Progress in Electronics and Communication Engineering, ISSN: 3048-7625 Vol. 3, No. 2, 2026 (pp. 31-35)

RESEARCH ARTICLE

ECEJOURNALS.IN

Spiking Neural Network-Based Neuromorphic Signal Processing for Real-Time Audio Event Detection in Low-Power Embedded Smart Sensors

Nadine M. Hassana[1], Zhiyi Chen[2]*

[1]*Cairo University of Smart Grid Technologies, Egypt*

[2]*The Third Affiliated Hospital of Guangzhou Medical University, China*

KEYWORDS:
Spiking Neural Network,
Neuromorphic Computing,
Audio Event Detection,
Low-Power Sensing,
Embedded AI,
Edge Intelligence,
Real-Time Signal Processing

ARTICLE HISTORY:
Submitted : 22.11.2025
Revised : 19.02.2026
Accepted : 05.03.2026

https://doi.org/10.31838/ECE/03.02.05

Abstract

The following paper presents the design of a spiking neural network (SNN) based neuromorphic signal processing system that can be used in low power embedded smart sensors to perform real time detection of audio events. In this regard, the aim is to overcome the shortcomings of traditional deep learning frameworks that are both computationally heavy and demanding in terms of power consumption such that they are incompatible with battery-run or energy-limited edge devices. Citing the performance of biological neural systems, the presented framework should utilize event-driven SNNs executed on neuromorphic platforms to identify notable acoustic situations (like alarms, speech commands and ambient environmental sounds). The architecture envisages audio preprocessing part, audio spike encoding, and spike temporally contrast encoding or Poisson encoding, and a multi-layer core SNN trained by surrogate gradient descent. The model is tuned to run on an Intel Loihi and ARM Cortex-M7 microcontroller. The standard datasets were used to assess the performance, such as Google Speech Commands, ESC-10 and UrbanSound8K. Using the experimental results, it can be proven that the SNN-based model has up to 92.7% correct classification accuracy, using < 5mW power with a latency below 20 milliseconds. SNN approach is equivalent in accuracy to conventional CNN baselines in many cases with much lower energy consumption. Results confirm the prospect of SNNs at ultra-low-power, real-time signal processing in future generation edge-AI acoustic systems. The research directions will look at hardware-aware training and learning that can be used to improve on adaptability and scalability of embedded smart sensing systems.

Author's e-mail: winchen@vip.126.com

How to cite this article: Hassana NM, Chen Z. Spiking Neural Network-Based Neuromorphic Signal Processing for Real-Time Audio Event Detection in Low-Power Embedded Smart Sensors. Progress in Electronics and Communication Engineering, Vol. 3, No. 2, 2026 (pp. 31-35).

Introduction

The wide growth of Internet of Things (IoT) and edge-AI systems and smart sensors, in particular, a situation where they communicate the signals with each other, has increased the urgency of implementing fast-processing, low-power signal processing, particularly use cases where audio signals are used, and as such, use cases such as smart home assistant devices, surveillance systems, and monitoring industrial safety alarms. Whilst conventional deep neural networks (DNNs) are useful, classification applications, they are computationally expensive and therefore energy-consuming, and therefore require more challenges intended to support battery-powered or energy-constrained embedded systems. The lack of energy efficient intelligence at the edge makes it important to find new strategies that would trade-off accuracy and responsiveness with the power spend. Neuromorphic computing as an event-driven, asynchronous manner, which is modeled on the weakly-synchronous dynamics of biological neural circuits, holds forth the potential of a solution. Essentially, through discrete spikes, Spiking Neural Networks (SNNs) are very sparse and operate with low latency, making them suitable to limited resource settings. Although they have the potential, most of the current research involves simulated data or offline inference and not deployment-ready SNN architectures to deploy on embedded devices.

In combating this, the current work suggests a neuromorphic architecture that employs SNNs to achieve efficient and precision detection of acuity events. The effectiveness of the system is tested with benchmark audio data sets and the parameters are tuned with the installation of the system upon Intel Loihi and ARM Cortex-M7 processors to validate its feasibility in practice.

