IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Dynamic- Structured Reservoir Spiking Neural Network in Sound Localization

Zahra Roozbehi[1], Ajit Narayanan[1]*, Mahsa Mohaghegh[2], and Samaneh-Alsadat Saeedinia[3]

[1]School of Electrical and Computer Engineering, Auckland University of Technology, Auckland, New Zealand
[2]Faculty of Design and Creative Technologies, Auckland University of Technology, Auckland, New Zealand
[3]Department of Electrical and Computer Engineering, Iran University of Science and Technology, Tehran, Iran

Corresponding author: Zahra Roozbehi (e-mail: zahra.roozbehi@aut.ac.nz).

,,

ABSTRACT Sound source localization is a critical problem in various fields, including communication, security, and entertainment. Binaural cues are a natural technique used by mammalian ears for efficient sound source localization. Spiking neural networks (SNNs) have emerged as a promising tool for implementing binaural sound source localization approaches. However, optimizing the topology and size of SNNs is crucial to reduce computational costs while maintaining accuracy. This paper proposes a real-time structure of a reservoir SNN (rSNN) called Adaptive-Resonance-Theory-based rSNN (ART-rSNN) for localizing sound sources in the time domain by integrating an energy-based localization method. The dataset used in this work is recorded by two different omnidirectional microphones from a real environment. The dataset includes various sound events such as speech, music, and environmental sounds. The proposed ART-rSNN architecture can dynamically adjust the location of its neurons to amplify estimated energy near the sound source, resulting in higher localization accuracy. Our proposed method outperforms several conventional and state of the art algorithms in terms of accuracy and is able to detect the front and back direction of azimuth angle. This work demonstrates the potential of dynamic neuron arrangements in SNNs for improving sound source localization in practical applications

INDEX TERMS Sound Localization, Spiking Neural Network, Dynamic Structure, ITD, Energy-Based Method, Adaptive Resonance Theory.

I. INTRODUCTION

IMAGINE we are blindfolded in a room, and we hear someone asking for help while the receiving sound is gradually diminishing or moving around. How is it that we can quickly detect where the sound may be coming from and how far away it is? Also, consider the environment is noisy. How do we manage to filter out the noise to still make a good guess as to the source of the sound? Sound Source Localization (SSL) stands as a skill of utmost importance in a varied range of applications, such as robotics, human-computer interaction (HCI), and virtual reality (VR). In the realm of robotics, SSL assumes a fundamental role in identifying the location of a sound source, especially in the presence of noise, which enables robots to identify and comprehend control commands. Within HCI, SSL proves to be an essential tool in isolating the speaker's voice from background noise, thus fostering clear communication in settings dominated by noise. In the context of VR, SSL is adopted to establish an immersive experience by localizing sound sources in the virtual environment and allowing users to perceive sound as if it were coming from a specific location. The problem of SSL has been approached by means of various techniques, which include traditional signal processing methods, machine learning algorithms, and biologically inspired models. These techniques encompass sound source localization modules, convolutional neural networks (CNNs), and recurrent neural networks (RNNs). Additionally, immune-based machine learning algorithms have been shown to increase accuracy and reliability, particularly in audio-visual approaches [1]. For sound classification, innovative techniques leverage Spiking Neural Network (SNN) encoding and spike pattern generation. [2] exploits the echo state SNN capability synergized with CNN classification methods, resulting in enhanced accuracy. Furthermore, [3] employs Convolutional Recurrent Neural Network (CRNN) methods incorporating Gammatone filtering and frequency-based approaches, yielding promising results. These multi-faceted methodologies showcase the evolution of SSL tech-

niques, embracing diverse technologies and demonstrating promising outcomes.

Traditional methods for a single SSL are based on time delay estimation and phase difference estimation using multiple microphones. These methods have limitations in noisy environments and require complex signal processing techniques [4]. In contrast, machine learning-based approaches for SSL, such as deep neural networks [5], support vector machines [6], and random forests [7], use large datasets to learn the relationship between the input sound signals and their corresponding source location. These approaches can handle noisy environments and do not require complicated signal-processing techniques. Instead, they solve the problem by feeding many examples of sound localization for the machine learning system to learn for itself how to localize the source. These approaches have shown promising results in the single SSL and have the potential to outperform traditional methods in terms of accuracy and computational efficiency. While these methods have shown promising results, they require a large amount of training data and computational resources.

To overcome the limitations of both traditional methods and machine learning-based approaches, researchers have turned to biologically inspired models for SSL. These models are inspired by the mammalian auditory system, which uses binaural cues, such as interaural time difference (ITD) and interaural intensity difference (IID), which together form the duplex theory of sound localization [8], [10]. Mammalian ears excel in sound localization, and researchers suggest that the human hearing system still outperforms machines in various auditory perception tasks [8]. One promising bio-inspired approach is the use of spiking neural networks (SNNs) that mimic the behavior of neurons in the human brain. SNNs are highly parallel and energy-efficient models that can process sensory information with high temporal precision and adapt to changing environmental conditions. Notably, these networks excel in energy reduction through event-triggering methods, employing spike encoding strategies. In simple terms, the spike-based coding in SNNs dictates that neurons activate solely in response to continuous spiking trains, utilizing all-or-none pulses (spikes) for information transmission. This coding strategy fosters sparseness in neuron activations, further enhancing the efficiency of SNNs [9].

Two popular types of SNNs are recurrent SNNs and reservoir SNNs. Recurrent SNNs have feedback connections that enable them to maintain temporal information and perform complex computations [11]. Reservoir SNNs have fixed random connections that create a dynamic system that can process input signals in a nonlinear manner [12]. Both types of SNNs have been used for SSL with promising results, but they also face challenges in optimizing their networks and achieving high accuracy [13], [14]. Numerous studies have explored Spiking Neural Network (SNN)-based methods for sound source localization, often leveraging interaural time difference (ITD) and interaural intensity difference (IID) cues and occasionally incorporating frequency features [15]–[18]. Some investigations have focused on low-frequency pure tone localization using delay lines [20], while others extended their scope to wider frequency ranges, achieving remarkable accuracy through medial superior olive (MSO) neurons [21]. Furthermore, a multi-tone phase coding ITD model demonstrated exceptional direction resolution [22], although hardware constraints led to the development of an energy-based method for enhanced practicality [23]. Additionally, these studies have explored diverse bio-inspired sound localization mechanisms, including spatiotemporal filtering and spiking nonlinearity [21].

To further improve SSL accuracy and reliability, recurrent neural networks (RNNs) have been developed for both static and dynamic scenes, capable of localizing events in full azimuth and elevation under matched and unmatched acoustic conditions, regardless of microphone arrays [24]. Additionally, a neuromorphic real-time sound tracking system was proposed, consisting of a neuromorphic auditory system with the aim of tracking high-frequency sounds in a biologically inspired way [25].

However, there are certain issues that arise when using SNNs for sound localization. Representing sound data inside the network with precise timing and providing significant information to allow the network to learn and analyze data accurately is a key challenge. Obtaining sufficient training data to teach the SNN how to localize sound accurately is also a challenge, which can be addressed by generating realistic simulations. However, creating an environment similar to the real world in simulations is also a challenge. Another issue is the computation time required by the SNN to process data, which can render it useless for certain tasks [28]. Additionally, SNNs are limited to single-tone frequency analyzing or narrow-band applications, which is another challenge that needs to be addressed.

Overall, there are still several key issues that need to be addressed before SNNs can be effective in real-world sound localization problems. These range from sensor calibration and obtaining realistic training data to data signal processing and computation time. However, with the right expertise, these issues can be managed successfully, allowing sound localization problems to be addressed faster, more accurately, and at a lower energy cost. To reduce computation costs, an important consideration for SNNs is the size of the network. A larger SNN may provide greater accuracy and complexity in its output, but it also requires higher amounts of memory and computational power [29]. To maximize the efficiency of the SNN, optimization techniques are used to determine the smallest network size necessary to achieve a desired performance level [30].

The optimization of an SNN initiates with the careful selection of an architecture tailored to the specific task. Various architectures, including recurrent and convolutional SNNs, present distinct strengths and weaknesses. Once the architecture is chosen, adjusting inter-neuronal weights becomes pivotal for optimizing the network size [31]. This process, often coupled with cost measures such as neuron count or parameter size aims to determine the most efficient network

size. Beyond weight adjustments, optimization extends to fine-tuning each neuron's threshold and learning rate. The threshold regulates neuronal activity, while the learning rate influences the formation of new connections, which is crucial for achieving accurate performance with minimal size [31].

The selection of an appropriate cost function is a critical step in sizing optimization. Cost functions dictate the trade-off between accuracy, memory usage, and computation time [32]. Studies such as like [33] emphasize the role of proper cost functions in enhancing noise rejection and reducing sensitivity, offering an effective approach to optimization. Depending on the optimization goal, an SNN may prioritize accuracy, leading to a larger size, or opt for low memory with reduced accuracy.

