In-Car Speech Enhancement Based on Source Separation Technique

Jeyasingh Pathrose¹, Mohamed Ismail M², Viswanathan Govindaraj³

¹Department of Electronics & Communication Engineering, B. S. Abdur Rahman Crescent Institute of Science & Technology, Chennai, India
²B. S. Abdur Rahman Crescent Institute of Science & Technology, Chennai, India
³Jasmine InfoTech Pvt Ltd, Chennai, India

Purpose: The purpose of this study was to investigate and analyze to increase the quality and intelligibility of speech in cars. The passenger dialogue inside the car, the sound of other equipment, and a wide range of interference effects are major challenges in the task of speech separation in-car environment.

Methods: Speech enhancement based on the source separation algorithm has been proposed to enhance the preferred speech signals using a microphone array inside a car. The proposed approach determines the signal direction in the time domain by utilizing the time difference of arrival (TDOA). TDOA signals are processed, and an adaptive least mean square method is used to determine the enhanced preferred signal.

Results: Experimental results show that the proposed approach yields an signal-to-noise ratio (SNR) of 7.4, perceptual evaluation of speech quality (PESQ) of 2.33 respectively. The proposed strategy outperforms existing methods in terms of PESQ, and SNR. The PESQ of the proposed method is 2.45%, and 5.22% better than the existing Independent Component Analyses, and Run length Source techniques.

Conclusion: Finally, compared to the existing methods, the suggested speech enhancement algorithm is more reliable and flexible, and it can properly identify the exact location of the sound.

Key Words: Car, Direction of arrival, Microphone array, Noise reduction, Speech enhancement.

INTRODUCTION

Speech enhancement (SE) is a pre-processing step in speech recognition that is also required to accommodate the growing demand for higher-quality speech. The speech signal is now used in a variety of systems including speaker identification, speech control, speech-to-text systems, voice over internet protocol accessibility of web applications, and interactive voice response system services. Voice recognition and other speaker activities (Benesty, 2018), interaction (Khonglah et al., 2019), sound aids (Li et al., 2018b), and coding of speech all require SE (Malathi et al., 2019). The SE is a difficult operation when the noisy signal is generated at a lower frequency (Dash et al., 2020). The quality of the voice signal should not be sacrificed while designing a speech signal-based system. However, speech signals can be damaged in practice due to a variety of disturbances such as echo, noise in the background, babbling noise babbling sound, and so on. Speech enhancement technology (Yang et al., 2016) can improve not just the signal-to-noise ratio (SNR) and audio perception of collected speech as well as the resilience of speech improvement and speaker verification systems. As a result, speech improvement in noisy contexts has gotten a lot of attention (Krause et al., 2021).

Speech intelligibility when utilizing in-vehicle speech applications has been impacted by engine noise and other noise sources, such as airflow from electric fans or automobile windows. Inside the car, the reflection of speech waves is employed to communicate particularly between the front and back seat passengers. In addition to in-car disturbances, the quality of speech communication

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
is generally poor. The speech signals are picked up by the microphone and the microphones are placed front seat headrest position.

Hands-free car kits and applications that recognize speech in the car now often use beamformer arrays or single-channel noise reductions. Microphone array processing focuses on speech improvement and localization, particularly in noisy or reverberant situations (Gannot et al., 2017). A microphone array is used in the car to increase voice communication quality (Tammen & Doclo, 2021). A microphone array may gather data in the spatial domain as well as the temporal and frequency domains. In this paper, two microphones are used for noise reduction. The source separation algorithm (SSA) method provides good noise reduction for frequencies when the noise elements of the microphone signals are uncorrelated. For practical microphone distances between 0.5 and 0.9 m. These correlations reduce the algorithm’s ability to suppress noise resulting in harmonic noise.

The main objective of the proposed system is to enhance the quality of speech in the car by using source separation-based adaptive least mean square (LMS). It is possible to handle multiple speech signals as well as a wide range of interfering effects. Separating as well as improving the signals as a result of the proposed method is beneficial to speech recognition in the future. Source separation is the technique of extracting signals from their combination as a source without an existing understanding of the mixing models.

The remainder of this work has been pre-arranged as follows. Section-2 summarizes the literature survey, and the problem about the array position inside the car is addressed in section 3. The proposed source separation for the car is given in section 4, outcomes are shown in section 5, and section-6 enfolds with a conclusion and future work.

MATERIALS AND METHODS

Literature survey

This section outlines the various investigations that have been carried out all year long to improve speech signaling. An overview of recent developments in the speech signal is given in this study.

