International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Enhancing Speech Quality: A Review of Noise Reduction Technologies

Shalu Pal[1], B. L. Pal[2]
[1]Computer Science & Engineering (Mewar University)Shalupal14@gmail.com
[2]Computer Science & Engineering (Mewar University)Blpal@mewaruniversity.co.in

Abstract- Speech enhancement is a critical area of research focusing on improving the intelligibility and quality of speech signals in the presence of background noise. This paper explores various filtering techniques employed for speech enhancement, particularly those aimed at noise reduction. We categorize filtering methods into two primary types: linear filtering and non-linear filtering. Linear techniques, including Wiener filtering and spectral subtraction, are discussed for their effectiveness in reducing stationary noise. In contrast, non-linear approaches, such as median filtering and adaptive filtering, are analyzed for their superior performance in more dynamic noise environments.

By evaluating the strengths and limitations of these methods, we provide a comprehensive overview of the current state of speech enhancement technologies. The findings highlight the significance of selecting appropriate filtering strategies based on the specific characteristics of background noise and the requirements of the target application. Overall, this study aims to contribute to the ongoing development of more effective speech enhancement solutions that can significantly improve communication quality in various environments.

Key words: Wiener filtering, Adaptive filtering, Spectral subtraction, Median filtering, Kalman filtering

1.INTRODUCTION

Your words may become more difficult to understand as a result. A clear speech pattern is one in which each word, sentence, and concept is expressed simply and clearly. The quality of speech transfer to the audience is referred to as speech clarity. It might be challenging to hear speech in a room that is reverberant and has distracting background noise. Imagine being surrounded by a lot of noise during a crucial online meeting. The washing machine is running, a fan is turned on, the dog is barking, children are playing, and construction is taking place.

Imagine being surrounded by a lot of noise during a crucial online meeting. You have to answer a call when the kids are playing, the dog is barking, the washing machine is running, a fan is on, and construction is taking place nearby. Most of the time, it is almost impossible to find a peaceful spot or to halt the noise. To enhance the quality of online meetings in these circumstances, we require specialized audio processing technologies that can eliminate background noise.

This is one of the best applications of speech enhancement.

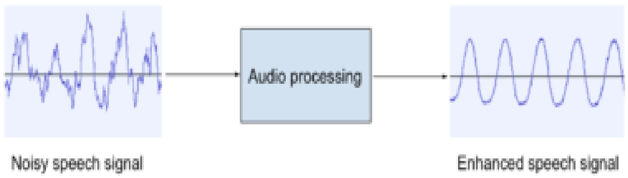

Fig-1: Speech Enhancement

It reduces or eliminates background noise to enhance the quality of a noisy spoken signal. Other words including noise cancellation (NC), noise reduction, noise suppression, and speech separation are occasionally used interchangeably with speech enhancement.

Applications for speech enhancement:

1. Voice communication via mobile devices, voice chats, conferencing apps, and other means. For speakers in noisy settings like restaurants, offices, or congested streets, SE algorithms enhance speech intelligibility.
2. Increasing the noise-robustness of various audio processing techniques. For example, speech augmentation can be used before a signal is sent to systems such as voice conversion, speaker identification, speech recognition, and emotion recognition.
3. Hearing aids. In noisy settings, speech may be

completely inaudible to people with hearing difficulties. Intelligibility is increased by reducing noise.

(A) SPEECH PRODUCTION

A change in air pressure creates a sound wave, which our ears process with the help of our brain to generate speech. Continuous articulator movements are the basis for speech, which is a complex process. The vocal tract is where speech begins, and the lips modulate the waves and different articulator movements of the tongue and teeth. Air waves leave the body through the mouth and nose to form speech. Human hearing ranges from 2 Hz to 20 kHz, while human speech ranges from 85 Hz to 8 kHz. Speech is composed of words, and each word is composed of phonemes. Since there are no restrictions on the quantity, speech is a dynamic process without any clearly defined components.Speech is composed of words, and each word is composed of phonemes. Since there are no restrictions on the quantity of units or words that can be employed, speech is a dynamic activity without any clearly defined components.

