

Systematic Review

# Speech Enhancement Algorithms: A Systematic Literature Review

Sally Taha Yousif <sup>†</sup> and Basheera M. Mahmood <sup>\*,†</sup> 

Department of Computer Engineering, University of Baghdad, Al-Jadriya, Baghdad 10071, Iraq;  
gs22.styousif@coeng.uobaghdad.edu.iq

\* Correspondence: basheera.m@coeng.uobaghdad.edu.iq

† These authors contributed equally to this work.

**Abstract:** A growing and pressing need for Speech Enhancement Algorithms (SEAs) has emerged with the proliferation of hearing devices and mobile devices that aim to improve speech intelligibility without sacrificing speech quality. Recently, a tremendous number of studies have been conducted in the field of speech enhancement. This study aims to map the field of speech enhancement by conducting a systematic literature review to provide comprehensive details of recently proposed SEAs. This systematic review aims to highlight research trends in SEAs and direct researchers to the most important topics published between 2015 and 2024. It attempts to address seven key research questions related to this topic. Moreover, it covers articles available in five research databases that were selected in accordance with the PRISMA protocol. Different inclusion and exclusion criteria have been performed. Across the selected databases, 47 studies met the defined inclusion criteria. A detailed explanation of SEAs in the recent literature is provided, with existing SEAs studied in a comparative fashion along with the factors influencing the choice of one over the others. This review presents different criteria related to the approaches utilized for signal modeling, the different datasets employed, types of transform-based SEAs, and the effectiveness of different measurements, among other topics. This study presents a systematic review of SEAs along with existing challenges in this field.

**Keywords:** speech enhancement; Wiener filter; MMSE; deep learning; noise; speech signal



Academic Editor: Frank Werner

Received: 6 February 2025

Revised: 28 April 2025

Accepted: 3 May 2025

Published: 6 May 2025

**Citation:** Yousif, S.T.; Mahmood, B.M. Speech Enhancement Algorithms: A Systematic Literature Review. *Algorithms* **2025**, *18*, 272. <https://doi.org/10.3390/a18050272>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Speech Enhancement Algorithms (SEAs) are computational methods designed to improve the acoustic properties of speech signals that have been degraded by added noise or other distortions, thereby improving human perception [1,2]. It is widely known that additive noise is the most prevalent and influential type of noise in real-world environments. Therefore, SEAs are designed to process noisy signals, restore clean speech, improve speech quality and intelligibility, mitigate noise pollution, and reduce listener fatigue [3]. In real-life scenarios, background noise often contaminates clean speech, resulting in noisy signals. Consequently, speech consistently includes unwanted degradation that creates a lower-quality signal for human listeners. This degradation leads to listener fatigue and significantly reduces the performance of speech recognition systems [4]. SEAs have been developed to process noisy signals and restore the original speech signal by effectively performing noise suppression to ultimately improve the perceived quality for human listeners [5]. Over the decades, several SEAs have been developed, significantly shaping modern audio processing technologies. In 1979, the spectral subtraction method was introduced [6]. This method can estimate the noise from the silent portions of a signal and subtract it from the spectrum. It helps to reduce background noise and improve speech

clarity, making it a foundational technique for many later advancements in telephony and audio applications. The Wiener filter, first developed in 1940 by Norbert Wiener, has been applied in the speech enhancement field to solve signal estimation problems for speech signals [7]. This statistical filtering method minimizes the mean squared error between the original and enhanced signal, leading to improved speech quality in noisy environments, and has found use in VoIP applications and hearing aids.

In the 1980s, Kalman filters began to find applications in speech processing. First proposed in the 1960s by Rudolf Kalman [8], the Kalman filter is a Bayesian recursive filter that can predict the optimal state of a speech signal, making it particularly useful in dynamically noisy environments such as live broadcasting and speech recognition systems. During the same period, Ephraim & Malah (1984) introduced an algorithm for estimating the Minimum Mean Square Error (MMSE) [9]. This algorithm uses statistical modeling to suppress noise while preserving speech intelligibility, capitalizing on the major importance of the short-time spectral amplitude. This method has become a key component in hearing aids and noise suppression features in modern smart devices [9]. The 1990s saw the emergence of subspace methods [10], which decompose the speech signal into distinct components, allowing speech to be separated from noise without significantly affecting clarity. These methods have been widely used in robust speech recognition and forensic audio analysis. Additionally, statistical modeling-based methods [11] that leverage Gaussian Mixture Models (GMMs) and Hidden Markov Models (HMMs) to predict clean speech based on previously recorded patterns have recently gained traction. These approaches have contributed to advancements in speech synthesis, assistive voice technologies, and speech-to-text conversion. These classical SEAs laid the groundwork for modern advancements, including deep learning-based speech enhancement techniques and Generative Adversarial Networks (GANs), which build upon these fundamental principles with more sophisticated data-driven approaches to noise suppression and speech quality improvement.

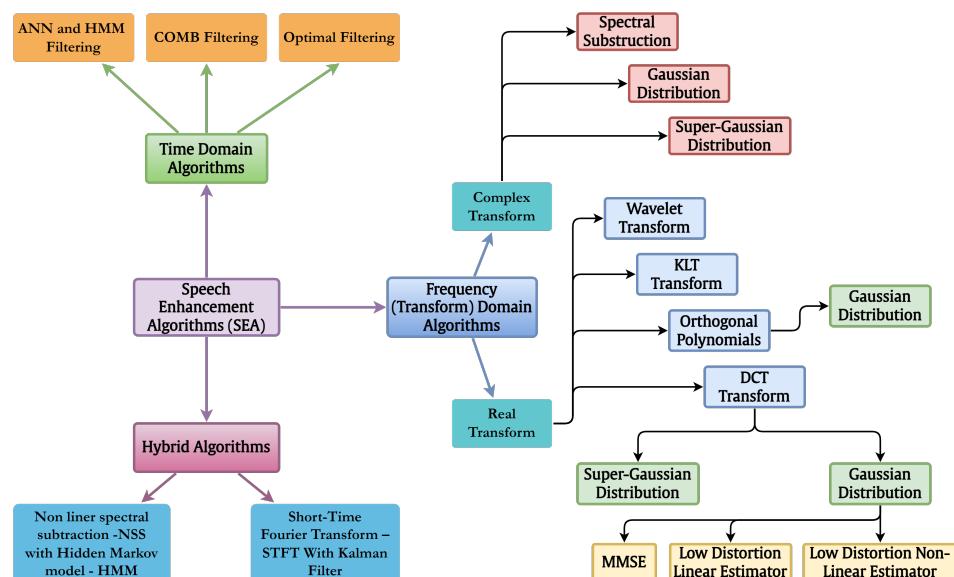
There are various criteria for classifying SEAs. Several studies have categorized SEAs based on the number of channels, distinguishing between single-channel and multi-channel approaches [12,13]. Other studies have classified SEAs into three basic categories according to different processing techniques: spectral-subtractive algorithms, methods based on statistical models, and optimization criteria [6]. The latter category includes Wiener filtering (WF) [14], Minimum Mean Square Error (MMSE) estimation [9], deep learning algorithms [15], and subspace algorithms [16].

A broad classification of SEAs can also be made based on processing domains, which can be divided into transform domain-based SEAs, time domain-based SEAs, and hybrid domain-based SEAs. Recently, there has been a growing trend toward speech enhancement models that combine the time and frequency (TF) domains [17–19]. In transform domain-based SEAs, degraded speech is processed in the transform domain, whereas temporal SEAs operate in the time domain. Each processing domain offers its own advantages and disadvantages [4]. Various discrete transforms have been used in SEAs, including the Discrete Fourier Transform (DFT) [9,20], Wavelet Transform (WT) [21], Discrete Cosine Transform (DCT) [22], Discrete Krawtchouk Transform (DKT), and Discrete Tchebichef Transform (DTT) [14]. It is worth mentioning that different types of DCTs are utilized to improve speech quality, such as by reducing noise, extracting essential features, and enabling efficient compression. Each type has specific advantages that make it suitable for particular applications. DCT Type-II (DCT-II) [23,24] is considered the most widely used type, especially in Mel-Frequency Cepstral Coefficients (MFCCs), which play a crucial role in speech recognition and enhancement. It is known for its excellent energy compaction, making it highly effective for speech compression and denoising. On the other hand,

the DCT Type-III (DCT-III) [25], also referred to as Inverse DCT (IDCT), is essential for reconstructing speech signals after enhancement. It is often used in filtering bank-based speech processing to restore speech quality after noise suppression. Another important variant is DCT Type-IV (DCT-IV) [26], which is commonly applied in sub-band coding and speech denoising. It provides a smoother spectral representation of speech signals, helping to reduce artifacts and improve overall clarity. Additionally, the Modified Discrete Cosine Transform (MDCT) [27] is widely used for audio compression and real-time speech applications. Overall, these DCT-based techniques have greatly contributed to speech enhancement by improving noise reduction, compression efficiency, and speech intelligibility. Their continued relevance in modern deep learning-based SEAs highlights their lasting impact on the field. Each category of SEA can be implemented using specific transforms. Figure 1 illustrates the different classes of SEA based on dedicated processing domains [1].

SEAs have garnered significant attention due to their wide range of applications. Numerous reviews and surveys have been conducted to examine different aspects of SEAs. For example, ref. [28] investigated various techniques for speech de-reverberation and enhancement in noisy and reverberant environments. Two primary backbone Deep Neural Networks (DNNs) were compared, one operating in the time domain and the other in the frequency domain. In [29], the authors conducted a literature survey that discussed challenges such as nonstationary noise and overlapping speech. These are common issues in real-world scenarios, where speech signals are often degraded. A comprehensive summary of speech enhancement techniques and their various applications was presented in [29], focusing on traditional methods such as adaptive filtering and Wiener filtering as well as modern approaches such as deep learning-based techniques. Today, SEAs are integrated into various fields and applied in many practical applications, including telecommunications, hearing aids, speech recognition, and audio restoration. This paper aims to offer valuable insights for researchers and engineers, helping them to choose the most suitable approach based on the specific challenges of their acoustic environments.

In particular, this systematic literature review is designed to assist speech researchers and academics by identifying critical research gaps in this field and highlighting promising directions for future studies. Through an in-depth analysis of speech enhancement methodologies, this review draws attention to the field's most impactful topics. Furthermore, it addresses recent challenges and uncovers current research gaps involving SEAs, providing valuable insights and guidance for future research trajectories.



**Figure 1.** General SEA classification based on processing domain.

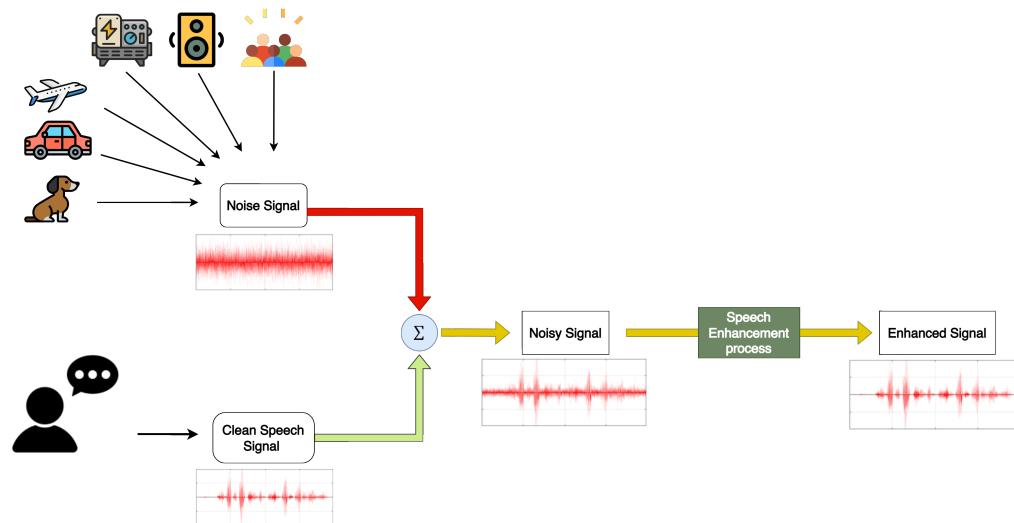
## 2. Background

Speech enhancement aims to improve the quality and intelligibility of speech signals. However, achieving better speech quality does not necessarily guarantee higher intelligibility, as these two criteria are independent. Most SEAs focus on noise reduction by improving the quality of the speech signal, often at the expense of reducing intelligibility. The quality of a speech signal refers to its clarity, intelligibility, and naturalness, which determine how well a listener or automated system can perceive and understand the speech. Quality is influenced by factors such as background noise, distortion, bandwidth, and temporal continuity [30]. High-quality speech signals exhibit minimal noise, reduced artifacts, and a well-preserved frequency spectrum, ensuring both naturalness and intelligibility [31]. In speech enhancement, various algorithms aim to improve quality by suppressing noise and distortions while preserving key speech components, making the resulting signals more suitable for applications such as telecommunication, speech recognition, and assistive hearing devices. The human ear plays a vital role in speech perception, making certain sound properties essential for speech quality. The ear is most sensitive to frequencies between 1 kHz and 4 kHz, where speech intelligibility is highest; thus, speech enhancement should preserve this range [32]. The ear also follows loudness perception patterns, meaning that speech should be balanced across frequencies in order to sound natural [33]. Additionally, the auditory system can distinguish rapid sound changes, meaning that avoiding distortion of key speech cues is crucial when enhancing speech signals [34]. Finally, binaural hearing helps with spatial awareness, enabling better speech perception in noisy environments [35]. The human voice itself can sometimes be an unclean source signal, even before any external noise or distortions are introduced; for instance, this can happen due to speech disorders, medical conditions such as dysphonia or Parkinson's disease, and stuttering, which can cause irregularities in voice production [36]. Emotional states such as whispering, shouting, or crying also affect speech clarity and make it harder to process [37]. Additionally, factors such as fatigue, dehydration, and aging can alter vocal quality, leading to hoarseness or instability [38]. In multilingual settings, accents and pronunciation variations may introduce further complexity [39]. Because of these natural variations, speech enhancement systems must adapt to intrinsic distortions in the human voice while preserving its unique characteristics. Interestingly, listeners can sometimes extract more information from a noisy signal than an enhanced one, especially when listening attentively; however, prolonged exposure to noisy signals can cause discomfort, prompting the development of methods that simultaneously enhance the quality and intelligibility of speech signals [40]. Deep learning techniques such as Convolutional Neural Networks (CNNs) have been employed to analyze and process distorted speech signals. These approaches enhance speech quality while facilitating clearer and more accurate speech inputs, benefiting tasks such as automatic speech recognition as well as applications in noisy environments.

