and K. Andersson, 'Early Detection of Parkinson's Disease from Micrographic Static Hand Drawings', in *Brain Informatics*, vol. 12960, M. Mahmud, M. S. Kaiser, S. Vassanelli, Q. Dai, and N. Zhong, Eds., in Lecture Notes in Computer Science, vol. 12960, Cham: Springer International Publishing, 2021, pp. 433–447. DOI: 10.1007/978-3-030-86993-9_39.

[9] B. Schoneburg, M. Mancini, F. Horak, and J. G. Nutt, 'Framework for understanding balance dysfunction in Parkinson's disease', *Movement Disorders*, vol. 28, no. 11, pp. 1474–1482, Sep. 2013, DOI: 10.1002/mds.25613.

[10] S. Del Din, A. Godfrey, C. Mazzà, S. Lord, and L. Rochester, 'Free-living monitoring of Parkinson's disease: Lessons from the field: Wearable Technology for Parkinson'S Disease', *Mov Disord.*, vol. 31, no. 9, pp. 1293–1313, Sep. 2016, DOI: 10.1002/mds.26718.

[11] J. E. McLennan, K. Nakano, H. R. Tyler, and R. S. Schwab, 'Micrographia in Parkinson's disease', *Journal of the Neurological Sciences*, vol. 15, no. 2, pp. 141–152, Feb. 1972, DOI: 10.1016/0022-510X(72)90002-0.

[12] M. Thomas, A. Lenka, and P. Kumar Pal, 'Handwriting Analysis in Parkinson's Disease: Current Status and Future Directions', *Movement Disord Clin Pract*, vol. 4, no. 6, pp. 806–818, Nov. 2017, DOI: 10.1002/mdc3.12552.

[13] A. W. A. Van Gemmert, H.-L. Teulings, and G. E. Stelmach, 'Parkinsonian Patients Reduce Their Stroke Size with Increased Processing Demands', *Brain and Cognition*, vol. 47, no. 3, pp. 504–512, Dec. 2001, DOI: 10.1006/brcg.2001.1328.

[14] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis, and M. Arnaoutoglou, 'Machine learning-based classification of simple drawing movements in Parkinson's disease', *Biomedical Signal Processing and Control*, vol. 31, pp. 174–180, Jan. 2017, DOI: 10.1016/j.bspc.2016.08.003.

[15] P. Sharma, S. Sundaram, M. Sharma, A. Sharma, and D. Gupta, 'Diagnosis of Parkinson's disease using modified grey wolf optimization', *Cognitive Systems Research*, vol. 54, pp. 100–115, May 2019, DOI: 10.1016/j.cogsys.2018.12.002.

[16] S. M. Ali et al., 'Wearable sensors during drawing tasks to measure the severity of essential tremor', *Sci Rep*, vol. 12, no. 1, p. 5242, Mar. 2022, DOI: 10.1038/s41598-022-08922-6.

[17] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa, and J. P. Papa, 'Deep Learning-Aided Parkinson's Disease Diagnosis from Handwritten Dynamics', in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Sao Paulo, Brazil: IEEE, Oct. 2016, pp. 340–346. DOI: 10.1109/SIBGRAPI.2016.054.

[18] J. Savalia, S. Desai, R. Geddam, P. Shah, and H. Chhikaniwala, 'Early-Stage Detection Model Using Deep Learning Algorithms for Parkinson's Disease Based on Handwriting Patterns', in *Advancements in Smart Computing and Information Security*, vol. 1759, S. Rajagopal, P. Faruki, and K. Popat, Eds., in Communications in Computer and Information Science, vol. 1759, Cham: Springer Nature Switzerland, 2022, pp. 323–332. DOI: 10.1007/978-3-031-23092-9_26.

[19] C. Taleb, L. Likforman-Sulem, C. Mokbel, and M. Khachab, 'Detection of Parkinson's disease from handwriting using deep learning: a comparative study', *Evol. Intel.*, vol. 16, no. 6, pp. 1813–1824, Dec. 2023, DOI: 10.1007/s12065-020-00470-0.

[20] G. D. Cascarano et al., 'Biometric handwriting analysis to support Parkinson's Disease assessment and grading', *BMC Med Inform Decis Mak*, vol. 19, no. S9, p. 252, Dec. 2019, DOI: 10.1186/s12911-019-0989-3.