Gentet et al.(2020) presented a speech enhancement algorithm it increases the SNR. The result showed that the technique has low-frequency noise and low computational complexity. Speech intelligibility optimization problem with a fixed perceived loudness restriction is a major drawback.

Alkaher & Cohen(2022) presented the dual microphone speech enhancement for enhancing speech communication in cars. The Pareto optimization decreases the overall speech distortion and relative gain reduction. The result demonstrates the dual-microphone system enhanced howling detection sensitivity. The drawback is that howling sounds may occur even before the speech reinforcement system reaches instability.

Lei et al.(2019) presented a wavelet analysis and blind source separation to enhance the performance of the voice control system. The experimental outcome demonstrates that the suggested technique successfully separates various speech signals in a demanding automotive setting without the need for prior information. Furthermore, the proposed method does not investigate in-vehicle speech recognition and the performance of vehicle speech management is low.

Wang et al.(2018) proposed an improved nonnegative matrix factorization (ImNMF)-based speech augmentation that was tested using the speaker verification system. The findings indicated that the recommended ImNMF can greatly enhance speech in noisy environments while also enhancing the speaker’s robustness in noisy environments like those of electric vehicles.

Tao et al.(2022) presented the enhanced speech enhancement and sound source localization algorithm which reduces microphone cost and also reduces the complexity. The dual-microphone sound algorithm effectively identifies the sound location, as well as the speech enhancement algorithm, is more resilient and adaptive than the previous method, according to experimental data. The expense and design requirements are the greatest drawbacks of this method.

Li et al.(2018a) presented the car speech enhancement method based on the distributed microphone. The dispersed microphone improves speech that has been distorted by noise in the car. The result showed that the suggested technique is more adaptable and it significantly improves the signal-to-noise ratio. Speech enhancement using distributed microphones is unusual.

Panda(2018) presented the stacked recurrent neural
network used to create a robust speech enhancement system. The traffic noise is canceled in the car speech recognition system. The simulated result shows that the suggested model has higher complexity with an optimum number of layers.

Qian et al. (2020) presented the car speech enhancement system based on a combination of a deep belief network and Wiener filtering. The deep belief networks parameters are optimized by using the quantum particle swarm optimization algorithm. The results of the experiment demonstrated that the suggested strategy may successfully reduce the original speech signal’s noise signal and improve the speech signal. Siegel et al. (2013) presented that fast learning is possible using echo state networks and concepts for a variety of applications, including speech recognition and detecting car driving actions.

Several works take into account the difficulties in the optimization of speech intelligibility. Unstable speech signal, the low performance of the microphone, cost and design requirements, and other factors were proposed, but none of them is successful in a speech enhancement system to overcome the above challenges this research proposed a novel source separation-based adaptive LMS and its detailed process is presented in the next section.

Problem formulation
Positioning a microphone array inside a car needs to obey some important requisites related to response quality, installation costs, and housekeeping easiness. Several works dealing with in-car speech separation and enhancement adopt two main microphone dispositions: microphones spread in the whole car interior. Although presenting some interesting results, this type of disposition presents some drawbacks that can difficult its adoption in commercial systems. First of all, spreading the microphones throughout the car interior make the receivers experiment with different levels of noise, produced by the horn sound, car engine sound, driver’s voice, the speeches of another person, and so on. This can lead to degradation in the beamforming response. Moreover, considering that the objective is to enhance the speech of the passenger, microphones are placed on the headrests of the front seats. The driver position, front-seat passenger position, and back seat passenger position are shown in Figure 1.

Let X and Y denote the number of source signals and microphones, respectively. The signals from the source are then referred to as

\[ s[t] = (s_1[t], s_2[t], ..., s_N[t])^T \]

The discrete time is denoted by t and the signals received can be labeled as

\[ x[t] = x_1[t], x_2[t], ..., x_M[t] \]

As a result of the delay and echoes between the source and the microphone, the combinations at the receiving end represent a more challenging mixing process known as convolution mixing.

\[ x_m[t] = \sum_{n=1}^{N} \sum_{d} a_{mn} s_n[t-d] + \text{noise}_m[t] \]

Where \( s_n[t] \) is the nth source signa \( x_n[t] \) is the mth microphone received signal and \( d \) is the discrete-time delay and reflects the source n to microphone m impulse response. Although these factors may fluctuate over time in practice, they are commonly considered to be stationary to simplify the model. The noise can be described as follows:

\[ N[t] = (n1[t], n2[t], ..., n_M[t]) \]
\[ x_m[t] = \sum_{n=1}^{N} \sum_{d} a_{mn} s_n[t-d] + \text{noise}_m[t] \]

