Transactions of the Indian National Academy of Engineering (2023) 8:535–550
https://doi.org/10.1007/s41403-023-00424-9

REVIEW ARTICLE

# Review of Emergency Vehicle Detection Techniques by Acoustic Signals

Krishnendu Choudhury[1,2] · Dalia Nandi[1]

Received: 13 March 2023 / Accepted: 31 August 2023 / Published online: 15 September 2023


## Abstract
The performance of emergency vehicle prioritization system is determined by its efficiency in reducing the response time of the vehicles detected. Emergency vehicle (EV) detection is possible through various methods, but the use of the acoustic signal in detection method is found to have an edge over the others. The present study reviewed various acoustic-based EV detection systems and their merits, which would provide a better understanding of techniques to be used for IoT system design with limited power and computational resources. It is found that EV siren detection accuracy decreases in low SNR conditions and is affected by the choice of features used in neural network (NN). It is observed that neural network-based system performance is better compared to other methods. Also, it is observed that network parameters of long short-term memory recurrent NN architecture are almost 150 time less as compared to other NN for similar detection accuracy. More research on acoustic-based systems is required to be done, for achieving high detection accuracy of 99% in low SNR condition of − 15 dB or below. Developing a universally deployable generalized model of a neural network in detecting EV siren and designing a system of low power with low computation requirements are the main goals which are analyzed in the present study. In detecting EV by acoustic-based method, the effects of noise, various signal domain features, environment, relative movement of source and detector, etc. have been studied here. The physical characteristics of the siren signal are analyzed and how these can be used in emergency vehicle detection systems are discussed. The present study analyzes the acoustic-based EV detection system into three major categories, namely digital signal processing-based systems, neural network-based systems, and statistical methods-based systems. For all the methods discussed in these categories, the research gaps are identified to indicate future research directions. Further, major challenges and future scope in the acoustic-based EV detection system are presented. Also, the new direction to increase the accuracy and decrease the latency of the EV detection system is discussed.

**Keywords** Emergency vehicle · Neural network · Wail and yelp · Acoustic signal detection · Background noise · Far-field detection

### Abbreviations

| | |
|---|---|
| ASR | Automatic speech recognition |
| PCEN | Per-channel energy normalization |
| MAP | Maximum a posteriori |
| CNN | Convolutional neural networks |
| K-NN | K-nearest neighbor |
| ANN | Artificial neural network |
| BRANN | Bayesian regularized artificial neural network. |
| GMM | Gaussian mixture models |
| SVM | Support vector machine |
| NN | Neural network |
| FFT | Fast Fourier transform |
| MDF | Module difference function |
| ADSR | Attack–decay–sustain–release |
| µPa | Micro-Pascal |
| SED | Sound event detection |
| ILD | Interaural level difference |
| ITD | Interaural time difference |
| EVP | Emergency vehicle priority |
| AV | Autonomous vehicle |
| EV | Emergency vehicle |

✉ Krishnendu Choudhury
krishnendu814u@gmail.com

Dalia Nandi
dalia@iiitkalyani.ac.in

1 Electronics & IT, Nehru Science Centre Mumbai, Dr E Moses Road Worli, Mumbai, Maharashtra 400018, India

2 Electronics and Communication, Indian Institute of Information Technology Kalyani, Kalyani Webel IT Park, Kalyani, India

| | |
|---|---|
| LCS | Longest common subsequence |
| LSTM-RNN | Long short-term memory recurrent neural network |
| MFCC | Mel-frequency cepstral coefficients |

## Introduction

The present path of automotive technology is evolving to bring full autonomy to vehicles in steps, which is called six levels of autonomy (Marchegiani and Newman 2022). In practical conditions, whether a manually driven car or an autonomous vehicle (AV) system, the challenge of detecting emergency vehicles beyond visual range under varied noisy environments will almost persist in the same nature. AVs are supposed to be mounted with various sensors to detect objects and other parameters in their environment to yield EVs (Sun et al. 2021). But manual drive vehicles do not have many sensory systems, for sound detection, message viewing, or image/ video viewing. To achieve accuracy, these sensor data are required to be analyzed by intelligent systems/machines in vehicles or through a distributed system like a wireless sensor network (Wang et al. 2017). In the current trend of AV, a vehicle may achieve complete autonomy if it has three systems viz. a visual system, an audio system and messaging system, and they are required to work together. Visual systems and audio systems are required for prompt reaction or reflex-based decisions in a situation, like, a sudden obstacle has come in the path of a vehicle. In the visual system, there has been a frequent referencing of a few challenges which are inevitable in a public road-like environment. The usual problem in a vision system is vision gets obliterated due to occlusion by other objects. The other problem is of lighting conditions that affect the camera to sense the objects within its field of vision. However, audio signals in a traffic environment are not restricted from reaching a sensor by any condition of the direct line of sight, nor they are affected by illumination conditions. This property of sound is very useful and, hence, is used widely in emergency services.

Emergency services are fire services, law and order services by police, emergency medical services by hospitals, etc. All of these services are provided by emergency vehicles which use lights and sirens to distinctly notify their presence in the vicinity of the traffic system. It is required that the services of the system shall have a low response time. For this, when an EV is heading to its destination, it shall receive a free corridor by prioritizing its call to cross the traffic signal. Many innovative solutions for EV priority are proposed, which can be categorized into five major types of emergency vehicle priority (EVP) systems based on their basic method of detection as shown in Table 1 (Mehendale et al. 2021). All of these EVP systems have certain stages, namely detection, extraction of information, analyzing information acquired, and then initiating the priority strategy. This priority system has the involved risk of collisions of an EV with other vehicles or pedestrians at the intersection (Mehendale et al. 2021; Lorenzo and Eilers 1991; Howard et al. 2011; Montoro et al. 2017). This responsibility can be imposed on the automatic detection systems, which are yet not accurate, and also on the auditory and visual sensory systems of detection in human beings. In object detection mechanism, vision or visual-based system accounts for 90% of input, in the traffic environment (Howard et al. 2011). Although emergency vehicles are distinct and equipped with visible warning devices, there are interferences that compete in drawing away attention. EV warning lights often get blocked when a big vehicle comes in front of the EV; thus, it becomes difficult for other vehicles to detect an EV approaching from the rear (Sun et al. 2021). In contrast to this, the sound of the siren is perceivable much before the EV is visible. It is observed that a visual- based EV detection system cannot alone provide accurate detection of EVs but the system can also improve manifold when an auditory detection mechanism is also used. An audio/sound-based system can provide a sense of an EV vehicle approaching from a particular direction and speed. Therefore, this feature of the auditory system can help motorists and pedestrians to detect the approaching emergency vehicle, thereby avoiding colliding with it. It is assumed pedestrians and drivers understand their responsibilities to yield to emergency vehicles.

