IEEE Sensors Letters VOL. 6, NO. 6, JUNE 2022

IEEE Sensors Council

Sensor signal processing

Spiking Neural Network-Based Radar Gesture Recognition System Using Raw ADC Data

Muhammad Arsalan[1,2*], Avik Santra[2**], and Vadim Issakov[1,2**]

[1] Infineon Technologies AG, 85579 Neubiberg, Germany
[2] Fakultät für Elektrotechnik, Informationstechnik, Physik, Technische Universität Braunschweig, 38106 Braunschweig, Germany
[*] Graduate Student Member, IEEE
[**] Senior Member, IEEE

Manuscript received February 21, 2022; revised April 4, 2022 and April 22, 2022; accepted April 28, 2022. Date of publication May 10, 2022; date of current version May 27, 2022.

Abstract—One of the main challenges in developing embedded radar-based gesture recognition systems is the requirement of energy efficiency. To facilitate this, we present an embedded gesture recognition system using a 60 GHz frequency modulated continuous wave radar using spiking neural networks (SNNs) applied directly to raw analog-to-digital converter (ADC) data. The SNNs are sparse in time and space, and event driven, which makes them energy efficient. In contrast to the previous state-of-the-art methods, the proposed system is only based on the raw ADC data of the target, thus avoiding the overhead of performing the slow-time and fast-time Fourier transforms (FFTs). Furthermore, the preprocessing slow-time FFT is mimicked in the proposed SNN architecture, where the proposed model processing speed of 112 ms advances the state-of-the-art methods by a factor of more than 2. The experimental results demonstrate that despite the simplification, the proposed implementation achieves recognition accuracy of 98.1%, which is comparable with the conventional approaches.

Index Terms—Sensor signal processing, gesture sensing, human–computer interface, neural engineering object (nengo), spiking neural networks (SNNs).

I. INTRODUCTION

Hand gesture sensing (HGS) has been an active research field over the last two decades because of its contactless nature and easy interfacing with machines [1], as compared with the click and touch devices. HGS contributes to many applications, such as TVs, smart home devices, automotive applications, and virtual reality. With the advancement in camera-based recognition systems, HGS is primarily dominated by camera-based systems [2], [3]. However, user privacy concerns and the requirement of environmental conditions (illumination, weather, etc.) limits the use of camera-based systems. Contrastingly, nonvision-based systems provide a cumbersome user experience due to their wearable nature.

Recently, contactless nonvision-based systems, such as radar-based systems, garnered a lot of attention because of their insensitive nature to the illumination conditions, invariance to hand occlusions, simpler signal processing pipeline, privacy-preserving features, ability to work within an enclosure, and their sensitivity to fine-grained gestures. There are two fundamental research directions to radar-based gesture systems: 1) building efficient miniature hardware for generating high-fidelity target data [4]–[6] and 2) the signal processing pipeline driven by deep learning to extract meaningful information from the target data of the user's intent [7]–[13]. Although the conventional deep neural networks (deepNets) approaches are quite promising in terms of gesture, detection, and recognition, energy consumption is still an issue making them unfavorable for portable devices. The majority of the energy in deepNets solutions is consumed by the multiply–accumulate (MAC) operations between layers [14]. Thus, researchers focus primarily on the reduction of MACs by using smaller networks, using pruning techniques and quantizing the weights [15], [16].

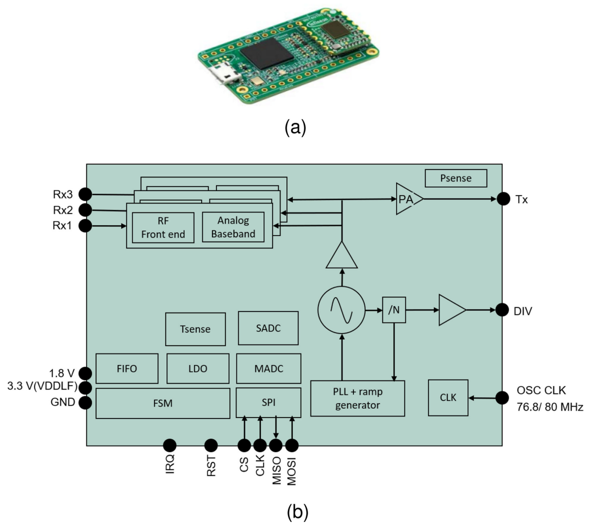
Contrary to traditional deepNets solutions, we propose a gesture system based on spiking neural networks (SNNs), where the information is encoded by spike timing, including latencies and spike rates. Furthermore, information is only transmitted when the neuron potential reaches a specific threshold, hence making the transmission sparse in nature with 1-bit activity, which substantially reduces the amount of data communicated between neurons. Moreover, the node accumulates information only when it receives a spike and the multipliers in MAC arrays are replaced with adders making the SNNs energy efficient. However, SNNs are hard to train due to the nondifferentiable transfer function hindering backpropagation, thus requiring a suitable learning mechanism for training the SNNs. There exists different spiking neuron models, which are defined at different levels of abstraction, but the leaky integrate-and-file (LIF) model is the most popular because of its biological plausibility and simple implementation requiring fewer computations (floating-point operations).