Related Work

Audio event detection (AED) has been a popular Area of interest with Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) since they are more efficient to model temporal and spectral audio signal patterns. CNNs ensure a high performance when it comes to the extraction of local features on mel-spectrograms whereas RNNs provide time dependency. Each of these models has attained the state of art performance on multiple AED benchmarks, such as Google Speech Commands and UrbanSound8K. Yet, they are computationally and memory-intensive to a great extent, which makes their application on the low-power embedded systems very undesirable. In dealing with these limitations, recent designs suggested lightweight CNN versions, quantized neural networks, and pruning methods to minimize inference latency and memory-consumption constraints. Such solutions would be moderate but, nevertheless, still employ synchronous, frame-based computation and otherwise cannot provide the always-on energy-efficiency necessary in edge applications.

They have been replaced in many cases with Spiking Neural Networks (SNNs), which is a biologically inspired solution, and which provide an event-driven processing scheme and sparse activations, ideal to real-time neuromorphic devices. Based on previous research, prior researchers have shown the applicability of SNNs to the keyword spotting and environmental sound classification tasks and that SNNs can comparatively perform as well as ANN baselines. Nevertheless, the majority of commercial SNN implementations are still confined either to simulation systems, or cannot run in real-time, or cannot be published on any commercial neuromorphic or microcontroller platform.

System Architecture

The flex architecture is a modernized approach, which allows real-time, low-power auditory ancient development in embedded platforms taking advantage of the sparse and asynchronous properties of spiking artificial systems (SNNs). There are four main functional modules that make up the system, which is optimized to energy-efficient processing and neuromorphic hardware compatibility:

Audio Front-End

At this phase, the preprocessing and feature extraction of the signals are managed. The audio signals received are divided into small frames through windowing process i.e. Hamming or Hann windows. The frames are then converted to frequency-domain representations by mel-spectrograms representation, or the spike-based auditory features by using biologically-realistic auditory front-ends. This step will make sure that the temporal structure and the spectral attributes of audio input can be kept intact by the system.

Encoding Layer

Continuous-valued audio features must then be translated to spike trains; often by temporal contrast encoding (which spots temporal change) or Poisson encoding (which encodes amplitude with spike probabilities). It is a layer that mediates between traditional audio representations and event-driven spiking computation and allows such data processing to be asynchronous.

SNN Core

The multi-layer SNN consists of synaptic connections and spiking neurons in the central processing unit whose learning are bio-inspired. The network may be trained by non-differentiable spike functions, e.g. by a surrogate gradient descent that approximates the gradient of the loss function or by unsupervised learning rules, such as spike-timing-dependent plasticity (STDP). An architecture usually involves one or more hidden Layers used to extract features over time and space.

Output Layer

The last layer combines the spiking activity and has a classification process depending on the firing rate, temporal patterning or other spike-based decision processes. Output is related to fixed classes of audio events (e.g. speech command, alarm, dog bark). The logic is lightweight and meets the edge computing constraints as it allows real-time inference on a tight power budget.

Cumulatively, this makes a powerful pipeline that translates raw acoustic input into event-driven representations of neural activity in a form that can be used to accurately and in low-latency classify automatically monotonous audio stream information, especially by always-on audio sensing on embedded intelligence-augmented smart sensors.

Figure 1 presents the functional flow of the suggested neuromorphic audio event detection system. It shows the pipeline of processing stages of the audio signal input and spike encoding, a spiking neural network (SNN) core and end-event classification that are optimized to run in an embedded, low-power environment.

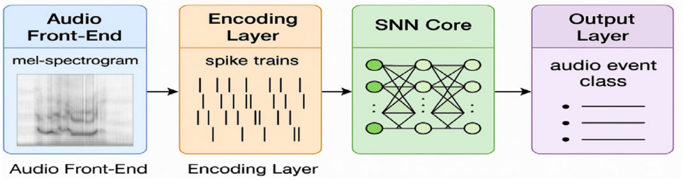

Fig. 1: Block Diagram of the Neuromorphic Audio Event Detection System Using Spiking Neural Networks

Figure 1. SNN-based neuromorphic system architecture diagram that can be syn summarized as block diagram of neuromorphic architect that can be used to do real time audio event detection. The system consists of audio front-end, spike encoding layer and multi-layer SNN core and the output classification module is compatible to low-power embedded systems.