### Different Approaches to Detect EV

There are various approaches to EV detection, one example solution to the said problem is using an image processing method. Here, to detect an EV, high-quality cameras and a high computational power processing unit are used, but the cost of the system can bother. Further, if an IoT-based application is to be implemented, the camera-based system with high resolution cannot be used, as it has a large power requirement. Considering various aspects, it comes out that an acoustic-based EVP system can serve better. However, to identify accurately an EV siren sound, acoustic-based systems have challenges in a random environment with background noise. But the other systems for EV detection (Mehendale et al. 2021) have demerits relatively more in comparison to sound-based detection systems which are pointed in Table 1

The motivation of the present literature is to cover all the different methods and challenges in acoustic-based approaches for detecting EVs. There is literature available on EV detection, but literature that sums up all the methods of EV siren detection is limited. In this study, we are presenting research done in understanding siren signal characteristics of the recommended frequency range of 1–4 kHz. This range is

**Table 1** EV detection methods other than acoustic system and their limitations

| SL no. | Detection category | Technology used | Additional gadget on vehicle | Limitation |
|---|---|---|---|---|
| 1 | Image processing based | Camera, Lidar | NO | a. Occlusion of EV by vehicle makes it difficult for recognizing the presence of an EV<br>b. Lighting conditions affect the accurate detection of objects. Hence, a camera-based system may fail to recognize EV |
| 2 | Line-of sight based | Infrared, RFID, Inductive-loop ultrasonic sensor | YES | a. Transmitter signal gets obstructed if an object is in between Transmitter and Receiver. Hence information about an EV may not reach all the nearby vehicles and receivers of the traffic controller<br>b. Maneuvering an EV near the sensor is a critical requirement |
| 3 | Wireless comm- unication based | Vanet, Zigbee | YES | a. Requirement of authentication at traffic signal intersection will take time and can cause response delay in the heavy traffic environment<br>b. As it is required to enter the network of the traffic controller, the network is at risk of hacking |
| 4 | GPS based | GPS module, GSM, Xbee | YES | a. Inaccuracy in positioning occurs when EV position is determined by only three visible satellites<br>b. Signal gets obstructed by large objects and are also affected due to atmospheric condition. Thus, there can be inaccuracy in positioning<br>c. In these systems, batteries drain out fast, which is not suitable for IoT applications |
| 5 | Node based | WSN, Timed Petri Nets, CAPRI, MAS, LQF-MWM Q-learning | YES | a. A node or network is vulnerable to hacking<br>b. Designed for low-speed application<br>c. Local storage capacity of a node is low. All the information are stored at some centralized location of the network. Hence, every time the central system is to be accessed for authentication or for any information related to the route to the destination for an EV. This will impact response time of system |

consistent with the peak sensitivity of human hearing (Lorenzo and Eilers 1991; Howard et al. 2011), it can be heard by pedestrians and people inside cars, and sensed by the roadside unit (RSU) sensors of the traffic management system. A microphone is utilized in a roadside unit (RSU), where the electrical characteristics of the magnetic material within the microphone play a vital role in achieving optimal performance. Among these characteristics, magnetic permeability stands out as a key factor, determining the material's capability to generate a magnetic field when subjected to an electric current or vice versa. A higher magnetic permeability allows for superior conversion between electrical and audio signals. Notably, cobalt ferrite is a type of ferrite material known for its intriguing magnetic properties, with permeability being one of its important attributes (Srinivasamurthy et al. 2018).

This range of 1–4 kHz is chosen due to the fact that high frequencies are not localizable, and sound energy below 1 kHz is wasted (Lorenzo and Eilers 1991; The Human Ear Hearing, Sound Intensity and Loudness Levels 2022). A siren sound detection in the free field is easier but in real traffic environments, challenges of detection are many, an understanding of which is essential while designing the system. For example, nowadays electromechanical sirens are largely being replaced by electronic devices. These pure electronic-based sirens' sounds may not achieve optimal power spectrum and, hence, will cause limitation of sound penetration. To be effective, the siren signal will have to compete with the masking noise generated on the road, the noise inside cars by radios, fans, etc., and problems due to modern sound insulation techniques(Howard et al. 2011; Angela 2013). Aforesaid and some other challenges are to be discussed in this study. This study also discusses and presents subcategories of acoustic-based systems. In this method, research has been done using various approaches like digital signal processing, analog electronics, neural network-based approach, and statistical methods for achieving an accurate detection of a siren sound. Some are listed in Table 2 with accuracy of detection; however, detection

**Table 2** Accuracy achieved in EV detection methods and the signal feature used

| Sl no. | Title of research article | year | Detection method | Signal feature used | Accuracy (%) |
|---|---|---|---|---|---|
| 1 | An automatic emergency signal recognition system for the hearing impaired | 2006 | Artificial neutral network | MFCC(Mel-frequency cepstral coefficients) | 99.0 |
| 2 | Recognition of the ambulance siren sound in Taiwan by the longest common subsequence | 2013 | Longest common sequence | Pitch frequency variation | 85.0 |
| 3 | Detection of alarm sounds in noisy environments | 2017 | Support Vector Machine | Pitch, zero crossing rate, short time energy, spectral flux, spectral roll-off, spectral centroid, spectral flatness, MFCC | 98.0 |
| 4 | Detection of ambulance and fire truck siren sounds using neural networks | 2018 | Long short-term memory recurrent neural network | MFCC | 93.8 |
| 5 | Emergency signal classification for the hearing impaired using multi-channel convolutional neural network architecture | 2019 | Multi-channel convolutional neural networks | Mel-Spectrogram | 88.2 |
| 6 | Detection of ambulance siren in traffic | 2019 | Bayesian regu- larized artificial neural network (BRANN) | MFCC, spectral centroid, spectral-flux, spectral roll-off and zero crossing rate | 99.0 |
| 7 | Acoustic-based emergency vehicle detection using convolutional neural networks | 2020 | CNN | Raw data, MFCC and log-Mel spectrogram | 98.2 |
| 9 | Parameter tuning for wavelet-based sound event detection using neural networks | 2021 | ANN | Wavelet transform | 97.0% |
| 10 | Emergency vehicles audio detection and localization in autonomous driving | 2021 | CNN and regression | MFCC, log-Mel spectrogram and raw data | 89.6 |

accuracy for the value of measured SNR is not known for all the methods.

## Siren Signal of EV

To detect siren sound accurately, signal pattern of the EV siren should be properly analyzed. EV types are, ambulances, police van, or fire trucks; they all use siren sounds that are distinctly different in frequency and pattern from one another as per the regulation of a country. In Taiwan, for an ambulance the frequency of the first tone used is 650-750Hz and the second tone is 900- 1000Hz, while in Japan two tones 960 Hz and 770 Hz are repeated at every 1.3s. Likewise, different countries have different sets of frequencies and patterns in use for sirens (Tran and Tsai 2020). These sounds appear to be similar in type and are noisy, but each is for different purposes and situations (Tran and Tsai 2020; What are the Different Sounds a Police Siren Makes 2022). Good knowledge of the physical characteristics of siren sound can provide different clues to segregate it from background noises as used in sinusoidal model-based systems of siren detection (Tran and Tsai 2020) and to analyze a sound signal.

### Physical Characteristics of Siren Signal

An ambulance has different siren modes of which “wail” and the “yelp” are the two mostly used. The wail signal has a frequency sweep of 5.8 sec between 600 Hz and 1200 Hz. While yelp has a frequency sweep of 0.3 sec between the frequency 600 Hz and 1100 Hz. Figure 1 shows the spectrogram of siren modes (Montoro et al. 2017; Angela 2013). The favorable characteristics of siren signals include a sufficiently loud and wide frequency spectrum (1–4 kHz) to overcome ‘masking’ noise, a rapid rise in pitch, and rapid cycling time. (Howard et al. 2011; Montoro et al. 2017).