The main contributions of this letter are as follows.

1) We propose an end-to-end pipeline for a radar-based gesture sensing system from raw data using SNN.
2) Compared with the previous approaches [17]–[19], which work on Doppler images, the proposed approach is applied to the raw analog-to-digital converter (ADC) data, thus avoiding the overhead of performing slow-time and fast-time fast-time Fourier transforms (FFTs).
3) We proposed a novel SNN architecture where the signal preprocessing (slow-time FFT) is mimicked in SNN.
4) The overall end-to-end latency is substantially reduced (by a factor >2) compared with previous approaches [17], [18].

Corresponding author: Muhammad Arsalan (muhammad.arsalan@infineon.com).

Associate Editor: J. M. Corres.
Digital Object Identifier 10.1109/LSENS.2022.3173589


(a)

(b)

Fig. 1. (a) *Infineon's BGT60TR13 C* FMCW radar chipset. (b) Simplified block diagram of *Infineon's BGT60TR13 C* radar chipset.

II. PROPOSED RECOGNITION SYSTEM

A. Hardware

The frequency modulated continuous wave (FMCW) radar chipset *BGT60TR13 C* [20] by Infineon Technologies has been used in this work. The radar chipset and its simplified block diagram are shown in Fig. 1(a) and (b), respectively. The chipset embodies transmitters 1 transmit Tx path antenna and 3 receive path antennas, mixer, and ADC. It as an external phase-locked loop that controls the linear frequency sweep. A 80 MHz reference oscillator controls the loop with a frequency divider output pin and the finite state machine (FSM) is clocked by a reference clock running at 80 MHz [21]. The voltage-controlled oscillator is enabled by varying the tune voltage V_{tuned} from 1 to 4.5 V to generate a linear frequency sweep from 57 to 63 GHz. For memory readout serial peripheral interface and queued serial peripheral interface is embodied in the chipset allowing a maximum data transfer up to 200 Mb/s (4 × 50 Mb/s). For streaming out of the data, the FSM generate an interrupt request (IRQ) flag when the threshold set by host is met in the memory. The chipset is capable of transmitting a signal up to 6 GHz bandwidth where the transmitted signal and received signals are time domain multiplied and fed forward for further processing. In our experiment, we have used only single Rx channel.

B. Signal Processing

1) Moving Target Indication (MTI) Filtering: The raw ADC data are collected across a chirp (fast time) and arranged in rows along with the frame (slow time). In the FMCW radar, the reflection from stationary objects in the surroundings can subdue the hand reflection. Therefore, we apply MTI to suppress the reflection from these stationary objects and leakage. At each frame j, a moving average filter is applied to the fast time $F(j)$, given mathematically as

$$F(j) = F(j) - S(j-1) \tag{1}$$

$$S(j) = \alpha \times F(j) + (1-\alpha) \times S(j-1) \tag{2}$$

where α denotes forget factor set to 0.01, and $S(j)$ is the average weight of the current fast time $F(j)$ and previous MTI value $S(j-1)$. The value of $S(j)$ is set as 0 for the first-time step. The filtered fast-time data are fed to the target detection block.

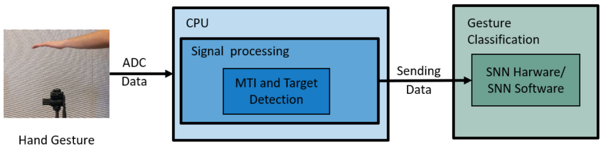


Fig. 2. Proposed signal processing chain.