The design of a speech enhancement architecture should involve multiple standards, including an appropriate speech model, addressing different noise types, intelligibility improvement, and quality enhancement. One of the most crucial factors in any design is the mitigation of artifacts, including Musical Noise (MN), which is a common artificial distortion generated during speech enhancement or noise reduction processes. MN manifests as isolated tonal or harmonic elements resembling random musical notes, which can disrupt the listening experience [9]. The process of speech degradation caused by background noise and the subsequent application of noise suppression techniques is depicted in Figure 2. As speech signals travel from source to destination, they often encounter additive background noise in uncontrolled environments [4]. Therefore, SEAs are an important optimization method for enhancing distorted signals and removing noise.

The critical challenge in developing highly effective SEAs lies in selecting suitable processing techniques for noise reduction and clarity improvement, along with accurately modeling the statistical properties of noise and speech. Techniques such as spectral subtraction and Wiener filtering are commonly employed for noise suppression. Additionally, the choice of an appropriate Probability Density Function (PDF) to model noise and speech signals significantly impacts the performance of such algorithms. There are different statistical PDFs for modeling these random signals, which can be stationary or nonstationary. For instance, different works have used Gaussian PDFs [9,30,41]. These are widely used due to their simplicity, but may not accurately represent real-world noise in nonstationary environments. Suppressing nonstationary noise is a more complex task than suppressing stationary noise. To address this, more advanced PDFs use Laplacian [20,42] or Gamma [43] distributions for more accurate noise modeling. Many studies have emphasized the importance of matching the PDF to the noise type in order to increase the algorithm's accuracy. Techniques such as Deep Neural Networks (DNNs), Kalman filtering, Non-negative Matrix Factorization (NMF), and MMSE estimation further leverage statistical models to refine speech signals, demonstrating the critical role of PDF selection in achieving optimal results.

Significant advancements have been made in recent decades thanks to the availability of open-source resources, techniques, and measurements for building and evaluating SEAs. This study seeks to review and analyze recent existing research on SEAs while providing a comprehensive overview of their concepts, development, and challenges.

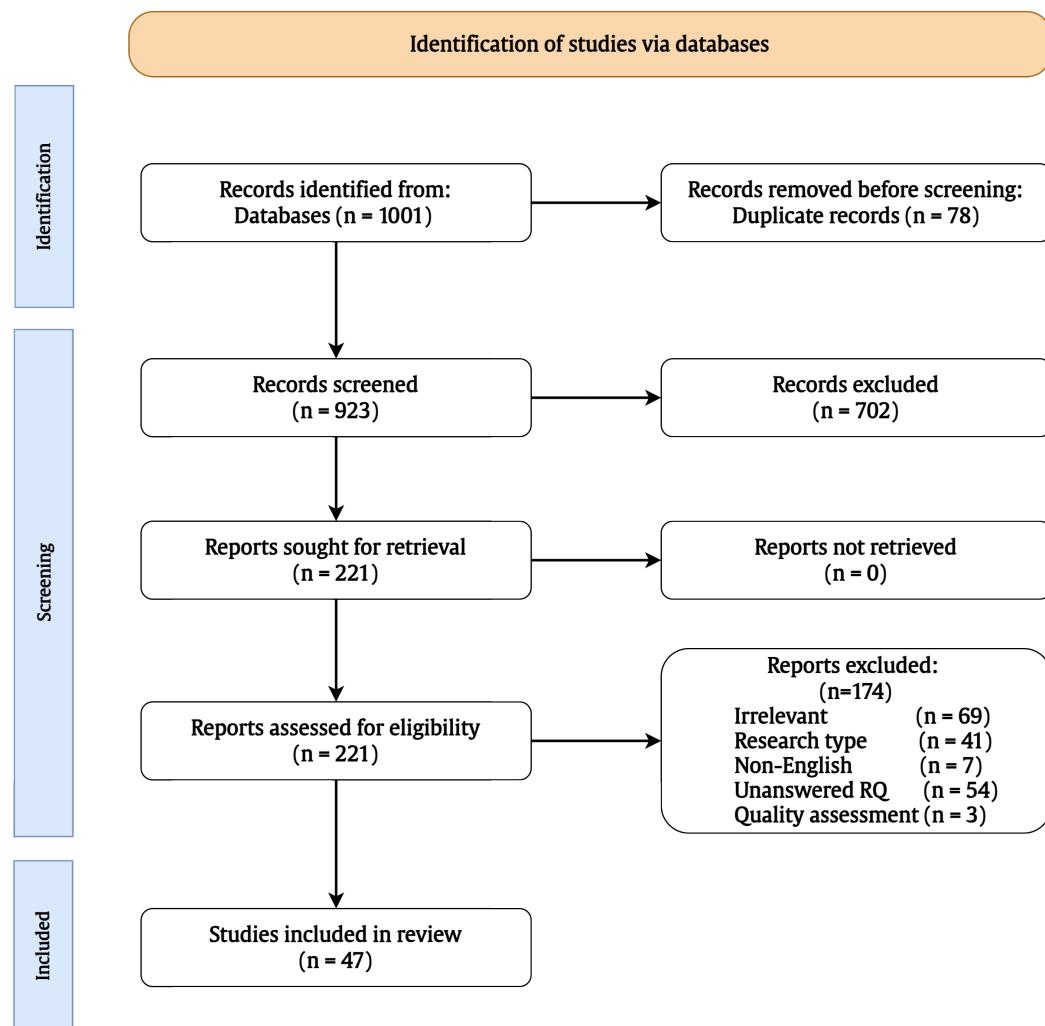


**Figure 2.** Degradation and denoising of a speech signal.

### 3. Method

The field of speech enhancement is of great interest due to its highly multidisciplinary nature and its relevance to many human-centric applications, including medical applications. Therefore, in this review we implement the idea of building a body of evidence and information on various SEA approaches. The survey conducted in this paper is divided into several stages, with each stage including specific elements of planning, conducting, and reporting to form the methodological framework. This systematic literature review is conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol [44,45], which is useful for visualizing the relationship between the stages of the literature review process. Figure 3 presents the PRISMA flowchart, which details the outcomes at each stage of the systematic review and visually summarizes the screening process while improving the quality and transparency of the SEA literature. First, the number of identified papers from the electronic database is recorded. The selec-

tion process of the systematic review is then carried out transparently by reporting the decisions made at each stage, including the number of studies. Reasons for including or excluding papers at the full-text stage are documented. The PRISMA flowchart, shown in Figure 3, enhances the reliability and reproducibility of reviews by enabling researchers to understand and evaluate the methodology and findings. The workflow of the methodology is initiated by identifying research questions, followed by formulating and outlining the search strategy and developing criteria for the inclusion or exclusion of studies. Lastly, the procedures for quality assessment and data extraction are implemented. Data are extracted to obtain essential information from the selected papers included in the review. In this review, the extraction process was performed independently by two reviewers to ensure greater accuracy and reduce bias. All stages of the review are detailed in the following sections.



**Figure 3.** PRISMA flowchart for study selection.

This work reviews the technologies used for speech enhancement processing. It explores the existing methods and techniques currently used in the field in addition to the potential challenges and limitations, providing insights about how methods for enhancing the quality of degraded speech can be improved. The following sections present the methodological steps and resources used to conduct this systematic review in comprehensive detail.

### 3.1. Research Questions (RQs)

Identifying and constructing the Research Questions (RQs) is the first and most essential process in any literature review. In this systematic literature review, RQs were developed by starting with the broad topic of Speech Enhancement Algorithms (SEAs), conducting preliminary research on specific issues related to SEAs, and narrowing down the focus within this field. The soundness of the potential questions was then evaluated by identifying the main challenges in existing SEAs. The following seven RQs were established to facilitate an in-depth examination of the field:

- RQ1:** What are the most commonly used recent research approaches in the field of SEAs?
- RQ2:** What are the main types of transform-based SEA used in the existing studies?
- RQ3:** What is the number of channels presented in the current works?
- RQ4:** What are the major models used to identify speech and noise signals?
- RQ5:** What are the well-known datasets used in existing papers?
- RQ6:** What are the most effective measurements for evaluating speech intelligibility and quality?
- RQ7:** What are the current limitations and challenges in the reviewed SEAs?

The construction of these RQs was formulated carefully. RQ1 aims to identify recent research topics and then conduct preliminary research on the field of SEAs, which is the subject of interest in this work. Then, the definition of essential issues that need to be addressed in this field is provided based on RQ2, RQ3, and RQ4. Specifically, RQ2 focuses on the recent types of transform-based SEAs, RQ3 highlights the number of channels presented in the current SEA, and RQ4 provides an overview of the major types of models for speech and noise signals. After that, RQ5 and RQ6 narrow the focus and scope of the research by demonstrating the most well-known datasets applied in the field and the most effective measures used for evaluating the intelligibility and quality of enhanced speech signals, respectively. Finally, RQ7 aims to investigate the current limitations and challenges in the reviewed studies on SEAs. The concept of RQ formulation has significant implications for our methodological process, and has been carefully implemented to guide researchers through these questions in order to provide broader benefits to the field of SEAs.

### 3.2. Search Strategy Based on Search Strings and Online Electronic Databases

After developing the research questions related to the SEA topic, the search strategy was performed by identifying specific categories and keywords that are commonly used in the related studies. The search process was then conducted across five electronic databases.

The first step in formulating a search strategy is to take the topic and break its terminology into discrete concepts (categories) and keywords. The following four concepts were formalized based on the literature review: Concept 1 is associated with the signal type (speech, noise, voice, and audio); Concept 2 is associated with the algorithm name (enhancement, improvement, noise reduction, and denoising); Concept 3 is associated with algorithm types (Wiener filtering, MMSE estimation, deep learning, spectral subtraction, and noise estimation); finally, Concept 4 is associated with the approach (system, technology, technique, method, and algorithm).

Then, a search string is formed by identifying primary keywords for each concept. Table 1 outlines the defined concepts along with their corresponding keywords. These keywords are commonly used in research studies on SEA and are aligned with the research questions. The categories and their corresponding keywords were selected based on an extensive literature review. As shown in the table, different keywords were defined for the four categories, which are expressed using similar or related terms. A thorough search was formalized based on the identification of alternative synonyms, acronyms, and spellings which include variations of the major term. The defined keywords were then organized un-

der the four distinct concepts and combined using the Boolean operator OR. Subsequently, the Boolean operator AND was applied to incorporate keywords across categories.

**Table 1.** Definitions of the four keyword groups.

Concept 1	Concept 2	Concept 3	Concept 4
Speech	Enhancement	Wiener Filtering	System
Voice	Improvement	MMSE Estimation or Noise estimation	Technology or Technique
Noise	Noise Reduction	Deep Learning	method
audio	Denoising	Spectral subtraction	Algorithm

As a result of the above procedure, search strings were created from the keywords within each concept and Boolean operator. Boolean operators have the ability to link together two or more search conditions, allowing more complex search logics to be specified. The implemented procedure was as follows:

- **Concept 1:** “Speech” OR “Voice” OR “Noise” OR “audio”
- **Concept 2:** “Enhancement” OR “Improvement” OR “Noise Reduction” OR “Denoising”
- **Concept 3:** “Wiener Filtering” OR “MMSE Estimation” OR “Deep Learning” OR “Spectral subtraction or Noise estimation”
- **Concept 4:** “System” OR “Method” OR “Technology or Technique” OR “Algorithm”.

The resulting string was structured as (Concept 1) AND (Concept 2) AND (Concept 3) AND (Concept 4). The search results were subsequently retrieved and stored using Mendeley Reference Management software. Then, the search string including the keywords based on the categories and Boolean operators (logical operations) was entered into the search box of the selected electronic library database or search engine. Afterwards, the search process was conducted by retrieving articles from the five electronic databases used for data collection. The five sources cover many aspects of the reviewed research area and provide effective search engines that are easy to use and appropriate for automatic search [46]. The list of libraries included SpringerLink, IEEE Xplore, ScienceDirect, ACM Digital Library, and Google Scholar. The aforementioned combinations of keywords based on Boolean expressions were used in the search process. The search engines mostly support logical operators and permit the use of parentheses to link multiple search conditions and define more complex logics. Parentheses are significant because they separate the specific elements of the search string in order to confirm that they are regarded as a group.