### Use of Siren Tone

In detecting EV, frequency domain analysis of acquired signal is performed to determine the presence or absence of siren tone. In this application using the Fast Fourier Transform (FFT), the fundamental frequency and harmonics of the acquired sound signal are extracted. To detect

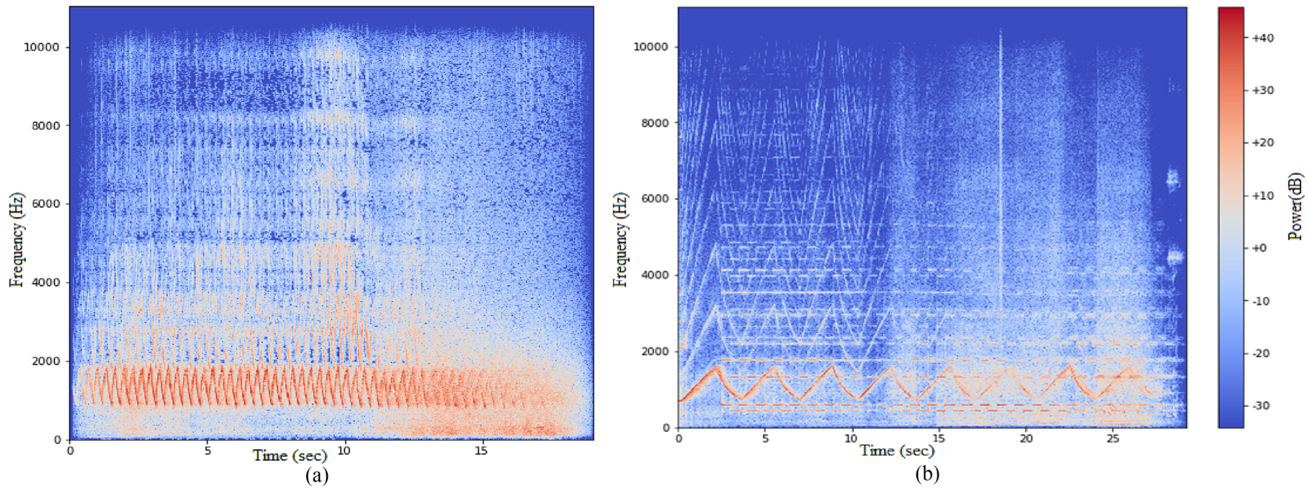


**Fig. 1** **a** Spectrogram of yelp **b** spectrogram of wail

the signal's fundamental frequency and its harmonics with more precision compared to the FFT, the popular method ARYule may be used (Montoro et al. 2017; Angione et al. 2016). As the source or receiver can come relatively closer or move farther away from each other, a shift in the apparent frequency of the siren sound can be observed due to the Doppler effect. This shift of frequency causes problems during signal analysis to determine the presence or absence of a siren tone. But it can certainly provide an idea of the direction and location of the EV siren.

### EV Localization and Its Arrival Direction

EV localization and direction of arrival is an important requirement in traffic management systems. Finding out the location of an EV in a lane can provide some "lead time" to adjust the traffic flow. This is helpful for the smooth passage of vehicles without any abrupt hitch in the traffic flow. To determine the location or arrival direction of EV, researchers have shown the use of methods like interaural time difference (ITD) (Sun et al. 2021; Shabtai and Tzirke 2019), the Doppler effect (Montoro et al. 2017), and the interaural level difference (ILD) (Montoro et al. 2017; Scola and Ortega 2010). The process used in ITD is the time difference in the arrival of a sound wave from a source to the two microphones, placed at fixed separation as shown in Fig. 2 (Scola and Ortega 2010). A sound signal has a finite velocity, so the time of arrival of emitted sound wave to a location will depend on the path traveled by sound wave to reach a distinct location. As for the case depicted in Fig. 2, the sound wave at mic2 will arrive 'Td 'time later in comparison to mic1. The time arrival differences can be calculated using the cross-correlation between signals arriving at two microphones. Once the time difference of arrival is found, the angular position of the sound source with respect to the center reference line can be computed by simple trigonometric calculation. If the angle value is positive, the source is on the right side and for the value negative, the source is on the left side. ITD method is good at low frequencies, i.e., below 1.5 kHz; in this range, time difference between two receivers is perceivable but at the higher frequencies, it is not. At the same time, ILD is perceivable at higher frequencies better in comparison to low frequencies. Hence, these together can provide cues for effectively localizing a sound source. Further detailed analyses are available in Scola and Ortega (2010). However, it has a requirement, i.e., from a source, two signals arriving at two receivers over different paths shall be noise free (Marchegiani and Newman 2022).

### EV Classification

EV services are of three broad types and their priority also varies. Hence, to meet emergency services within a response time, identifying the type of EV is very important. EV siren emits a sound that has a defined frequency and pattern. In order to classify EVs, spectral analysis of siren sounds using FFT, and an algorithm to detect the differences in cyclical patterns of the siren can be implemented (Tran et al. 2018). Hence, by fixing the frequency range for each different EV service type, EVs can be differentiated and this can provide intelligence to the system in prioritizing the EVs. (Tran and Tsai 2020; Ebizuka et al. 2019).

## Challenges of Acoustic-Based Systems

In acoustic-based event detection, it is essential to determine the time position of the occurring events in the sound received. This is useful in improving the performance of a

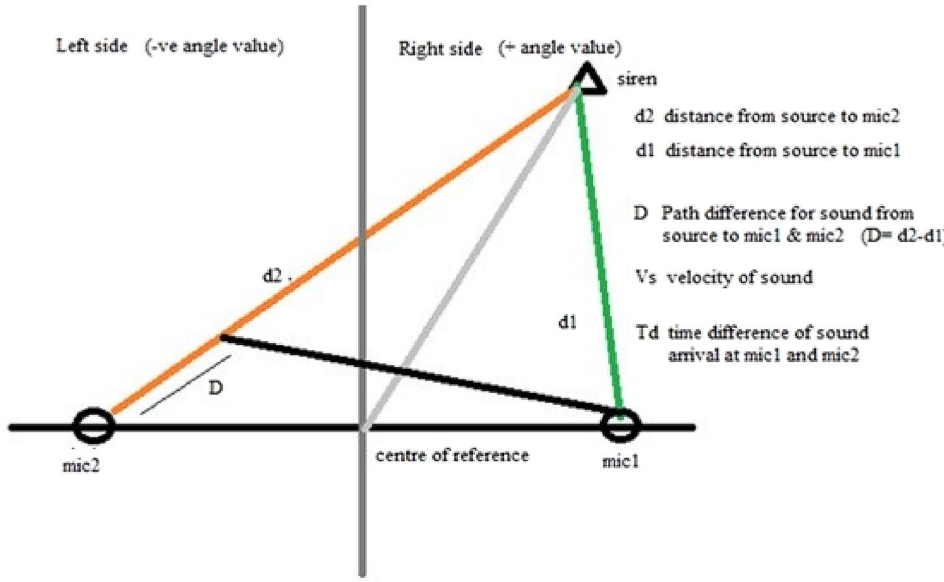


**Fig. 2** Diagram depicting microphone location and method for locating source siren

neural network-based system, as only those sound segments are to be analyzed, instead of analyzing the entire segment of the received sound. This method is like keyword spotting and can be employed for the detection of non-speech sounds like EV siren sounds in the automatic event detection system. We have a good understanding of siren sound patterns but still, the sound event detection (SED) system has multiple challenges that need to be addressed as discussed below.