TABLE 1. System Parameters Summary

Parameter	Symbol	Value
Number of ADC samples	NTS	64
Number of Chirps	N_c	1
Chirp Time	T_c	32 μs
Number of Transmit Antennas	N_{TX}	1
Number of Receive Antennas	N_{RX}	1
Total Bandwidth	B	5 GHz
Frame Time	T_f	75.476 ms
Azimuth Antenna Field of View	Θ_{FOV}	70°

2) Target Detection: Thresholding is used to perform target detection and selection on the filtered fast-time data. The threshold is determined by using the mean value of the fast-time data with a scaling factor. For example, the threshold Γ at frame j is given by

$$\Gamma_j = \beta \times \sum_{n=1}^{N_b} F_n(j) \tag{3}$$

where β is the scaling factor set to 3, chosen empirically for the tradeoff between false positive and probability of detection. n is index along range bins and N_b is the number of range bins. Once a moving target is detected, the filtered raw ADC data are accumulated and fed into SNN for gesture recognition.

C. System Design

Fig. 2 shows the experimental setup for recording hand gestures. A gesture is made in front of a 60 GHz radar chipset configured with the system (and derived) parameters given in Table 1.

The radar chipset is connected to a PC through a USB device to store the ADC data. Gesture recording is automated and starts as soon as the hand enters the field of view of the radar and the frame count starts. Each gesture is recorded and labeled for 32 consecutive frames with a large interclass variance. Our dataset contains some gestures of high variance, and gestures with shorter number frames (less than 32) are appended with zeros chirp values.

III. SPIKING NEURAL NETWORKS

Although inspired by the biological nervous systems, artificial neural networks (ANNs) still are unable to capture the complex neurocomputational properties of the biological neurons. To fill in this gap, the neuromorphic community has introduced the third generation of ANNs known as SNNs. SNNs, in contrast to the ANNs, capture more closely the functionality of the nervous system by taking into account not only the spatial but also temporal aspects of the input data for the construction of the computational model. The sparse and asynchronous communication in SNNs allows data processing in a massively parallel fashion [22]. Moreover, the low power consumption, fast inference, and event-driven information processing of SNNs make them a suitable candidate for efficient implementation of deep neural networks/machine learning tasks where energy efficiency is a

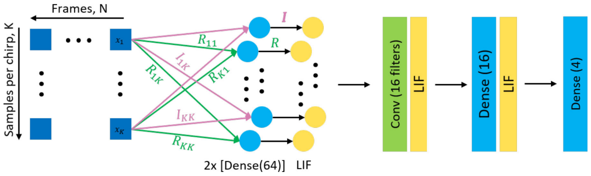


Fig. 3. Proposed SNN architecture.

prime requirement. The following sections describe the experiment and methods applied in this letter for the analysis of SNNs.

A. Neural Engineering Object (Nengo) Simulator

Nengo is a neural network tool for simulating large-scale neural systems. It has shown applications in cognitive science, psychology, artificial intelligence, and neuroscience [23]. For simulating deepNets, Nengo offers NengoDL [24] that allow simple merging of TensorFlow library, hence providing usage to rich features, such as convolutional layers. Nengo is based on a neural engineering framework to develop spiking neuron models with enormous applications in machine learning and deep learning, such as image classification [25], action selection [26], inductive reasoning [26], speech production [27], motor control [28], and planning with problem solving [29].

B. Proposed SNN Architecture

The proposed SNN architecture is shown in Fig. 3. Since our main objective is to make the model resource efficient in terms of computation and energy, our model uses LIF neurons. To construct the proposed model, we use NengoDL [24], which offers a differential approximation of the firing rate of the LIF neurons, allowing the training to be done in a conventional backpropagation manner. We use SoftLIF [30] activation (an approximation to LIF) with a multiclass cross-entropy function as an objective function.

The first layer of the SNN mimics the discrete Fourier transform (DFT). Since the DFT is a linear transformation and can be represented using two successive multiplications, each DFT dimension can be represented by a single dense layer. Let the input data have $N_s \times N_f$ dimensions, where N_s is the total number of samples per chirp and N_f is the total number of frames. Since the DFT is complex, the layer has $2 \times N_s$ nodes to compute both the real and imaginary values. The weights of the layer are calculated using the DFT trigonometric equation

$$D_k = \sum_{l=0}^{L-1} X_l \left[\cos\left(\frac{2\pi}{L}kl\right) - i\sin\left(\frac{2\pi}{L}kl\right)\right] \qquad (4)$$

where k and l take values between 0 and $N_S - 1$. In matrix form, (4) can be written as follows:

$$\boldsymbol{D} = (\boldsymbol{W_R} + i\boldsymbol{W_I})\boldsymbol{X} \qquad (5)$$

where $\boldsymbol{D}$ is the result of the transform, $\boldsymbol{X}$ is the input vector, and $\boldsymbol{W_R}$ and $\boldsymbol{W_I}$ are the real and imaginary coefficients.