However, the details of the search step based on the defined search string sometimes need to be carefully checked to ensure the best results depending on the keyword combinations and the database being used. The complete documentation of the search strategy, including its details, is illustrated in Table 2, which provides the name of the online electronic database (digital library), the date of the search, the combination(s) of keywords based on Boolean operators, and the number of studies retrieved (with and without filter). Therefore, the result is a documentary report containing all the details related to the search strategy, which helps in tracking the subsequent research phases.

**Table 2.** Documentation of the search strategy with details of the online electronic databases.

Survey Sources	Data Access	Keywords	No. of Papers Without Filter	No. of Papers with Filter (by Selected Years) 2015–2024
ACM Digital Library	27/11/2024	(“Enhancement” OR “noise reduction” OR “denoising”) AND (“wiener filtering” OR “deep learning”)	40	38
IEEE Xplore	26/11/2024	(“speech”) AND (“noise reduction”) AND (“deep learning” OR “wiener filtering”)	45	24
Springer Link	27/11/2024	(“Speech” OR “Voice” OR “Noise” OR “audio”) AND (“Enhancement” OR “Improvement” OR “Noise Reduction” OR “Denoising”) AND (“Wiener Filtering” OR “MMSE Estimation” OR “Deep Learning” OR “Spectral subtraction”) AND (“System” OR “Tool” OR “Technology’ or Technique” OR “Algorithm”)	68	36
Science Direct	26/11/2024	(“Enhancement”) AND (“Wiener Filtering”) AND(“Algorithm”)	467	260
Google Scholar	26/11/2024	(“Speech” OR “Noise”) AND (“Enhancement” OR “Noise Reduction”) AND (“Wiener Filtering” OR “MMSE Estimation” OR “Deep Learning”) AND (“System” OR “Tool” OR “Technology or Technique” OR “Algorithm”)	366	216

### 3.3. Study Selection Based on Inclusion and Exclusion Criteria

Selection criteria aim to identify studies that provide direct evidence on the research topic. To minimize the potential for bias, selection criteria should be specified during the protocol definition phase and can be refined during the search process [47]. In this systematic review we identified target study papers based on specific formalized criteria. These criteria, collectively known as the eligibility criteria, should be specified carefully when searching for papers, as they help to narrow down the scope of the relevant topic. The criteria used for exclusion and inclusion set the boundaries of this paper, and were evaluated based on the details mentioned in Table 3. The first excluding step in this stage was to remove redundant papers, followed by screening the papers against the keywords and formulated research questions following the steps in [46,48]. Any paper that was not expected to provide an accurate answer to the research questions was excluded. Next, papers were assessed against the inclusion and exclusion criteria based on the title, abstract, and full text. Included studies were from peer-reviewed journals, conferences, and workshop papers. In the event of multiple versions of a selected paper, only the most recently updated version was included; all other versions were excluded. Table 2 shows that the search string retrieved a total of 986 articles. In addition, fifteen papers were added from other surveys and overviews. Of 1001 papers, 78 were removed due to duplication, leaving 923 papers for further review. Subsequently, the retained papers were checked by authors based on the title, abstract, and keywords of the papers. The authors were considered the reviewers, and the number of retrieved records was 221. To assess the actual relevance of the studies, selection was performed on all potentially relevant studies by applying the set of inclusion and exclusion criteria. The criteria were applied to the retrieved records in order to assess them and make decisions about which retained papers were reviewed. The adopted exclusion and inclusion criteria are illustrated in Table 3.

**Table 3.** Criteria used in the exclusion and inclusion stage.

Inclusion Criteria	Exclusion Criteria
Papers that are relevant to the topic of SEA.	Papers that are not relevant to the topic of SEA.
Papers published during the period 2015–2024	Papers published before 2015.
Scientific journals or conference papers.	Papers that were not submitted to scientific journals or conferences.
Papers written in English language	Papers written in a language other than English.
Papers answering one or more of the research questions.	Papers that did not answer any of the research questions.
Papers that were more than three pages and that were not duplicates.	Papers that were duplicate papers or not available.

Based on [47,49], two researchers evaluated each paper and reached agreement about whether to include or exclude it. In cases of disagreement, the matter was discussed and resolved based on the predefined criteria. The applied criteria further reduced the number of research papers to 50.

### 3.4. Quality Assessment (QA) Rules

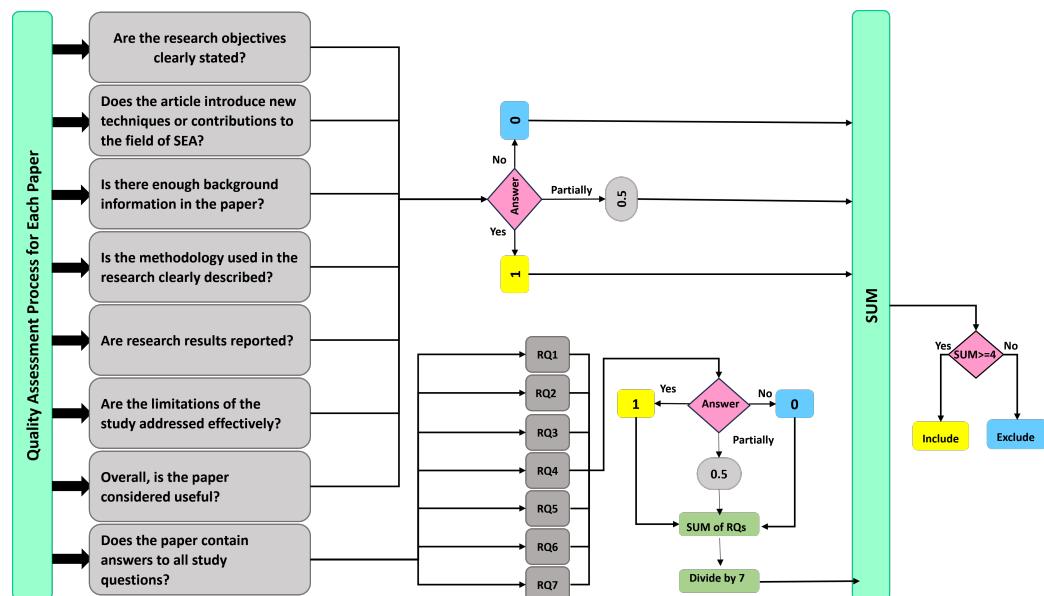
In any systematic literature review, applying Quality Assessment (QA) rules is considered the last step in identifying the final list of candidate papers for inclusion. QA rules were applied to evaluate the quality of the research papers in accordance with the set research questions. Given the lack of standard empirically grounded QA rules suitable for use in such study designs, we implemented a scoring system to evaluate the quality of the papers eligible for inclusion in our systematic review based on the rules provided in [50]. Following the study selection process, we applied eight additional criteria in order to improve the results. QA rules were applied to each nominated paper in order to assess the quality of its contents and select only those studies most relevant to our SEA topic. Using the procedure suggested in [47], we checked the quality of each candidate paper for inclusion based on the following QA checklist:

- QA1.** Are the research objectives clearly stated?
- QA2.** Does the article introduce new techniques or contributions to the field of SEA?
- QA3.** Is there enough background information in the paper?
- QA4.** Is the methodology used in the research clearly described?
- QA5.** Does the paper contain answers to the research questions?
- QA6.** Are research results reported?
- QA7.** Are the limitations of the study addressed effectively?
- QA8.** Overall, is the paper considered useful?

Each question had three possible answers, with points being scored in the following way:

1. The study was marked with “Yes” if the answer to the QA item was positive, where “Yes” = 1.
2. The study was marked with “No” if the answer to the QA item negative, where “No” = 0.
3. The study was marked with “Partial” if it only partially answered the QA item, where “Partial” = 0.5.

The procedure for quality assessment is illustrated in Figure 4. The QA score of the research studies was achieved by weighing their quality alongside the QA questions; a paper was selected if it had a quality score equal to or greater than four. The results are presented in Table 4. Then, for each research study, the total score was computed based on a predefined threshold. If the total score was greater than or equal to four, the candidate study was included; on the other hand, if the score was less than four, the candidate study was excluded. Based on this procedure, three studies were excluded after implementing the quality assessment.



**Figure 4.** Flowchart of the quality assessment procedure.

In the QA procedure, studies were scored based on the degree to which specific criteria were met. The steps are summarized as follows. First, the questions used for the assessment procedure were defined. A scale was then determined to assign ranks to the papers based on the list of quality assessment questions. The total score value was obtained after summing all the weights provided based on these questions. Afterwards, a threshold of four was set, with papers excluded if this threshold was not met. On this basis, three papers were excluded.

The final number of retained studies was 47. The excluded papers are highlighted in purple in Table 4. All details of the QA answers for the retained papers are shown in Table 4.

The full workflow of our systematic review after all methodological steps is depicted in Figure 5.

**Table 4.** Characteristics of the candidate papers for inclusion based on the QA questions.

Ref	Publication Title	Year	Type	QA Questions (QA) [1, 2, 3, 4, 5, 6, 7, 8]	Sum
[3]	Speech Enhancement Algorithm Based on Super-Gaussian Modeling and on Super-Gaussian Modeling and Orthogonal Polynomials	2019	Journal	[½, 1, 1, 1, 0, 1, 1, ½]	6
[1]	Low-Distortion MMSE Speech Enhancement Estimator Based on Laplacian Prior	2017	Journal	[1, ½, ½, 1, 1, ½, 0, 1]	5.5
[15]	Deep Learning-Based Speech Enhancement Algorithm Using Charlier Transform	2023	Conference	[0, 1, 0, ½, 0, 1, 1, 1]	4.5

**Table 4.** Cont.

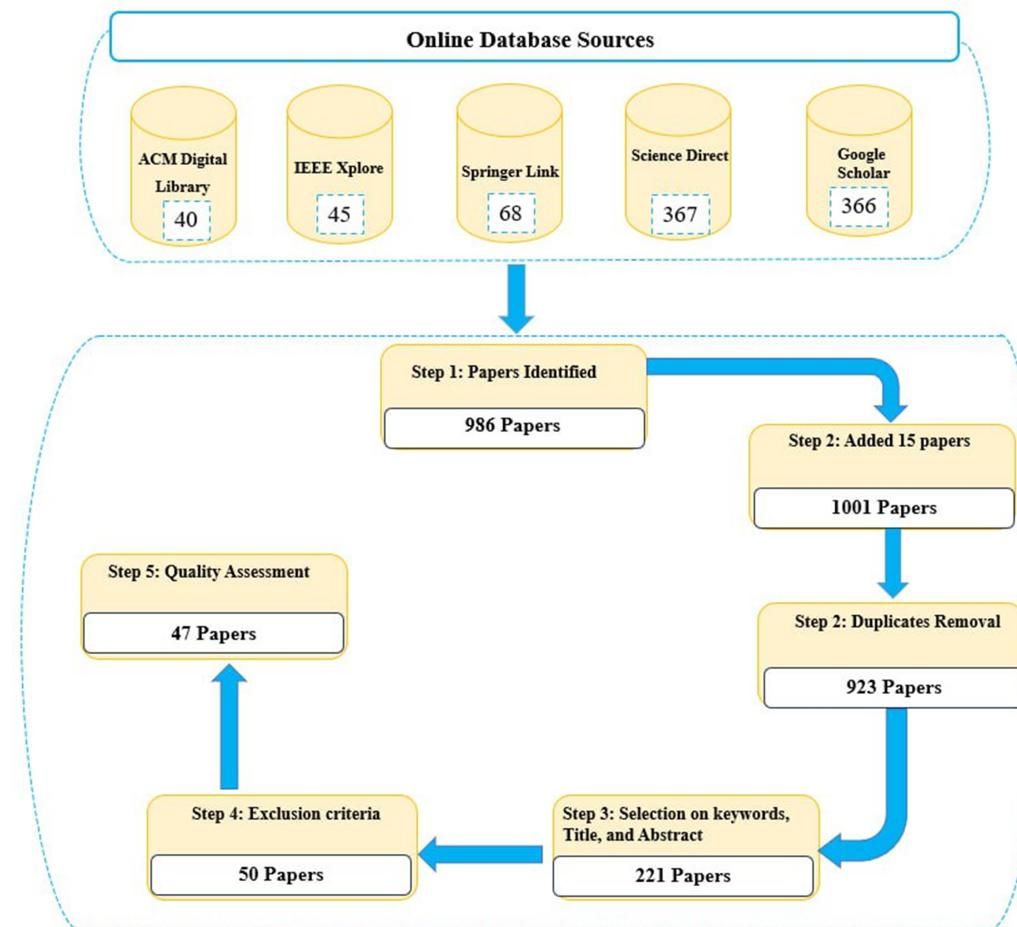
Ref	Publication Title	Year	Type	QA Questions (QA) [1, 2, 3, 4, 5, 6, 7, 8]	Sum
[51]	A Joint Learning Algorithm for Complex-Valued T-F Masks in Deep Learning-Based Single Channel Speech Enhancement Systems	2019	Conference	[1, ½, 0, 1, 1, 0, ½, 1]	5
[52]	Speech enhancement using deep neural network based on mask estimation and harmonic regeneration noise reduction for single channel microphone	2022	Journal	[0, ½, ½, 1, ½, 1, 0, ½]	4
[53]	Deep Noise Tracking Network: A Hybrid Signal Processing/Deep Learning Approach to Speech Enhancement	2018	Conference	[1, 1, ½, 1, 1, ½, ½, ½]	6
[54]	A novel Long Short Term Memory(LSTM)-based speech preprocessor for speaker diarylation in realistic mismatch conditions	2018	Journal	[1, ½, ½, 1, 1, ½, ½, 1]	5.5
[55]	Speech enhancement using a reduced complexity MFCC-based Deep Neural Network	2017	Journal	[½, 1, 1, 1, 1, ½, ½, ½]	6
[56]	Speech enhancement with phase sensitive mask estimation using a novel hybrid neural network	2021	Journal	[1, ½, ½, 1, 1, ½, ½, 1]	5.5
[57]	Adaptive spectral subtraction to improve quality of speech in mobile communication	2018	Workshop	[½, 1, 1, 1, ½, ½, ½]	6
[58]	Model-based speech enhancement for hearing aids	2018	Journal	[½, 0, 0, ½, 1, ½, 0, 0]	2.5
[59]	An NMF-based MMSE Approach for Single Channel Speech Enhancement Using Densely Connected Convolutional Network	2021	Journal	[1, 1, 1, ½, 1, 1, ½, ½]	6.5
[60]	Deep MMSE: A deep learning approach to MMSE-based noise power spectral density estimation	2020	Journal	[½, 0, ½, ½, 1, ½, 1, 0]	4
[61]	Deep learning for minimum mean-square error approaches to speech enhancement	2019	Journal	[½, 1, ½, 1, 0, 1, 0, ½]	4.5
[62]	Investigation of Single Channel Speech Enhancement: A Comparative Study	2024	Conference	[½, ½, 1, 1, ½, 0, 1, ½]	5
[63]	Speech enhancement by using deep learning algorithms	2024	Conference	[0.4, ½, ½, 1, 0, ½, 1, ½]	4.5
[64]	DNN-based speech enhancement using soft audible noise masking for wind noise reduction	2018	Conference	[½, 1, 1, ½, 1, ½, 0, 1]	5.5
[65]	Siamese Hybrid RNN Model for Speech Enhancement: A Novel Approach for Noise Reduction in Speech Signals	2023	Conference	[1, 1, 1, 1, 1, ½, 1, 1]	7.5
[66]	Design of multichannel wiener filter for speech enhancement in hearing aids and noise reduction technique	2016	Conference	[½, 1, ½, 1, 1, ½, ½, 1]	6
[67]	Reduction of noise for speech signal enhancement using Spectral Subtraction method	2016	Journal	[1, 1, 1, 1, 1, 1, 1, ½]	7.5
[68]	Speech Enhancement Algorithms for Wind Noise Reduction	2022	Conference	[½, 1, ½, ½, 1, 1, 1, ½]	6
[69]	Oblique Projection and Cepstral Subtraction in Signal Subspace Speech Enhancement for Colored Noise Reduction	2017	Workshop	[1, 0, 0, 1, ½, 1, 0, 1]	4.5