Balancing the threshold, learning rate, network weights, and cost function constitutes a complex but vital process. This meticulous tuning ensures the creation of an SNN with the most suitable size and performance for the given task [34]. Neuron models proposed in [34] expedite response speed by dynamically regulating neuron membrane conductivity based on spiking activity and external input. Other optimization approaches, such as adjusting firing thresholds [35] and [36], focus on reducing latency. While numerous studies have achieved significant advancements, challenges persist in mitigating computation costs, selecting optimal network sizes, and refining application-based learning laws.

Addressing the mentioned issues, this paper proposes a new recurrent SNN architecture to detect and localize a single sound event. The main contribution of this study is developing a self-modified architecture of SNN to SSL, in broadband frequency ranges. This structure uses machine learning methods to identify the event, and then track the sound source in a dynamic indoor environment. The reservoir structure of the proposed SNN as a kind of recurrent Neural network architecture is efficiently able to jointly detect and track the sound source [29] due to taking advantage of fast learning at low training cost and amenability to hardware implementation [30]. In this regard, this study tailors a reservoir-SNN (rSNN) structure in order to investigate the impact of several interconnection parameters on the performance of sound event localization. The superiority of the newly designed rSNN architecture can be expressed as follows:

- Integration of Energy-based and ITD cues to increase accuracy and determine distance as well as azimuth angle
- Addressing possible false negative azimuth angle estimation in the proposed algorithm
- Self-modification of the network size and the spatial position of the neurons
- Fusing spatial and temporal features of the proposed rSNN to localize the sound source

In addition to comparing the proposed strategy with three conventional and well-established methodologies, namely Energy-Based, GCC-PHAT, and Music Algorithms, this paper also scrutinizes it alongside a recent conventional STDP-based SNN method [21], and LS-SVM [37]. This paper investigates how modifying the network size and arrangement can speed up the convergence of the proposed rSNN for sound source localization. This is conducted by comparing the proposed algorithm within two fixed and dynamic structures. The paper is organized as follows: Section 2 describes the materials and methods, Section 3 presents the proposed novel rSNN architecture and learning algorithm, and Section 4 represents simulation results. Section 5 evaluates the role of dynamic structure, and finally, Section 6 concludes.

II. MATERIALS AND METHODS

A. ART-RSNN ARCHITECTURE

The overall structure of the presented Art-rSNN architecture consists of three modules, the input, rSNN, and the output. Figure 1 indicates the overarching steps in the Art-rSNN method. It reveals that the proposed sound localization strategy is composed of five main stages, data acquisition, encoding, input to the observed neuron, mapping received information to the size-growing hidden network, and finding the maximum potential neuron place as the estimation of the sound source.

Due to the linearity of the sound physical laws, the received sound is a linearly filtered version of the audio, corresponding to the location of the sensors and sources, as well as the acoustical environment. In the proposed structure, inspired by binaural hearing, there should be considered at least two sensors in the environment to specify the synchrony patterns, by means of a pair of location-specific filters related to the sensors.

Figure 1 depicts that sound is first encoded based on a desired spike detection algorithm, here is BSA, and then both the audio signal and the spike sequence code are input to the proposed structure which is composed of two groups of observed and hidden neurons. The observed neurons directly receive data from the environment, and the hidden neurons do not receive direct outside input and train based on the observed neurons' activities. In this structure, the location of the maximum-energy hidden neuron estimates the sound source location. As shown in Fig. 2, the first network size is directly relevant to the number of sensors in the under-study environment, and here we have considered that the minimum possible sensor quantity is two. In the proposed recurrent network, each neuron position in the initial arrangement is matched to the locations of the sensors. Clarifying the issue, the given figure indicates how the network grows and the new neurons are generated. Figure 2 indicates that at the initial state, architecture embarks on its work by the number of observed neurons as same as the number of sound sensors. The number of hidden neurons can be considered as the minimum possible number, for example, zero. Then, in each estimation epoch by the rSNN, a new hidden neuron will be generated according to the estimated location of the sound source. Then by considering the small-world technique, the role of newly generated neurons improves the learning quality of the proposed structure. In the new configuration, a threshold

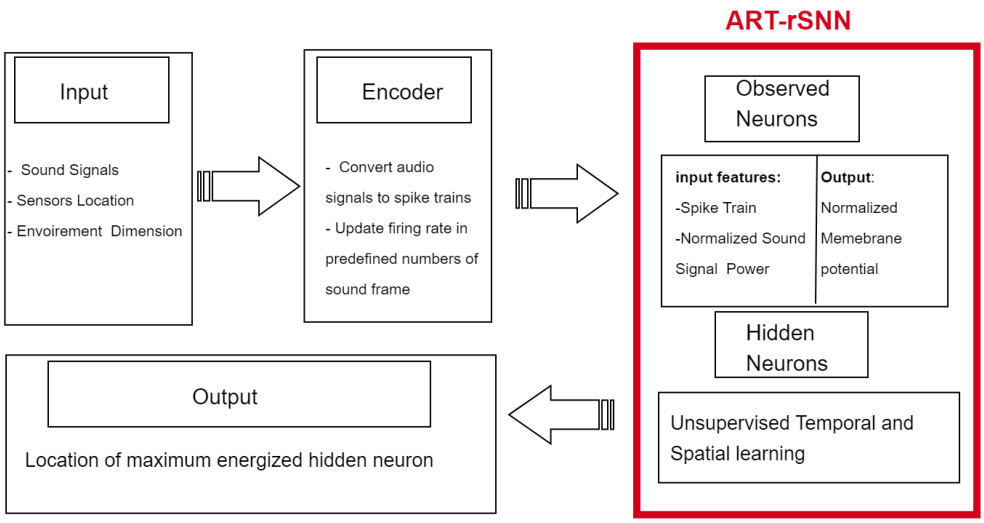

FIGURE 1. Main Stages and ingredients of the proposed Method

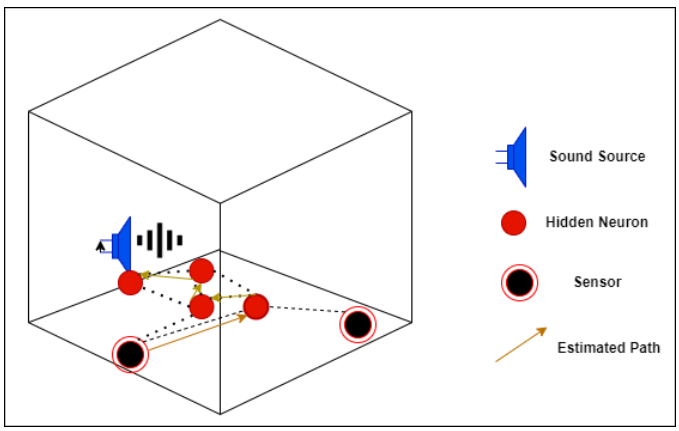

FIGURE 2. Changing the network size approach

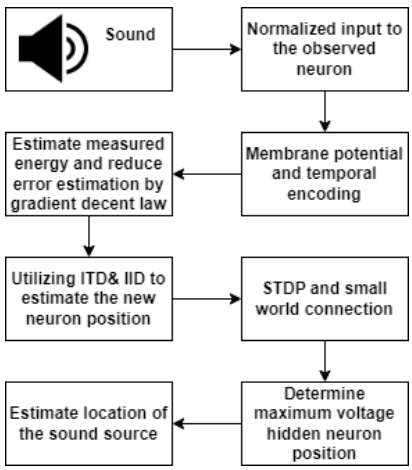

FIGURE 3. Art-rSNN sound localization process pipeline

of a minimum required energy is considered to eliminate the neurons that receive low-power signals. This threshold limits the networks' connectivity. The active neighborhood area is detected based on the neurons' interaction criterion. In this regard, this paper proposes a new dynamical structure inspired by the resonance theory in neural networks. The pipeline of the proposed methodology is illustrated in Fig. 3. According to the given pipeline in Fig. 3, the incoming sound signal is normalized and temporally encoded. The membrane potential of the observed neuron is also normalized and the weights of the rSNN are updated so that normalized membrane potential outputs track the normalized measured data in order to reduce energy estimation error. The ITD and IID cues are intended between two high-potential neurons to estimate the new position of the generated hidden neuron by the time that the error estimation is reduced to a desired value. At last, the location of the highest potential hidden neuron is the final estimation of the sound source. The architecture of the proposed method is demonstrated in the following figure. To extend the proposed idea in a mathematical model, this paper uses the concept of Adaptive Resonance Theory (ART) as a biologically plausible theory of how a brain learns to consciously attend, learn, and recognize patterns in a constantly changing environment. The next section develops Art-rSNN.