Fig-2: Audio Stream of Continuous Speech

(B)SPEECH RECOGNITION

The technology utilized in IVR settings and other task-oriented voice recognition applications is the main topic of this section. It is sometimes described as speaker independent and perhaps possessing a large vocabulary (hundreds of thousands of words). Speech recognition can be thought of as the process of converting a continuous-time audio stream into a series of distinct words or components, such phonemes. But because speech signals vary from speaker to speaker and within speakers, the conversion process is not straightforward. Speech signals are not composed of discrete, well-formed sounds, but rather of a series of "target" sounds, some of which are fairly short, with transitions in between.

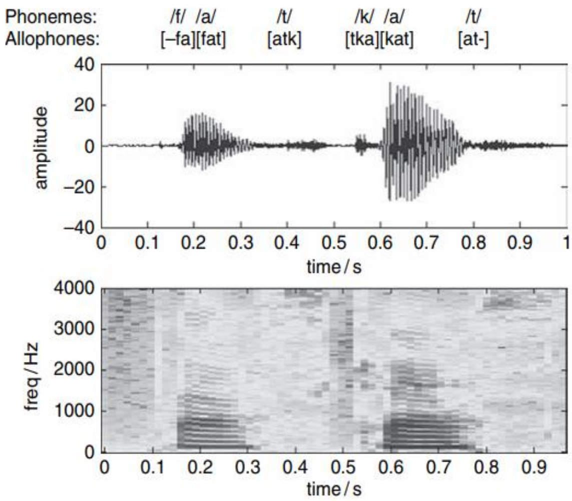

Fig-3: Time-series (upper) and time-frequency (spectrogram) (lower)

Characteristics Of Speech Recognition System

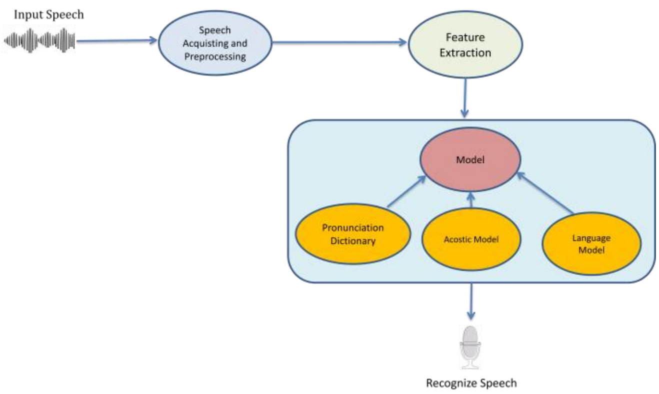

Fig-4: Basic components of a speech recognise

The following are basic terms used in speech recognition systems:

•Phonemes: The fundamental unit of language sound

•Graphemes: fundamental units of text

•Utterances: A spoken word or words, a sentence, or even several sentences that convey a single idea are called utterances. The ASR system is divided into the following categories according to the kinds of utterances it can recognize.

•Isolated words: This kind of voice recognition software may identify several words that are separated by a pause or a silent condition in addition to recognizing a single word. The "Listen/not-Listen" requirements of these ASR systems require the speaker to pause in between sentences.

• Connected words: This method is similar to identifying individual words, but it enables users to identify words with little to no space between them.

• Continuous speech: Although there are now a number of obstacles to the development of such a system, it enables users to converse regularly.

• Spontaneous speech: Spontaneous speech is the natural sound that occurs naturally around us and isn't practiced or recorded. Alexa, Siri, and Cortana are a few examples of these systems.

2. Noise Definition:

In a communication system, Unwanted or undesirable signals that are randomly added to the signal that is really carrying information or mixed with a voice signal during speech signal synthesis or transmission are known as noise in a communication system. The initial signal that is being sent from one end to the other is thereby disrupted. The phrase can also refer to signals that are random (unpredictable) and don't provide any useful information, as noises in seismic data or sonar images, even if they aren't interfering with other signals or may have been intentionally produced as comfort noise.Inother words, noise is a signal that conveys information about its origins and the surrounding environment.

For example, background voice conversations in a crowded setting may interfere with the ability to hear a desired speech or conversation, while the sound of the 0061 car engine provides information about the engine's condition and efficiency.

. Here, there are two distinct kinds of noises:

1-Based on life:

There are several forms of noise in our lives, but generally speaking, there are four recognised sorts of noise:

• Continuous noise: a sound produced continuously, such as by machinery that operates nonstop.