## Far-Field Sound Event Detection Problems

EV siren sounds are detected by the transducers like the microphone. The performance of the sound detection system degrades as the distance between the source and receiver keeps increasing (Wang et al. 2017; Mesaros et al. 2021). Because when the siren sound wave-front propagates through a medium it is affected by spreading, absorption, ground reflection, and so on. In a real environment, the sound source can be located at various distances from the microphone, resulting in the acquisition of signals with different intensities and signal-to-noise ratio. The intensity of the audio event acquired by the microphone is attenuated due to various factors some of which are as given below (Howard et al. 2011; Montoro et al. 2017).

a. The sound level in the far-field setup is reduced by 6 dB for each doubling of distance from the source.
b. Due to the atmospheric absorption during the propagation of the sound wave.
c. The sound reflected by the ground surface interfering with the sound propagating directly from the source.
d. Due to the presence of barriers.

## Background Noise Problem

Acoustic-based siren signal detection has an advantage over other methods of EV detection. But in a real-world environment, background noise impacts the performance of the system in the extraction and detection of the siren signals (Sun et al. 2021). These background noises are a mix of multiple sounds of cars, animals, wind, birds, construction site noise, park background noise, crowd noise, car horns, fountain sounds, rain noise, schoolyard noise, etc. (Fatimah et al. 2020). These are varying acoustic qualities and have a non-stationary behavior. The open environment can have any combination of these background noises which changes SNR; thus, it has the propensity to mask desired siren signal (Mesaros et al. 2021). Knowing the masking threshold, the signal level is kept at 6–10 dB above the masked thresholds, to ensure 100% detectability of EV siren (Howard et al. 2011). Also due to the complexity of these individual noises, the a priori model of background noise is difficult in a road environment. To reduce the impact of background noise, one out of many solutions is to first identify and extract unique acoustic events in the input audio file, then extract various properties of these events (Mehendale et al. 2021). Another typical problem associated with audio analysis systems is that it is related to the duration of the sound events of interest which are of varying lengths. Hence, uncertainty in occurrence and uncertainty of the valid length of acquisition time makes its analysis complex.

## Vehicle in Motion Causes a Problem Due to Doppler Shift in Frequency

A relative motion between the source and the receiver creates an apparent shift in the frequency emitted by the source, i.e., frequency of sound at the receiver either goes up or down than the actual transmitted frequency (Montoro et al. 2017). This predominant phenomenon is due to the Doppler Effect, which varies the received frequency up to 1.178 times the emitted frequency of an ambulance, moving at a speed of 100 km/h (Miyazakia et al. 2013). This shift in transmitted frequency can cause the siren sound frequency to move in the noise frequency range of the urban environment (Sun et al. 2021). Addressing this issue is essential for improving the accuracy of the system. A shift in the siren's sound frequency up to +45 Hertz is considerable, which is sufficient for a speed difference of 80 km/h between the emergency vehicle and the car.

### Soundproof Vehicle

Modern cars cabin is designed to insulate their inside environment from outside traffic noise. But this luxury muffs up the sound of siren's, so it becomes difficult to hear any alarm sound within a car (Howard et al. 2011; Fatimah et al. 2020). It has been estimated that closed cars can attenuate sound up to 30 dB; hence for siren sound to be audible, it has to be maintained at the sound pressure level of 72 dB relative to the reference value of 20micro Pascal ($\mu$Pa) (Howard et al. 2011). Also from prior research, it is observed that this insulation impacts the audibility distance of siren sound. It is seen that a siren sound source moving at a speed of 60 km/hr. and having an intensity of 115 dB, is heard only within a distance of 100 meters from a moving car. When the air conditioning is turned on and the radio is at a minimum audibility level, inside a vehicle, the siren's sound audibility rapidly gets reduced to a distance of 50 meters from the car (Palecek and Cerny 2016).

### Hearing-Impaired People are Allowed to Drive in Most of the World

Sound events are distinct form of alerting sensation to Ears. And it is useful for person driving vehicle or inside a vehicle to maneuver or control. With improvement in hearing aid, people with hearing impairment are now being allowed to drive vehicle. But, it has been found that drivers do not use their hearing aids, as because these gadgets amplify the background noise of the traffic environment when heard, it becomes too loud and distracting (Carmel et al. 2017; Rudzyn et al. 2007). Hence, it is necessary to provide a visual indication to them for any EV en route in the direction of the vehicle (Bubar et al. 2020). The process or methods covered under the study has been majorly impacted by the problems/ challenges mentioned above. Researchers have tried to overcome these challenges through their solutions which are broadly described in the subsequent sections.

## Different EV Detection Methods Based on Acoustic Signal

In a traffic environment, there are multiple types of sound/ acoustic signals differing in intensity and duration. Every acoustic signal has time domain and frequency domain features that can be extracted and are distinctively recognizable. The EV siren tone is having similarity to the musical tone. Hence time domain analysis of music are applicable to siren tone. In a siren tone, in the beginning, there is often a sudden increase in energy; this short phase is called the attack phase of the tone. After the attack phase, the sound of the tone stabilizes called decay phase and reaches a steady phase with almost a periodic pattern. This is the third phase, also called the sustain phase, which makes up most of the duration of a tone. Here, the energy almost remains constant or slightly decreases. In the final phase of the tone, also called the release phase, the musical tone fades away (Müller 2015).

The envelope of the sound waveforms for two different musical instruments Piano and Violin are plotted in Fig. 3 for comparison. It has been observed that the waveforms are distinctly different from each other although they are playing at the same frequency. Comparing the attack phase of a piano with the violin, it is clear that the piano has a shorter duration of the attack phase, i.e., it reaches its peak intensity more rapidly compared to the violin. In piano, the decay phase of the waveform envelope is relatively shorter, and its intensity decreases continually over a longer duration of time which is its sustain phase. But in violin, there is no decay phase, the envelope of waveform directly moves to sustain phase and has an oscillating pattern. In the release phase, envelope of the waveform of the piano follows the same amplitude gradient as in sustain phase and is of shorter duration compared to the violin which is comparatively steeper. Thus, time domain analysis of the envelope of the waveform can be used to differentiate between different acoustic signals generated by different sources.

This sound envelope, i.e., attack–decay–sustain–release (ADSR) model is used for distinguishing between two sound sources. Similarly, there are different features of Audio signals, which are categorized at the level of abstraction, temporal scope, signal domain, etc. In the signal domain, we use time domain features like amplitude energy, RMS energy, ZCR, etc. (Müller 2015). For any particular acoustic signal detection, it is required to understand the signal's both time and frequency features and it has to be interpreted properly in a traffic environment. In a real-world environment, different sound sources are present, emitting sounds of varied tempos and intensities. Extracting the desired siren signal, i.e., wail, yelp, or