B. ADAPTIVE RESONANCE THEORY

The theory states that resonance regulates learning in neural networks with feedback (recurrence). Thus, it is more than a neural network architecture or even a family of architectures. Through the dynamic creation of recognition categories for encoding distinct input samples, an ART module is capable of self-adjusting the scale of its recognition field, in terms of the number of committed nodes, with respect to the complexity of the problem domain. Its fast commitment mechanism

and capability of learning at a moderate speed guarantee high efficiency. However, given a dataset, the scale of the ART recognition field (i.e., the number of output clusters) depends on a global threshold parameter called vigilance. While in principle, one could control ART's recognition representation by fine-tuning the vigilance parameter, in practice, suggesting an appropriate vigilance value requires prior knowledge of the scale and the distribution of the problem data set, which is unlikely to be available [23].

In addition, while sound travels through the air, acoustic energy is emitted omnidirectionally from the sound source. The strength of a sound source diminishes at a rate inversely proportional to the square of the distance. The tradition of this algorithm is given as follows [16]:

$$y_i(t) = \zeta_i \frac{S(t)}{|r_i - r(t)|^{\alpha}} + \varepsilon_i(t) \tag{1}$$

Equation (1) describes the relation of measured signal $y_i(t)$ on the ith sensor with $S(t)$ as the actual sound energy, recorded from a 1-meter distance from the sound source. ζ_i is the gain factor of the ith acoustic sensor. r_i (sensor location) and $r(t)$ (unknown location) indicate the coordinate of the ith sensor node and sound source at time t. Each variable is a vector with two additional variables (when in a two-dimensional (2-D) plane). ϵ_i is the measurement noise, modeled by zero-mean Gaussian Noise. When there are m sensor nodes, the value of α is the path loss exponent which is considered $\alpha = 2$. Regarding both deterministic and metaheuristic algorithms, all observations from the multiple sensors are aggregated as an estimator of $r(t)$, where the solution of the localization problem is the argument (pair of coordinates) that minimizes the expression.

$$\hat{r}(t) = \arg\min_{r} \sum_{i=1}^{m} \frac{1}{\sigma_{\xi_i}^2} \left(y_i - \zeta_i \frac{S(t)}{|r_i - r(t)|^{\alpha}} \right)^2 \tag{2}$$

where $\frac{1}{\sigma_{\xi_i}^2}$is the variance of acoustic gain factor. The estimator in (2) is highly nonconvex, with singularities in each sensor's coordinates, several suboptimal solutions, and saddle regions. All the enumerated features make the problem very challenging in the field of numerical optimization, making it a good candidate in the context of regression and ANNs. Recurrent spiking neural networks have shown promise in addressing optimization problems due to their ability to process spatial and temporal data effectively. The rSNN architecture, as a type of Recurrent SNNs, leverages SNNs and embodies a Liquid State Machine architecture, which is instrumental in tackling complex problems [26]. In the design and architecture of rSNNs, leaky integrate-and-fire (LIF) neurons are often chosen as the spiking neuron models due to their capacity to generate diverse spike patterns with a logical time cost [27]. The membrane potential V in the LIF neurons evolves according to a specified equation, contributing to the adaptability and efficiency of the proposed neural network model [27].

Incorporating the LIF model enhances the computational capabilities of rSNNs, allowing for the representation of intricate temporal and spatial patterns in the context of the optimization problem under consideration [27]. The membrane potential V evolves according to the equation [25]:

$$\frac{dV_i(t)}{dt} = \frac{1}{\tau_m} \left(-V_i(t) + I_i(t) \right) \tag{3}$$

Where τ_m denotes the membrane decaying time constant. $I_i(t)$ is the synaptic current.

III. PROPOSED ART-RSNN METHOD

This paper introduces a novel structure of a Liquid State Machine (LSM), which is a type of reservoir spiking neural network capable of generating new hidden neurons. The proposed structure leverages a small-world connection strategy to achieve its functionality. The architecture and construction details of the LSM are inspired by the principles of reservoir computing and spiking neural networks. The LSM consists of randomly connected liquid layers and readout layers, allowing for the modification of weights. This design enables the generation of complex dynamics akin to the brain and facilitates real-time task processing. Figure 4 indicates the proposed architecture.

In the initial states, only m observable neurons that receive input signals, measured signals of microphones/sensors, are regarded. The main goal is estimating the real energy of the signal by approximating y as the neuron output value. The algorithm of the proposed strategy is given below and the proof of updating weights in(5) is presented in Appendix A.

Art-rSNN Algorithm

1: Initialization:
W_{ij}, ϕ_i ,τ ,A , ϵ , and τ_m to be the Non-Zero arbitrary values.
Set β to a value between 0 and 1 Set V_j to random values.
Set I_j to zero value.
Set Constant Parameter, c=342

Art-rSNN Algorithm: Continue

2: update Synaptic currents:

$$I_i(t) = W_{ij} V_j \exp\left(-c\Delta t_s/\tau\right) \tag{4}$$

3: Update Observed Neurons' Weights:

$$\Delta W_{ij} = 0.5 \left(\tanh\left(y_o\right) - \tanh\left(V_i\right)\right) \left(1 - \tanh\left(V_i\right)^2\right) \times Vj \exp\left(-c\frac{t_i - t_j}{\tau}\right) \tag{5}$$

4: Update Hidden Neurons' Weights:

$$W_{ij} = \begin{cases} AVje^{\frac{\Delta t_{sij}}{\tau}} & \Delta t \geq 0, \quad i,j \in Ni \\ 0 & \Delta t < 0, \quad \text{or} \quad i,j \notin Ni \end{cases} \tag{6}$$

5: Estimate Hidden neuron position:

$$\kappa = \frac{V_i}{V_j + \text{sgn}\left(V_j\right)\varepsilon} \tag{7}$$

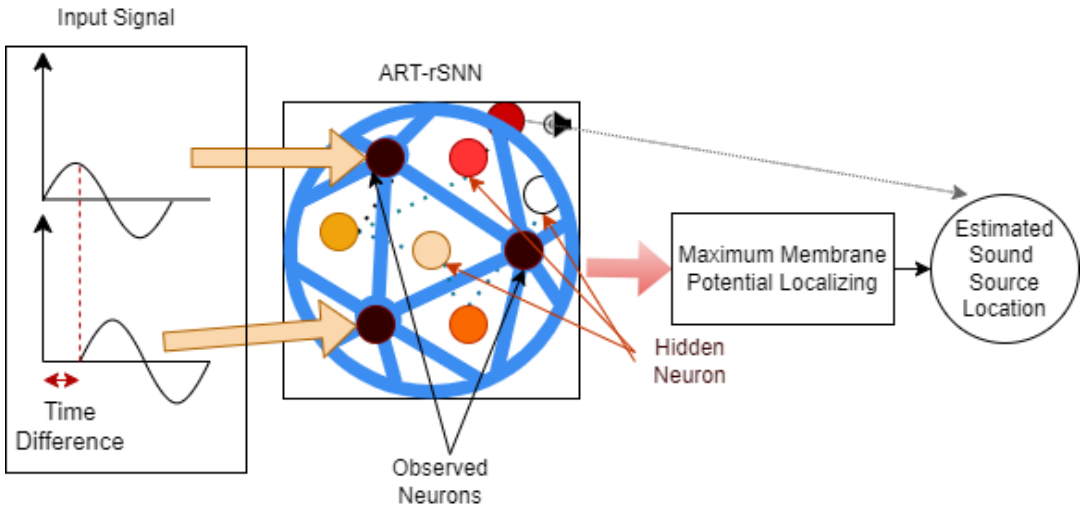

FIGURE 4. Architecture of the proposed Art-rSNN sound localization strategy

6: Generate a New Hidden Neuron i, and Determine Its Voltage Potential equal to the Maximum Voltage of all Neurons:

$$V_i = \max(V) \text{ and } V_i \geq V_j > V_k, \varepsilon > 0 \quad k \neq i, j \quad (8)$$

7: Calculate the distance of the new neuron From the Maximum Voltage Neuron:

$$\text{d}=\begin{cases} \frac{c\kappa\Delta t_{sij}}{1-\kappa} & \text{if } V_i \neq V_j \quad d_{ij} = \| \text{Pos}_i - \text{Pos}_j \|_2 \\ \frac{d_{ij}}{2} & \text{else} \end{cases} \quad (9)$$

8: Determine the location of the new neuron position:

$$\text{Pos}\, V_{\text{new}} = \text{Pos}\, V_i + \begin{bmatrix} d\cos(\theta) \\ d\sin(\theta) \end{bmatrix} \quad (10)$$