**Table 4.** Cont.

Ref	Publication Title	Year	Type	QA Questions (QA) [1, 2, 3, 4, 5, 6, 7, 8]	Sum
[70]	Performance of single-channel speech enhancement algorithms on Mandarin listeners with different immersion conditions in New Zealand English	2024	Conference	[0, ½, ½, 1, ½, ½, 1, ½]	4.5
[71]	Single Channel Speech Enhancement: Using Wiener Filtering with Recursive Noise Estimation	2016	Conference	[½, ½, 1, 0, ½, 1, 0, ½]	4
[72]	On DCT-based MMSE estimation of short time spectral amplitude for single channel speech enhancement	2023	Journal	[½, 1, ½, 1, 0, ½, 0, 1]	4.5
[73]	Single-channel speech enhancement using improved progressive deep neural network and masking-based harmonic regeneration	2022	Journal	[½, ½, ½, 1, 1, ½, 1, 0]	5
[74]	Single-channel speech enhancement with correlated spectral components: Limits-potential	2020	Journal	[0, ½, 1, ½, 0, 0, 0, ½]	2.5
[75]	Speech Enhancement using Spectral Subtraction-type Algorithms: A Comparison and Simulation Study	2015	Conference	[1, 1, 1, 1, 1, ½, 1, 1]	7.5
[76]	Speech enhancement with noise estimation and filtration using deep learning models	2023	Journal	[1, 1, 1, 1, ½, 1, 1, 1]	7.5
[77]	Speech enhancement system using deep neural network optimized with Battle Royale Optimization	2024	Workshop	[1, 1, 1, 1, 1, ½, 1, 1]	7.5
[78]	Drone audition: Audio signal enhancement from drone embedded microphones using multichannel Wiener filtering and Gaussian-mixture based post-filtering	2016	Conference	[½, 1, 1, ½, ½, ½, 1, 1]	5.5
[79]	Fast distributed multichannel speech enhancement using novel frequency domain estimators of magnitude-squared spectrum	2017	Conference	[1, 1, 1, 1, 1, 1, 1, ½]	7.5
[80]	Two-Stage Single-Channel Speech Enhancement with Multi-Frame Filtering	2023	Conference	[½, 1, 0, ½, 1, ½, 0, ½]	4.5
[81]	Wavelet based Noise Reduction Techniques for Real Time Speech Enhancement	2018	Workshop	[½, ½, 0, 0, ½, 0, 0, 0]	1.5
[82]	Noise Reduction Algorithm for Speech Enhancement	2023	Conference	[1, 1, ½, 0, 0, ½, 1, 0, ½]	4.5
[83]	Speech Enhancement And Noise Reduction In Forensic Applications	2023	Conference	[1, ½, ½, ½, 1, ½, 0, 0]	4
[84]	Deep bi-directional long short-term memory-based speech enhancement for wind noise reduction	2017	Conference	[1, 1, ½, 1, 1, 1, ½, 1]	7
[85]	Design of matrix wiener filter for noise reduction and speech enhancement in hearing aids	2016	Conference	[1, 1, 1, 1, ½, 1, 1, 1]	7.5
[86]	A review of deep learning techniques for speech processing	2023	Journal	[1, 1, 1, ½, 1, 1, 1, 1]	7.5
[87]	A Novel Single Channel Speech Enhancement Based on Joint Deep Neural Network and Wiener Filter	2015	Conference	[1, 1, 1, 1, 1, 1, 1, 1]	8
[88]	Empirical Evaluation of Hybrid Filtering: An approach for Speech Enhancement	2015	Conference	[1, 1, 1, 0, 1, 1, ½, 1]	6.5
[89]	Laplacian-Gaussian Mixture Based Dual-Gain Wiener Filter for Speech Enhancement	2016	Conference	[½, ½, ½, ½, 1, ½, 0, 1]	4.5

**Table 4.** Cont.

Ref	Publication Title	Year	Type	QA Questions (QA) [1, 2, 3, 4, 5, 6, 7, 8]	Sum
[90]	Speech Enhancement Using A Critical Point Based Wiener Filter	2017	Journal	[1, 1, 1, 1, 1, ½, 1, 1]	7.5
[91]	Speech Denoising Using Transform Domains in the Presence of Impulsive and Gaussian Noises	2017	Journal	[½, 1, ½, 1, 1, ½, ½, 1]	6
[92]	On the Importance of Super-Gaussian Speech Priors for Machine-Learning Based Speech Enhancement	2017	Journal	[½, 1, 1, ½, ½, ½, 1, 1]	5.5
[93]	DNN-BASED AR-WIENER FILTERING FOR SPEECH ENHANCEMENT	2018	Conference	[1, 1, 1, 1, 1, 1, 1, 1]	8
[94]	REAL-TIME SPEECH ENHANCEMENT USING EQUI-LIBRIATED RNN	2020	Conference	[1, 1, 1, 1, 1, 1, 1, 1]	8
[95]	Speech enhancement using modified wiener filter based MMSE and speech presence probability estimation	2020	Journal	[1, 1, 1, 1, 1, 1, 1, 1]	8
[96]	Deep neural networks for speech enhancement in complex-noisy environments	2019	Journal	[1, 1, 1, 1, 1, 1, 1, 1]	8
[97]	Speech Enhancement Algorithm Based on a Hybrid Estimator	2021	Journal	[1, 1, 1, 1, 1, 1, 1, 1]	8

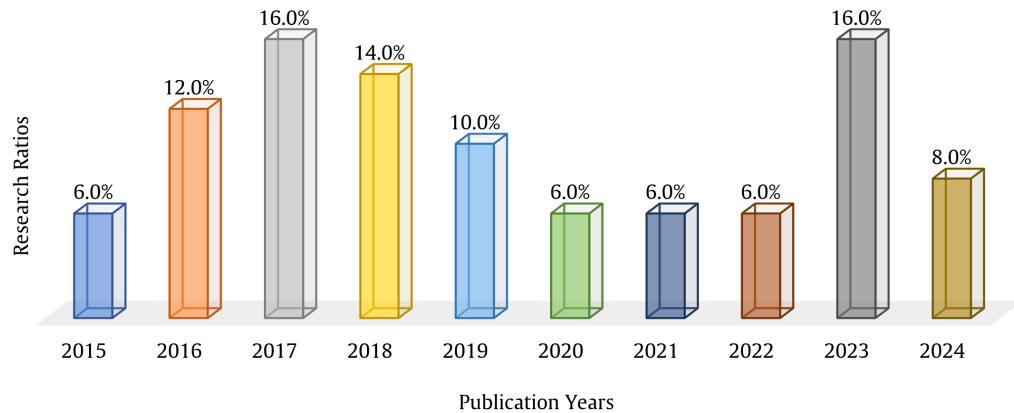
**Figure 5.** The full paper selection procedure.

## 4. Results and Discussion

The aim of this stage is to discuss and accurately record information from the candidate studies. To reduce the chance of bias, the data extraction process is clearly defined. Various types of algorithms are used for speech enhancement, in which multiple transforms and techniques have been implemented. To cover the trends in the candidate articles, an overview of these studies according to their year of publication is provided based on the 50 papers retained after the exclusion step. The year-wise distribution of the collected studies is displayed in Table 5. For more clarity and to provide a better understanding of trends in the SEA research area, the number of related research studies per year along with their percentage ratio is presented in Figure 6.

**Table 5.** Year-wise breakup of the collected studies.

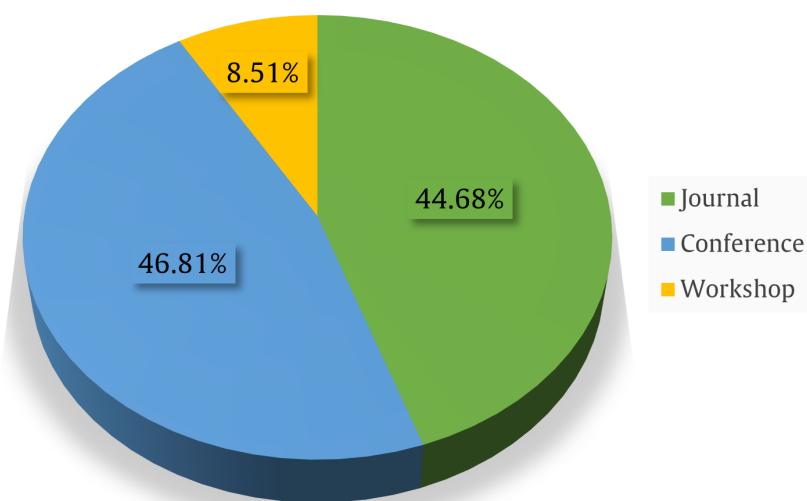
2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
[84]	[66]	[55]	[53]	[51]	[74]	[56]	[52]	[65]	[62]
[87]	[67]	[69]	[54]	[60]	[94]	[97]	[68]	[72]	[63]
[88]	[71]	[79]	[57]	[3]	[95]	[59]	[73]	[76]	[70]
	[78]	[1]	[58]	[96]				[80]	[77]
	[85]	[84]	[64]	[61]				[82]	
	[89]	[90]	[93]					[15]	
	[91]	[81]						[83]	
	[92]								[86]



**Figure 6.** Distribution of research trends related to SEA from 2015 to 2024.

Figure 7 shows the collected papers divided according to their types and sources, with the average provided for each type of publication. A final total of 47 studies were retained. As presented in Figure 7, 44.68% of the studies were published in journals, 46.81% were presented at conferences, and 8.51% were presented in workshops.

Table 6 shows that the majority of the 47 papers collected for the full review were indexed in Scopus and Web of Science. Specifically, 42 papers (89.4%) were indexed in Scopus and 39 papers (83.0%) in the Web of Science database, with a smaller subset of eight papers (17.0%) indexed in the Emerging Web of Science. This indicates that the majority of the collected literature is well-represented in established indexing platforms, which ensures a robust foundation for our systematic review. Overall, the dataset reflects a strong emphasis on credible and widely recognized sources. In this section, the research questions are answered in detail to achieve the objectives accurately and clearly.