[21] J. Faouzi, H. Janati, K. K. LEE, T. Carryer, R. Yurchak, and AvisP, 'johannfaouzi/pyts: Release of version 0.10.0'. Zenodo, Dec. 09, 2019. DOI: 10.5281/ZENODO.3568218.

[22] F. Chollet, 'Xception: Deep Learning with Depthwise Separable Convolutions', 2016, DOI: 10.48550/ARXIV.1610.02357.

[23] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep Residual Learning for Image Recognition', 2015, DOI: 10.48550/ARXIV.1512.03385.

[24] A. G. Howard *et al.*, 'MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications', 2017, DOI: 10.48550/ARXIV.1704.04861.

[25] H.-Y. Kim, 'Analysis of variance (ANOVA) comparing means of more than two groups', *Restor Dent Endod*, vol. 39, no. 1, p. 74, 2014, DOI: 10.5395/rde.2014.39.1.74.

[26] F. Pedregosa et al., 'Scikit-learn: Machine Learning in Python', 2012, DOI: 10.48550/ARXIV.1201.0490.

[27] P. E. McKnight and J. Najab, 'Mann-Whitney U Test', in *The Corsini Encyclopedia of Psychology,* 1st ed., I. B. Weiner and W. E. Craighead, Eds., Wiley, 2010, pp. 1–1. DOI: 10.1002/9780470479216.corpsy0524.

[28] B. M. Cesana, 'What p-value must be used as the Statistical Significance Threshold? P<0.005, P<0.01, P<0.05 or no value at all?', *BJSTR*, vol. 6, no. 3, Jul. 2018, DOI: 10.26717/BJSTR.2018.06.001359.

[29] M. San Luciano et al., 'Digitized Spiral Drawing: A Possible Biomarker for Early Parkinson's Disease', *PLoS ONE*, vol. 11, no. 10, p. e0162799, Oct. 2016, DOI: 10.1371/journal.pone.0162799.

[30] G. Cordasco et al., 'Gender Identification through Handwriting: an Online Approach,' *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Mariehamn, Finland, 2020, pp. 000197-000202, DOI: 10.1109/CogInfoCom50765.2020.9237863.

Attila Zoltán Jenei was born in Debrecen, Hungary in 1995. He graduated from the Budapest University of Technology and Economics as a Biomedical Engineer (Master's Degree, 2020). Since January 2020, he has been a department engineer and Ph.D. student at the Laboratory of Speech Acoustics, Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Information Technology. His research focuses on diagnostic support for Parkinson's disease with non-invasive medical data. He participated in the Student Research Societies of Budapest University of Technology and Economics and was awarded in 2017 and 2019. From 2021, he is the Vice President, and from 2023, the President of the Department of Engineering Sciences in the National Association of Doctoral Students.

Dávid Sztahó is a research fellow at the Budapest University of Technology and Economics. He completed his MSc studies in 2008 in Infromatics engineering. Since 2018 he is the head of the Laboratory of Speech Acoustics. He completed his PhD studies at the Doctoral School of Computer Science of the Budapest University of Technology and Economics, in the field of emotion recognition by speech signal. He received his PhD degree in 2014. His research interests include: speech technology, speech acoustics, speaker recognition and verification for forensic purposes, biomarker analysis by artificial intelligence, computer analysis of EEG signals.

István Valálik MD., Ph.D., MSc. neurosurgeon and head physician of the Department of Neurosurgery at St. John's Hospital, Budapest, honorary associate professor at the University of Debrecen. In 2011 he defended PhD thesis "CT-guided stereotactic thermolesion and deep brain stimulation in the treatment of Parkinson's disease", in 2015 MSc in Health Services Management. His scientific interest focused on movement disorders, MR-tractography-based surgical planning, psychiatric surgery, acoustic and motion analysis. He developed a planning software for stereotactic brain surgery and portable neuro-navigation system. Since 2010 he is acting in the Executive Committee of the European Society for Stereotactic and Functional Neurosurgery (www.essfn.org). In 2013 he was awarded by the Hungarian Academy of Sciences for the book "Stereotactic and Functional Neurosurgery". In 2019 he participated in mission of successful surgical separation of Bangladeshi craniopagus twins.

Guidelines for our Authors

Format of the manuscripts

Original manuscripts and final versions of papers should be submitted in IEEE format according to the formatting instructions available on

https://journals.ieeeauthorcenter.ieee.org/
Then click: "IEEE Author Tools for Journals"
- "Article Templates"
- "Templates for Transactions".