9: Determine the sign of Azimuth angle:

$$f(\varphi) = \alpha_l(\varphi + \sin(\varphi)) - \alpha_l(\text{sgn}(\varphi)\pi - 2\varphi)|\beta\sin(\varphi)|, \quad \alpha_l = \frac{d_{ij}}{2c}, \quad 0 < \beta < 1 \quad (11)$$

$$\theta = \begin{cases} \cos^{-1}\left(\frac{c\Delta t_{sij}}{d \cdot d_{ij}}\right) & \text{if } |\Delta t_{sij} - f(\varphi)| < |\Delta t_{sij} - f(-\varphi)|, \\ & \varphi = \cos^{-1}\left(\frac{c\Delta t_{sij}}{d \cdot d_{ij}}\right) \\ -\cos^{-1}\left(\frac{c\Delta t_{sij}}{d \cdot d_{ij}}\right) & \text{else} \end{cases} \quad (12)$$

Δt_{ij} is the difference of spike time in neuron i and neuron j, and ϵ is a constant parameter. β is 0.001. Equations (7) to (11) describe how a new neuron is generated. κ is the ratio of neuron potential i and j. W_{ij} is the synaptic weight between neurons i and j. This parameter indicates that the received energy by neuron i how much is stronger than jth neuron. d denotes the sound source distance from the reference neuron. Equation (14) calculates the new position of the neuron. f(φ) in (10) indicates if the source is located at the front or back. Equation (11) describes the azimuth angle calculation formula. Figure 5 indicates the graphical abstract of the

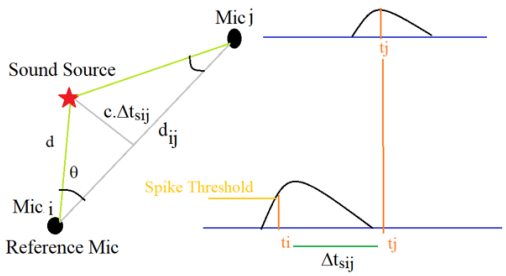

FIGURE 5. Sinusoidal relation between parameters based on time delay

proposed method. The power of the signals is considered directly relevant to the membrane potentials of the neurons in the proposed structure. To evaluate the performance of the proposed method, we implement the proposed strategy on a real database, including two omnidirectional sensor data. The data utilized for assessing the suggested approach was captured by the researchers within their laboratory and subjected to pre-processing prior to being inputted into their proposed architecture alongside other comparable methodologies, fairly. Then, we compare the results with several well-known methods.

The proposed algorithm's practicality is highlighted by its innovative use of Spiking Neural Networks (SNNs) for event triggering, offering an energy-efficient solution. Integrated into a 2D spatiotemporal SNN framework, it processes signals' magnitudes, reducing preprocessing time compared to other techniques. While time-domain methods alone might sacrifice accuracy, the algorithm's swift localization strategy excels in tracking moving sound sources, surpassing current deep learning methods in speed. Notably, the algorithm's reliance on online learning laws eliminates the need for extensive datasets or pre-training. These enhancements distinctly showcase the algorithm's practicality and detail effectiveness in sound source localization

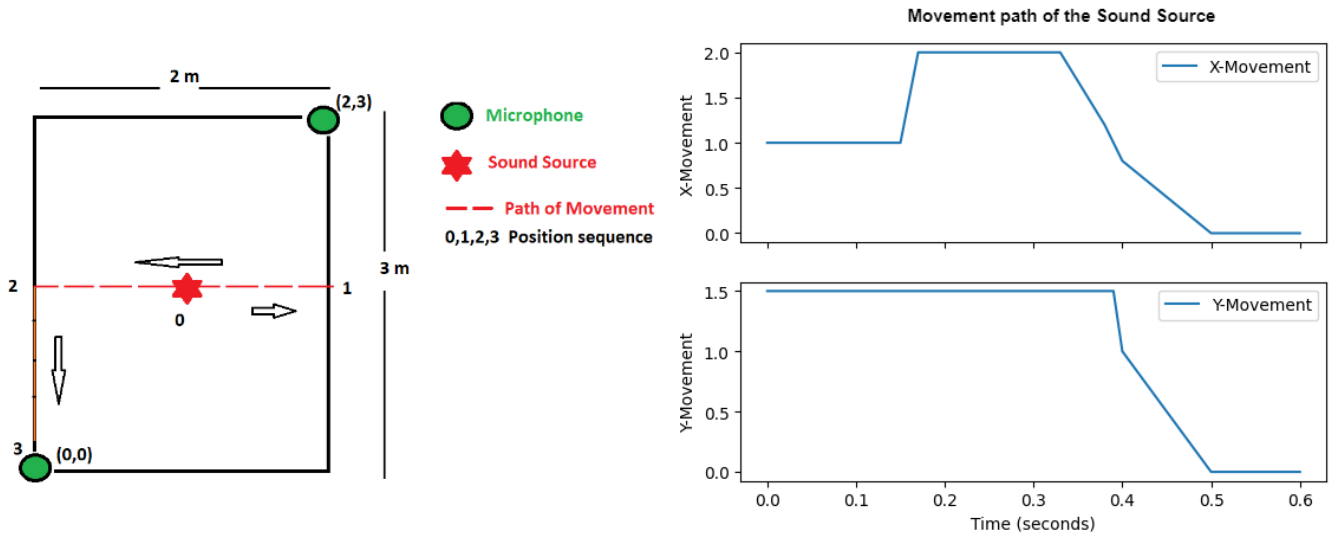

FIGURE 6. The Sound Source and Microphones arrangements and movement path of the sound source in 2D x- y axes

IV. SIMULATION RESULTS

In this section, we analyze and compare several methods of DOA, TDOA, and IMID to localize sound sources. Python 3.10 is used to analyze the data in this study. The utilized dataset is described in the following section.

A. DATASET

We evaluated the proposed approach on two datasets: our recorded data and the L3DAS22 multi-channel speech enhancement challenge dataset. The first dataset comprises four couples of recorded signals, including periodic noisy clapping sounds at positions (0,0), (0,1.5), (1,1.5), and (1,2). Two sensors are located on the ground at positions (0,0) and (2,3). The mean of the background noise is 0.42 (W) with a standard deviation of 5 (kW), and the signal-to-noise ratio ranges between 5 to 8 dB. Two omnidirectional microphones were used to record the audio. The noises are mainly generated by vehicle movements, approximately lower than 30 dB. The area under study is 2*3 m, located in a larger 12 m² area equipped with furniture and negligible reverberation. The microphone's Z-axis is zero. A sound source is considered in this record, which moves linearly along a 2D environment. The environment arrangement and the sound source's x and y movement paths are indicated in Fig.6. The second dataset is provided by the L3DAS22 Task 2. It is split into three subsets, consisting of 600, 150, and 150 30-second-long audio recordings for the train, validation, and test splits, respectively. There are 14 types of sound events selected from the FSD50K dataset. The maximum number of overlapping sound events is three, but here we utilized one overlapping and 4 classes of the sound events, 'writing, knock, Drawer open and close, cupboard open and close'. The room impulse response (RIR) is sampled in an office room with dimensions around 6 m (length) by 5 m (width) by 3 m (height). FOA microphone arrays are placed in the center of the room, with the position of the FOA microphone arrays set to be the origin of the coordinates. The signals recorded in a real environment are depicted in Fig. 8. Real datasets are typically favored over synthetic datasets due to their broader range of inputs and improved representation of real-world scenarios. Despite the benefits of artificial datasets, including the ability to generate large training datasets without manual data labeling and the alleviation of privacy concerns, the techniques employed to train with synthesized datasets may not be equipped to handle the uncertainties inherent in real environments. Additionally, synthetic data is difficult to validate for its accuracy, and it does not copy the original content exactly [38], [39] As shown in Fig. 7, Signal Noise Ratio (SNR) is low, and there is background noise in both recorded signals; So, filtering is necessary to clean the data. The first 10 seconds of the recorded signals include only a single clap hand audio signal in position (0,1). We use a band-pass filter to remove the background noise of the recorded audio. The utilized filter is Butterworth, 5 degrees with bandpass 400 Hz – 1000 Hz. Figure 8 reveals 1-second filtered signals, recorded by two sensors 1 and 2.