The output of the first layer is then fed to convolutional layer of filter size 3 and a total number of 16 filters with stride 2. This is followed by LIF as a nonlinearity function, which converts the output into spikes. Next is a fully connected layer with 16 neurons appended with LIF as an activation function. Finally, we have a classification layer with four neurons as the model output.

The softLIF activation is replaced with LIF to make the network a spiking one during the testing phase. The weights and biases are extracted from the trained model and used in connections of LIF neurons. To obtain an accurate measure of the spiking neurons over time, the test inputs are presented multiple times/steps to the network.

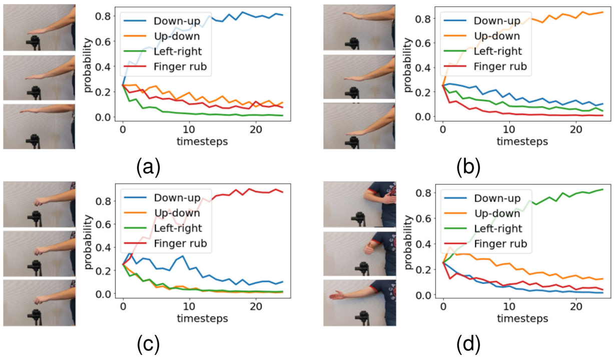


Fig. 4. Examples of four different gestures (a) down–up, (b) up–down, (c) finger rub, and (d) left–right, along with the SNN model class firing probability over time.

TABLE 2. Classification Accuracy of the Proposed SNN Versus Other SNN and Conventional ANN Models

Approaches	Input type	Input Size	Accuracy	Latency (ms)
LSTM	Range over time	64 × 32	96.9 %	311
2DCNN-LSTM	Range over time image	64 × 32	97.6 %	290
TCN	Range over time	64 × 32	98.9 %	960
1DCNN-LSTM	Range over time	64 × 32	97.2 %	315
SNN [17]	Range over time image	64 × 32	98.5 %	258
SNN [18]	Range over time image	64 × 32	97.5 %	229
Proposed SNN	ADC	64 × 32	98.1%	112

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Dataset

We used the dataset from [17] and [18] for our experiments. The dataset contains 7819 swipes of four different hand gestures:

1) down–up (moving hand in up direction);
2) top–down (moving hand in down direction);
3) left–right (moving the hand towards right); and
4) rubbing two fingers, as shown in Fig. 4.

The dataset is divided into 80% training dataset and 20% testing dataset, respectively.

B. Results

The performance of the proposed system is evaluated using classification accuracy. The system achieves a similar average accuracy of 98.1% as is achieved by the state-of-the-art models [17], [18] and counterpart deepNets, such as long short-term memory (LSTM), 1-D convolution LSTM (1DCNN-LSTM), 2-D convolution LSTM (2DCNN-LSTM), and temporal convolutional networks (TCN) over random trials, as given in Table 2. Furthermore, it can be seen that our proposed architecture has significantly low latency, as compared to the state-of-the-art [17], [18] and its counterpart deepNets.

C. Discussion

This letter presents an SNN-based gesture sensing system. In contrast to conventional deepNets, SNNs not only focus on the static values but also on the occurrence times of these static values [31]. SNNs are energy efficient because they use fewer neurons as compared to artificial neural networks. Moreover, SNNs are fast, scalable, and hardware friendly, and hence very cost-effective. Compared to previous SNNs solutions, the proposed solution only uses the raw data of the target and avoids the necessity of the slow-time and fast-time FFT operations. Moreover, in contrast to previous approaches, which use range over time images of 64 × 32 dimensions with 32 chirps per frame, the proposed system relies only on a single chirp per frame, which reduces the computation cost with a similar level of accuracy for classifying four gestures, as given in Table 2. This performance of the proposed system is attributed to the use of spatio-temporal information encoding by SNN, exploiting the network dynamics for learning. Moreover, the end-to-end (inference) latency of our proposed SNN is more than 2× faster than the state-of-the-art SNN methods. The latency here is the time taken by the system to infer after the gesture has been performed. Fig. 4 shows the examples of four different gesture samples and their firing choice by the model. It can be seen that after a few time steps the SNN starts firing for a sample from the correct classes. This is because presenting each image for longer allows integrating spikes over a longer time period and thus achieves better accuracy.