Proposed speech enhancement method
This paper describes a speech enhancement technique that is based on source separation and adaptive LMS to improve passenger voice commands derived from a signal containing...
varied interferences and different speech sounds. To begin, the microphone captures the necessary signals, which are then passed to the source separation, which separates the denoised signals and removes the permutation ambiguity. The frequency domain is used in the separation process. As a result, the convolutive mixes are converted to the frequency domain using short-time Fourier transform (STFT). Finally, we use inverse short-time Fourier transform (ISTFT) to convert the unmixed signals to the time domain. The adaptive LMS receives the source separation outputs. Figure 2 depicts the specific procedure.

**Pre-processing method**

Two microphones are categorized by shape and number of microphones used. An array microphone of a circular shape with two microphones has been used in the analyzed method.

Let us consider the input source matrix given below,

\[
X(n) = \begin{bmatrix}
X_{1,1} & X_{1,2} & \cdots & X_{1,n} \\
X_{2,1} & X_{2,2} & \cdots & X_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
X_{N,1} & X_{N,2} & \cdots & X_{N,n}
\end{bmatrix}
\]

\[X(n) = \begin{bmatrix}
X_{1,1} & X_{1,2} & \cdots & X_{1,n} \\
X_{2,1} & X_{2,2} & \cdots & X_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
X_{N,1} & X_{N,2} & \cdots & X_{N,n}
\end{bmatrix} \times W_n \tag{7}
\]

\[
W_n: \text{window function coefficients, } X(n): \text{input source mixture.}
\]

Considering that the proposed technique is intended to process voice signals only, the windowed signal bandwidth is limited to 300 Hz to 4.2 kHz. Butterworth bandpass filters with fixed cut-off frequencies are used to combine input sources. \(X(n)\) is a matrix representing filtered input sources.

\[
X(n) = \begin{bmatrix}
X_{1,1} & X_{1,2} & \cdots & X_{1,n} \\
X_{2,1} & X_{2,2} & \cdots & X_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
X_{N,1} & X_{N,2} & \cdots & X_{N,n}
\end{bmatrix} \times h(n) \tag{8}
\]

Where, \(h(n)\): filter coefficients.

**Time direction of arrival (TDOA)**

The input direction is calculated by measuring the time taken by each microphone to receive the source. When signals are received at two microphones that are physically apart, the cross-correlation function can be used to describe the time difference between them as follows:

\[
R_{ij}(\tau) = \sum_{n=0}^{N-1} x_i[n] x_j[n - \tau] \tag{9}
\]

the two microphones are denoted by \(i\) and \(j\). \(x_i[n]\) and \(x_j[n]\) are the signals received at microphones \(i\) and \(j\) respectively; \(n\) is the time-sample index, and \(\tau\) is the signal correlation lag. The signals’ STFT can be expressed as,
In-Car Speech Enhancement

\[ R_i(\tau) = \frac{1}{N} \sum_{k=0}^{N-1} X_i(k)X_j(k) \times e^{j2\pi k\tau} \]  \[ \text{[10]} \]

where, \( X_i(k), X_j(k) \): FFT of \( x_i(n), x_j(n) \), \( N \): number of FFT points, FFT: fast fourier transform.

The anticipated time difference between two signals is represented by Eq [11], based on the cross-correlation output and the time arguments that match the outputs’ highest peak.

\[ \text{Delay} = \arg\max(R_i(\tau)) \]  \[ \text{[11]} \]

The source signal and the instant at the microphone are separated in time by Eq [12],

\[ \tau_{\text{delay}} = \frac{\text{Delay}}{C} \]  \[ \text{[12]} \]

where, \( C \): speed of sound.

As a result, \( \theta \) can be used to estimate the direction of the input source.

\[ \theta = \sin^{-1}\left( \frac{\tau_{\text{delay}}}{d} \right) \]  \[ \text{[13]} \]

Where, \( d \): separation between the two microphones.

The input source’s direction is given in Eq [13]. Figure 3 shows the flow diagram of the proposed SSA algorithm.