V. EVALUATION OF THE ROLE OF DYNAMIC STRUCTURE IN RSNNS

In this section, the performance of an RSNN is evaluated with two different real sample data in an environment. Sound sources are respectively recorded at (0,1.5) and (1,1.5), considering the location of Mic2, namely (0,0) as the reference node. Figure 9 (a,b,c) depicts the proposed architecture how localizes a sound source, located at (1,1.5). In the fixed structure, hidden neurons are randomly arranged, and the location of the neuron with the higher membrane potential is considered as the best estimation of sound sources. With the aim of integrating spatial data, instead of a mere time difference cue, the STDP updating law is modified according

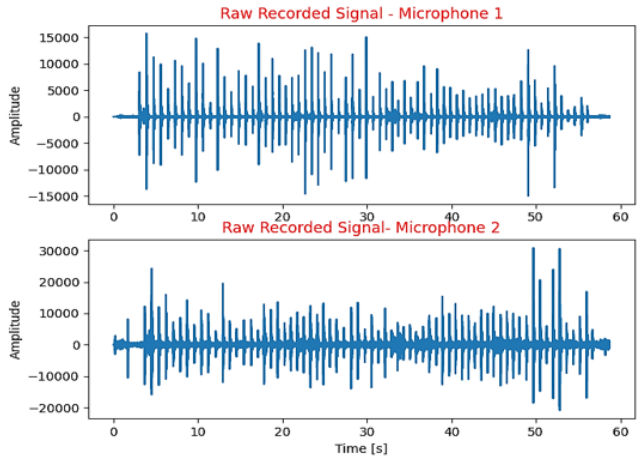

FIGURE 7. Recorded Raw signals by microphones 1 and 2 which are respectively located at (2,3) and (0,0). The upper figure is the recorded signal by microphone 1 and the lower figure indicates the recorded signal by microphone 2

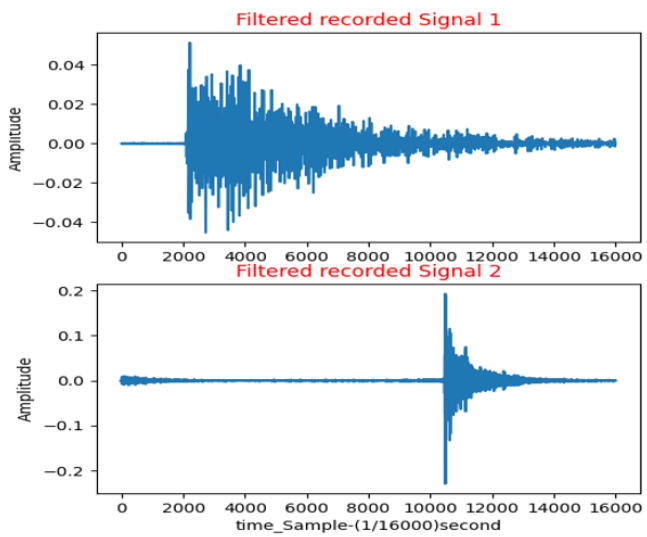

FIGURE 8. Filtered signals, in the upper figure, signal1 is recorded at position(2,3) and the lower figure indicates the signals which are recorded at (0,0). The signals are filtered by a Butterworth band pass filter. 1 second of the recorded signal is depicted

to the following equation.

$$w_{ij}(t) = AV_j\, e^{-\left(\frac{d_{ij}}{c_0}\right)}, \quad c_0 > 0, \quad A > 0 \qquad (13)$$

VI. EVALUATION OF THE ROLE OF DYNAMIC STRUCTURE IN RSNNS

In this section, we analyze the performance of an RSNN with a fixed and dynamic structure. The tuning law is chosen based on the acoustic velocity in the environment. A sample arrangement of the fixed structure RSNN is indicated in Fig. 11.

As shown in Fig. 11 hidden neurons are considered as well as two I/O neurons. Received energy is predicted by the observed neurons, which the relevant Mean Square Error (MSE) is illustrated in Fig. 12. Figures 12, and 13 illustrate that the calculated MSE converges to a specific value which denotes that the proposed estimator is biased; therefore, we should have normalized the input, properly to have an unbiased estimation. Although MSE quickly converges to its steady-state value, this question still arises what happens if the number of hidden neurons increases? To respond to this question, we raise the number of hidden neurons to 100.

TABLE 1. Network Parameters

Parameter	Value
τ	0.01
A	100
Train	70%
Test	30%
c_0	342

Figure 13 indicates the MSE of sound energy prediction by the 100 hidden neurons. Higher convergence speed is clearly detectable in Fig.13; however, utilizing the time-process function of the time library in Python 3.10 indicates a logarithmical increase of computational time cost from approximately 0.024 to 0.079 seconds at each iteration in the same processor. Calculating the processing time of the first fixed structure and the second larger structure indicates that although the iteration numbers of the smaller network are higher, the incremental time process of the smaller network is not significantly much more than the larger one, while their accuracy is almost the same. Therefore, knowing how much we can increase the network size can reduce computational costs. So comparing the computational time cost and convergence speed of fixed and dynamic structures, figures 11, 12, and 13 indicate that although the computational cost of the proposed strategy is not much lower than the fixed one with 10 hidden neurons, possibly due to integrating the ART section computation costs to the ART-SNN method, the precision of sound localization has increased. The simulated network parameters are given in Table 1.

We compare some well- and conventional SSL algorithms, namely known energy-based, MUSIC, GCC-PHAT, LS-SVM, and conventional SNN, to better understand their performance in at least 5 sample examples. Table 2 compares the proposed method with two conventional sound localizing methods for sound source steady-state error averages and the standard deviations for the four mentioned recorded data. The given Table 2 indicates the superiority of the proposed method in localizing the sound source with only two sensors in comparison to both SNN-based and non-SNN-based approaches. Furthermore, it seems that sensitivity to the sensors' arrangements in both MUSIC and GCC-PHAT algorithms should have triggered the higher error in sound localizing.

To assess our proposed method on the L3DAS22 dataset, we employed two key metrics: Accuracy and Mean Error at 20 degrees (ER_{20}). The results are visually presented in

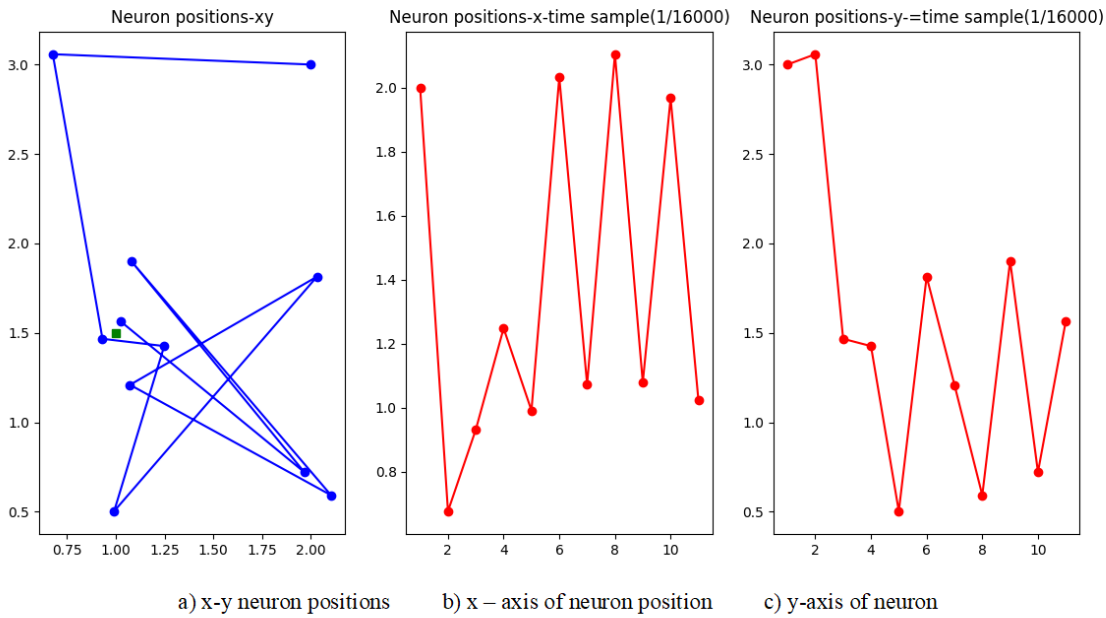

FIGURE 9. a) Tracking high potential neuron position progress in the environment, b) x-time sample data of the neuron position, created by the proposed dynamic architecture, c) y-time sample data of the neuron position, created by the proposed dynamic architecture- Sound Source (green square) is located at (1,1.5)

TABLE 2. Comparison of SSL algorithms on the recorded data

Item	MAE(deg)	MDE(m)	SD	RT
Energy-Based	-	(0.7 to 4)	1.2 (m)	Yes
MUSIC	33	-	15(deg)	No
GCC-PHAT	27	-	12.02(deg)	No
LS-SVM	15	-	10 (deg)	No
Conventional SNN	15	-	15(deg)	Yes
Fixed-rSNN	7.28	0.42	0.5(m)	Yes
ART-rSNN	3.4	0.38	0.322(m)	Yes

MAE: Mean of Direction Angle Error of Estimation
MDE: Mean of Distance Error of Estimation
SD: Standard Deviation of Error of Estimation
RT: Real-Time Applicability of the methods.

