Design and Implementation of Environmental Warning Systems for Drivers with Hearing Impairments

Tasripan, Hendra Kusuma*, and Alfian Nur Rafli Huzaini

ABSTRACT

Hearing impairment encompasses conditions where individuals experience difficulties in hearing, categorized into deafness and hard of hearing. The employment opportunities provided by online taxi companies in Indonesia, particularly for motor or car drivers, offer a sense of independence to those with hearing impairments. However, driving necessitates environmental awareness, especially of auditory signals. Hence, there’s a need for a surrounding environment warning system to convert crucial sounds into visual alerts, focusing on sounds like sirens and railway crossing warnings. This study introduces a warning system for drivers with hearing impairments, utilizing Recurrent Neural Network and Long Short Term Memory (RNN-LSTM) methods, implemented on a Raspberry Pi. The system, comprising a Raspberry Pi 4, microphone, LCD TFT, and LEDs, processes captured siren sounds into text alerts on the LCD and blinking LED signals. Testing the system with different sound recognition durations—2, 3, and 4 seconds—yielded accuracies of 78%, 82%, and 91%, respectively. Results indicate that prediction accuracy is influenced by the duration of sound recognition. Future research could explore enhancements with higher RAM Raspberry Pi 4 and more sensitive microphones to reduce noise.

Keywords: Hearing impairment, Raspberry Pi, RNN-LSTM, Sound recognition.

1. INTRODUCTION

According to the World Federation of the Deaf (WFD) in 2019 [1], there are 70 million deaf individuals globally, yet only 2% have access to education through sign language. In Indonesia, annually, over 5000 babies are born with hearing loss, with approximately 4 in every 1000 Indonesians being deaf, and around 40 million experiencing some form of hearing impairment. Employment opportunities in sectors like transportation, provided by some online taxi companies in Indonesia, have notably eased job seeking for the deaf, particularly for car drivers. However, driving requires environmental awareness, especially auditory sensitivity.

This research develops a warning system for drivers with hearing impairments, aimed at assisting deaf drivers in detecting ambient sounds such as ambulance sirens, vehicle horns, or train sounds, enabling them to be aware of their surroundings through the device.

2. LITERATURE REVIEW

Recurrent Neural Networks (RNNs) are designed to handle sequential data, overcoming the limitations of traditional neural networks by learning from temporal sequences. Through recurrent processing, where the network’s output serves as its subsequent input, RNNs excel in tasks that require understanding temporal context, such as speech recognition and language translation. They are particularly effective in audio signal classification, recognizing patterns over time in sound recordings to differentiate between speech and music. By analyzing audio frames sequentially, RNNs extract key features like frequency and amplitude, enabling accurate audio classification by capturing the essence of the temporal context in audio signals.

While Recurrent Neural Networks (RNNs) effectively process sequential data, such as speech, their capability to retain information is generally limited to short-term dependencies. To address scenarios requiring the analysis
of both short-term and long-term dependencies within longer data sequences, Long Short-Term Memory (LSTM) networks, a variant of RNNs, are employed. LSTMs are engineered to preserve information over extended periods through a complex structure consisting of four inter-connected layers: the cell state, forget gate, input gate, and output gate as shown in Fig. 1. The cell state carries information across hidden layers to subsequent LSTM layers, maintaining the network’s memory. The forget gate, equipped with a sigmoid activation function, decides which information to retain or discard from the cell state. The input gate utilizes two neural networks—one with a sigmoid activation and another with tanh activation—to select and process new input data for inclusion in the cell state. Lastly, the output gate generates the final output vector. This intricate arrangement allows LSTMs to learn and remember long-term dependencies in input sequences, making them a powerful tool for speech classification tasks and other applications requiring the analysis of extensive sequential data.

3. RELATED WORKS

Scarpiniti et al. [3] developed a deep Recurrent Neural Network (DRNN) leveraging Long Short-Term Memory (LSTM) cells for classifying audio recordings from construction sites. Their DRNN model processes a variety of spectral features, including Mel-Frequency Cepstral Coefficients (MFCCs), Mel-scaled spectrograms, chroma features, and spectral contrast, achieving a notable accuracy of 97% on their testing dataset, outperforming alternative approaches.

Yu et al. [4] introduced a Bidirectional Recurrent Neural Network (Bi-RNN) enhanced with an attention mechanism, exploring both serial and parallel attention models for analyzing Short-Time Fourier Transform (STFT) spectrograms. Their findings showed that the parallel attention approach significantly improved performance, surpassing the serial model in effectiveness.