Source separation (direction classification, masking, and reconstruction)

The spatial filter will function as the radial bins to effectively isolate the desired input signal from the interference source. A circular array representing the sound source direction from two places defines the radial bins. There are two radial bins in all, with each bin denoting a 30-degree angle.

\[ R_n = [r_1, r_2, r_3, \ldots, r_{12}] \]  \[ \text{[14]} \]

Whereas \( r_1 \) stands for 0 degree, \( r_2 \) for 30 degrees, and \( r_{12} \) for 330 degrees. You can manually set the radial bin status or use the DOA of the input source. Radial bins are masked based on the input source’s DOA. The \( r_2 \) bin is the only one that is veiled if the angle from the DOA is 60 degrees. In the analyzed approach, the input signal is divided into desirable and interfering sources to obtain the intended signal. The intended and interference sources’ radial bins are provided below.

\[ R_n(\text{desired}) = [r_1, r_2, r_3, \ldots, r_{12}] \]  \[ \text{[15]} \]

\[ R_n(\text{interference}) = [r_1, r_2, r_3, \ldots, r_{12}] \]  \[ \text{[16]} \]

There are 12 different categories for the circular array microphone’s speech directions. Phase angle and array geometry characteristics are used to calculate each time-frequency bin’s direction. The directions of each time-frequency (T-F) bin are categorized based on the spoke direction. The definition of the direction vector for the input T-F bins,

\[ Z_n = [i_1, i_2, \ldots, i_n] \]  \[ \text{[17]} \]

where, \( n \): number of samples per frame.

By using radial bin status to mask the direction vector,

Figure 3. Flow chart of the proposed methodology. DOA: direction of arrival, T-F: time-frequency, LMS: least mean square, PI: speaker 1.
one may determine the masking coefficient of desired and interfering signals.

\[ W_{n}(\text{desired}) = Z_{n} \times R_{n}(\text{desired}) \] \[ W_{n}(\text{interference}) = Z_{n} \times R_{n}(\text{interference}) \]

By combining the input signal and the desired and interference signals' masking coefficients, the desired and interference signal is produced.

The desired and interference signals are separated by,

\[ Y_{n}(\text{desired}) = X_{i}(k) \times W_{n}(\text{desired}) \] \[ Y_{n}(\text{interference}) = X_{i}(k) \times W_{n}(\text{interference}) \]

to transform a frequency domain signal into a time domain signal, the inverse STFT is used. To further improve them, the adaptive LMS filter is fed both the desired and interference signals that were successfully reconstructed.

Adaptive LMS filter

An adaptive filter is applied to the model based on the input and output signals. By measuring the difference between the input signal and the background noise, the adaptive filter changes the filter coefficients automatically.

The parameters for the adaptive LMS filter are, \( u(n) \): the sources of interference signal include noise and other elements, \( d(n) \) desired signal, \( y(n) \): enhanced desired signal, \( e(n) \): error signal between \( u(n) \) and \( y(n) \).

It is possible to develop adaptive filters using both infinite impulse response and finite impulse response (FIR). The adaptive algorithm based on FIR is used by the method under study. An interference source is employed in an adaptive filtering process along with the desired direction source \( d(n) \) to reduce noise in the desired signal. Iteratively reducing the error signal \( e(n) \) is possible by changing the FIR filter coefficients.

\[ y(n) = k(n)^{T} \times w(n) \]

Where a representation of input filter vector \( k(n) \) is, \( k(n) = [u(n), u(n-1), \ldots, u(n-N+1)]^{T} \).

The filter coefficients vector, \( w(n) \) can be expressed as, \( w(n) = [w_{s}(n), w_{t}(n), \ldots, w_{e_{c}}(n)]^{T} \).

The error signal is calculated by using Eq [23],

\[ e(n) = d(n) - y(n) \]

the filter coefficients are updated by using Eq [24],

\[ w(n+1) = (1-\mu_{c}).w(n) + \mu.e(n).k(n) \]

an improved desired signal is produced by comparing the desired source and adaptive filter output. The desired source automatically lowers its levels of distortion and noise.

RESULTS

In this section, the experiment details and the simulation outcomes are given and analyzed. Figure 4 depicts a typical situation for the operation of speech enhancement in the car. There is a speech signal as well as noise signals in the car environment. Noise sources include ambient noises and interfering speech, whereas desirable sources include speech itself. The NOISEX-92 database’s recording of the noise source signal from a moving car Rohith & Bhandarkar(2021). The interference source speaker is situated precisely 180