Fig. 14. Accuracy, calculated as the percentage of correct predictions among the total, serves as a comprehensive indicator of the system's overall performance. A higher accuracy percentage signifies better alignment between predicted and true sound source locations. The accompanying chart also illustrates the Mean Error at 20 degrees, providing insights into the average angular deviation between predicted and true angles. This metric offers a nuanced evaluation, emphasizing the system's accuracy specifically at the critical angle of 20 degrees. The bar chart collectively provides a comprehensive view of our sound localization system's effectiveness, facilitating interpretation and comparison under various conditions.

In Fig. 14, we present a comparative analysis of sound localization results achieved by the CRNN, ART-rSNN, ResNet-Conformer, and SRP-PHAT methods. The plot showcases the mean error at 20 degrees for each method and their corresponding accuracy values. Notably, our proposed method, ART-rSNN, exhibits a lower mean error at 20 degrees compared to the other methods, indicating superior performance in terms of localization precision. The accuracy of our method stands out, showing results nearly identical to CRNN, with only a marginal 0.01 decrease in accuracy compared to CRNN. Furthermore, our method outperforms ResNet-Conformer and SRP-PHAT in accuracy.

These outcomes affirm the efficacy of incorporating Mag features in the L3DASS dataset for sound localization. Specifically, our method achieves an accuracy of 69.8%, slightly below the 70.3% accuracy achieved by CRNN, while maintaining a notable advantage in computational efficiency. The proposed ART-rSNN method demonstrates a calculation time, approximately one-tenth that of deep learning methods like CRNN and ResNet-Conformer. This significant reduction in computation time not only attests to the computational efficiency of our approach but also positions it as a promising solution for real-time applications where speed is crucial. In summary, the results presented in Fig. 9 underscore the favorable trade-off between accuracy and computational efficiency offered by our ART-rSNN method when compared to existing state-of-the-art techniques in sound localization.

VII. CONCLUSION

This paper proposes a new rSNN architecture with a dynamic network arrangement that can modify network size to increase the performance of compromise between accuracy and network structure. The proposed network initializes by the possible smallest size and grows gradually based on the

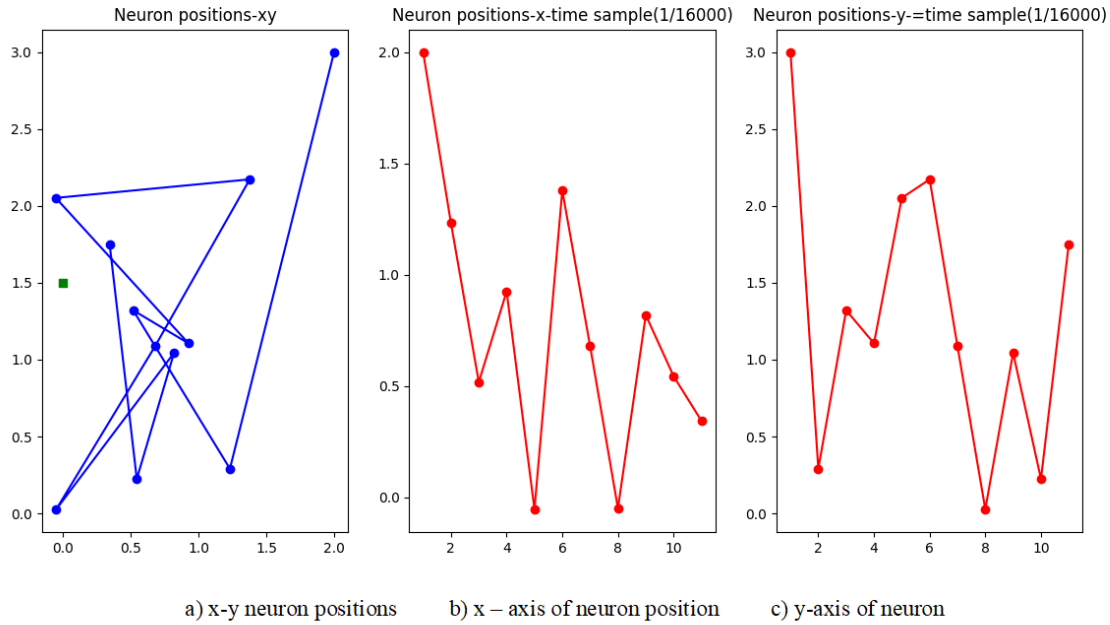

FIGURE 10. a) Tracking high potential neuron position progress in the environment, b) x-time sample data of the neuron position, created by the proposed dynamic architecture, c) y-time sample data of the neuron position, created by the proposed dynamic architecture- Sound Source (green square) is located at (0,1.5)

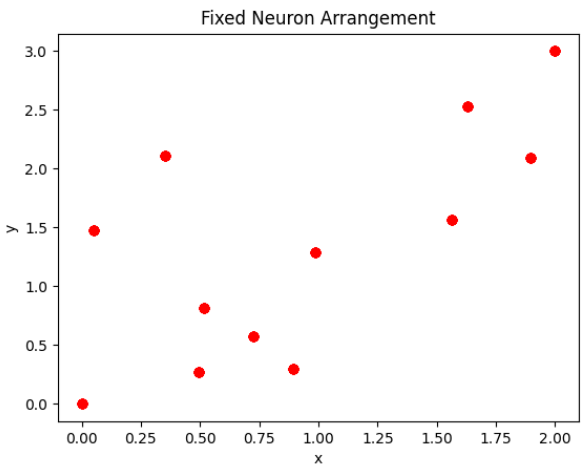

FIGURE 11. Fixed Neuron arrangement with 10 hidden and 2 I/O neurons (located at (0,0) and (2,3))

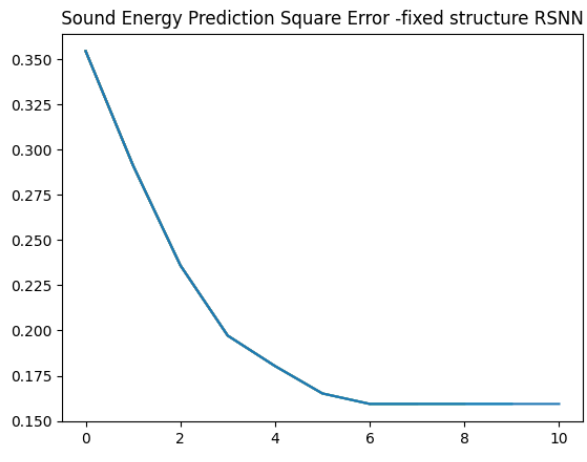

FIGURE 12. Sound Energy Prediction MSE, calculated by the fixed structure RSNN. The Y-axis indicates MSE amplitude and the X-axis represents the iteration number

error of estimation. The proposed system works based on the encoding procedure's threshold to provide an event trigger-based approach. These features can enhance the ability of the new architecture to be utilized in event-trigger sound source localization so that the neurons in different positions are activated based on the target path trajectory. The proposed method is investigated by a fixed structure network and four other conventional algorithms: Energy–Based by Normal and random search distribution strategy and GCC-PHAT, MUSIC algorithms, and a conventional STDP-based SNN and LS-SVM. Results indicate that the proposed ART-rSNN method is able to converge to the target location in a few iteration numbers with a higher estimation accuracy rather than the fixed structure SNNs and the other classic methods. Furthermore, in our comprehensive evaluation of the L3DAS22 dataset for 2D sound source localization, a comparative analysis with CRNN, ResNet-Conformer, and SRP-PHAT reveals the superior performance of the proposed ART-rSNN method. Despite the higher speed of our approach

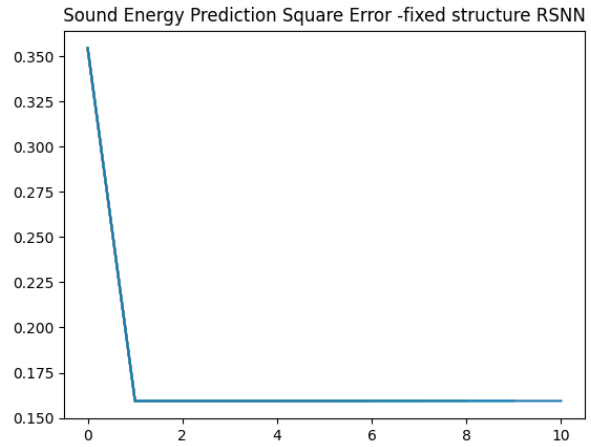

FIGURE 13. MSE of Sound Energy prediction Fixed Structure RSNN method with 100 hidden neurons. The Y-axis indicates MSE amplitude and the X-axis represents the iteration number

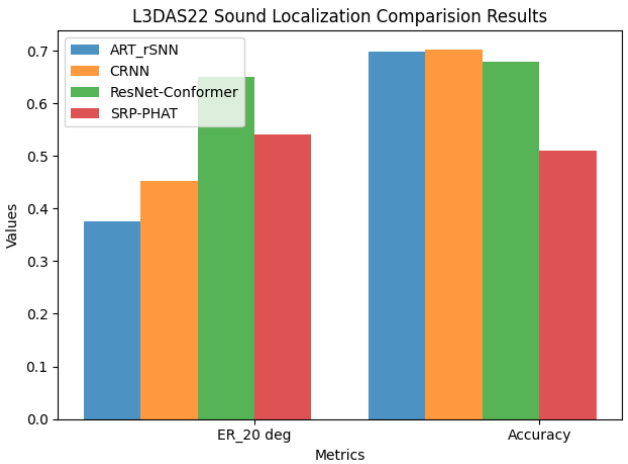

FIGURE 14. Mean Error and Accuracy Comparison on L3DAS22 in 2D Sound Source Localization Task

compared to state-of-the-art deep learning methods, our system exhibits lower mean error and nearly identical accuracy. These results underscore the efficiency of our system in achieving precise sound source localization with reduced computational demands.