**Figure 7.** Percentage of papers per source type.

**Table 6.** Indexing platforms for collected papers.

Indexing Platform	Number of Papers	Percentage of Total Papers (47)
Scopus	42	89.4%
Web of Science	39	83.0%
Emerging Web of Science	8	17.0%

#### RQ1. What are the most commonly used recent research approaches in the field of SEAs?

Speech enhancement processes aim to recover the desired clean speech of the damaged signal, with different approaches depending on the type of degradation and noise signal. Practically, this process remains challenging, specifically when dealing with high noise levels, nonstationary noise, and reverberation. Therefore, multiple approaches have emerged in recent years. Different types of SEA have been adopted in the reviewed papers. It is worth noting that deep learning has recently attracted the attention of many researchers. The reviewed studies on SEAs investigate varied topics, including statistical based approaches, deep learning-based approaches, and hybrid methods that incorporate different approaches such as conventional speech enhancement and deep learning techniques. Specifically, topics such as phase-sensitive masks, harmonic regeneration, and time-frequency mask estimation have become popular. Hybrid approaches combining techniques such as Wiener filtering with neural networks are commonly used to improve voice intelligibility and lower noise in complex situations such as mobile communication, hearing aids, and drones. Among the 47 reviewed studies, several cover deep learning [51,53]. Nine of the retained studies were dedicated to DNNs [52,55,64,73,76,77,80,92,96], while other studies presented CNNs for use in SEAs. CNNs are a class of deep neural architecture consisting of one or more pairs of alternating convolutional and pooling layers [86]. In speech enhancement, long-term context is important for improving and understanding speech signals and processing continuous noise. RNNs can retain long-term context via modules such as LSTMs, making them better suited to handling long or complex signals. CNNs work well with short contexts, but require deeper networks or additional techniques such as dilated convolutions in order to simulate long-term effects. Therefore, RNNs were used in several of the reviewed studies. The SEA presented in [15] utilizes a CNN and a Fully-Connected Neural Network (FCNN). Another study used a novel LSTM-based speech preprocessor for speaker diarization in a realistic mismatch scenario [54]. To address the nonstationarity of noise, one work used a Recurrent Neural Network (RNN)-based speech

enhancement system to reduce wind noise [84], while another used an equilibrated RNN for real-time enhancement [94]. It is noteworthy that several studies were dedicated to hybrid speech enhancement algorithms. These dedicated hybrid forms of deep learning included RNN(LSTM)+CNN [56] and CNN+RNN [86] varieties. A different fundamental concepts using deep learning for speech processing was presented [86]. These methods show that deep learning is becoming more sophisticated and adaptable in solving speech enhancement problems.

Another major approach is based on statistical models and optimization criteria, placing the problem of speech enhancement within a statistical and estimation framework. Among the retained studies, 18 out of 47 cover statistical approaches such as Wiener filtering (eight studies) [66,71,78,85,89–91,95], MMSE estimation (six studies) [1,3,68,72,79,97], or spectral subtraction (three studies) [57,67,75]. Spectral subtraction is a classical speech enhancement method that estimates the noise spectrum during speech pauses and subtracts it from the noisy spectrum to obtain the estimated clean signal. This process can also be performed by multiplying the noisy spectrum using the gain function and combining it with the phase of the degraded speech signal. The final study in this category uses a wavelet denoising approach [82]. Of the 47 reviewed papers, ten are dedicated to hybrid forms of speech enhancement: Wiener Deep Neural Network (WDNN) [87,93]; MMSE+DL [60]; MMSE+Densely Connected Convolution Network (DCCN) [59]; DL+MMSE [61]; spectral subtraction with Wiener filter, CNN, and GNN [63]; hybrid Wiener filter for 1D-2D [88]; oblique projection and cepstral subtraction [69]; Mel-Frequency Cepstral Coefficients (MFCC)+DNN [83]; and WF generalized subspace+deep complex U-Net [70]. A summary of the main types of algorithms used in SEAs is provided in Table 7.

**Table 7.** Main types of algorithms used in SEAs.

Algorithm Name	Algorithm References	Advantages	Disadvantages
Deep Neural Networks (DNN)	[52,55,64,73,76,77,80,92,96]	Adaptable to complex noise; data-driven; handles non-linear relationships.	Requires large datasets; computationally intensive; prone to overfitting.
Convolutional Neural Networks (CNN)	[15,86]	Effective for local feature extraction; scalable; translation-invariant.	Limited long-term context; struggles with temporal dynamics.
RNN/LSTM	[54,84,94]	Captures sequential dependencies; handles long-term noise patterns.	Slow training; vanishing/exploding gradients; computationally heavy.
Hybrid (RNN + CNN)	[56,86]	Combines spatial (CNN) and temporal (RNN) modeling; robust for dynamic noise.	Complex integration; difficult to optimize; high resource demands.
Wiener Filtering	[66,71,78,85,89–91,95]	Mathematically simple; effective for stationary noise.	Assumes noise statistics are known; poor for non-stationary noise.
MMSE Estimation	[1,3,68,72,79,97]	Minimizes mean square error; statistically robust.	Computationally intensive; requires accurate noise estimation.
Spectral Subtraction	[57,67,75]	Simple implementation; fast; historically proven.	Introduces “musical noise” artifacts; relies on accurate silence detection.
Wavelet Denoising	[82]	Multi-resolution analysis; handles non-stationary noise.	Parameter-sensitive; limited effectiveness for high-frequency noise.
Hybrid (Wiener + DNN)	[87,93]	Combines model-based and data-driven approaches; improves generalization.	Requires careful hyperparameter tuning; increased complexity.

**Table 7.** Cont.

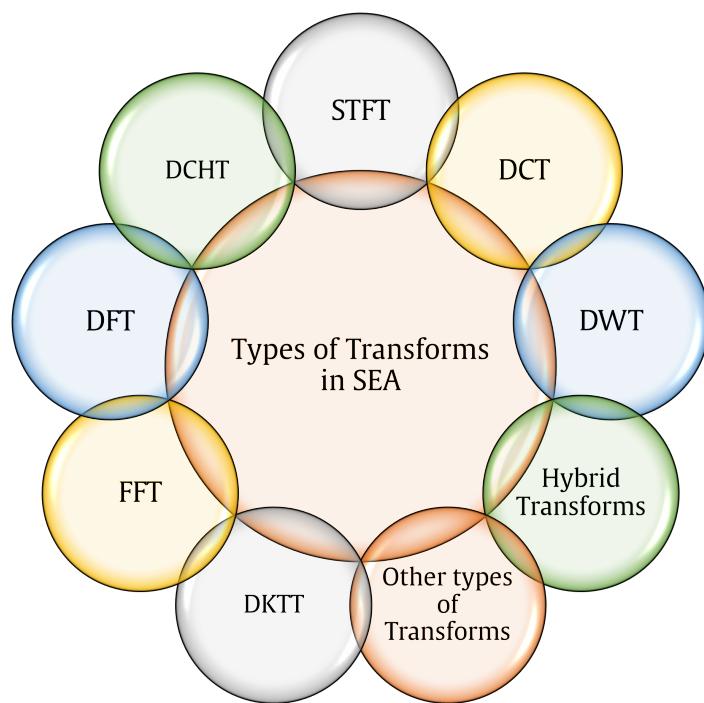
Algorithm Name	Algorithm References	Advantages	Disadvantages
Hybrid (MMSE + DCCN) [59,61]		Balances error minimization and deep feature learning; robust performance.	High computational cost; integration challenges.
Hybrid (Spectral Subtraction + CNN/GNN) [63]		Leverages traditional and deep learning methods; flexible.	Complex implementation; risk of over-smoothing.
Hybrid (1D-2D Wiener Filter) [88]		Handles multi-dimensional noise; adaptable to varied scenarios.	Limited validation in dynamic environments.
Oblique Projection + Cepstral Subtraction [69]		Reduces reverberation; preserves speech quality.	Computationally demanding; niche applicability.
MFCC + DNN [83]		Leverages spectral features (MFCC) with deep learning; improves intelligibility.	Depends on MFCC accuracy; limited to spectral domain.
Deep Complex U-net [70]		Processes complex spectral data; effective for phase-sensitive tasks.	High memory usage; requires specialized training.

#### RQ2: What are the main types of transform-based SEA used in the existing studies?

The main kinds of transform-based SEA adopted by the reviewed findings are indicated in this section. It is worth mentioning that the processing of speech enhancement can be applied in the time domain and frequency domain. However, discrete transforms have been recognized as a very useful tool in signal processing such as speech enhancement. In this approach, speech is viewed in transform domains, resulting in a massive shift in terms of robust ability to analyze the components of speech signals [98]. Discrete transforms have different properties such as energy compaction and localization, which can be used to perform transform analysis in various practical applications. The significance of transforms in sequential data processing raises the possibility of using them in SEA research. Studies dedicated to SEAs include different types of transforms used in this field. As can be observed from Figure 8, the main types of transforms used in the reviewed papers are as follows: nine studies used the Short-Time Fourier Transform (STFT) [53,60,75,78,80,84,92,94,96]; five studies used the Fast Fourier Transform (FFT) [57,66,67,90,93]; and three studies utilized the Discrete Cosine Transform (DCT) [1,72,89]. In another technique presented by [3,97], the authors used the Discrete Krawtchouk–Tchebichef Transform (DKTT), a powerful transform that has high energy compaction and localization properties, to handle speech signal coefficients. Finally, one paper was dedicated to the Discrete Fourier Transform (DFT) [56], another one used the Deep Complex Hybrid Transform (DCHT) [15], and [82] used the Wavelet Transform (WT). Additionally, some of the reviewed papers used more than one transform: FFT+DCT [86], STFT+FFT [71], DFT+STFT [61], DFT+STFT [51], DCT+DTCWPT+DWT [95], STFT+DCT [63], DFT+DCT+DWT [88], STFT+WSST (Wavelet Synchro-Squeezing Transform) [91], DFT+STFT [79], and STFT+FFT [87]. The main types of transforms used in SEAs are listed in Table 8 along with their features.

**Table 8.** Main types of transforms used in SEAs.

Transform	Transform References	Key Features
Short-Time Fourier Transform (STFT)	[53,60,75,78,80,84,92,94,96]	Time-frequency analysis using overlapping windows; effective for non-stationary signals.
Fast Fourier Transform (FFT)	[57,66,67,90,93]	Computationally efficient spectral analysis; widely used for frequency-domain processing.
Discrete Cosine Transform (DCT)	[1,72,89]	High energy compaction; useful for compression and noise reduction.
Discrete Krawtchouk-Tchebichef Transform (DKTT)	[3,97]	High energy localization; robust for speech signal coefficient analysis.
Discrete Fourier Transform (DFT)	[56]	Fundamental frequency-domain analysis; represents signals via sinusoidal components.
Deep Complex Hybrid Transform (DCHT)	[15]	Combines complex transforms with deep learning for hybrid time-frequency processing.
Wavelet Transform (WT)	[82]	Multi-resolution analysis; effective for non-stationary noise and transient signals.
FFT + DCT	[86]	Combines FFT's speed with DCT's energy compaction for enhanced signal analysis.
STFT + FFT	[71,87]	Integrates time-frequency (STFT) and spectral (FFT) analysis for advanced processing.
DFT + STFT	[51,61,79]	Merges DFT's spectral precision with STFT's time-frequency resolution.
DCT + DTCWPT + DWT	[95]	Combines DCT, Dual-Tree Complex Wavelet Packet Transform (DTCWPT), and Discrete Wavelet Transform (DWT) for multi-level analysis.
STFT + DCT	[63]	Links STFT's time-frequency analysis with DCT's energy compaction for feature extraction.
DFT + DCT + DWT	[88]	Hybrid approach for multi-domain analysis of complex signals.
STFT + WSST	[91]	Combines STFT with Wavelet Synchro Squeezing Transform (WSST) for enhanced time-frequency resolution.



**Figure 8.** The main types of transforms used in SEAs.

#### RQ3: What is the number of channels presented in the current works?

The reviewed papers presented several main approaches that improve the performance of SEA in noisy environments based on the channel number. To answer this question, we focused on selected SEA studies that examined single channel-based speech enhancement and multi-channel-based speech enhancement. Single-channel SEA uses a single-channel speech signals; in this approach, signals can be collected by a single microphone. On the other hand, multi-channel SEA works on multiple channel signals, which helps to improve the degraded signal. The single-channel scenario is more common in real situations [80]. The majority of studies focused on single-channel speech enhancement processing due to its simplicity and widespread use [1,51,53,54,56,57,59–64,67,69–72,80,84,86–88,92,94,95]. Multi-frame filtering techniques are often used for single-microphone scenarios in combination with other methods, such as Wiener filtering or machine learning [86]. Few of the retained studies addressed multi-channel approaches in SEA. These mostly addressed multi-channel scenarios based on different methods, including Wiener filtering, subspace projection, and multi-channel Wiener filters for spatial noise reduction [66,78,79,85].

#### RQ4: What are the major models used to identify speech and noise signals?

Modeling of speech and noise signals has always been an issue due to environmental randomness. The evaluated publications used range of models for noise and speech signals. These models describe the variety of approaches used by researchers to address issues related to speech enhancement, such as Wiener filter and MMSE. As presented in the existing SEAs based on statistical approaches, the Probability Density Functions (PDF) introduced to model speech and noise transform coefficients are assumed to be either Gaussian or super-Gaussian functions. Because of these assumptions, the aforementioned PDF types provide high allowable mapping patterns for the speech and noise component distribution; however, this is not always the case in real situations. The most widely used model to represent noise signal is Gaussian, although it cannot represent the true situation due to the high number of different noise characteristics in various environments [42]. Both speech and noise are assumed to obey random processes and are treated as random variables; thus, the basic models that represent speech or noise signals are Gaussian [79,87,91,93,95],

Laplacian [1,97], super-Gaussian [3,92], and Gaussian Mixture Model (GMM) [78]. Other reviewed works were based on hybrid models: GMM+HMM [86], Gaussian (noise) and Laplacian (clean speech) [89], Gaussian+Laplacian (clean speech) and Gamma (noise) [72], GMM+HMM [54], and Gaussian+Gamma [60].

