An optimized fixed equalizer for speech enhancement

He Qi · **Siow Yong Low** · **Ka Fai Cedric Yiu**

the date of receipt and acceptance should be inserted later

Abstract This paper proposed a new method to design the fixed single-channel equalizer parameters based on Perceptual Evaluation of Speech Quality (PESQ) and the Short-Time Objective Intelligibility (STOI). The proposed equalizer does not require the need for noise estimation nor a voice activity detector. The fixed weightings are obtained by a training process containing a set of training data. A multi-objective optimization problem was formulated with the PESQ and STOI as the objective measures, which results in a set of candidate solutions via the Pareto optimality. Experimental results demonstrate improvements in PESQ and STOI, and the trade-off between these two measures is well-defined in a set of Pareto optimal solutions to allow flexible adjustment.

Keywords Equalizer · Speech enhancement · STOI · PESQ

1 Introduction

Speech is the primary input for any voice communication system. However, in the real environment, noise contamination is inevitable, which degrades the voice communication systems' performance. This problem gives rise to the popularity of speech enhancement, which aims at improving the quality or the intelligibility of speech in the wrong environments. Speech enhancement is applied in various applications such as hearing aids, assistive listening devices, speech recognition systems, and voice communication systems.

The problem of speech enhancement can be viewed as an adaptive equalization design problem. In essence, the equalization process involves a known digital filter and desired system modeled by a know

This paper is supported by RGC Grant PolyU. 152245/18E and PolyU Grant G-UAHF.
The second author is supported by the Fundamental Research Grant Scheme (FRGS/1/2019/TK04/USMC/02/1) under the auspices of the Malaysian Ministry of Higher Education (MOHE).
He Qi
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, PR China
E-mail: qivicky.he@connect.polyu.hk

Siow Yong Low
University of Southampton Malaysia (UoSM), Iskandar Puteri, Johor, Malaysia
E-mail: sy.low@soton.ac.uk

Ka Fai Cedric Yiu
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, PR China
E-mail: macyiu@polyu.edu.hk

digital filter. Equalization is an established audio processing technique to compensate for the spectral balance of an audio signal [23]. The equalization typically consists of a computationally efficient weighting to emphasize or deemphasize the signal in a frequency band such that the overall quality of the signal is improved.

The quality improvement here refers to the frequency compensation for frequency balance. A natural extension to that will be a noise-based equalization system to enhance speech in noisy environments. Some prior work related to this includes the perceptual headphone equalization for reducing ambient noise [45]. However, these approaches require a reference microphone and a dynamic equalizer.

The research of audio equalization is abundant, and there are various types of equalizers [11, 16, 42, 48, 13]. The most powerful and flexible one is the parametric equalizer, which allows users to add a peak or a notch in the audio spectrum at an arbitrary location [11, 52]. A peak can help with a better complex mix, or add coloration to an instrument's sound, while notches can attenuate unwanted noise and reducing feedback. According to the order of filters, it can be divided into low-order graphic and high-order graphic. Another common form of equalizers is graphic equalizer, consisting a set of filters [1, 24, 28]. The graphic equalizer is a popular device for sound enhancement and can independently adjust the gain of multiple frequency regions in the audio signal. Though the low order parametric and graphic equalizers are well-established with low computational cost and minimal latency, they do have a limited precision of noise control. However, some extension methods such as high-order parametric equalizers or convolution-based graphic equalizers can shape the spectral content of a signal, obtaining a higher precision of noise control, at the expense of higher computational load and latency. There are many other approaches to equalization, such as the matched equalizer and optimal design techniques [49, 30] and matching analog prototypes [4, 37], to name but a few.

Another widely used class of equalizers can be categorized as spectral subtractive methods [18]. It spectrally subtracts the noise estimate from the noisy spectrum envelope, which could effectively eliminate the stationary background noise. This method is a nonlinear and stand-alone noise suppression algorithms and used to reduce unwanted broadband noise [10, 16]. Based on the original method, many improved algorithms have been proposed, such as spectral over-subtraction [7], non-linear spectral subtraction [33] and the iterative method [42]. Non-linear spectral subtraction assumes that noise affects spectral components unequally and uses the frequency dependent factor subtraction for different types of noise. In addition, this subtraction process takes into account the frequency dependent signal to noise ratio (SNR), the idea being to apply a minimal subtraction factor in high SNR regions at frame level, which makes the process incoherent or nonlinear [33]. Multiband spectral subtraction divides the speech spectrum into various orverlapping bands and each band applies independent subtraction, which can be used in both time and frequency domain by using suitable windows [35, 31]. The Winer filter produces the minimum mean square error estimate of the desired signal [6, 26]. However, the Wiener short-time spectral amplitude (STSA) estimator is derived from the optimal minimum mean square error (MMSE) signal spectral estimator, not the optimal spectral amplitude estimators under the assumed statistical model and criterion. More STSA estimators are proposed over the years, which is based on modeling speech and noise spectral components as statistically independent Gaussian random variables [18]. However, any spectral estimator erroneous estimate will result in noise artefacts and this mismatch results in poor performance and a voice activity detector (VAD) is needed to detect speech [2, 22, 43].