While these findings demonstrate the efficacy of the proposed method, it is essential to address challenges associated with high-speed moving sound sources in real-time implementations. Future research endeavors could focus on refining the architecture to effectively handle multiple sound sources, broadening the applicability of the proposed neural network beyond single-source localization scenarios. In conclusion, the proposed ART-rSNN architecture exhibits promising capabilities, marking a significant advancement in sound source localization techniques, particularly in the context of 2D localization on the L3DAS22 dataset, where it outshines deep learning counterparts in both speed and accuracy.

VIII. APPENDIX

To prove an updating law for a supervised learning strategy for LIF neurons, we start with a cost function based on the error of energy estimation from normalized recorded sound signal power inputs and normalized membrane voltage of LIF neurons. The weights are then updated based on this cost function. The proof will involve demonstrating that the updating law leads to a decrease in the cost function over time, indicating that the network is learning to estimate the energy of the sound signal more accurately. Let's consider the cost function as follows:

$$J = \frac{1}{2} E^T E$$
$$E = \left(\tanh\left(y_{s_{N\times 1}}\right) - \tanh\left(V_{o_{N\times 1}}\right)\right)_{N\times 1} \quad (14)$$

Vo: Observed Neuron

Equation (13) describes the square error on normalized energy estimation, and the concept of energy-based methods is integrated into ITD via the formula provided in the text. However, the search results do not provide any additional information on ITD or how it is related to the cost function and energy estimation :

$$V = \exp\left(-c\Delta t_s/\tau\right) I \rightarrow \quad (15)$$

$$I = \frac{1-c}{\tau} \exp\left(-c\Delta t_s/\tau\right) V \rightarrow I = W_{ij} \exp\left(-c\Delta t_s/\tau\right) V \quad (16)$$

c: sound speed

$$\Delta t_s = \text{input spike time - neuron spike time} \quad (17)$$

where W_{ij} is the synaptic weight between neurons i and j and updated based on Spike Time Dependent Plasticity (ST DP) laws for hidden neurons: ST DP:

$$W_{ij} = \begin{cases} AV_j e^{\frac{\Delta t_{sij}}{\tau}} & \Delta t \geq 0, \quad i,j \in Ni \\ 0 & \Delta t < 0, \quad \text{or} \quad i,j \notin Ni \end{cases} \quad (18)$$

,

$$\Delta t_s = t_i - t_j$$

Where t_i and t_j are spike times of the i_{th} and j_{th} neurons, respectively. Ni is the Neighbourhood of the neurons in the small word connections, A is the maximum synaptic weight, $\Delta t = t_i - t_j$ is the spike time difference, and τ is the time constant. τ is the time constant of synaptic plasticity law.
Our proposed methodology is rooted in an energy-based framework, with a central focus on the manipulation of Leaky Integrate-and-Fire (LIF) neuron voltages, which play a pivotal role in our approach. To facilitate comprehension, we draw an analogy between the behavior of observed neurons and input-output (I/O) entities, akin to the functionality of loudspeakers. This analogy is substantiated by the inherent

resemblance between the LIF neuron model and the dual loudspeaker lumped model.

In the context of our algorithm, the behavior of the LIF neuron aligns seamlessly with the dual circuit of a loudspeaker's lumped element model. Specifically, the equation representing the lumped model of a loudspeaker captures the dynamics of the electrical circuit, reflecting the LIF neuron's ability to adjust its voltage in response to incoming signals. This logical connection between the LIF model and the dual loudspeaker lumped model forms the foundation of our energy-based framework. Neuron Voltage adjustment, akin to tuning a loudspeaker, facilitates dynamic adaptation to signals. Our energy-based law systematically enhances simulation by tuning LIF neuron parameters, crucial for optimizing the algorithm's performance.

To calculate the error boundary on the evaluated dataset, we integrate principles from the loudspeaker lumped model. This serves as a reference for analyzing spatial characteristics and quantifying the maximum error in distance estimation, providing insights into the algorithm's spatial accuracy.

$$L_c\frac{di}{dt} + Ri + Bl\cdot\frac{dx}{dt} = E(t) \tag{19}$$

,

$$V + \frac{R}{C}\frac{dV}{dt} + Bl\cdot\frac{dx}{dt} = E(t) \tag{20}$$

$$\frac{R}{C}\frac{dV}{dt} = \underbrace{-Bl\cdot\frac{dx}{dt} + E(t)}_{I} = -V + I \rightarrow LIF \text{ Model}$$

and $I \propto$ Sound Energy

$$I = \tau\frac{dV}{dt} + V \rightarrow V = k\cdot e^{\frac{\Delta t}{\tau}}\cdot I \rightarrow I = W\cdot V\cdot e^{-\frac{\Delta t}{\tau}} \tag{21}$$

$$y_{\text{sound\_Energy}} = \frac{S}{|d-d_s|^2} + \varepsilon \approx \frac{S}{|d|^2} = W\cdot V\cdot e^{-\frac{\Delta t}{\tau}}$$

Taylor Expansion:

$$\left\{e^{-\frac{\Delta t}{\tau}} = 1/\exp\left(\frac{\Delta t}{\tau}\right)\right\} \approx \frac{1}{1+\frac{\Delta t}{\tau}+\frac{1}{2}\left(\frac{\Delta t}{\tau}\right)^2} \approx \frac{1}{d^2}$$

if $\frac{1}{\tau}$ = sound wave speed = c $\rightarrow$ we expect that $c\Delta t \approx d$ in the best Δt calculation, (TDOA) Under this assumption, the Error boundary is calculated as follows:

$$|Er| : \left|d^2 - \frac{1}{2}\left(d^2+2d+2\right)\right| = \left|\frac{1}{2}\left(d^2-2d-2\right)\right| \tag{22}$$

$$= \frac{1}{2}\left|(d-1)^2 - 3\right|$$

According to our dataset, the maximum of d is 3, So:

$$|Er| \leq 0.5 \tag{23}$$

After calculating the error boundary, which was found to be 0.5 meters in our evaluation, we gained valuable insights into the spatial accuracy of our proposed algorithm on the L3DAS22 dataset. This measure signifies the maximum allowable deviation between the estimated and actual distances, providing a critical metric for assessing the reliability of our method. The demonstrated accuracy reinforces the robustness of our algorithm and its potential applicability in real-world scenarios where precise sound source localization is essential.

ACKNOWLEDGMENT

As authors, we would like to express our sincere appreciation and deepest gratitude to Auckland University of Technology for providing us with the opportunity to conduct our research.

REFERENCES

[1] L. Ngo, J. Cha, and J.-H. Han, "Deep neural network regression for automated retinal layer segmentation in optical coherence tomography images," IEEE Transactions on image processing, vol. 29, pp. 303-312, 2019.

[2] A. Zhang, W. Zhu, and J. Li, "Spiking echo state convolutional neural network for robust time series classification," IEEE Access, vol. 7, pp. 4927-4935, 2018.

[3] K. Rosero, F. Grijalva, and B. Masiero, "Sound Events Localization and Detection Using Bio-inspired Gammatone Filters and Temporal Convolutional Neural Networks," IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023.

[4] B. Rafaely, Fundamentals of Array Signal Processing. John Wiley & Sons, 2015.

[5] Y. Wu, R. Ayyalasomayajula, M. J. Bianco, D. Bharadia, and P. Gerstoft,"Sslide: Sound source localization for indoors based on deep learning," in 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021, pp. 4680–4684.

[6] L. Wang, Y. Wang, G. Wang, and J. Jia, "Near-field sound source localization using principal component analysis–multi-output support vector regression," in International Journal of Distributed Sensor Networks, vol. 16, no. 4, p. 1550147720916405, 2020.

[7] M. A. Pillai et al., "Acoustic source localization using random forest regressor," in 2019 International Symposium on Ocean Technology (SYMPOL), 2019: IEEE, pp. 191-199.

[8] J. C. Middlebrooks, "Sound localization," Handbook of Perception and Cognition, vol. 4, pp. 411–484, 1996.