Gan [5] proposed an innovative RNN framework incorporating a channel attention mechanism specifically for music feature classification. Combining Gated Recurrent Units (GRUs) and Bidirectional LSTM (Bi-LSTM) with an attention system, this model assigns differential attention across various segments of the RNN output. This strategy enables a more refined understanding of musical characteristics, leading to a classification accuracy of 93.1% on the GTZAN dataset and an AUC score of 92.3% on the MagnaTagATune dataset, indicating superior performance compared to other evaluated models.

4. METHODOLOGY

This system is an integration of both hardware components and software solutions. On the hardware side, the setup is designed to capture audio signals which are then processed by a Raspberry Pi 4. The processed information is visually represented through text alerts and corresponding LED light indications, based on the sound’s classification. The key hardware elements include a Raspberry Pi 4, a microphone for sound capture, an LCD TFT screen for displaying text alerts, and LED lights for visual signals. Software-wise, the system employs the RNN-LSTM method to analyze the audio data collected by the microphone. The outcomes of this analysis are then presented on the LCD TFT screen. A comprehensive view of the system’s architecture and workflow is detailed in the block diagram depicted in Fig. 2.

The microphone captures sound which is then relayed to the Raspberry Pi 4, serving as the central processing unit and controller of the system. The Raspberry Pi 4 processes audio data pre-trained by an RNN LSTM model during a prior training phase. Once processed, sounds are categorized into three distinct classifications: Traffic, Ambulance, and Fire Truck. These classifications are then displayed on a TFT LCD screen directly connected to the Raspberry Pi 4. In addition to visual alerts on the screen, the system also activates blinking LED lights, offering an immediate visual warning to the driver, enhancing situational awareness and safety.

4.1. Hardware Design

In this phase, a detailed design and plan will be developed, outlining the implementation strategy for the system.
and specifying the necessary components. The system’s architecture will incorporate a Raspberry Pi 4, a TFT LCD, and a microphone, as illustrated in Fig. 2. The TFT LCD will be directly connected to the Raspberry Pi 4, while the microphone will be integrated into the system via a USB port. This setup aims to ensure seamless communication between the hardware components, facilitating the effective capture, processing, and display of audio signals as visual alerts.

The components outlined in Fig. 3 are arranged and designed as depicted in Fig. 4. This design illustrates the
interconnections and layout of the system’s hardware, showcasing how the Raspberry Pi 4, microphone, TFT LCD, and LED lights are integrated to function cohesively for the intended application.

4.2. Software Design

The software design in this study encompasses the process of capturing incoming audio inputs, which are subsequently analyzed and transformed into visual outputs displayed on the TFT LCD screen. We utilized the Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) architecture for this purpose.

LSTM, an advancement over traditional RNN, is particularly adept at retaining information over extended periods, making it well-suited for tasks requiring memory of long-term dependencies. In this research, the LSTM model was trained to identify and classify three specific classes. The detailed architecture and source code of the LSTM model implemented in this study are illustrated in Fig. 5, showcasing the framework’s intricacies and how it processes the sound data to achieve the desired classification.

4.3. The Dataset

The dataset pivotal to this research comprises audio samples each with a duration of three seconds. This uniform duration is crucial for the feature extraction process, ensuring consistency across all samples. The dataset serves as a foundational element in data preprocessing and feature extraction stages, subsequently facilitating the training and classification of distinct audio characteristics. The objective of this classification is to accurately identify and
distinguish between different sounds, specifically tailored to provide auditory warnings for individuals with hearing impairments.

This dataset includes three categories of sounds: Traffic, Ambulance, and Fire Truck, aimed at alerting the deaf. A significant portion of these audio samples was recorded directly using the speaker device, while others were sourced from various online platforms. Each collected audio sample is stored in a dedicated folder in .wav format. The choice of .wav format ensures the retention of high-quality audio without the loss associated with compression, making it ideal for preserving the integrity of sound data for the process of Audio Signal Feature Extraction.

4.4. Sound Classification Learning Process

The learning process for sound recognition, or training phase, is designed to develop a predictive model that functions as the core intelligence of the system on the Raspberry Pi 4. This model processes audio inputs captured by a microphone and classifies them into one of three predefined categories. These categories encompass the sounds of an ambulance siren, a fire truck siren, and general road traffic noise. The ultimate goal is for this model to accurately identify and categorize these sounds, enabling the system to provide timely and relevant alerts based on the environmental audio cues detected. A block diagram shown in Fig. 6 detailing this end-to-end learning process illustrates how the trained model will analyze and interpret sound inputs on the Raspberry Pi 4, turning raw audio into actionable classifications.