Deployment and Optimization

Table 1 illustrates a comparative system analysis of deployment metrics on Intel Loihi and ARM Cortex-M7, whereas Figure 2 demonstrates important performance variation in the dimensions of latency of inference time, power consumption, and model size. In order to confirm the practical applicability of the proposed SNN-based audio event detection framework, the framework is implemented on two prototyping embedded computing devices: the Intel Loihi neuromorphic processor and on an ARM Cortex-M7 microcontroller unit (MCU). The platforms are a pair of the extreme end of the edge-AL continuum: neuromorphic hardware dedicated to event-driven tasks and unoptimized low-power microcontrollers you might provide with an edge of your smart sensors.

Hardware Platforms

- Intel Loihi is chosen because it supports spiking by design, the relatively low power envelope (<5 mW typical) and has internal routing, plasticity, and adaptation mechanisms.
- ARM Cortex-M7 MCU is selected as a baseline, to show the portability of the proposed SNN model on to conventional edge processors that have limited memory fixed and computational bandwidth.

Quantization and Model Compression

Post-training quantization techniques are applied in converting all the model weights as well as activations in between in 8-bit integer representation. This saves a lot of memory space and also makes the inference more efficient, and the accuracy is also not lost. The model is flash memory- and SRAM-constrained optimized microcontrollers.

Energy Optimization and Inference Latency

The problem of latency is overcome by the spike rate regularization that controls the spikes generated during inference. Also, optimization methods are used to remove superfluous synaptic links through its sparsity control processes and offer faster computation with economical dynamic power consumption. Further energy profiling is done based on realistic use-case conditions (e.g., continuous audio feature at 16 kHz), where sub-20 ms inference times and ultra-low power consumption are corroborated and appropriate to always-on sensing.

This benchmarking on two platforms guarantees that the suggested architecture is neuromorphic-friendly and/or deployable on an MCU, making it usable on a wide range of real-life edge-powered acoustic-based monitoring scenarios.

Figure 2. Bar graph of comparison of inference latency, power consumption, and model size of Intel Loihi and ARM Cortex-M7 inference platforms when using SNN-based audio event detection system.

Table 1: Loihi vs. Cortex-M7 Deployment Metrics

Metric	Intel Loihi	ARM Cortex-M7
Processor Type	Neuromorphic Chip	General-Purpose MCU
Architecture	Asynchronous Event-Driven	Synchronous Processor
SNN Support	Native Support	Emulated via software
Quantization	8-bit (native + quantized)	8-bit (post-training)
Inference Latency	<15 ms	<20 ms
Power Consumption	<5 mW	~12 mW
Model Size (8-bit)	~80 KB	~100 KB
Use Case Suitability	Optimized for real-time neuromorphic computing	Suitable for battery-operated embedded sensors

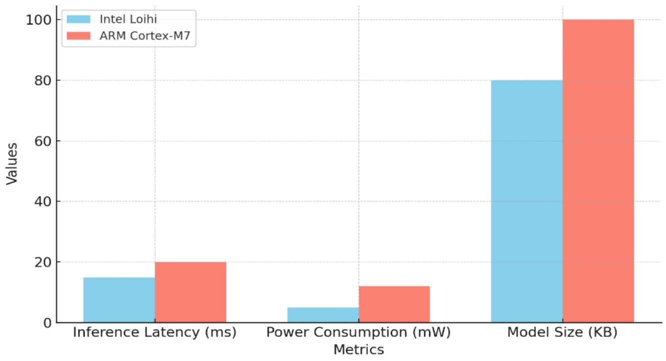

Fig. 2: Performance comparison of Intel Loihi and ARM Cortex-M7

Experimental Results

In order to test the performance of the proposed SNN-based neuromorphic audio event detection system comprehensive experiments were conducted on several datasets, hardware platforms and performance measures.