[9] A. Zhang, X. Li, Y. Gao, and Y. Niu, "Event-driven intrinsic plasticity for spiking convolutional neural networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 5, pp. 1986-1995, 2021.

[10] R. O. Schmidt, Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propag., vol. 34, no. 3, pp. 276–280, 1986.

[11] Y. Xu, J. Du, L.-R. Dai, and C. Lee, "A regression approach to sound localization in reverberant environments," IEEE/ACM Trans. Audio, Speech, Language Process., vol. 25, no. 3, pp. 676–688, 2017.

[12] Y. Chen, X. Zhang, L. Li, and L. Ma, "Sound source localization based on improved support vector machine," in Proc. of the 11th World Congress on Intelligent Control and Automation, 2014, pp. 5363–5368.

[13] J. Yang, W. Chen, and Y. Zheng, "A sound source localization method based on random forests," in Proc. of the IEEE International Conference on Robotics and Biomimetics, pp. 401-406, 2015.

[14] H. Lu and Q. Zhou, "Recurrent neural networks for sound localization in both static and dynamic scenes," IEEE Access, vol. 7, pp. 36307-36317, 2019.

[15] Y. Ding and X. Bao, "A bio-inspired sound localization method using synchronous spiking patterns," IEEE Transactions on Cybernetics, vol. 51, no. 9, pp. 4251-4260, 2021.

[16] X. Chen, H. Xu, M. Lu, and Y. Zhou, "A low-complexity and energy-efficient sound source localization method based on acoustic energy decay model," IEEE Sensors Journal, vol. 21, no. 5, pp. 5459-5469, 2021.

[17] T. Zhang, Z. He, and H. Wang, "A bio-inspired spiking neural network model for sound source localization," Neurocomputing, vol. 363, pp. 39-49, 2019.

[18] Z. Zhang, J. Gao, Y. Ma, and S. Liu, "A bio-inspired auditory system with attention-based sound localization for robot audition," IEEE Transactions on Cognitive and Developmental Systems, vol. 12, no. 3, pp. 343-355, 2020.

[19] S. A. Saeedinia, M. R. Jahed-Motlagh, A. Tafakhori, and N. Kasabov, "Design of MRI structured spiking neural networks and learning algorithms for personalized modelling, analysis, and prediction of EEG signals," Scientific Reports, vol. 11, no. 1, p. 12064, 2021.
[20] K. Voutsas and J. Adamy, "A biologically inspired spiking neural network for sound source lateralization," IEEE Transactions on Neural Networks, vol. 18, no. 6, pp. 1785-1799, Nov. 2007.
[21] B. Glackin, J. A. Wall, T. M. McGinnity, L. P. Maguire, and L. J. McDaid, "A spiking neural network model of the medial superior olive using spike timing dependent plasticity for sound localization," Frontiers in Computational Neuroscience, vol. 4, p. 18, 2010.
[22] Z. Pan, M. Zhang, J. Wu, J. Wang, and H. Li, "Multi-tone phase coding of interaural time difference for sound source localization with spiking neural networks," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 2656-2670, 2021.
[23] S. D. Correia, S. Tomic, and M. Beko, "A feed-forward neural network approach for energy-based acoustic source localization," Journal of Sensor and Actuator Networks, vol. 10, no. 2, p. 29, 2021.
[24] S. Adavanne, "Sound event localization, detection, and tracking by deep neural networks," 2020.
[25] E. C. Escudero, F. P. Peña, R. P. Vicente, A. Jimenez-Fernandez, G. J. Moreno, and A. Morgado-Estevez, "Real-time neuro-inspired sound source localization and tracking architecture applied to a robotic platform," Neurocomputing, vol. 283, pp. 129-139, 2018.
[26] W. Zhang and P. Li, "Composing Recurrent Spiking Neural Networks using Locally-Recurrent Motifs and Risk-Mitigating Architectural Optimization," arXiv preprint arXiv:2108.01793, 2021.
[27] G. Bellec et al., "A solution to the learning dilemma for recurrent networks of spiking neurons," Nature communications, vol. 11, no. 1, p. 3625, 2020.
[28] K. Roy, A. Jaiswal, and P. Panda, "Towards spike-based machine intelligence with neuromorphic computing," Nature, vol. 575, no. 7784, pp. 607-617, 2019.
[29] F. Denk, S. D. Ewert, and B. Kollmeier, "On the limitations of sound localization with hearing devices," The Journal of the Acoustical Society of America, vol. 146, no. 3, pp. 1732-1744, 2019.
[30] F. Grondin and J. Glass, "SVD-PHAT: A fast sound source localization method," in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019: IEEE, pp. 4140-4144.
[31] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," Wireless Communications and Mobile Computing, vol. 2017, 2017.
[32] M. U. Liaquat, H. S. Munawar, A. Rahman, Z. Qadir, A. Z. Kouzani, and M. P. Mahmud, "Localization of sound sources: A systematic review," Energies, vol. 14, no. 13, p. 3910, 2021.
[33] A. Zhang, Y. Niu, Y. Gao, J. Wu, and Z. Gao, "Second-order information bottleneck based spiking neural networks for sEMG recognition," Information Sciences, vol. 585, pp. 543-558, 2022.
[34] A. Zhang, Y. Han, Y. Niu, Y. Gao, Z. Chen, and K. Zhao, "Self-evolutionary neuron model for fast-response spiking neural networks," IEEE Transactions on Cognitive and Developmental Systems, vol. 14, no. 4, pp. 1766-1777, 2021.
[35] A. Zhang, J. Shi, J. Wu, Y. Zhou, and W. Yu, "Low Latency and Sparse Computing Spiking Neural Networks With Self-Driven Adaptive Threshold Plasticity," IEEE Transactions on Neural Networks and Learning Systems, 2023.
[36] Y. Chen, Y. Mai, R. Feng, and J. Xiao, "An adaptive threshold mechanism for accurate and efficient deep spiking convolutional neural networks," Neurocomputing, vol. 469, pp. 189-197, 2022.
[37] H. Chen and W. Ser, "Acoustic source localization using LS-SVMs without calibration of microphone arrays," in 2009 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1863-1866, 2009.
[38] T.-H. Tan, Y.-T. Lin, Y.-L. Chang, and M. Alkhaleefah, "Sound source localization using a convolutional neural network and regression model," Sensors, vol. 21, no. 23, p. 8031, 2021.
[39] N. Yalta, K. Nakadai, and T. Ogata, "Sound source localization using deep learning models," Journal of Robotics and Mechatronics, vol. 29, no. 1, pp. 37-48, 2017.

ZAHRA ROOZBEHI has a degree in Ba Honar Kerman University and an M.Sc. degree from Alzahra University. She is a Ph.D. student researching the spatio-temporal behavior of the brain. Her expertise includes data analysis, mathematics, neuroscience, and research tool implementation. She enjoys generating new ideas and finding solutions. Her colleagues describe her as motivated and resourceful. She is currently working on mimicking the mammalian hearing system and applying the model in technology. Her interests include data analytics, computational neuroscience, brain-inspired modeling, fractal geometry, and spiking neural networks.

AJIT NARAYANAN received his B.Sc. (Hons.) degree in communication science and linguistics from the University of Aston, Birmingham, U.K. in 1973, and his Ph.D. degree in philosophy from the University of Exeter, Philosophy, Exeter, U.K. in 1976. He is presently a Professor at the Auckland University of Technology, School of Engineering, Computer Science, and Mathematics. He has been a lecturer, professor, and dean at universities in the United Kingdom before arriving in New Zealand in 2007. He has published over 100 articles and conducted reviews for various journals and conferences on artificial intelligence and its applications in the medical field. His research interests include artificial intelligence, nature-inspired computing, machine learning, computational statistics, and machine ethics.

MAHSA MCCAULEY is a Senior Lecturer and Director of Women in Tech in AUT's School of Computer, Engineering, and Mathematical Sciences. She is a well-recognised leader in AI and machine learning. She is also the founder of the charitable trust She Sharp, a women's technology networking and learning group, where she works to encourage young New Zealand girls to consider what a career in technology offers. She was named the Emerging Leader category winner in the 2013 Westpac Women of Influence Awards and was one of ten finalists for the 2018 Kiwibank New Zealander of the Year. In 2019 she was the Champion Award winner of the YWCA Equal Pay awards, and in 2020 presented the Massey University Distinguished Alumni Award.

SAMANEH-ALSADAT SAEEDINIA earned a Bachelor's degree from Imam Khomeini International University and a Master's degree from Iran University of Science and Technology. Her areas of interest are intelligent control and modeling, neuroscience, and neural computations. She won third prize in the Kharazmi Scientific Competition in Tehran Province in 2006. In 2018, she published a book on electrical instruments. Currently, she is pursuing a Ph.D. at IUST.