5. Result and Discussion

This section presents the outcomes of implementing, testing, and evaluating the devices developed for this research. It covers the tool’s realization, test findings, and an analysis of these results. The focus is on assessing the system’s practical performance, identifying its strengths and areas for improvement, and evaluating its effectiveness in aiding drivers with hearing impairments.

5.1. Device Implementation

The system device can be powered by the vehicle’s electrical system or alternatively by a power bank. The setup, featuring a Raspberry Pi 4 and a TFT LCD connected directly, along with a microphone attached via USB, has been successfully implemented. For optimal visibility, the device should be positioned within the car where the driver can easily see it, ensuring immediate awareness of any alerts issued. The recommended placement of the device within the vehicle is illustrated in the image in Fig. 7.

5.2. Device Testing

The evaluation of this device involved testing with siren sounds sourced from both the YouTube platform and live highway sirens. Data samples were collected 100 times, segmented by duration variables of 2, 3, and 4 seconds, to assess the device’s recognition capabilities under different conditions. The test samples were categorized into four groups: ambulance sirens, fire truck sirens, other assorted sirens, and road traffic noises. Specifically, the testing protocol included 30 samples each for ambulance and fire truck sirens, and 40 samples for traffic sounds, which comprised 20 miscellaneous siren sounds and 20 instances of road noise.

In the tests conducted for the 2-second duration, the results were as follows: For the ambulance siren sound, there were 23 accurate predictions and 7 inaccurate ones, resulting in a success rate of 77%. Similarly, the fire truck siren sound testing also yielded 23 correct predictions and 7 inaccuracies, achieving a 77% accuracy rate. The road traffic sound tests showed a better performance with 26 correct predictions and only 4 incorrect ones, leading to an
87% accuracy rate. The distribution of these percentages is visually represented in the histogram shown in Fig. 8.

In the conducted tests for the 3-second and 4-second time variables, the following results were observed: For the 3-second duration, the ambulance siren tests yielded 22 correct and 8 incorrect predictions (73% accuracy), fire siren tests had 24 correct and 6 incorrect predictions (80% accuracy), and road traffic sound tests resulted in
28 correct and 2 incorrect predictions (93% accuracy). Meanwhile, for the 4-second duration, ambulance siren sound tests produced 25 correct and 5 incorrect predictions (83% accuracy), fire truck siren sound tests showed 27 correct and 3 incorrect predictions (90% accuracy), and road traffic sound tests again resulted in 28 correct and 2 incorrect predictions, maintaining a 93% accuracy rate. The outcomes of the 3-second and 4-second sound recognition tests are depicted in Figs. 9 and 10.

The testing, conducted 90 times for each of the 2-second, 3-second, and 4-second time variations, yielded cumulative data for each variation as detailed in Fig. 11.

5.3. Discussion

Among the three tested time variations, the lowest prediction accuracy occurred at the 2-second interval, achieving a 78% accuracy rate. Conversely, the 4-second interval demonstrated the highest prediction accuracy, reaching 91%. Notably, road traffic sounds yielded the highest prediction percentages. The lower accuracy rates for siren sounds can be attributed to the similarity between the sounds of fire engine sirens and ambulance sirens, as well as between police sirens and railroad crossing barriers, which occasionally led to incorrect predictions.

6. Conclusion

This research project aims to develop a device that aids drivers with hearing impairments by accurately identifying environmental sounds, specifically focusing on ambulances and fire trucks, and delivering timely alerts. The device intends to provide these warnings through visual cues, such as text displays on a screen and LED blinking, to ensure clear and immediate communication to the driver.

Testing has shown that duration significantly impacts prediction outcomes, with longer test durations yielding higher success rates. Specifically, employing a 4-second time variable during testing achieved the highest accuracy, reaching 91%.

For future development, this research can be advanced by utilizing a Raspberry Pi equipped with more than 1 GB of RAM, alongside a microphone that offers greater sensitivity to sound and enhanced noise reduction capabilities. Additionally, expanding the dataset to include more varied samples of ambulance and fire engine siren sounds would refine the training process and yield more precise results.

Acknowledgment

We wish to thank all the drivers with hearing impairments who contributed to the success of this study by participating in the testing of the system.

Conflict of Interest

Authors declare that they do not have any conflict of interest.

References