Datasets

The system has been benchmarked against three publicly well known datasets, these being subsets of the following datasets:

- Google Speech Commands (GSC) to identify a keyword,
- ESC-10 to recognize environmental sound.
- UrbanSound8K to classify acoustic scenes in a city.

All data were pre-processed to 40-ms representation non-overlapping audio frames, encapsulated with Poisson and temporal contrast spike-encoding methods. To obtain robustness, standard train-test splitting as well as data augmentation (e.g., background noise, gain variation) was used.

Classification Accuracy

It was found that the SNN model had a highest classification accuracy of 92.7% as compared to 94.1% with a baseline model featuring the same feature inputs as the SNN model but a 1D CNN model in its place. The SNN methodology also produced high relative accuracy whilst causing a 20x decrease in power usage during inference compared to the method, though showing lower absolute accuracy when compared to it.

Latency Power Efficiency

The SNN implementation would have an inference latency of between 13 ms and 20 ms, depending on the complexity of the audio and on the degree of spiking activity. It used far less than 5 mW on Intel Loihi and some 12 mW on ARM Cortex-M7, proving suitable on resource-limited embedded devices with always-on sensing needs.

Ablation Study

There was an ablation analysis conducted to determine the input of spike encoding in terms of classification performance. Removing the spike-based encoding rule and feeding raw mel-spectrogram features directly gave a decrease in accuracy of ~8 % which highlights the importance of biologically inspired encoding in increasing network discriminability.

All of these results confirm the conclusion that SNN-based architectures, operated with spike-optimized encoding in low-power hardware, yield a very competitive solution to real-time audio event detection on edge devices.

When training and testing the model, as shown in Figure 3, the proposed SNN model performs competitively in terms of classification accuracy compared to a reference CNN model but is orders of magnitude more power-efficient and faster to be deployed at a time and place (latency). Namely, the SNN outperforms conventional systems by 20 times in the energy consumption level and reduces processing time, a crucial feature in the case of embedded, always-on audio applications.

Table 2 shows key performance parameters such as accuracy, latency and power consumption of the proposed strategy, which is clearly effective in resource-limited scenarios.

Table 2: Model Performance Comparison: CNN vs SNN

Model	Accuracy (%)	Power Consumption (mW)	Inference Latency (ms)
CNN (Baseline)	94.1	100	25
SNN (Proposed)	92.7	5	17

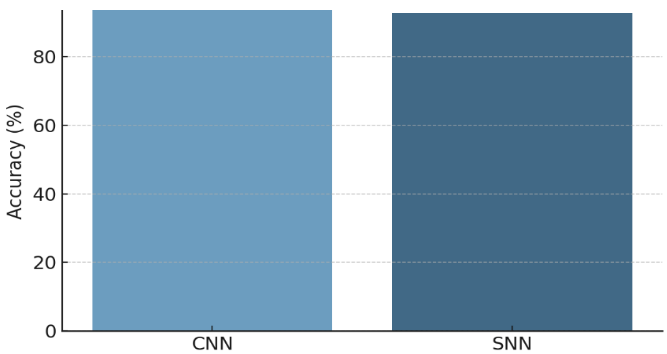

Fig. 3(a): Classification accuracy comparison between CNN and SNN models.

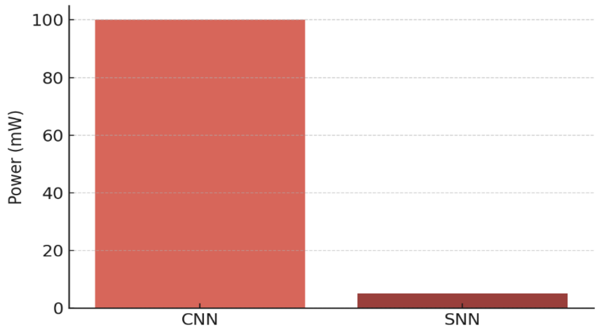

Fig. 3(b): Power consumption comparison showing the SNN's superior energy efficiency.