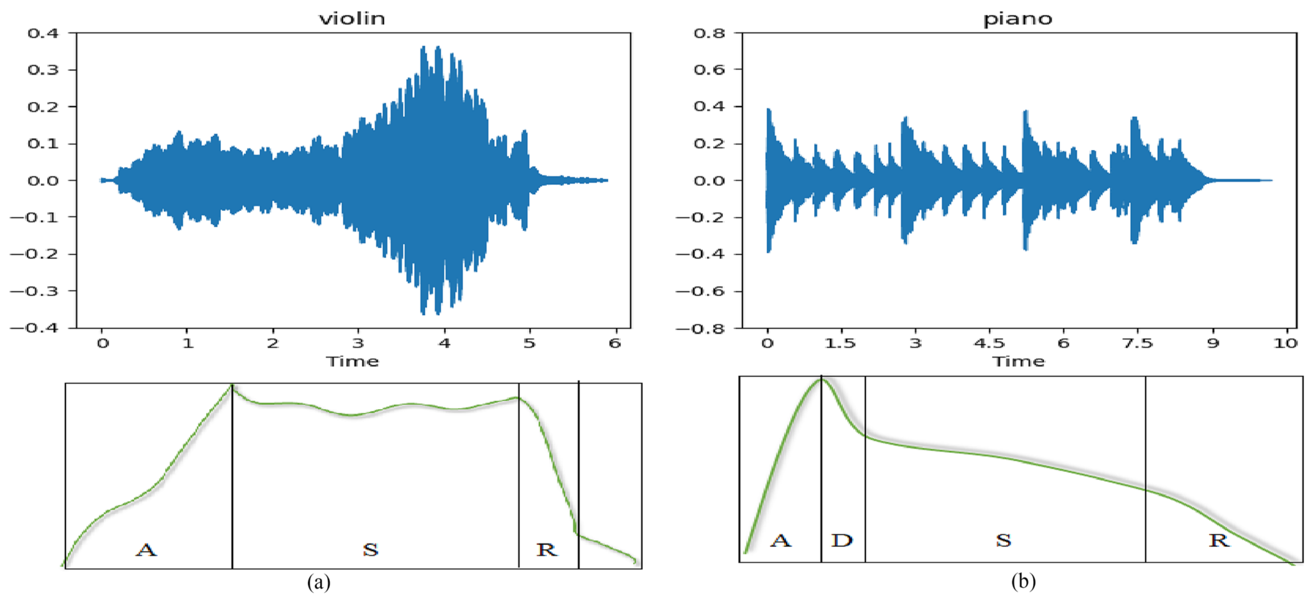


**Fig. 3** ADSR phase of **a** violin and **b** piano

Hi–Lo sound in the presence of other undesired sounds is a complex task. In a siren sound detection system, a sound is detected through transducer-like microphones; here, the signal is received as an electrical signal. On the received signal, audio processing techniques/ methods are applied to extract various features that can be used in detecting EV. Researchers have used various techniques for detecting EV siren sounds and they can be categorized broadly as follows.

## Digital Signal Processing-Based System

The two qualities that are distinctive of a typically pure siren signal is the frequency content and the periodic repetition in it. In the siren signal, there are a number of spectral components of which the lowest is the dominant frequency, i.e., fundamental frequency. Other spectral components may have harmonic or in-harmonic relation to the fundamental. Here, the task is to detect the dominant frequency of the signal. In this work, the spectral analysis gives insight into the presence or absence of a siren signal by detecting the dominant frequency. For this, methods like STFT or FFT are used. But due to the presence of many harmonics as a significant component, these methods can fail in the detection of the siren signal. There are other methods/tools, used alone or in combination in digital signal processing-based system implementation, like adaptive filter, autocorrelation, convolution, spectrogram, MFCC, wavelet, MDF, etc.

Siren signals are sudden sound events, characterized as an instantaneous change in the signal. To capture these changes, wavelets are found to be extremely useful and efficient as they are localized in nature and, therefore, can be used to detect sharp transitions or contrasts in a signal (Rane et al. 2019). It is verified that the siren sound in a real field environment is accompanied by sources of undesired background noise, like Brownian noise. Hence. to segregate the foreground event, i.e., siren signal, one may have a requirement to scale the signal gain by an autoregressive filter which is then followed by the application of dynamic range compression (Ick and McFee 2021). This amplifies the foreground sound and suppresses background noise. Research has shown that the spectrum of natural sound is large; hence, to detect alarm sound, it is proposed that the pitch frequency of the under-test signal is to be evaluated and then contrasted with the pre-characterized alarm frequency (Ick and McFee 2021). Also, finding the periodicity in a sound signal using autocorrelation and then classifying the signal by an ML-based classifier to detect a siren sound has also been used. A method like frequency matching to find the longest basic sequence to detect an alarm or siren signal has also been researched (Palecek and Cerny 2016).

In other methods, like pitch-based detection of a siren, the audio signal is sampled at some sample rate. The samples that are obtained are divided into a number of windows of equal length/size called fragments. For each window, the pitch is determined using MDF (module difference function). For this, ‘L’ lag is determined as similar to the autocorrelation method. L is chosen where MDF is minimum and the relation between pitch, sampling rate, and lag is used for estimating the pitch of the received signal. A list of pitches is determined over the total samples considered with its probability values which are then considered for the decision of the presence of an emergency signal (Jonnadula and Khilar 2021; Meucci et al. 2008). MDF is mentioned to be less computationally costly than autocorrelation. A system implemented using this method can face design complexity issues when for high accuracy with high sampling rate is required and as well desired signal has to be acquired for a larger duration. Some of the other works are tabulated below in Table 3 to show the method of detection in DSP-based acoustic systems and the gaps/new approaches identified.

## Neural Network-Based Systems

These systems are implemented using standalone or a combination of neural networks like RNN, CNN, BRANN, MLP, SVM, etc. A neural network-based system has five major stages. This begins with the signal acquisition, and next is pre-processing of this acquired signal, from this pre-processed signal, the feature is extracted. Then comes employing methods to select features which are contributing to the classification task. Then, the selection of features is used for training and then the model is generated for siren detection.

The first stage of the system is to acquire a sound signal using audio sensors or a transducer like a microphone to detect the presence of an abnormal sound. Here, to increase the accuracy and decrease the cost of computation, a band-pass filter is used, which restricts the frequency of the acquired signal between 500 and 2000 Hz frequency (Angione et al. 2016; Fatimah et al. 2020; Dobre et al. 2017). In the next step, the signal may be pre-processed to reduce the noise. The next stage is for the classification task which has sub-stages (Marchegiani and Posner 2017), like computing the pre-processed signals features. However, for the task at hand, all the features shall not be relevant to the architecture of the neural network (NN) selected. Hence, to reduce the dimensionality, a statistical method like the Kruskal–Wallis test may be used to select those features which contribute to the classification task (Fatimah et al. 2020). The classification problem is usually a binary problem, where the two classes to be classified are siren sound and non-siren sound (Sun et al. 2021; Fatimah et al. 2020).

For an audio detection process in neural network (NN)-based method like use of long short-term memory recurrent neural network (LSTM RNN) (Tran et al. 2018; Lilja et al. 2017), usually, two approaches have been followed. One approach is the use of a classifier trained with the signal feature (Sun et al. 2021; Padhy et al. 2019; Schröder et al. 2013). In this approach, one may employ a method to extract useful features of a signal in the time domain and /or in the frequency domain. It is understood that aggregated features of signals are beneficial for acoustic-based EV detection. Thus, combinations or variations of features are used to train artificial neural networks as a classifier or used in shallow learning algorithms, like support vector machine (SVM) (Marchegiani and Newman 2022; Sun et al. 2021; Mehendale et al. 2021; Carmel et al. 2017; Raval and Christopher 2021), Gaussian mixture models (GMM) (Rudzyn et al. 2007), and k-nearest neighbors (K-NN) (Fatimah et al. 2020; Rudzyn et al. 2007). After the classifier is trained, it is used in the real field for testing an unknown set of signal features to detect the presence or absence of a siren signal. After satisfactory test performance, the classifier is used as a model for the implementation of the detector system (Sun et al. 2021; Carmel et al. 2017; Schröder et al. 2013).

The other approach is the use of a deep neural network. In this approach, the network architecture uses the raw data of the signal waveform instead of any extracted feature. This network is like CNN, which learns the feature from raw data automatically while being trained with the data set (Tran and Tsai 2020; Chavdar et al. 2020). And after the training, it is used to detect the siren signal.