• Intermittent noise is defined as a noise level that fluctuates quickly. A passing train could be the cause of this.

•Impulsive noise: typically associated with the construction and demolition sectors. These deafening noises. Impulsive noises are often produced by construction equipment and explosions.

• Low-frequency noise: This type of noise permeates every aspect of our daily environment. Low-frequency noise is all around us, whether it's the roaring of enormous diesel engines or the low background hum of a nearby nuclear power plant.

2- According to signal processing (colored noise):

Signal processing (coloured noise) states that a variety of noises are carried by signals during generation or transmission, including additive noise (white noise, additive white Gaussian noise, black noise, Gaussian noise, pink noise or flicker noise, brownian noise, contaminated Gaussian noise, power-law noise, Cauchy noise, and multiplicative noise), quantization noise, poisson noise, shot noise, transient noise, burst noise, phase noise, background noise, comfort noise, and electromagnetically induced noise.

3. LINEAR AND NON-LINEARFILTERING

In audio processing, speech improvement and noise reduction are essential jobs, especially for enhancing clarity in hearing aids, telecommunication, and many other applications. In order to achieve effective improvement, filtering procedures are essential. These methods fall into two general categories: linear and nonlinear filtering. Let's investigate both: The output of a linear filter is a linear combination of the input samples after applying a linear operation to the incoming voice signal.

The following are some of the main features of linear filtering:

(i) Homogeneity and Additivity: Linear filters meet the requirements of superposition, which means that the filter's reaction to the whole of its inputs is equal to the sum of its responses to each input separately.

(ii) Finite Impulse Response (FIR) Filters: These filters are made with a set of coefficients that specify their behavior and have a finite period response.

(iii) Indefinite Impulse Response (IIR) Filters: These filters use a feedback mechanism that lets previous outputs affect subsequent outputs, enabling an indefinite length response.

(iv) Modeling Speech and Noise**:**These algorithms rely on statistical models to represent the characteristics of both the desired speech signal and the interfering noise.

(v) Estimating Signal-to-Noise Ratio (SNR):A crucial step is estimating the a priori SNR, which describes the ratio of speech power to noise power, as this information is used to determine how much noise to suppress.

(vi) Noise Reduction Techniques**:**Once the models and SNR are estimated, various techniques are employed to reduce or remove the noise, such as spectral subtraction, Wiener filtering, or other noise reduction methods

A. Wiener Filtering: A popular method for improving speech, the Wiener filter filters away noise in order to reduce the mean-squared error between the estimated and original speech signals. It suppresses the noise while maintaining the speech signal by first estimating the noise spectrum and then applying a spectral gain function.Wiener filtering is an effective approach for estimating noisy signals using secondary microphone channels and has a broad range of applications in the areas of speech and audio signal processing. The Wiener filtering technique has also been applied to noise cancellation in automotive environments. To enhance the target speech quality, noise introduced into the pilot channel is evaluated by a filter and eliminated. It uses the correlation between the noise in secondary microphones and the distorted speech in the pilot channel for this estimation.

B. Spectral subtraction:The spectral subtractionis the family of different variants of the spectral subtraction method such as spectral over-subtraction, multi-band spectral subtraction, Wiener filtering, iterative spectral subtraction, and spectral subtraction based on perceptual properties. Thus, the principle of the spectral subtractive-type algorithms is to estimate the short-time spectral magnitude of the speech by subtracting estimated noise from the noisy speech spectrum or by multiplying the noisy spectrum with gain functions and to combine it with the phase of the noisy speech. A. Spectral over-subtraction in this algorithm, two additional parameters are introduced in the spectral subtraction method: over-subtraction factor, and noise spectral floor to reduce the remnant noise.

C. Adaptive Filtering: The filter is referred to as an adaptive filter if its weights self-adjust in accordance with a predetermined rule or an optimal algorithm. The signals are processed based on a predetermined criterion, which is typically correlation or estimated mean squared error (MSE). One could think of this process as an adaptive filtering. A self-regulating system that uses recursive algorithms for processing is called an adaptive filter. In order to modify some previously assumed filter parameters under the influence of incoming signals, an error signal is generated based on a comparison of the initial input and training. Adjusting the filter parameters continues until the condition reaches a stable state. Regarding the use of adaptive filters for speech noise reduction, they can provide. Adaptive filters can provide the greatest results when it comes to speech noise reduction applications. The noise's cause is somewhat comparable to the signal that is generated at random, and it is always exceedingly challenging to assess its statistics. The fixed filter's design is an entirely unsuccessful phenomenon that causes the noisy signal to change continuously while speaking. In the context of information in the noise cancellation process, some signals change very quickly, necessitating the use of self-regularized algorithms with the ability to converge quickly. System identification, channel equalization, and signal enhancement through noise reduction are a few uses for adaptive filtering.