#### RQ5: What are the well-known datasets used in existing papers?

Datasets are essential to advancing the development of SEA, as they provide results with scale, robustness, and confidence. They are a fundamental tool in speech analytics that provide the data from which analysts extract significant information, and can be used to conduct and identify trends in speech enhancement research. Speech signals under clean and noisy conditions are usually taken from different well-known databases to build a strong foundation of the proposed work. Based on the reviewed papers, some studies were adopted from the same datasets. The datasets adopted in the 47 included studies are listed in Table 9 along with some brief details.

**Table 9.** Summary of datasets used in SEA research based on the retrieved articles.

Dataset Name	References Dataset	Used	Description	Dataset's Reference and/or Link
CHiME-5 dataset	[86]		A collection of recordings made in a home environment. It contains 6.5 h of audio from multiple microphone arrays. It is designed to test the systems performance in noisy environments.	[99], link: <a href="https://paperswithcode.com/dataset/chime-5">https://paperswithcode.com/dataset/chime-5</a> (accessed on 28 December 2024).
CHiME-4 dataset	[53,80]		The corpus consists of real data and simulated data. Real data is recorded in 4 real noisy environments and uttered by actual talkers. Where, the simulated one has been generated by artificially mixing clean data with real-world backgrounds noise.	<a href="https://www.chimechallenge.org/challenges/chime4/index">https://www.chimechallenge.org/challenges/chime4/index</a> (accessed on 28 December 2024)
NOISEX-92 Dataset	[1,3,15,51,53,56,92,93]		It contains 15 common types of noise in real-world environments, with a length of about 4 min for each. These are highly non-stationary noises.	[100], link: <a href="http://svr-www.eng.cam.ac.uk/comp.speech/Section1/Data/noisex.html">http://svr-www.eng.cam.ac.uk/comp.speech/Section1/Data/noisex.html</a> (accessed on 29 December 2024)
DREGON Dataset	[78]		The dataset contains both clean and noisy in-flight audio recordings continuously annotated with the 3D position of the target sound source using an accurate motion capture system.	[101], link: <a href="http://dregon.inria.fr/datasets/dregon/">http://dregon.inria.fr/datasets/dregon/</a> (accessed on 29 December 2024)
TIMIT Dataset	[1,3,15,51,56,60,61,64,78,84,86,87,89,92,93,95,97]		It is a standard dataset used for evaluation speech systems. It consists of 630 speakers of 8 dialects of American English each reading 10 phonetically rich sentences.	[102], link: <a href="https://paperswithcode.com/dataset/timit">https://paperswithcode.com/dataset/timit</a> (accessed on 29 December 2024)
NOIZEUS Dataset (simulation)	[57,71,75,79,88,95]		The noisy database contains 30 IEEE sentences (produced by three male and three female speakers) corrupted by eight different real-world noises at different SNRs. The noise was taken from the AURORA database.	[103], link: <a href="https://ecs.utdallas.edu/loizou/speech/noizeus/">https://ecs.utdallas.edu/loizou/speech/noizeus/</a> (accessed on 30 December 2024)
LibriSpeech Dataset	[61,63,86]		It is a collection of about 1000 h of spoken English speech generated from public domain audiobooks. It is widely used in speech processing research and features high audio quality and clean transcription.	[104], link: <a href="https://www.openslr.org/12">https://www.openslr.org/12</a> (accessed on 30 December 2024)
TSP Dataset	[69,72]		It contains over 1400 utterances spoken by 24 speakers (half male, half female). The data was recorded in an anechoic room and includes the original samples (48 kHz sampling rate), and also the data filtered and subsampled to different sample rates.	[105], link: <a href="https://www.mmsp.ece.mcgill.ca/Documents/Data/">https://www.mmsp.ece.mcgill.ca/Documents/Data/</a> (accessed on 30 December 2024)

Datasets mentioned in the reviewed papers include CHiME-II and CHiME-III, which were used in [51] to compare the performance of algorithms in various noise environments. CHiME-4 was used to evaluate the one-channel test in [80]. Aurora-4 was used in [51] to assess system performance in various noisy environments. Other datasets adopted in the reviewed studies include SPANZ [70], BKB [70], VCTK Corpus, which provides diverse English accents for speech enhancement tasks [61], QUT [61], VoiceBank-DEMAND [94], IEEE Speech [96], RSG-10(Noise), which provides random signal generation noise [60], WSJO Corpus, AMI Corpus, and ADOS Corpus [54]. These datasets are frequently used in SEA research. They offer a variety of settings for the testing and training phases of speech enhancement algorithms. Synthetic noise datasets are often used to assess model performance and replicate difficult acoustic conditions. In future studies, larger and more varied datasets could be used to improve robustness and generalizability.

#### RQ6: What are the effective measurements for evaluating speech intelligibility and quality?

It is necessary to assess SEA performance based on effectiveness metrics that provide meaningful and reliable procedures. The analyzed studies show that two principal criteria are used to measure the goodness of speech signals, namely, quality and intelligibility. Speech quality deals with speech clarity, nature of distortion, and amount of background noise, while speech intelligibility deals with the percentage of words that can be clearly understood. Good quality does not always guarantee high intelligibility, as they are independent of each other. Therefore, most SEAs improve quality at the expense of reducing intelligibility [40,106]. In general, there are two categories of speech quality assessment, namely, subjective and objective measures. Objective measurements are preferred, since they are dependent on a mathematical notion. Many metrics are commonly used in SEA evaluation, and every metric has its own strengths and weaknesses. Perceptual Evaluation of Speech Quality (PESQ) is considered the most important objective measure based on the reviewed studies. A total of 30 out of 47 studies used PESQ, constituting a majority of the reviewed papers. Table 10 illustrates the most commonly used measurements, their abbreviations, a brief description, their percentage of use in the reviewed works, and which studies used them. The measurements mentioned in Table 10 can be considered the most widely used in the field of speech enhancement.

**Table 10.** Measurements and their descriptions.

Measurements	Description	Percentage	References
PESQ (Perceptual Evaluation of Speech Quality)	A very important objective measure. It predicts the subjective Mean Opinion Score, and its score is mapped to a MOS-like scale that is regarded an accurate estimation measure available recently.	(30 out of 47) = 63%	[1,3,51,53,56,60,61,63,64,69–72,75–80,84,86–88,91–97]
SNR (Signal to Noise ratio)	This is a commonly used metric for evaluating SEA that measures the ratio of speech signal power to noise signal power. A higher SNR ratio indicates better performance.	(11 out of 47) = 23%	[3,57,63,64,66,67,71,75,88–90]
STOI (Short-Time Objective Intelligibility)	Establishes high correlations with intelligibility of noisy and time-frequency weighted noisy speech.	(12 out of 47) = 25%	[51,60,61,63,70,72,77,78,80,86,93,96]
The composite measures, which are: a- SIG (Signal Distortion), b- BAK (Background Noise), c- OVL (Overall Quality).	A set of objective quality measures used to assess speech quality and overall effectiveness of SEA. SIG: to measure signal distortion. BAK: to measure noise distortion. OVL: to measure overall quality.	(3 out of 47) = 6%	[3,94,97]

**Table 10.** *Cont.*

Measurements	Description	Percentage	References
MOS-LQO (Mean Opinion Score - Listening Quality Objective).	It utilizes a spectro-temporal measure of similarity between a reference signal and a test signal to produce a MOS-LQO score that is ranged from 1 (the worst) to 5 (the best).	(2 out of 47) = 4%	[61,63]
SI-SNR (scale-invariant SNR).	It evaluates speech enhancement and the level of distortion in the processed signal is measured by comparison with the reference signal.	(2 out of 47) = 4%	[63,70,80]
SegSNR (segmental signal-to-noise ratio)	Measures the frame-based segmental signal-to-noise ratio (SNR), and averages SNR measurements over short frames. It is a reasonable measure of speech quality.	(12 out of 47) = 25%	[1,15,56,72,76,79,88,91–93,95–97]
SDR (Signal-to-Distortion Ratio) and its version SI_SDR (Scale-Invariant SDR)	It is used to measure signal quality in comparison to its distorted or noisy version.	(5 out of 47) = 10%	[51,53,63,80,84]
MOS (The subjective Mean Opinion Score)	It is a well-known speech quality assessment in terms of human perception is the mean opinion score of human subjective ratings.	(2 out of 47) = 4%	[63,88]

#### RQ7: What are the current limitations and challenges in the reviewed SEAs?

This section reviews the state-of-the-art approaches and previous findings related to speech enhancement problems based on the retained papers. Regardless of the successful SEA applications in various fields, many gaps and limitations still exist that should be addressed. The main concern of this research is focused on finding the significant issues that still need to be handled. The significant limitation that has been addressed in the retained papers is focusing on randomness of the signal, which has a great impact on subsequent speech and noise modeling. The nonstationarity of signals affects the statistical modeling of speech and noise signals [1,78,97]. Therefore, the noise reduction performance of many SEAs depends on noise estimation and signal modeling, especially in complex environments [61,83,96]. This issue is also related to another challenge caused by the surrounding environment in the form of simultaneous conversations and overlapping speech. This problem is termed the “cocktail party problem”, and happens when two or more people are talking at the same time [48].

Data scarcity is another issue addressed in the current research. Using shared datasets is a cost-effective and feasible way to gradually advance a research field, allowing results to be analyzed and improved. The limited public availability of real-world noisy datasets for robust model training can affect the performance of SEAs in general [56,94]. Typically, SEAs work well for enhancing speech quality but poorly for enhancing intelligibility. Different articles show a tradeoff between quality and intelligibility when seeking to improve the overall performance and accuracy of the entire SEA system; therefore, some works tend to enhance both properties together [1,3,51]. SEAs based on single-channel microphone environments aim to enhance the magnitude component in the time–frequency (TF) domain. On the other hand, the phase component of the input noisy signal is used without any processing. Accordingly, sound quality is damaged due to the mismatch between the estimated magnitude and the unprocessed phase spectra, causing perceptually disturbing artifacts that are sometimes referred to as phase mismatches [51,56,72]. In general, SEAs based on neural networks aim to learn a noisy-to-clean transformation based on a supervised learning principle. Nevertheless, the trained networks may not be effective at handling the signal, and types of noise that were not present in the training data can cause

domain mismatch between the training and test sets [107]. Domain mismatch decreases model performance in testing environments that differ from the training data [51,76,80,92]. Many recent works aim to effectively mitigate this problem [107]. In addition, the high computational complexity of deep learning models means that algorithms based on these models often require significant computational resources [51,76,80,92]. Although DL-based SEAs have made significant strides, challenges such as speech distortion and artifacts persist. These problems diminish perceived auditory quality as well as the accuracy of speech enhancement systems, particularly when employing lightweight models [108]. Other issues include phonetic distortion [61,83,92], musical noise [69], and colored noise [69,96]. Therefore, it is important to address the balance between speech distortion and noise reduction, deal with speech and noise modeling in real situations, and increase quality and intelligibility simultaneously. The main challenges of SEA are shown in Figure 9. It should be noted that this research focused on papers published between 2015 and 2024, providing an overview of recent challenges and trends in SEA research.



**Figure 9.** Main challenges in SEA research.

## 5. Conclusions

This study presents a systematic review of current research on SEAs, with a focus on papers examining different topics in the field. Through a systematic review, we have aimed to highlight current research trends around SEAs and guide researchers to the most important topics published between 2015 and 2024. A total of 47 conference papers, articles, and workshop studies were retained for review from five research databases: Springer Link, Science Direct, IEEE Xplore, ACM Digital Library, and Google Scholar. Our research methodology included several stages: following a brief introduction and background on SEAs, we formulated our research questions, search strategy, and quality assessment items. Afterwards, we analyzed the selected studies based on their applicability to the research questions. Statistics on the most frequently used datasets, discrete transforms, and measurements in the reviewed papers are provided. The findings of this study illustrate that speech enhancement algorithms represent an emerging research field, with an increasing amount of work conducted over the years. Numerous studies focus on deep learning, along with a

number of papers focusing on statistical approaches. Several of these studies applied more than one SEA technique. The publication trends in SEA research are detailed along with the main challenges and current research gaps. Future directions of SEA research are explored as well. It is expected that this review will help researchers and provide a comprehensive systematic review of recently published SEA studies.

**Author Contributions:** Conceptualization, B.M.M.; methodology, S.T.Y. and B.M.M.; formal analysis, S.T.Y.; investigation, S.T.Y. and B.M.M.; resources, S.T.Y. and B.M.M.; writing—original draft preparation, S.T.Y. and B.M.M.; writing—review and editing, B.M.M. and S.T.Y.; visualization, B.M.M. and S.T.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article.