The neural network-based approach is highly dependent on data quality as well as quantity for better accuracy. Also, for training a neural network, labeled data are required. In acoustic-based event detection systems, these data are acquired under various scenarios such as in different traffic locations, weather conditions, sirens of different types, overlapped sirens, different levels of noise, their collection distances, and in the Doppler effect. This wide variation results in the non-uniformity of data, hence labeling the data becomes a difficult task. For a NN to achieve accuracy, the dataset should be rich and cover almost all the real environment scenarios. By the use of ensemble architecture, the accuracy of the NN-based system can be increased (Tran and Tsai 2020). And in achieving this, same data are fed to two or more different neural network architectures, and the output of all these NN is used for inferring the final output or decision. It has been observed researchers for the implementation of acoustic event detection have found features like MFCC, Log-Mel spectrogram, etc. from the sample of the audio signal. The entire signal is divided into short window frames and for each frame, the feature is then extracted or computed. Features extracted for example are ZCR, energy, entropy, spectral centroid, spectral spread spectral entropy, etc. extracted. The extracted features can be in both the time and frequency domains to train NN like Bayesian regularized artificial neural network (BRANN) (Rane et al. 2019). Shown in Table 4 are a few works in the field of EV detection using NN.

## Statistical Methods-Based System

To detect a sound event, a system requires a lot of data. These data quantity are huge, hence to make any intelligent and accurate inference out of the data is difficult. But statistical feature can be obtained from these data which can reduce

**Table 3** DSP-based acoustic system implementation and gaps/new approach identified

| Ref | Title and year | Method of detection | New approach/gaps |
|---|---|---|---|
| Shabtai and Tzirkel (2019) | Detecting the direction of emergency vehicle sirens with microphones. 2019 | Microphone array | Methods need to be explored for effectively and accurately determining whether the objects are moving or stationary on the front, side, and rear of vehicles, using the Doppler effect |
| Fatimah et al. (2020) | An automatic siren detection algorithm using the Fourier decomposition method and MFCC | FDM and the use of Kurtosis energy and variance for each sub band | Accuracy achieved for what value of SNR is not mentioned. Hence, system performance under background noise needs to be tested |
| Palecek and Cerny (2016) | Emergency horn detection using embedded systems | Spectral center of gravity | How to receive the good quality of siren sound under bad weather conditions |
| Meucci et al. (2008) | A real-time siren detector to improve the safety of guides in the traffic environment. 2008 | Use of module difference function for pitch detection | The method needs to be explored for determining the siren sound, by taking an advantage of the long duration of the siren signal as compared to other sound sources present in the traffic environment |
| Supreeth et al. (2020) | Identification of ambulance siren sound and analysis of the signal using statistical method. 2020 | Sound device module, fast Fourier transform (FFT), and Normalization | Method required to distinguish between music and siren when likely to have almost same frequency |
| Kiran and Supriya (2017) | Siren detection and driver assistance using modified minimum mean square error method. 2017 | Fast Fourier transform (FFT) | It is required to develop a circuit system or algorithm which can differentiate between short- and long-duration signals, even in noisy environment |
| Dobre et al. (2017) | Improved low computational method for siren detection | Adjusting the resonant frequency of the band-pass filter | The gap of the proposed solution is that it must use two branches for processing the signal: one specific to the wail signal and another for the yelp signal |

**Table 4** NN-based acoustic system and gaps identified

| Ref. | Title and year | Method of detection | New approach/gaps |
|---|---|---|---|
| Marchegiani and Newman (2022) | Listening for Sirens: Locating and Classifying Acoustic Alarms in City Scenes. 2018 | Convolutional neural net- work (CNN) | a. Acoustic event detection performance to be tested when more than one event occurred<br>b. Performance of event detection system in the presence of other noise sources to be tested<br>c. Semantic segmentation is a critical requirement for denoising<br>d. How to increase classification rate beyond 94% to be checked |
| Sun et al. (2021) | Emergency Vehicles Audio Detection and Localization in Autonomous Driving 2021 | Convolutional neural networks (CNN) | a. Need for a system in which the recall rate is the same at any distance of the source from the AV<br>b. This method takes 1.5 s of audio signals at a 48 k sampling rate from every channel. How to achieve high performance at a lower sampling rate and reduce the complexity of data computation to be seen<br>c. How to increase the model performance of source detection beyond 70 m |
| Tran and Tsai (2020) | Acoustic-Based Emergency Vehicle Detection Using Convolutional Neural Networks 2020 | Convolutional neural networks (CNN) | a. Accurate detection of EV sound is one requirement in ITMS for determining the priority of EV. But, how this same network will perform to give three outputs simultaneously, i.e., siren detected, location of EV, and direction of arrival of EV to be tested<br>b. In IoT applications, low- power design is a critical requirement. Thus roadside units are designed with limited hardware resources for low computational needs. As here, it is proposed to use 2 CNN blocks, how a large computational resource requirement to be handled in an IOT scenario, needs to explore |
| Rane et al. (2019) | Detection of Ambulance Siren in Traffic, 2019 | Bayesian regularized artificial neural network. (BRANN) | Neural networks can be trained for more sources of sound events recorded in Indian traffic conditions. And neural networks have to experiment for their increase in accuracy of detecting target signal in Indian conditions. Along with traffic sounds, other traffic noises like horns, and engine sounds can be relegated and their contributions can be measured showing BRANN training to have the lowest MSE |
| Jonnadula and Khilar (2021) | Comparison of Various techniques for Emergency Vehicle Detection using audio Processing. 2021 | Artificial neural network (ANN) | Depth of hidden layer vs. accuracy improvement need to be tested |

**Table 4** (continued)

| Ref. | Title and year | Method of detection | New approach/gaps |
|---|---|---|---|
| Marchegiani and Posner (2017) | Leveraging the Urban Soundscape: Auditory Perception for Smart Vehicles 2017 | K-nearest neighbor (K-NN) | a. Need for development of place dependent soundscape models<br>b. Multimodal detection and recognition methods to be developed |
| Padhy et al. (2019) | Emergency signal Classification for the Hearing Impaired using Multichannel Convolutional Neural Network Architecture, 2019 | Multi-channel convolutional neural networks (CNN) | Use of recurrent convolutional neural networks and binary classification task into a multi-class classification for predicting the type of the emergency signal (police siren, ambulance siren, fire alarm, etc.) |

the task of handling huge data and thereby reduces system complexity and cost. These systems are implemented using alone or a combination of minimum mean square error, maximum a posteriori (MAP) adaptation, maximum likelihood linear regression, per-channel energy normalization (PCEN), spectral centroid, etc. In the statistical approach, the spectral and temporal feature of the segment is extracted. And from the spectral and temporal features so obtained, statistical parameters like mean, standard deviation, covariance, etc. are determined for the segmented frames of signal. These are then used as models for the neural network to be trained for classification or used with a mathematical approach, i.e., rule based to detect a siren signal. A few research works in this field are mentioned in Table 5 along with the identified gap.

## Future Scope in Acoustic-Based Signal Detection

Over four decades researches are being carried on finding out effective methods to detect EV siren sound. But their scope of implementation in a real environment is still awaiting due to accuracy and performance limitations. This may be due to the fact these methods do not encompass all real-world situations. Some of those are discussed previously in this study, which may affect the accuracy and performance. For example, the need of considering the background noise model or the effect of frequency shift, etc. Reviewing previous research works in this field, some future scopes are identified as given below.

a. For the detection of the siren tone, it is required to segregate the frequency of the siren sound. A system may be implemented which uses a bank of adaptive filters, and the central frequency of the filter may adapt based on the in-harmonic coefficient of the signal. To implement this, a calibration system to adjust the filters' center frequency to the desired frequency is required. Research is required to develop an ANN model which will provide the coefficient of in-harmonic to which the filters center frequency will be shifted.
b. Siren signal has time domain features as mentioned below. It is to be examined to see if these may be used in an NN-based system to determine the presence of a siren signal and its modes.