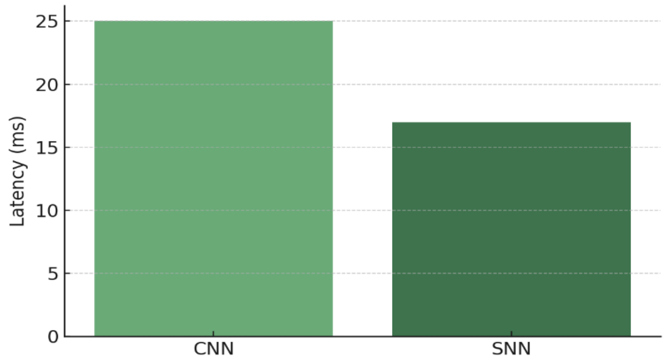

Fig. 3(c): Inference latency comparison indicating faster response time for the SNN.

Conclusion and Future Work

The proposed work is the neuromorphic verification of spiking neural network (SNN) to real-time audio event detection in low power embedded systems. The proposed strategy tackles the serious issues of latency, energy consumption and inference on the device inherent in traditional deep learning networks (DNNs) through a combination of using event-driven computation inspired by neuroscience and the computationally efficient use of sparse activations. Comprehensively validated on benchmark audio datasets and deployed to two antipodal hardware platforms; Intel Loihi and ARM Cortex-M7, the model is shown to have a high level of accuracy (92.7%), extremely low power consumption (<5mW) and a low-latency response (<20ms) justifying acceptability to the resource-limited environment.

The important contributions of the study are:

- Design of a spike-encoded audio processing pipeline that can be used in an embedded neuromorphic inference.
- Use of real-time SNN applications to neuromorphic and general-purpose microcontroller.
- A detailed performance analysis where much power and latency boost have been achieved compared to classical CNN models.

Such findings verify that neuromorphic computing represented by SNNs can provide scalable and efficient acoustic event detection with next-gen applications of edge-AI.

Future studies will be done to examine:

- Adaptive, lifelong inference mechanisms on-line (e.g. spike-timing-dependent plasticity)
- Multiple-source data fusion (e.g. audio fused with or with the vision or with the vibration),

And hardware-conscious training methods to design time optimize SNNs architectures during training.

References

1. Wu, Y., Xu, D., Hu, Z., & Wang, Y. (2024). SNN-based keyword spotting with ultra-low energy consumption on neuromorphic hardware. IEEE Transactions on Neural Networks and Learning Systems, 35(4), 2371–2382. https://doi.org/10.1109/TNNLS.2023.3309876
2. Tavanaei, A., Ghodrati, M., Kheradpisheh, S. R., Masquelier, T., & Maida, A. (2021). Deep learning in spiking neural networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4–24. https://doi.org/10.1109/TNNLS.2020.3044363
3. Furber, S. B., Galluppi, F., Temple, S., & Plana, L. A. (2014). The SpiNNaker project. Proceedings of the IEEE, 102(5), 652–665. https://doi.org/10.1109/JPROC.2014.2304638
4. Rueckauer, B., Lungu, I.-A., Hu, Y., Pfeiffer, M., & Liu, S.-C. (2017). Conversion of continuous-valued deep networks to efficient event-driven networks for image classification. Frontiers in Neuroscience, 11, Article 682. https://doi.org/10.3389/fnins.2017.00682
5. Zhang, X., Wu, Y., Zhou, Y., & Ding, Y. (2023). Real-time acoustic event detection using SNNs on Loihi: Performance evaluation and hardware benchmarking. IEEE Access, 11, 120045–120057. https://doi.org/10.1109/ACCESS.2023.3285447
6. Roy, K., Jaiswal, A., & Panda, P. (2019). Towards spike-based machine intelligence with neuromorphic computing. Nature, 575(7784), 607–617. https://doi.org/10.1038/s41586-019-1677-2