To determine its energy efficiency, we examine the energy consumption per classification. We used the hardware metrics of the μBrain chip defined in [32] and mathematically given as

$$E_c = N_{\text{spikes}} \times E_{\text{spikes}} + \delta T \times P_{\text{leakage}} \tag{6}$$

where E_c is the energy consumed per classification, N_{spikes} is the maximum number of spikes during classification, $E_{\text{spikes}} = 2.1$ pJ is the energy per spike, $P_{\text{leakage}} = 73\,\mu$W is the static leakage power, and δT is the inference time. Assuming δT to be 28 ms, the energy consumption per classification of the proposed system is $E_c = 2.05\mu$J.

V. CONCLUSION

In this letter, we present a novel SNN-based gesture-sensing system using 60 GHz frequency modulated continuous wave radar. In contrast to the state-of-the-art methods that operate on range over time images, the proposed method only works with raw ADC data, thus avoiding the overhead of slow-time and fast-time FFTs, which makes the model more than 2× faster in end-to-end processing. The evaluation of our proposed SNN architecture on four gestures has a similar level of accuracy performance in comparison to the state-of-the-art SNNs, thus making the proposed system favorable for low-latency and low-power-embedded implementations. In the future, we aim to implement SNNs for more and complex gestures for real-time inference.

ACKNOWLEDGMENT

This work was supported in part by the Electronic Components and Systems for European Leadership Joint Undertaking under Grant 826655 (Tempo) and in part by the German Federal Ministry of Education and Research (BMBF) through the KI-ASIC project under Grant 16ES0992K.

REFERENCES

[1] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for driver's hand-gesture recognition," in *Proc. 11th IEEE Int. Conf. Workshops Autom., Face, Gesture Recognit.*, 2015, pp. 1–8.

[2] A. Malima *et al.*, "A fast algorithm for vision-based hand gesture recognition for robot control," in *Proc. IEEE 14th Signal Process. Commun. Appl.*, 2006, pp. 1–4.

[3] D.-S. Tran *et al.*, "Real-time hand gesture spotting and recognition using RGB-D camera and 3D convolutional neural network," *Appl. Sci.*, vol. 10, no. 2, 2020, Art. no. 722.

[4] V. Lammert *et al.*, "A 122 GHz ISM-band FMCW radar transceiver," in *Proc. German Microw. Conf.*, 2020, pp. 96–99.

[5] V. Issakov, A. Bilato, V. Kurz, D. Englisch, and A. Geiselbrechtinger, "A highly integrated D-band multi-channel transceiver chip for radar applications," in *Proc. IEEE BiCMOS Compound Semicond. Integr. Circuits Technol. Symp.*, 2019, pp. 1–4.

[6] J. Rimmelspacher, R. Ciocoveanu, G. Steffan, M. Bassi, and V. Issakov, "Low power low phase noise 60 GHz multichannel transceiver in 28 nm CMOS for radar applications," in *Proc. IEEE Radio Freq. Integr. Circuits Symp.*, 2020, pp. 19–22.

[7] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz, "Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4207–4215.

[8] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, 2015.

[9] A. Santra and S. Hazra, *Deep Learning Applications of Short Range Radars*. Norwood, MA, USA: Artech House, 2020.

[10] Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic continuous hand gesture recognition using FMCW radar sensor," *IEEE Sensors J.*, vol. 18, no. 8, pp. 3278–3289, Apr. 2018.

[11] M. Arsalan and A. Santra, "Character recognition in air-writing based on network of radars for human-machine interface," *IEEE Sensors J.*, vol. 19, no. 19, pp. 8855–8864, Oct. 2019.

[12] M. Arsalan, A. Santra, K. Bierzynski, and V. Issakov, "Air-writing with sparse network of radars using spatio-temporal learning," in *Proc. 25th 25th Int. Conf. Pattern Recognit.*, 2021, pp. 8877–8884.

[13] M. Arsalan, A. Santra, and V. Issakov, "Radar trajectory-based air-writing recognition using temporal convolutional network," in *Proc. 19th IEEE Int. Conf. Mach. Learn., Appl.*, 2020, pp. 1454–1459.