- Median Filters: A popular nonlinear filter for reducing noise. It substitutes the median value of nearby samples for each sample. This effectively eliminates spikes and impulse noise without noticeably softening edges.
- Wiener-like Adaptive Filters: Adaptive filters that use nonlinear algorithms may use decision-based strategies to evaluate the contributions of previous samples according to their attributes.

- Wavelet Transform: Wavelet-based techniques can produce improved speech signals by offering varying degrees of information for various frequency components and enabling nonlinear coefficient adjustment.

D. Kalman Filtering:A recursive approach called Kalman filtering models the speech and noise as distinct stochastic processes in order to estimate the clean voice signal. Although it can be sensitive to the precision of the assumed signal and noise models, it has been used in a variety of noise reduction tasks. The Kalman is also known as a mathematical technique since it works via a prediction and correction mechanism. In filter order to estimate the desired variables in a way that minimizes the error between the measured and original data statistically, Kalman filter combines all of the available data, including measured data, system knowledge, and measurement equipment. The Kalman filter is typically applied to reduce white noise. Nevertheless, other techniques were created to apply the Kalman approach to lessen the colored noises as well.

In situations where the noise characteristics are well understood and can be accurately predicted, linear filters might be more suitable. In more complicated situations when speech signals are heavily distorted by different kinds of noise or where conventional assumptions regarding linearity are not true, nonlinear filters typically perform exceptionally well.

4. AI and ML Techniques for Background Noise Reduction

Recent years have seen tremendous advancements in the domains of machine learning (ML) and artificial intelligence (AI), leading to the development of practical solutions for difficult problems such as background noise removal in audio signals. Because these tactics use large datasets and sophisticated algorithms to find intricate patterns and linkages, they are more flexible and generalizable than traditional approaches.

1. Deep learning approach

Given an arbitrary noisy signal consisting of arbitrary noise and speech signals, create a deep learning model that will reduce or entirely remove the noise signal while preserving the speech signal without any audible distortion.

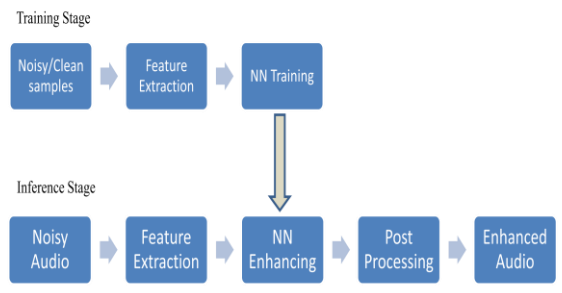

Fig-5: SE using deep learning

Let's go over the main steps of this approach.

(i) Training data: The fields of machine learning (ML) and artificial intelligence (AI) have advanced tremendously in recent years, which has resulted in the creation of workable solutions for challenging issues like audio signal background noise removal. Compared to traditional methods, these strategies are more adaptable and generalizable since they make use of extensive datasets and complex algorithms to identify complex patterns and connections.

- No background noise should be audible in a clean speech dataset.Training voices and noises should be diverse to help the model generalize on unseen voices and noises

- Samples from high-quality microphones are preferred since they provide for greater flexibility in data augmentations. **Feature extraction**: An example of reasonable feature extraction is an audio spectrogram or spectrogram-based features like Mel-frequency cepstral coefficients (MFCCs), which is a time and frequency representation of the signal that reflects the human auditory system's response.