![Figure 4](image-url)
In-Car Speech Enhancement

Table 1. Parameters used in the source separation method

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numbers of sources</td>
<td>2</td>
</tr>
<tr>
<td>Source classifications</td>
<td>Source of speech</td>
</tr>
<tr>
<td>Number of mics</td>
<td>2</td>
</tr>
<tr>
<td>Sampling rate</td>
<td>8 kHz</td>
</tr>
<tr>
<td>Size of the FFT window</td>
<td>512</td>
</tr>
<tr>
<td>Adaptive LMS filter step size</td>
<td>0.01</td>
</tr>
<tr>
<td>Filter length</td>
<td>32</td>
</tr>
</tbody>
</table>

FFT: fast fourier transform, LMS: least mean square

The algorithm receives data from the two microphones as input. The analysis uses one sample input. The required interference speech is used as the input source, and MATLAB is used to display the findings. The parameters that were used in the SSA are listed in Table 1.

Test setup

Two speakers namely, speaker 1 (SP1) and SP2 are used for providing the desired source data and interference data. Both microphones are placed on a rear unit setup of a car at a distance of 0.1 m from the speakers. The microphone array is placed at the center of the rear unit. The SP1 is set as the desired source and the SP2 is set as an interference source. The real time experimental setup in car is depicted in Figure 5.

Simulation results for 512 buffer size processing

In this simulation, we assume that the two microphone receivers, namely, M = 2. Let’s take into account the fact that degrees away from the desired source speaker, which is angled at 90 degrees to the device. Totally 50 sets of data were collected with different combinations of input sources at different angles. Microphone receivers include Mic 1, Mic 2, and so on.

The real-time data collection experimental setup in the car is shown in Figure 5.

Figure 5. Real-time data collection experimental setup in the car. SP: speaker.

Figure 6. Time domain representation for 512 buffer size processing.
omnidirectional microphones operate. Thus, in simulation studies, it is sufficient to ensure the mixed matrix’s invertibility rather than specify the precise locations or arrangements of the microphones.

**Figure 7.** Frequency domain representation for 512 buffer size processing.

**Figure 8.** Time domain representation for 4,096 buffer size processing.
In-Car Speech Enhancement

The frequency of car sound is primarily distributed in 50–500 Hz with no apparent regularity across time. Figure 6 shows the speech signals in the time domain.

Figure 7 shows the speech and different kinds of speech mixed signals in the time-frequency domain respectively. Rear seat passenger speech with background music, rear seat passenger speech, speech mixture of rear seat passenger, and one of the rear speech passengers have a wide frequency distribution. They can provide a great deal of information about speech characteristics.

Simulation results for 4,096 buffer size processing

More simulations are carried out with various receiving signals to validate the proposed method’s flexibility. Assume that $M = 4$ and that the receiving end consists of two microphones. The attenuation of the speech signal is significantly affected by the distance between the source and the microphone. However, due to the limited interior

![Figure 9. Frequency domain representation for 4,096 buffer size processing.](image)

**Table 2.** Comparison of results of SNR, SDR, SAR, and PESQ of SSA method with existing methods at different source combinations of input

<table>
<thead>
<tr>
<th>Combination of sources</th>
<th>Desired source</th>
<th>Methods for comparison</th>
<th>SNR (dB)</th>
<th>SDR (dB)</th>
<th>SAR (dB)</th>
<th>PESQ</th>
<th>SNR (dB)</th>
<th>SDR (dB)</th>
<th>SAR (dB)</th>
<th>PESQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rear seat passenger speech with background music</td>
<td>Rear seat passenger speech</td>
<td>Proposed SSA</td>
<td>9.2</td>
<td>8.2</td>
<td>9.9</td>
<td>2.7</td>
<td>9.6</td>
<td>8.7</td>
<td>9.5</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MMSS</td>
<td>5.2</td>
<td>5.5</td>
<td>6.1</td>
<td>1.9</td>
<td>5.5</td>
<td>5.3</td>
<td>5.9</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MVDR</td>
<td>4.4</td>
<td>4.1</td>
<td>5.8</td>
<td>1.6</td>
<td>4.2</td>
<td>3.9</td>
<td>5.1</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ICA</td>
<td>3.8</td>
<td>2.8</td>
<td>3.9</td>
<td>1.1</td>
<td>3.6</td>
<td>2.8</td>
<td>4.2</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DSB</td>
<td>2.8</td>
<td>2.2</td>
<td>3.1</td>
<td>0.9</td>
<td>2.1</td>
<td>2</td>
<td>3.2</td>
<td>0.8</td>
</tr>
<tr>
<td>Speech mixture of rear seat passengers</td>
<td>One of the rear seat passenger’s speech</td>
<td>Proposed SSA</td>
<td>9.6</td>
<td>8.7</td>
<td>9.5</td>
<td>2.6</td>
<td>9.6</td>
<td>8.7</td>
<td>9.5</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MMSS</td>
<td>5.2</td>
<td>5.5</td>
<td>6.1</td>
<td>1.9</td>
<td>5.5</td>
<td>5.3</td>
<td>5.9</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MVDR</td>
<td>4.4</td>
<td>4.1</td>
<td>5.8</td>
<td>1.6</td>
<td>4.2</td>
<td>3.9</td>
<td>5.1</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ICA</td>
<td>3.8</td>
<td>2.8</td>
<td>3.9</td>
<td>1.1</td>
<td>3.6</td>
<td>2.8</td>
<td>4.2</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DSB</td>
<td>2.8</td>
<td>2.2</td>
<td>3.1</td>
<td>0.9</td>
<td>2.1</td>
<td>2</td>
<td>3.2</td>
<td>0.8</td>
</tr>
</tbody>
</table>