Length of the manuscripts

The length of papers in the aforementioned format should be 6-8 journal pages.
Wherever appropriate, include 1-2 figures or tables per journal page.

Paper structure

Papers should follow the standard structure, consisting of *Introduction* (the part of paper numbered by "1"), and *Conclusion* (the last numbered part) and several *Sections* in between.
The Introduction should introduce the topic, tell why the subject of the paper is important, summarize the state of the art with references to existing works and underline the main innovative results of the paper. The Introduction should conclude with outlining the structure of the paper.

Accompanying parts

Papers should be accompanied by an *Abstract* and a few *Index Terms (Keywords)*. For the final version of accepted papers, please send the short cvs and *photos* of the authors as well.

Authors

In the title of the paper, authors are listed in the order given in the submitted manuscript. Their full affiliations and e-mail addresses will be given in a footnote on the first page as shown in the template. No degrees or other titles of the authors are given. Memberships of IEEE, HTE and other professional societies will be indicated so please supply this information. When submitting the manuscript, one of the authors should be indicated as corresponding author providing his/her postal address, fax number and telephone number for eventual correspondence and communication with the Editorial Board.

References

References should be listed at the end of the paper in the IEEE format, see below:

a) Last name of author or authors and first name or initials, or name of organization
b) Title of article in quotation marks
c) Title of periodical in full and set in italics
d) Volume, number, and, if available, part
e) First and last pages of article
f) Date of issue
g) Document Object Identifier (DOI)

[11] Boggs, S.A. and Fujimoto, N., "Techniques and instrumentation for measurement of transients in gas-insulated switchgear," IEEE Transactions on Electrical Installation, vol. ET-19, no. 2, pp.87–92, April 1984. DOI: 10.1109/TEI.1984.298778

Format of a book reference:

[26] Peck, R.B., Hanson, W.E., and Thornburn, T.H., Foundation Engineering, 2nd ed. New York: McGraw-Hill, 1972, pp.230–292.

All references should be referred by the corresponding numbers in the text.

Figures

Figures should be black-and-white, clear, and drawn by the authors. Do not use figures or pictures downloaded from the Internet. Figures and pictures should be submitted also as separate files. Captions are obligatory. Within the text, references should be made by figure numbers, e.g. "see Fig. 2."
When using figures from other printed materials, exact references and note on copyright should be included. Obtaining the copyright is the responsibility of authors.

Contact address

Authors are requested to submit their papers electronically via the following portal address:

https://www.ojs.hte.hu/infocommunications_journal/about/submissions

If you have any question about the journal or the submission process, please do not hesitate to contact us via e-mail:

Editor-in-Chief: Pál Varga – pvarga@tmit.bme.hu

Associate Editor-in-Chief:
József Bíró – biro@tmit.bme.hu
László Bacsárdi – bacsardi@hit.bme.hu

FAREWELL TO TAMÁS GÁBOR CSAPÓ

"Rejoice with those who rejoice, and weep with those who weep."
Romans 12:15

I met Tamás Gábor Csapó in the spring of 2006, when we taught the course Speech Information Systems to the whole class of the then 5-year computer engineering course. He passed with distinction and contacted me saying that he was interested in the subject and would like to work on it. In the autumn, he presented a TDK (student research) paper on the machine implementation of prosodic variation, which won him 1st prize at BME VIK and 1st place at the 2007 OTDK (nationwide student conference) in Computer Science. In 2008, me and Mark Fék were the advisors of his successfully defended MSc thesis. The challenge has not been solved since then and is still a subject of research. In autumn 2008 he started his PhD studies at BME TMIT.