1 Gradient of the siren signal during rise and fall across the line at half of the peak value of the signal. The pattern for wails and yelp siren sounds can be distinguished using these features from other signals in the traffic environment.
2 Time elapsed between two successive peaks of the yelp or wail pattern in a siren signal.

**Table 5** Statistical method-based acoustic system and gaps identified

| Ref No | Title and year | Method of detection | New approach/gaps |
|---|---|---|---|
| Fatimah et al. (2020) | An automatic siren detection algorithm using Fourier decomposition method and MFCC. 2020 | An automatic siren detection algorithm using Fourier decomposition method and MFCC. 2020 | a. Least number of feature with higher accuracy for practical realizable real-time application need to be achieved<br>b. The accuracy achieved by the system for what value of SNR has not been mentioned |
| Palecek and Cerny (2016) | Emergency horn detection using embedded systems. 2016 | Spectral centroid, correlation | a. There is a requirement of the designing transducer/microphone to work in bad weather conditions as well to reduce the impact of reflected sounds on the transducer<br>b. Properly selection of frame length |
| Supreeth et al. (2020) | Identification of ambulance siren sound and analysis of the signal using statistical method. 2020 | Short-time Fourier transform and statistical measure | In filtering method, amplitude lesser than 0.8 V is discarded. This may remove the desired signal frequency. When a siren signal is far off, the SNR may become very low |
| Park and Trevino (2013) | Automatic Detection of Emergency Vehicles for Hearing Impaired Drivers. 2013 | Measuring the means and variances of reflection coefficients and using linear predictive coding | The algorithm is not fully proved and may result in false alarms. Hence, how to handle false alarm or to minimize it need to be checked |
| Dubin and Zietz (2004) | Personal Alerting Technology for people with Hearing Difficulty: Pilot Study. 2004 | Mahalanobis distance for classification of alert signals | Efficiency of the system in low SNR requires to be examined |

3 The number of rise and fall changes of the siren signal over a fixed length of time after a SED.

c. Effect on siren tone causes shifting of the original signal at the receiver. Thus, accuracy in the detection of a sound signal decreases. Hence, a method of reducing the impact of the Doppler effect has to be explored or researched.

d. Automatic speech recognition (ASR) technique can be utilized for siren detection also. These techniques perform well in relatively clean conditions, but their robustness remains a major challenge. Experiments are to be carried for finding the use of ASR techniques that will make the system robust not only to the various kinds of noise interference but also to the varying loudness (or sound level)

e. Sound models of a traffic signal are different for different geographical regions. Hence, a sound model for traffic signals in Indian scenarios may be developed and has a scope of research. This will help to train the neural network for which field data collection may not be required for any new NN architecture training.

f. Siren signals are to be acquired and processed to determine their presence. For this, the use of multiple microphones simultaneously can improve the detection accuracy by enhancing the target signal.

**New directions:** The authors aim is to increase the overall accuracy of the system and to reduce the EV response time by the application of IoT, machine learning and 5G technology. However foremost requirements are detection of EV presence, its type, and location for prioritizing EV in a traffic route. In acoustic-based EV detection systems, non-stationary noises and the harmonics of the siren signal can create problem. The use of adaptive line enhancer (Aviva Atkins February 2020) can improve the desired signal by reducing non-stationary noise and harmonic noise from an audio signal. There is scope of optimizing the filter based on choice of filter parameter, i.e., its order 'M' and learning rate 'm'.

In case of siren detection, only a few frequencies are to be detected, so employing the Goertzel algorithm is much more convenient than radix 2 DFT (When to Use the Goertzel Algorithm Instead of the FFT 2022; The Goertzel Algorithm 2022). Due to the Doppler effect as the target frequency is shifted, Goertzel module may not always detect the presence of a siren signal accurately. So, the detection system can use Goertzel filter banks in parallel, where each bank is having different parametric values (sample rate, block size) with the aim to capture the Target frequency with some positive or negative deviations. Thus, it is possible to get a pattern of magnitude which may be used to predict the presence or absence of target frequency with better accuracy. The data obtained from the Goertzel filter bank can be given to a trained logistic regression classifier to predict the presence or absence of target frequency, which can determine EV in the Traffic and send this information immediately to the ITMS using IoT infrastructure. Once the information of EV detection is received, ITMS can broadcast it to all the vehicles in the lane.

## Conclusions

The present study reviewed various EV detection methods but the main focus is on the acoustic signal-based system as it has several advantages like object detection features beyond visual range. The analysis of siren tone signal characteristics is made for EV detection, localization as well as for the classification. The acoustic signals-based detection system has many challenges, like far-field SED, background noise, Doppler shift in frequency, presence of soundproof vehicles, and hearing-impaired drivers which are discussed in the present study. It has been observed from the present study that, to detect sudden event occurrence, a good frequency resolution has to be achieved for detecting low-frequency signals and a good time resolution has to be achieved for high-frequency signals. The acoustic-based detection system is categorized into digital signal processing system, statistical system, and neural network system. The gaps/scope for further works are listed for all the EV detection methods. For every detection system, it is noticed that data acquisition method is the same, and noise suppression are to be done at a prior.

The use of signal processing methods like FFT and filtering techniques to detect the presence of siren signals are demonstrated here. Increasing the accuracy and lowering the detection time is still desired, and the methods like segmenting the total received signal, use of spectral center of gravity can be explored. Further to increase accuracy, the detection systems also can use raw data and feature sets in combination to train the neural network.

The EV detection systems implemented so far are shown to have used a pure digital signal processing method, a pure neural network-based method, and /or a combination of both. There are some implementations where statistical features and rule-based methods are also used for the classification of siren signals. Overall, the current scope of research revolves around employing signal processing techniques to enhance received signal quality and then use of neural networks or deep learning for the classification of received sound signals. The system's accuracy can be improved by fine-tuning the network architecture and enhancing model generalization through large datasets from various regions. Moreover, reducing non-stationary background noises by implementing ALE filters, followed by the use of Goertzel filter banks in the received signals, can further enhance detection accuracy.

Due to the fact that the availability of proper datasets is very limited to train a model, sufficient diversity of the training data is expected, which may lead to good generalization of the trained models. It is shown in a few of the research studies that a model with more hidden layers in NN increases accuracy in detection. Thus, more research on the aforementioned topics needs to be carried out for accurate and cost-effective system development.

**Author Contributions** Conceptualization and writing original draft: KC; Supervision, review and editing: DN. Authors reviewed the final manuscript. The author(s) read and approved the final manuscript.

**Funding** No funding received.

**Availability of Data and Materials** The datasets, figure used during and/or analyzed during the current study are available from the open source research paper of author available.

## Declarations

**Conflict of Interest** The authors declare that they have no competing interests.

**Ethical Approval and Consent to Participate** Not applicable.

**Consent for Publication** Not applicable.

## References

Angela DP (2013) Emergency vehicle siren noise effectiveness. Master Thesis. University of Windsor Ontario, Canada.