space, it is found that the speech signal attenuation affected by changing the microphone location is low. Because of this, even if the microphone location in the car varies, the signal-to-interference ratio (SIR) at each receiving end is set to be the same.

The time-domain signal of mixed signals with interference from music and noise is shown in Figure 8 for each of the two microphones.

The time-domain signals produced by the suggested approach is more resemble the original signals than the SSA outcome in Figure 9. Additionally, the spectrograms have minimal residual noise and frequency loss.

Comparison of test results

The objective separation criteria, MATLAB toolbox, is used to assess how well the SSA approach separates data Rohith & Bhandarkar(2021). Various voice combinations with the desired signal are captured in the data sets. Each combination delivers the voice signal in the direction of the selected source, while the noise signal is in the opposite direction. A comparison is presented in Table 2 based on existing techniques, input source combination, and desired signal direction. The SNR, signal to distortion ratio (SDR), signal to artefact ratio (SAR), and perceptual evaluation of speech quality (PESQ) findings of the proposed SSA method are compared to those of existing techniques at various source combinations of input.

According to a result comparison, the SSA method is superior to the current approaches. There is a small variation between how well minimum variance distortionless response (MVDR) and minimum mean square spectral amplitude estimator (MMSS) performed. In comparison to existing approaches, the performance findings show that the SSA method provides good SDR, SNR, PESQ, and SAR at each input source combination. The proposed test configuration is not well suited for noisy mixtures, which frequently include voice signals that have been reverberantly mixed with numerous competing background noise sources. Additionally, the SSA-based proposed speech improvement includes improved noise reduction. However, as the number of microphones rises, processing power rises as well, making it very impossible to deploy on Hearing Aid Devices or smartphones.

DISCUSSIONS

In the car environment, a range of interference effects and varied speech signals provide a major difficulty for the operation of the system. In this paper, a proposed adaptive LMS and SSA algorithm is used to improve the speech signal. The direction of the input source is obtained using the TDOA between the microphones. The input signals are masked and reconstructed to get the desired interference signals. The separated signals are adaptively filtered and enhanced. The proposed SSA method is implemented and validated using MATLAB with different combinations of input source mixtures, which has yielded the expected results. Simulation results show that the proposed technique can successfully separate various speech signals without the need for prior information in the car system. The efficiency of the system is then demonstrated using a two-microphone simulation. The signal quality parameters of the input signal to the enhanced desired signal are SNR of 1.91 dB and PESQ of 2.6 are achieved.

Ethical Statement

My research guide reviewed and ethically approved this manuscript for publishing in this Journal.

Acknowledgements

The authors would like to thank the reviewers for all of their careful, constructive and insightful comments in relation to this work.

Declaration of Conflicting Interests

This paper has no conflict of interest for publishing.

Funding

N/A

Author Contributions

Conceptualization: Jeyasingh Pathrose, Data curation: Mohamed Ismail M, Formal analysis: Viswanathan Govindaraj, Funding acquisition: Viswanathan Govindaraj, Investigation: Mohamed Ismail M, Methodology: Jeyasingh Pathrose, Project administration: Viswanathan Govindaraj, Resources: Viswanathan Govindaraj, Software: Mohamed Ismail M,
Supervision: Jeyasingh Pathrose, Validation: Viswanathan Govindaraj, Visualization: Viswanathan Govindaraj, Writing—original draft: Mohamed Ismail M, Writing—review & editing: Jeyasingh Pathrose, Approval of final manuscript: all authors.

REFERENCES