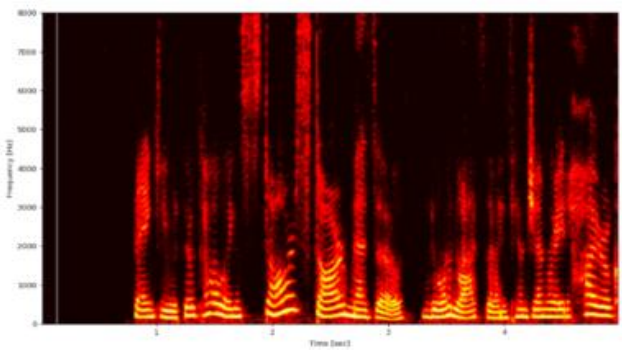

Fig-6: Speech spectrogram

2. Neural Network: Almost every kind of neural network architecture can be adjusted to improve

voice. Then, in order to use image processing methods like convolutional networks, we treat spectrograms as images. Additionally, audio can be represented as sequential data, in which case recurrent neural networks may be a suitable option. This is especially true for long short-term memory units (LSTM) and gated recurrent units (GRU).

(i) Training: In order to recover the speech spectrum from the noisy/corrupted input, the model "learns" generic patterns of clean speech spectrums and noise spectrums during the training stage. Following the training stage, the model can be used for inference; it takes noisy audio input, extracts features, sends it to the neural network, obtains the clean speech features, and, during post-processing, recovers the clean speech signal in the output. Research indicates that deep learning-based speech enhancement models outperform conventional methods and exhibit notable noise reduction, both for stationary and non-stationary noises.

5. Conclusion

Since voice enhancement is impacted by various forms of noise and there are numerous methods and techniques for noise reduction, noise reduction is an intriguing and challenging problem to address. An overview of numerous noise reduction techniques is examined in this the publication, and their results are contrasted with those of other researchers over a range of parameters.

The findings demonstrate that it functions effectively at both low and high SNR levels and in both colored and additive white Gaussian noise (AWGN). Next is the adaptive kalman, which performs better in white Gaussian (WGN) noise, albeit in some situations, performance varies as the noise becomes useful. The wiener filter, which functions best when the noise is stationary, comes next. The adaptive filter algorithm LMS follows, which is good for low cost, complexity, and boosting the SNR in various noises and color noise. However, the RLS, FEDS, and FAP approaches converge more quickly than the LMS. Next on the list of effective techniques is spectral subtraction, which has the drawback of using noisy phases and producing residual noise. The usage of noisy phases that result in a roughness in the voice quality is the drawback of spectral subtraction, which is on the list of effective techniques but also creates residual noise.

Additionally, neural network methods are rather good because they are slower than filtering methods. The deep learning de-noising method is the most complex and takes a long time to complete, but the results are poor. The best neural network method is ADALINE. The hardest noise to remove from a speech signal is non-stationary noise, and the hardest noise is real-world unknown natural noise (mixed noise), followed by white Gaussian noise (WGN) and colored noise.

6. FUTURE WORK

In our future work, we aim to explore the application of adaptive Wiener filtering as a promising advancement in speech enhancement techniques. Adaptive Wiener filtering offers several benefits that can significantly improve speech quality and intelligibility in varying acoustic environments. Here are some key points outlining its usefulness:

1. Dynamic Noise Adaptation: Adaptive Wiener filtering adjusts itself in real-time based on the characteristics of the incoming audio signal and the surrounding noise environment. This adaptability ensures that the filter can effectively respond to changing conditions, resulting in more accurate noise reduction.

2. Improved Speech Quality: By continuously refining its parameters, adaptive Wiener filtering can better distinguish between speech and background noise, leading to enhanced clarity and quality of the processed speech. This is particularly valuable in scenarios with fluctuating noise levels, such as urban environments or crowded settings.

3. Enhanced Performance in Non-Stationary Environments: Many real-world situations involve non-stationary noise sources that vary over time. Adaptive Wiener filtering is well-suited for such conditions, as it can track these variations and modify its response accordingly, offering superior performance in comparison to static filtering methods.

4. Customization with Machine Learning: By integrating machine learning environments.

5. Broader Applications: The potential applications of adaptive Wiener filtering extend beyond traditional speech enhancement. It can be utilized in telecommunications, hearing aids, and assistive listening devices, making it a techniques, we can train adaptive Wiener filters that learn from historical data to anticipate and react to specific noise patterns. This could further enhance the filter's effectiveness in diverse and challenging acoustic versatile tool for improving communication in various contexts.