[14] V. Sze, Y. -H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," in *Proc. IEEE*, vol. 105, no. 12, pp. 2295–2329, Dec. 2017.

[15] E. Hayashi *et al.*, "Radarnet: Efficient gesture recognition technique utilizing a miniature radar sensor," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2021, pp. 1–14.

[16] M. Scherer, M. Magno, J. Erb, P. Mayer, M. Eggimann, and L. Benini, "TinyRadarNN: Combining spatial and temporal convolutional neural networks for embedded gesture recognition with short range radars," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10336–10346, Jul. 2021.

[17] M. Arsalan, M. Chmurski, A. Santra, M. El-Masry, R. Weigel, and V. Issakov, "Resource efficient gesture sensing based on FMCW radar using spiking neural networks," in *Proc. IEEE MTT-S Int. Microw. Symp.*, 2021, pp. 78–81.

[18] M. Arsalan, A. Santra, M. Chmurski, M. El-Masry, G. Mauro, and V. Issakov, "Radar-based gesture recognition system using spiking neural network," in *Proc. 26th IEEE Int. Conf. Emerg. Technol. Factory Automat.*, 2021, pp. 1–5.

[19] A. Safa, A. Bourdoux, I. Ocket, F. Catthoor, and G. G. E. Gielen, "On the use of spiking neural networks for ultralow-power radar gesture recognition," *IEEE Microw. Wireless Compon. Lett.*, vol. 32, no. 3, pp. 222–225, Mar. 2022.

[20] 60 GHz-Infineon technologies. [Online]. Available: https://www.infineon.com/cms/en/product/sensor/radar-sensors/radar-sensors-for-iot/60ghz-radar/bgt60tr13c/

[21] S. Trotta *et al.*, "2.3 SOLI: A tiny device for a new human machine interface," in *Proc. IEEE Int. Solid- State Circuits Conf.*, 2021, pp. 42–44.

[22] N. Kasabov, K. Dhoble, N.Nuntalid and G. Indiveri, "Dynamic evolving spiking neural networks for on-line spatio- and spectro-temporal pattern recognition," *Neural Netw.*, vol. 41, pp. 188–201, 2013.

[23] T. Bekolay *et al.*, "Nengo: A python tool for building large-scale functional brain models," *Front. Neuroinform.*, vol. 7, 2014, Art. no. 48.

[24] NengoDL, [Online]. Available: https://www.nengo.ai/nengo-dl/

[25] J. A. K. Ranjan, T. Sigamani, and J. Barnabas, "A novel and efficient classifier using spiking neural network," *J. Supercomput.*, vol. 76, pp. 6545–6560, 2019.

[26] V. Senft, T. C. Stewart, T. Bekolay, C. Eliasmith, and B. J. Kröger, "Reduction of dopamine in basal ganglia and its effects on syllable sequencing in speech: A computer simulation study," *Basal Ganglia*, vol. 6, no. 1, pp. 7–17, 2016.

[27] B. J. Kröger, T. Bekolay, and C. Eliasmith, "Modeling speech production using the neural engineering framework," in *Proc. 5th Conf. Cogn. Infocommun.*, 2014, pp. 203–208.

[28] K. E. Friedl, A. R. Voelker, A. Peer, and C. Eliasmith, "Human-inspired neurorobotic system for classifying surface textures by touch," *IEEE Robot. Automat. Lett.*, vol. 1, no. 1, pp. 516–523, Jan. 2016.

[29] J. Knight, A. R. Voelker, A. Mundy, C. Eliasmith, and S. Furber, "Efficient spinnaker simulation of a heteroassociative memory using the neural engineering framework," in *Proc. Int. Joint Conf. Neural Netw.*, 2016, pp. 5210–5217.

[30] E. Hunsberger and C. Eliasmith, "Training spiking deep networks for neuromorphic hardware," 2016, *arXiv:1611.05141*.

[31] N. Kasabov, *Time-Space, Spiking Neural Networks and Brain-Inspired Artificial Intelligence*. Berlin, Germany: Springer, 2018.

[32] J. Stuijt, M. Sifalakis, A. Yousefzadeh, and F. Corradi, "μbrain: An event-driven and fully synthesizable architecture for spiking neural networks," *Front. Neurosci.*, vol. 15, May 2021, Art. no. 538.