**Acknowledgments:** The authors would like to thank the University of Baghdad for their general support.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
BAK	Background Noise
CNN	Convolutional Neural Network
DCCN	Densely Connected Convolution Network
DCHT	Deep Complex Hybrid Transform
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DKT	Discrete Krawtchouk Transform
DKTT	Discrete Krawtchouk–Tchebichef Transform
DNN	Deep Neural Network
DTCWPT	Discrete-Time Continuous Wavelet Packet Transform
DTT	Discrete Tchebichef Transform
FCNN	Fully Connected NN
FFT	Fast Fourier Transform
GMM	Gaussian Mixture Model
GNN	Generative Neural Network
HMM	Hidden Markov Model
KLT	Karhunen–Loeve Transform
LSTM	Long Short-Term Memory
MFCC	Mel-Frequency Cepstral Coefficients
MMSE	Minimum Mean Square Error
MN	Musical Noise
MOS	Mean Opinion Score
MOS-LQO	Mean Opinion Score Listening Quality Objective
NMF	Non-negative Matrix Factorization
NSS	Nonlinear Spectral Subtraction
OVL	Overall Quality
PDF	Probability Density Function
PESQ	Perceptual Evaluation of Speech Quality
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

QA	Quality Assessment
RNN	Recurrent Neural Network
RQs	Research Questions
SDR	Signal-to-Distortion Ratio
SEA	Speech Enhancement Algorithm
SEAs	Speech Enhancement Algorithms
SegSNR	Segmental Signal-to-Noise Ratio
SIG	Signal Distortion
SI-SNR	Scale-Invariant SNR
SNR	Signal-to-Noise Ratio
STFT	Short-Time Fourier Transform
STOI	Short-Time Objective Intelligibility
TF	Time–Frequency
WDNN	Wiener Deep Neural Network
WF	Wiener Filtering
WSST	Wavelet Synchro-Squeezing Transform
WT	Wavelet Transform

## References

- Mahmmod, B.M.; Ramli, A.R.; Abdulhussian, S.H.; Al-Haddad, S.A.R.; Jassim, W.A. Low-Distortion MMSE Speech Enhancement Estimator Based on Laplacian Prior. *IEEE Access* **2017**, *5*, 9866–9881. [[CrossRef](#)]
- Berouti, M.; Schwartz, R.; Makhoul, J. Enhancement of speech corrupted by acoustic noise. In Proceedings of the ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing, Washington, DC, USA, 2–4 April 1979; Volume 4, pp. 208–211.
- Mahmmod, B.M.; Ramli, A.R.; Baker, T.; Al-Obeidat, F.; Abdulhussain, S.H.; Jassim, W.A. Speech enhancement Algorithm based on super-Gaussian modeling and orthogonal polynomials. *IEEE Access* **2019**, *7*, 103485–103504. [[CrossRef](#)]
- Mahmmod, B.M.; Ali, T.M. Speech Enhancement: A Review of Various Approaches, Trends, and challenges. In Proceedings of the 2024 17th International Conference on Development in eSystem Engineering (DeSE), Khorfakkan, United Arab Emirates, 6–8 November 2024.
- Vihari, S.; Murthy, A.S.; Soni, P.; Naik, D.C. Comparison of speech enhancement algorithms. *Procedia Comput. Sci.* **2016**, *89*, 666–676. [[CrossRef](#)]
- Boll, S. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust. Speech Signal Process.* **1979**, *27*, 113–120. [[CrossRef](#)]
- Nasir, R.J.; Abdulmohsin, H.A. A Hybrid Method for Speech Noise Reduction Using Log-MMSE. *Iraqi J. Sci.* **2025**, *66*, 860–875. [[CrossRef](#)]
- Paliwal, K.; Basu, A. A speech enhancement method based on Kalman filtering. In Proceedings of the ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, TX, USA, 6–9 April 1987; Volume 12, pp. 177–180.
- Ephraim, Y.; Malah, D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Trans. Acoust. Speech Signal Process.* **1984**, *32*, 1109–1121. [[CrossRef](#)]
- Hu, Y.; Loizou, P.C. A subspace approach for enhancing speech corrupted by colored noise. In Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, USA, 13–17 May 2002; Volume 1, p. I-573.
- Gales, M.J.; Young, S.J. Robust continuous speech recognition using parallel model combination. *IEEE Trans. Speech Audio Process.* **1996**, *4*, 352–359. [[CrossRef](#)]
- Zhang, W.; Benesty, J.; Chen, J. Single-channel noise reduction via semi-orthogonal transformations and reduced-rank filtering. *Speech Commun.* **2016**, *78*, 73–83. [[CrossRef](#)]
- Naik, D.C.; Murthy, A.S.; Nuthakki, R. A literature survey on single channel speech enhancement techniques. *Int. J. Sci. Technol. Res.* **2020**, *9*.
- Jassim, W.A.; Paramesran, R.; Zilany, M.S.A. Enhancing noisy speech signals using orthogonal moments. *IET Signal Process.* **2014**, *8*, 891–905. [[CrossRef](#)]
- Jerjees, S.A.; Mohammed, H.J.; Radeaf, H.S.; Mahmmod, B.M.; Abdulhussian, S.H. Deep Learning-Based Speech Enhancement Algorithm Using Charlier Transform. In Proceedings of the 2023 15th International Conference on Developments in eSystems Engineering (DeSE), Baghdad & Anbar, Iraq, 9–12 January 2023; pp. 100–105.
- Hu, Y.; Loizou, P.C. A generalized subspace approach for enhancing speech corrupted by colored noise. *IEEE Trans. Speech Audio Process.* **2003**, *11*, 334–341. [[CrossRef](#)]

17. Li, J.; Li, J.; Wang, P.; Zhang, Y. DCHT: Deep Complex Hybrid Transformer for Speech Enhancement. *arXiv* **2023**, arXiv:2310.19602.
18. Gnanamanickam, J.; Natarajan, Y.; Sri Preethaa, K.R. A hybrid speech enhancement algorithm for voice assistance application. *Sensors* **2021**, *21*, 7025. [CrossRef] [PubMed]
19. Yu, H.; Zhu, W.P.; Ouyang, Z.; Champagne, B. A hybrid speech enhancement system with DNN based speech reconstruction and Kalman filtering. *Multimed. Tools Appl.* **2020**, *79*, 32643–32663. [CrossRef]
20. Chen, B.; Loizou, P.C. A Laplacian-based MMSE estimator for speech enhancement. *Speech Commun.* **2007**, *49*, 134–143. [CrossRef] [PubMed]
21. Ambikairajah, E.; Tattersall, G.; Davis, A. Wavelet transform-based speech enhancement. In Proceedings of the Fifth International Conference on Spoken Language Processing, Sydney, Australia, 30 November–4 December 1998.
22. Soon, Y.; Koh, S.N.; Yeo, C.K. Noisy speech enhancement using discrete cosine transform. *Speech Commun.* **1998**, *24*, 249–257. [CrossRef]
23. Rao, K.R.; Yip, P. *Discrete Cosine Transform: Algorithms, Advantages, Applications*; Academic Press: Cambridge, MA, USA, 2014.
24. Huang, X.; Acero, A.; Hon, H.W.; Reddy, R. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*; Prentice Hall PTR: Hoboken, NJ, USA, 2001.
25. Ahmed, N.; Natarajan, T.; Rao, K.R. Discrete cosine transform. *IEEE Trans. Comput.* **2006**, *100*, 90–93. [CrossRef]
26. Britanak, V.; Yip, P.C.; Rao, K.R. *Discrete Cosine and Sine Transforms: General Properties, Fast Algorithms and Integer Approximations*; Elsevier: Amsterdam, The Netherlands, 2010.
27. Princen, J.; Johnson, A.; Bradley, A. Subband/transform coding using filter bank designs based on time domain aliasing cancellation. In Proceedings of the ICASSP’87. IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, TX, USA, 6–9 April 1987; Volume 12, pp. 2161–2164.
28. Wang, H.; Pandey, A.; Wang, D.L. A systematic study of DNN based speech enhancement in reverberant and reverberant-noisy environments. *Comput. Speech Lang.* **2025**, *89*, 101677. [CrossRef]
29. Chhetri, S.; Joshi, M.S.; Mahamuni, C.V.; Sangeetha, R.N.; Roy, T. Speech Enhancement: A Survey of Approaches and Applications. In Proceedings of the 2023 2nd International Conference on Edge Computing and Applications (ICECAA), Namakkal, India, 19–21 July 2023; pp. 848–856.
30. Botchev, V. Speech enhancement: Theory and practice. *Comput. Rev.* **2013**, *54*, 604–605.
31. Quackenbush, S.R.; Barnwell, T.P.; Clements, M.A. *Objective Measures of Speech Quality*; Prentice Hall: Hoboken, NJ, USA, 1988. Available online: <https://books.google.iq/books?id=Dj1BAQAAIAAJ> (accessed on 5 February 2025).
32. Moore, B.C. *An Introduction to the Psychology of Hearing*; Brill: Leiden, The Netherlands, 2012. Available online: <https://books.google.iq/books?id=LM9U8e28pLMC> (accessed on 5 February 2025).
33. Fletcher, H.; Munson, W.A. Loudness, its definition, measurement and calculation. *Bell Syst. Tech. J.* **1933**, *12*, 377–430. [CrossRef]
34. Plomp, R. Rate of decay of auditory sensation. *J. Acoust. Soc. Am.* **1964**, *36*, 277–282. [CrossRef]
35. Blauert, J. *Spatial Hearing: The Psychophysics of Human sound Localization*; MIT Press: Cambridge, MA, USA, 1997.
36. Ramig, L.O.; Sapir, S.; Fox, C.; Countryman, S. Changes in vocal loudness following intensive voice treatment (LSVT<sup>®</sup>) in individuals with Parkinson’s disease: A comparison with untreated patients and normal age-matched controls. *Mov. Disord. Off. J. Mov. Disord. Soc.* **2001**, *16*, 79–83. [CrossRef]
37. Scherer, K.R. Vocal affect expression: A review and a model for future research. *Psychol. Bull.* **1986**, *99*, 143. [CrossRef]
38. Titze, I.R.; Martin, D.W. Principles of Voice Production. *J. Acoust. Soc. Am.* **1998**, *104*, 1148. [CrossRef]
39. Flege, J.E. Second language speech learning: Theory, findings, and problems. *Speech Percept. Linguist. Exp. Issues Cross-Lang. Res.* **1995**, *92*, 233–277.
40. Al Banna, T.H. Hybrid Speech Enhancement Method Using Optimal Dual Filters and EMD Based Post Processing. Master’s Thesis, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, 2008.
41. Soon, I.Y.; Koh, S.N. Low distortion speech enhancement. *IEE Proc.-Vision Image Signal Process.* **2000**, *147*, 247–253. [CrossRef]
42. Abutalebi, H.R.; Rashidinejad, M. Speech enhancement based on  $\beta$ -order mmse estimation of short time spectral amplitude and laplacian speech modeling. *Speech Commun.* **2015**, *67*, 92–101. [CrossRef]
43. Martin, R. Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors. In Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, USA, 13–17 May 2002; Volume 1, p. I-253.
44. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Prisma Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann. Intern. Med.* **2009**, *151*, 264–269. [CrossRef]
45. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef]
46. Liao, B.; Ali, Y.; Nazir, S.; He, L.; Khan, H.U. Security analysis of IoT devices by using mobile computing: A systematic literature review. *IEEE Access* **2020**, *8*, 120331–120350. [CrossRef]