In the meantime, he has also been involved in teaching and our projects ranging from basic research to applications. In 2014, he spent six months at Indiana University as a Fulbright scholar with his family, where he was motivated to study articulation using ultrasound. After returning home, he defended his PhD thesis and was one of the initiators of the Lingual Articulation Research Group at ELTE, led by Alexandra Markó, in collaboration with BME (MTA-Lendület 2016-21). In 2017, he won his first OTKA (Hungarian National Science Program) grant on Articulatory Movement-based Speech Generation (2017-22). In 2022, he won another OTKA grant on Articulation and Brain Signal Analysis for Speech-based Brain-machine Interface (2022-26). Simultaneously, he became the Area Editor for Neural Speech Technology at the Infocommunications Journal. He has also built close relationships with colleagues at Szeged University. He has played a key role in winning and implementing our national (e.g. National Lab for Artificial Intelligence and National Lab for Infocommunications) and international (e.g. H2020, AAL, Horizon Europe) proposals. He has also contributed creatively to the development of our industrial applications. By the age of 39, he has published nearly 180 papers, with more than 320 independent citations. He has fulfilled the publication requirements for the degree Doctor of the Hungarian Academy of Sciences (MTA). Around Christmas, I encouraged him to start preparing his habilitation and MTA doctoral thesis.

Tamás was also open and supportive towards the students. Together we consulted Mohammed Al-Radhi, one of the first Stipendium Hungaricum scholarship holders at BME VIK, who has since become a valued colleague. Tamás was the supervisor of three PhD students in 2024.

Tamás was not only a great computer scientist, but also a great community and network builder. His open, relaxed and friendly nature and his deep faith in God made it easy to connect with him. We were honoured to attend their wedding in 2010 and followed with interest the growth of their family with four children. During COVID, he and his family started a new life in the countryside. It was good to hear his enthusiastic reports about the renovation of the house. In the summer of 2023, he organised a small international conference called Moonshine in his village. He was also involved in the ENFIELD Network of Excellence, which was launched in September 2023. On 25 January 2024, he still sent me an excellent research project plan.

It was a bolt from the blue that on 31 January 2024, his earthly journey came to an end and he moved to his heavenly home. Neither our closer nor our more distant colleagues were aware of the spiritual burdens Tamás was carrying. What led him to this point remains an eternal mystery. The lesson that remains with us is to try to look out for each other and support those around us. His wife and children can count on our solidarity and support.

2024. 02. 11.

On behalf of BME TMIT and Smartlabs,
Géza Németh, Head of SmartLabs

(The English translation was generated with the help of Deepl from the Hungarian original)

SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS

Who we are

Founded in 1949, the Scientific Association for Infocommunications (formerly known as Scientific Society for Telecommunications) is a voluntary and autonomous professional society of engineers and economists, researchers and businessmen, managers and educational, regulatory and other professionals working in the fields of telecommunications, broadcasting, electronics, information and media technologies in Hungary.

Besides its 1000 individual members, the Scientific Association for Infocommunications (in Hungarian: Híradástechnikai és Informatikai Tudományos Egyesület, HTE) has more than 60 corporate members as well. Among them there are large companies and small-and-medium enterprises with industrial, trade, service-providing, research and development activities, as well as educational institutions and research centers.

HTE is a Sister Society of the Institute of Electrical and Electronics Engineers, Inc. (IEEE) and the IEEE Communications Society.

What we do

HTE has a broad range of activities that aim to promote the convergence of information and communication technologies and the deployment of synergic applications and services, to broaden the knowledge and skills of our members, to facilitate the exchange of ideas and experiences, as well as to integrate and harmonize the professional opinions and standpoints derived from various group interests and market dynamics.

To achieve these goals, we…

- contribute to the analysis of technical, economic, and social questions related to our field of competence, and forward the synthesized opinion of our experts to scientific, legislative, industrial and educational organizations and institutions;
- follow the national and international trends and results related to our field of competence, foster the professional and business relations between foreign and Hungarian companies and institutes;
- organize an extensive range of lectures, seminars, debates, conferences, exhibitions, company presentations, and club events in order to transfer and deploy scientific, technical and economic knowledge and skills;
- promote professional secondary and higher education and take active part in the development of professional education, teaching and training;
- establish and maintain relations with other domestic and foreign fellow associations, IEEE sister societies;
- award prizes for outstanding scientific, educational, managerial, commercial and/or societal activities and achievements in the fields of infocommunication.

Contact information

President: **FERENC VÁGUJHELYI** • *elnok@hte.hu*

Secretary-General: **GÁBOR KOLLÁTH** • *kollath.gabor@hte.hu*

Operations Director: **PÉTER NAGY** • *nagy.peter@hte.hu*

Address: H-1051 Budapest, Bajcsy-Zsilinszky str. 12, HUNGARY, Room: 502

Phone: +36 1 353 1027

E-mail: *info@hte.hu*, Web: *www.hte.hu*