In summary, the exploration of adaptive Wiener filtering for speech enhancement holds great promise, particularly when incorporated with machine learning and AI techniques. Our future research will focus on optimizing this method to achieve superior results in noise reduction and speech intelligibility, thereby enhancing user experience in real-world applications.

References

1. Milind U. Nemade, Satish K. Shah" Speech Enhancement Techniques: Quality vs. Intelligibility" International Journal of Future Computer and Communication, Vol. 3, No. 3, June 2014
2. S.ChinaVenkateswarlu , Dr.K.Satya Prasad , Dr.A.SubbaRami Reddy" Improve Speech Enhancement Using Weiner Filtering" Global Journal of Computer Science and Technology Volume 11 Issue 7 Version 1.0 May 2011
3. Michelle Gutiérrez-Muñoz, Marvin Coto-Jiménez" An Experimental Study on Speech Enhancement Based on a Combination of Wavelets and Deep Learning" Computation 2022, 10, 102.
4. Yariv Ephraim"Statistical-Model-Based Speech Enhancement System" vol. 80, No. 10,October 1992
5. https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1118&context=eesp
6. Martin Holters, Udo Zölzer" GRAPHIC EQUALIZER DESIGN USING HIGHER-ORDER RECURSIVE FILTERS" Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06), Montreal, Canada, September 18-20, 2006
7. Navneet Upadhyay, Rahul Kumar Jaiswala" Single Channel Speech Enhancement: using Wiener Filtering with Recursive Noise Estimation" Procedia Computer Science 84 (2016) 22 – 30
8. Navneet Upadhyaya,Abhijit Karmakarb" Speech Enhancement using Spectral Subtraction-type Algorithms: A Comparison and Simulation Study" Procedia Computer Science 54 (2015) 574 – 584
9. P. Renevey, A. Drygajlo" Detection of reliable features for speech recognition in noisy conditions using a statistical criterion" in Proc. Workshop Consistent Reliable Acoust. Cues Sound Anal., Aalborg, Denmark, Sep. 2001
10. Pankaj Bactor, Anil Garg," Different Techniques for the Enhancement of the intelligibility of a Speech Signal", International Journal of Engineering Research and Development, 2012
11. (https://electrovoice.com/media/downloads/ev_pabible-05-add04-eq_and_equalizers-1980.pdf)
12. R. McAulay; M. Malpass"Speech enhancement using a soft-decision noise suppression filter" IEEE Volume: 28, Issue: 2, April 1980
13. Soha A. Nossier, Julie Wall, Mansour Moniri, Cornelius Glackin" Two-stage deep learning approach for speech enhancement and reconstruction in the frequency and time domains" 2022 International Joint Conference on Neural Networks (IJCNN)
14. Book by Philipos C. Loizou"Speech Enhancement"
15. Soha A. Nossier, Julie Wall, Mansour Moniri, Cornelius Glackin, Nigel Cannings "Two-stage deep learning approach for speech enhancement and reconstruction in the frequency and time domains" IEEE 2022 International Joint Conference on Neural Networks (IJCNN).
16. Robert J.Mcaualy, Marilyn L. Malpass" Speech Enhancement Using a Soft-Decision Noise Suppression Filter" IEEE , VOL. ASSP-28, NO. 2, APRIL 1980
17. C. Ramesh Kumar, M. Sandhiya, B. Sreeman Kumar, H. Shyam, M. Yuvaraj" Speech Enhancement Using Modified Wiener Filtering

With Silence Removal" 2023 3rd International Conference on Intelligent Technologies (CONIT)

18. Ruqaya Jamal Nasir, Husam Ali Abdulmohsin" Review: Noise Reduction Techniques for Enhancing Speech" 2024, Vol. 65, No. 10, pp: 5798-5818

19. Sunita Dixit, 2Dr. MD Yusuf Mulge"Review on Speech Enhancement Techniques"Vol.3 Issue.8, August- 2014, pg. 285-290

20. Tayseer M. F. Taha, Ahsan Adeel, Amir Hussain"A Survey on Techniques for Enhancing Speech"Volume 179 - No.17, February 2018

21. Kris Hermus, Patrick Wambacq, and Hugo Van hamme"A Review of Signal Subspace Speech Enhancement and Its Application to Noise Robust Speech Recognition"Volume 2007, Article ID 45821, 15 pages