47. Keele, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; Technical Report; Keele University: Keele, UK; University of Durham: Durham, UK, 2007.
48. Dhouib, A.; Othman, A.; El Ghoul, O.; Khribi, M.K.; Al Sinani, A. Arabic Automatic Speech Recognition: A Systematic Literature Review. *Appl. Sci.* **2022**, *12*, 8898. [[CrossRef](#)]
49. Stapic, Z.; López, E.G.; Cabot, A.G.; de Marcos Ortega, L.; Strahonja, V. Performing systematic literature review in software engineering. In Proceedings of the Central European Conference on Information and Intelligent Systems. Faculty of Organization and Informatics Varazdin, Varaždin, Croatia, 19–21 September 2012; pp. 441–447.
50. Kmet, L.M. *Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields*; Alberta Heritage Foundation for Medical Research: Edmonton, AB, Canada, 2004. Available online: <https://coilink.org/20.500.12592/3rfzhd> (accessed on 5 February 2025).
51. Lee, J.; Kang, H.G. A Joint Learning Algorithm for Complex-Valued T-F Masks in Deep Learning-Based Single-Channel Speech Enhancement Systems. *IEEE/ACM Trans. Audio Speech Lang. Proc.* **2019**, *27*, 1098–1109. [[CrossRef](#)]
52. Md Jamal, N. Speech Enhancement Using Deep Neural Network Based on Mask Estimation and Harmonic Regeneration Noise Reduction for Single Channel Microphone. Ph.D. Thesis, Universiti Tun Hussein Onn Malaysia, Parit Raja, Malaysia, 2022.
53. Nie, S.; Liang, S.; Liu, B.; Zhang, Y.; Liu, W.; Tao, J. Deep Noise Tracking Network: A Hybrid Signal Processing/Deep Learning Approach to Speech Enhancement. In Proceedings of the 19th Annual Conference of the International Speech Communication, INTERSPEECH 2018, Hyderabad, India, 2–6 September 2018; pp. 3219–3223. Available online: [https://www.isca-archive.org/interspeech\\_2018/nie18\\_interspeech.html](https://www.isca-archive.org/interspeech_2018/nie18_interspeech.html) (accessed on 5 February 2025).
54. Sun, L.; Du, J.; Gao, T.; Lu, Y.D.; Tsao, Y.; Lee, C.H.; Ryant, N. A novel LSTM-based speech preprocessor for speaker diarization in realistic mismatch conditions. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 5234–5238.
55. Razani, R. *Speech Enhancement Using a Reduced Complexity MFCC-Based Deep Neural Network*; McGill University (Canada): Montreal, QC, Canada, 2017.
56. Hasaninezhad, M.; Ouyang, Z.; Zhu, W.P.; Champagne, B. Speech enhancement with phase sensitive mask estimation using a novel hybrid neural network. *IEEE Open J. Signal Process.* **2021**, *2*, 136–150. [[CrossRef](#)]
57. Purushotham, U.; Suresh, K. Adaptive spectral subtraction to improve quality of speech in mobile communication. *Int. J. Commun. Networks Distrib. Syst.* **2018**, *21*, 297–314. [[CrossRef](#)]
58. Kavalekalam, M.S. Model-Based Speech Enhancement for Hearing Aids. Ph.D. Thesis, Aalborg University, Aalborg, Denmark, 2018. Available online: <https://vbn.aau.dk/da/publications/model-based-speech-enhancement-for-hearing-aids> (accessed on 5 February 2025).
59. Li, X.; Bao, C.; Cui, Z. An NMF-based MMSE Approach for Single Channel Speech Enhancement Using Densely Connected Convolutional Network. In Proceedings of the 2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Xi'an, China, 17–19 August 2021; pp. 1–5.
60. Zhang, Q.; Nicolson, A.; Wang, M.; Paliwal, K.K.; Wang, C. DeepMMSE: A deep learning approach to MMSE-based noise power spectral density estimation. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 1404–1415. [[CrossRef](#)]
61. Nicolson, A.; Paliwal, K.K. Deep learning for minimum mean-square error approaches to speech enhancement. *Speech Commun.* **2019**, *111*, 44–55. [[CrossRef](#)]
62. Ma, S.; Lv, Z.; Du, F.; Zhang, X. Investigation of Single Channel Speech Enhancement: A Comparative Study. In Proceedings of the 2024 4th International Conference on Neural Networks, Information and Communication (NNICE), Guangzhou, China, 19–21 January 2024; pp. 364–370.
63. Cui, J. Speech Enhancement by Using Deep Learning Algorithms. Ph.D. Thesis, University of Southampton, Gelang Patah, Malaysia, 2024.
64. Bai, H.; Ge, F.; Yan, Y. DNN-based speech enhancement using soft audible noise masking for wind noise reduction. *China Commun.* **2018**, *15*, 235–243. [[CrossRef](#)]
65. Saha, B.; Vasundhara, R., V.D. Siamese Hybrid RNN Model for Speech Enhancement: A Novel Approach for Noise Reduction in Speech Signals. In Proceedings of the 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India, 10–11 August 2023; pp. 1676–1680. [[CrossRef](#)]
66. Modhave, N.; Karuna, Y.; Tonde, S. Design of multichannel wiener filter for speech enhancement in hearing aids and noise reduction technique. In Proceedings of the 2016 Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, India, 19 November 2016; pp. 1–4. [[CrossRef](#)]
67. Saldanha, J.C.; Shruthi, O.R. Reduction of noise for speech signal enhancement using Spectral Subtraction method. In Proceedings of the 2016 International Conference on Information Science (ICIS), Kochi, India, 12–13 August 2016; pp. 44–47. [[CrossRef](#)]
68. Rakshitha, A.H.; R, S.; Chikkamath, S.; Desai, S.D. Speech Enhancement Algorithms for Wind Noise Reduction. In Proceedings of the 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 27–29 May 2022; pp. 1–6. [[CrossRef](#)]

69. Surendran, S.; Kumar, T.K. Oblique Projection and Cepstral Subtraction in Signal Subspace Speech Enhancement for Colored Noise Reduction. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 2328–2340. [[CrossRef](#)]
70. Zhang, Y.C.; Hioka, Y.; Hui, C.T.J.; Watson, C.I. Performance of single-channel speech enhancement algorithms on Mandarin listeners with different immersion conditions in New Zealand English. *Speech Commun.* **2024**, *157*, 103026. [[CrossRef](#)]
71. Upadhyay, N.; Jaiswal, R.K. Single Channel Speech Enhancement: Using Wiener Filtering with Recursive Noise Estimation. *Procedia Comput. Sci.* **2016**, *84*, 22–30. [[CrossRef](#)]
72. Shi, S.; Paliwal, K.; Busch, A. On DCT-based MMSE estimation of short time spectral amplitude for single-channel speech enhancement. *Appl. Acoust.* **2023**, *202*, 109134. [[CrossRef](#)]
73. Ping, H.; Yafeng, W. Single-channel speech enhancement using improved progressive deep neural network and masking-based harmonic regeneration. *Speech Commun.* **2022**, *145*, 36–46. [[CrossRef](#)]
74. Mowlaei, P.; Stahl, J.K.W. Single-channel speech enhancement with correlated spectral components: Limits-potential. *Speech Commun.* **2020**, *121*, 58–69. [[CrossRef](#)]
75. Upadhyay, N.; Karmakar, A. Speech Enhancement using Spectral Subtraction-type Algorithms: A Comparison and Simulation Study. *Procedia Comput. Sci.* **2015**, *54*, 574–584. [[CrossRef](#)]
76. Kantamaneni, S.; Charles, A.; Babu, T.R. Speech enhancement with noise estimation and filtration using deep learning models. *Theor. Comput. Sci.* **2023**, *941*, 14–28. [[CrossRef](#)]
77. Kumar Shukla, N.; Shajin, F.H.; Rajendran, R. Speech enhancement system using deep neural network optimized with Battle Royale Optimization. *Biomed. Signal Process. Control* **2024**, *92*, 105991. [[CrossRef](#)]
78. Manamperi, W.N.; Abhayapala, T.D.; Samarasinghe, P.N.; Zhang, J.A. Drone audition: Audio signal enhancement from drone embedded microphones using multichannel Wiener filtering and Gaussian-mixture based post-filtering. *Appl. Acoust.* **2024**, *216*, 109818. [[CrossRef](#)]
79. Tu, J.; Xia, Y. Fast distributed multichannel speech enhancement using novel frequency domain estimators of magnitude-squared spectrum. *Speech Commun.* **2015**, *72*, 96–108. [[CrossRef](#)]
80. Lin, S.; Zhang, W.; Qian, Y. Two-Stage Single-Channel Speech Enhancement with Multi-Frame Filtering. *Appl. Sci.* **2023**, *13*, 4926. [[CrossRef](#)]
81. Ravi, B.R.; Deepu, S.P.; Ramesh Kini, M.; Sumam, D.S. Wavelet based Noise Reduction Techniques for Real Time Speech Enhancement. In Proceedings of the 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 22–23 February 2018; pp. 846–851. [[CrossRef](#)]
82. Kumar, S.P.; Daripelly, A.; Rampelli, S.M.; Nagireddy, S.K.R.; Badishe, A.; Attanthy, A. Noise Reduction Algorithm for Speech Enhancement. In Proceedings of the 2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT), Karaikal, India, 25–26 May 2023; pp. 1–5.
83. Bharti, S.; Jha, P.; Arora, M.; Kumar, A. Speech Enhancement And Noise Reduction In Forensic Applications. In Proceedings of the 2023 26th Conference of the Oriental COCOSDA International Committee for the Co-Ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA), Delhi, India, 4–6 December 2023; pp. 1–5. [[CrossRef](#)]
84. Lee, J.; Kim, K.; Shabestary, T.; Kang, H.G. Deep bi-directional long short-term memory based speech enhancement for wind noise reduction. In Proceedings of the 2017 Hands-free Speech Communications and Microphone Arrays (HSCMA), San Francisco, CA, USA, 1–3 March 2017; pp. 41–45. [[CrossRef](#)]
85. Modhave, N.; Karuna, Y.; Tonde, S. Design of matrix wiener filter for noise reduction and speech enhancement in hearing aids. In Proceedings of the 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 20–21 May 2016; pp. 843–847. [[CrossRef](#)]
86. Mehrishi, A.; Majumder, N.; Bharadwaj, R.; Mihalcea, R.; Poria, S. A review of deep learning techniques for speech processing. *Inf. Fusion* **2023**, *99*, 101869. [[CrossRef](#)]
87. Han, W.; Zhang, X.; Min, G.; Zhou, X. A novel single channel speech enhancement based on joint Deep Neural Network and Wiener Filter. In Proceedings of the 2015 IEEE International Conference on Progress in Informatics and Computing (PIC), Nanjing, China, 18–20 December 2015; pp. 163–167.
88. Mergu, R.R.; Dixit, S.K. Empirical evaluation of hybrid filtering: An approach for speech enhancement. In Proceedings of the 2015 International Conference on Information Processing (ICIP), Pune, India, 16–19 December 2015; pp. 139–144.
89. Wei, J.; Ou, S.; Shen, S.; Gao, Y. Laplacian-Gaussian mixture based dual-gain wiener filter for speech enhancement. In Proceedings of the 2016 IEEE International Conference on Signal and Image Processing (ICSIP), Beijing, China, 13–15 August 2016; pp. 543–547.
90. Lu, M.; Zhou, X.; Jaber, N.; Hua, K.; Ali, M. Speech enhancement using a critical point based Wiener Filter. In Proceedings of the 2017 Advances in Wireless and Optical Communications (RTUWO), Riga, Latvia, 2–3 November 2017; pp. 175–179.
91. Liu, H.; Zhang, R.; Zhou, Y.; Jing, X.; Truong, T.K. Speech denoising using transform domains in the presence of impulsive and gaussian noises. *IEEE Access* **2017**, *5*, 21193–21203. [[CrossRef](#)]
92. Rehr, R.; Gerkmann, T. On the importance of super-Gaussian speech priors for machine-learning based speech enhancement. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *26*, 357–366. [[CrossRef](#)]

93. Yang, Y.; Bao, C. DNN-based AR-Wiener filtering for speech enhancement. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 2901–2905.
94. Takeuchi, D.; Yatabe, K.; Koizumi, Y.; Oikawa, Y.; Harada, N. Real-time speech enhancement using equilibrated RNN. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 851–855.
95. Bolisetty, V.V.; Yedukondalu, U.; Santiprabha, I. Speech enhancement using modified wiener filter based MMSE and speech presence probability estimation. *Int. J. Inform. Commun. Technol.* **2020**, *9*, 63–72. [CrossRef]
96. Saleem, N.; Khattak, M.I. Deep Neural Networks for Speech Enhancement in Complex-Noisy Environments. *Int. J. Interact. Multimed. Artif. Intell.* **2020**, *6*, 84–90. [CrossRef]
97. Mahmmud, B.M.; Abdulhussain, S.H.; Naser, M.A.; Alsabah, M.; Mustafina, J. Speech enhancement algorithm based on a hybrid estimator. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1090*, 12102. [CrossRef]
98. Mahmmud, B.M.; bin Ramli, A.R.; Abdulhussain, S.H.; Al-Haddad, S.A.R.; Jassim, W.A. Signal compression and enhancement using a new orthogonal-polynomial-based discrete transform. *IET Signal Process.* **2018**, *12*, 129–142. [CrossRef]
99. Barker, J.; Watanabe, S.; Vincent, E.; Trmal, J. The fifth’CHiME’speech separation and recognition challenge: Dataset, task and baselines. *arXiv* **2018**, arXiv:1803.10609.
100. Varga, A.; Steeneken, H.J. Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Commun.* **1993**, *12*, 247–251. [CrossRef]
101. Strauss, M.; Mordel, P.; Miquet, V.; Deleforge, A. DREGON: Dataset and methods for UAV-embedded sound source localization. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1–8.
102. Garofolo, J.S.; Lamel, L.F.; Fisher, W.M.; Pallett, D.S.; Dahlgren, N.L.; Zue, V.; Fiscus, J.G. *TIMIT Acoustic-Phonetic Continuous Speech Corpus*; Linguistic Data Consortium: Philadelphia, PA, USA, 1993. Available online: <https://catalog.ldc.upenn.edu/LDC93S1> (accessed on 5 February 2025).
103. Hu, Y. Subjective evaluation and comparison of speech enhancement algorithms. *Speech Commun.* **2007**, *49*, 588–601. [CrossRef]
104. Panayotov, V.; Chen, G.; Povey, D.; Khudanpur, S. Librispeech: An asr corpus based on public domain audio books. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 19–24 April 2015; pp. 5206–5210.
105. Kabal, P. *TSP Speech Database*; Database Version 2; McGill University: Montreal, QC, Canada, 2002.
106. Hussein, H.A.; Hameed, S.M.; Mahmmud, B.M.; Abdulhussain, S.H.; Hussain, A.J. Dual Stages of Speech Enhancement Algorithm Based on Super Gaussian Speech Models. *J. Eng.* **2023**, *29*, 1–13. [CrossRef]
107. Frenkel, L.; Goldberger, J.; Chazan, S.E. Domain adaptation for speech enhancement in a large domain gap. In Proceedings of the INTERSPEECH 2023, Dublin, Ireland, 20–24 August 2023; pp. 2458–2462.
108. Guan, H.; Dai, W.; Wang, G.; Tan, X.; Li, P.; Liang, J. Reducing Speech Distortion and Artifacts for Speech Enhancement by Loss Function. In Proceedings of the 25th Annual Conference of the International Speech Communication, INTERSPEECH 2024, Kos Island, Greece, 1–5 September 2024; pp. 1730–1734.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.