Angione F, Novak C, Imeson C, Lehman A, Merwin B, Pagliarella T, Samardzic N, D'Angela P, Ule H (2016) Study of a low-frequency emergency siren in comparison to traditional siren technology. Proc Mtgs Acoust 29(1):030008

Aviva Atkins February (2020) Speech enhancement based on adaptive line enhancer. Master Thesis. Israel Institute of Technology

Bubar A, Eckstein B, Ell A, Hilts E, Martin S, Powell T, Miguel Cruz A, Rios Rincon A (2020) Emergency siren detection technology and hearing impairment: a systematized literature review. Disabil Rehabil 18:295

Carmel D, Yeshurun A, Moshe Y (2017) Detection of alarm sounds in noisy environments. In: IEEE 25th European Signal Processing Conference (EUSIPCO), pp 1839–1843

Chavdar M, Gerazov B, Ivanovski Z, Kartalov T (2020) Towards a system for automatic traffic sound event detection. In: IEEE 28th Telecommunications Forum (TELFOR), pp 1–4

De Lorenzo RA, Eilers MA (1991) Lights and siren: a review of emergency vehicle warning systems. Ann Emerg Med 20(12):1331–1335

Dobre RA, Negrescu C, Stanomir D (2017) Improved low computational method for siren detection. In: IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME), pp 318–323

Dubin S, Zietz S (2004) Personal alerting technology for people with hearing difficulty: pilot study. In: IEEE Proceedings of the 30th Annual Northeast Bioengineering Conference, pp 202–203

Ebizuka Y, Kato S, Itami M 2019 Detecting approach of emergency vehicles using siren sound processing. In: IEEE Intelligent Transportation Systems Conference (ITSC), pp 4431–4436

Fatimah, B, Preethi A, Hrushikesh V, Singh A and Kotion H.R 2020 An automatic siren detection algorithm using Fourier decomposition method and MFCC. In: IEEE 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp 1–6

Howard CQ, Maddern AJ, Privopoulos EP (2011) Acoustic characteristics for effective ambulance sirens. Acoust Aust 39(2):43–53

Ick C, McFee B 2021 June. Sound event detection in urban audio with single and multi-rate PCEN. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 880–884

Jonnadula EP, Khilar PM (2021) Comparison of various techniques for emergency vehicle detection using audio processing. In: Gupta BB (ed) Cloud security. CRC Press, Boca Raton, pp 64–75

Kiran SL, Supriya M (2017) Siren detection and driver assistance using modified minimum mean square error method. In: IEEE International Conference on Smart Technologies for Smart Nation (SmartTechCon), pp 127–131

Lilja AP, Raboshchuk G, Nadeu C (2017) A neural network approach for automatic detection of acoustic alarms. In: Biosignals, pp 84–91

Marchegiani L, Posner I (2017) Leveraging the urban soundscape: Auditory perception for smart vehicles. In: IEEE International Conference on Robotics and Automation (ICRA), pp 6547–6554

Marchegiani L, Newman P (2022) listening for sirens: Locating and classifying acoustic alarms in city scenes. IEEE Trans Intell Transp Syst 23(10):17087–17096

Mehendale N, Patel R, Mange S, Mulik S (2021) Review of emergency vehicle priority systems. SSRN. https://doi.org/10.2139/ssrn.3859424

Mesaros A, Heittola T, Virtanen T, Plumbley MD (2021) Sound event detection: a tutorial. IEEE Signal Process Mag 38(5):67–83

Meucci F, Pierucci L, Del Re E. Lastrucci L and Desii P (2008) A real-time siren detector to improve safety of guide in traffic environment. In: IEEE 16th European Signal Processing Conference, pp 1–5

Miyazakia T, Kitazonoa Y, Shimakawab M (2013) Ambulance siren detector using FFT on dsPIC. In: Proceedings of the 1st IEEE/IIAE International Conference on Intelligent Systems and Image Processing, pp 266–269

Montoro L, Drastrup JK, Johan M (2017) Automatic detection of emergency cars. Master Thesis, Aalborg University.

Müller M (2015) Music Representations, Fundamentals of music processing: Audio, analysis, algorithms, applications. Springer (**Audio Representation**)

Padhy S, Tiwari J, Rathore S, Kumar N (2019) Emergency signal classification for the hearing impaired using multi-channel convolutional neural network architecture. In: IEEE Conference on Information and Communication Technology, pp 1–6

Palecek J, Cerny M (2016) Emergency horn detection using embedded systems. In: IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI), pp 257–261

Park SW, Trevino J (2013) Automatic detection of emergency vehicles for hearing impaired drivers. Texas a&m Univ-Kingsv EE/CS Dep MSC 192:637–665

Rane D, Shirodkar P, Panigrahi T, Mini S (2019) Detection of ambulance siren in traffic. In: IEEE International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), pp 401–405

Raval P, Christopher J (2021) Parameter tuning for wavelet-based sound event detection using neural networks. International conference on computational intelligence in music, sound, art and design (Part of EvoStar). Springer, Cham, pp 235–247

Rudzyn B, Kadous W, Sammut C (2007) Real time robot audition system incorporating both 3D sound source localization and voice characterization. In: Proceedings of IEEE International Conference on Robotics and Automation, pp 4733–4738

Schröder J, Goetze S, Grützmacher V, Anemüller J (2013) Automatic acoustic siren detection in traffic noise by part-based models. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp 493–497

Scola CF, Ortega MDB (2010) Direction of arrival estimation: a two microphones approach. Master Thesis. Blekinge Institute of Technology.

Shabtai NR, Tzirke E (2019) Detecting the direction of emergency vehicle sirens with microphones. In: EAA Spatial Audio Signal Processing Symposium, pp 137–142

Srinivasamurthy KM, Angadi VJ, Kubrin SP, Matteppanavar S, Kumar PM, Rudraswamy B (2018) Evidence of enhanced ferromagnetic nature and hyperfine interaction studies of Ce-Sm doped Co-Ni ferrite nanoparticles for microphone applications. Ceram Int 44(15):18878–18885

Sun H, Liu X, Xu K, Miao J, Luo Q (2021) Emergency vehicles audio detection and localization in autonomous driving. arXiv preprint arXiv:2109.14797

Supreeth H V, Rao S, Chethan KS, Purushotham U (2020) Identification of Ambulance Siren sound and Analysis of the signal using statistical method. In: IEEE International Conference on Intelligent Engineering and Management (ICIEM), pp 198–202

The Goertzel Algorithm (2022). https://courses.cs.washington.edu/courses/cse466/12au/calendar/Goertzel-EETimes.pdf. Accessed 06 Jan 2022

The Human Ear Hearing, Sound Intensity and Loudness Levels (2022). https://courses.physics.illinois.edu/phys406/sp2017/Lecture_Notes/P406POM_Lecture_Notes/P406POM_Lect5.pdf. Accessed 03 March 2022

TranYVT, Tsai YC, Ho W (2018) Detection of ambulance and fire truck siren sounds using neural networks. In: Proceedings of 51st Research World International Conference, pp 9–14

Tran VT, Tsai WH (2020) Acoustic-based emergency vehicle detection using convolutional neural networks. IEEE Access 8:75702–75713

Wang Y, Getreuer P, Hughes T, Lyon RF, Saurous RA (2017) Trainable frontend for robust and far-field keyword spotting. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 5670–5674

'What are the Different Sounds a Police Siren Makes (2022). https://www.extremetacticaldynamics.com/blog/what-are-the-different-sounds-a-policesiren-makes/. Accessed 10 April 2022

When to Use the Goertzel Algorithm Instead of the FFT? (2022).https://medium.com/geekculture/the-great-unknown-the-goertzel-algorithm-492681f30866. Accessed 02 Feb 2022

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.