Furthermore, many novel methods are explored based on the classical techniques incorporate with the machine learning and deep learning technique [27, 21, 14]. Two popular applied model are the convolutional neural network (CNN) [32, 55] and the long short-term memory of recurrent neural network (RNN) [53], which comprise of the transformation function to convert the spectral features of the clean signal and noisy signal. Research shows that CNN works well for extracting the time-frequency features and has good performance in speech intelligibility test. However, this kind of methods has high requirements on computational power and the size of training data, which would be hard to applied in some simple products without a powerful computer hardware.

Drawing inspiration from equalization, this paper investigates the use of a fixed equalizer for speech enhancement applications. The proposed fixed equalizer bypasses the need for any noise estimation, VAD or complex deep learning system. In fact, the equalizer will just have a fixed gain with the aim of improving the quality and the intelligibility of speech for different types of noise and signal to noise ratios (SNR). To the best of the authors' knowledge, there is no investigation thus far on the use of fixed equalizer for speech enhancement applications.

The equalizer is proposed to optimize the objective measures from a set of training data with different noise type and SNRs. Two of the most widely used objective measures in speech enhancement are Perceptual Evaluation of Speech Quality (PESQ) and the Short-Time Objective Intelligibility (STOI). An improvement in PESQ gives an overall improvement in speech quality and STOI provides an indication of how much speech intelligibility is improved [19,29]. We pose these two measures as evaluation criteria in a multi-objective optimal problem setting. The Pareto optimality is then used to produce a set of candidate solutions for the multi-objective problem. Here, the multi-objective function is solved by converting it into a single objective function via a linear combination. Then the sequential quadratic programming algorithm is proposed to solve the set of constrained problems. Experimental results with various real-world noise show that the trained fixed equalizer improve both the PESQ and STOI scores. The results also extend the trade-off between the STOI and PESQ measures in a form of Pareto optimal solutions, which enables a flexible trade-off choice between the two measures. However, there is a limitation with the proposed method as it is based on prior training. The system will not perform to expectation if there is a big mismatch between the noise encountered and the training. To make the system more robust, different learning algorithms can be applied to modelling errors and uncertainties, such as PD-type iterative learning control [51], type-2 fuzzy (IT2F) approach [54].

2 Signal Model

Let $s(n)$ be the speech signal and $v(n)$ the noise signal. The noisy signal can be denoted as

$$x(n) = s(n) + v(n). \tag{1}$$

The purpose of the fixed single channel speech enhancement is to recover the clean signal from the polluted speech signal. It involves estimating the target signal spectrum by multiplying fixed gain multipliers to each of short-term spectral spectrum. By applying short-time Fourier Transform (STFT), the noisy signal is transformed from time domain to frequency domain with a set of short time spectral components, and then a frequency-domain filter could be applied. By applying short-time Fourier Transform (STFT), the noisy signal is transformed from time domain to frequency domain with a set of short time spectral components and then a frequency-domain filter could be applied. The STFT is given as

$$X(i,k) = \sum_{n=0}^{N-1} x(n)h(n-kR)e^{-jin} \tag{2}$$

where $h(n)$ is the window function with a hop size of R and length N and k is time index, which represents a short-time section of speech signal $x(n)$ at time k and $i \in [i_0, \cdots, i_{N-1}]$. Denote the gain multipliers as $w(i)$. The enhanced coefficient of output signal $y(k)$ in frequency domain could be presented as

$$Y(i,k) = w(i) \cdot X(i,k) \tag{3}$$

Then, the enhanced spectrum is converted back to time domain by Inverse Short-time Fourier Transformation (ISTFT), which is given as

$$y(n) = ISTFT[Y(i,k)] \tag{4}$$

Clearly, the main task is to find a fixed set of weighting $w(i)$, which will equalize out the noise and preserve the desired speech. To determine equalizer weighting, we firstintroduce two objective measures. Then, the equalizer weighting problem can be transformed into an optimization problem by setting the $\mathbf{w} = [w_0, \cdots, w_{N-1}]$ as the decision vector and the measures as the objective functions.

3 Equalizer Design

3.1 Preliminary

Generally speaking, the assessment of speech quality includes subjective and objective evaluation. Subjective evaluation relies on listeners' subjective listening tests, which could be quite accurate but costly and time-consuming. Objective evaluation measures the numerical distance between the reference signal and the processed signals, predicting the speech quality with high correlation [34]. Two of the most popular objective measures for speech enhancement are PESQ and STOI, which are closely related to the human auditory perception and widely used in speech enhancement as evaluation criteria. PESQ is a traditional evaluation, applied cognitive modeling and computing disturbance between the separated and clean speech. STOI is a latest popular measure, computing the correlation of short-time temporal envelopes.

3.1.1 Perceptual evaluation of speech quality (PESQ)

PESQ is an automated computation algorithm developed by the International Telecommunication Union (ITU) for speech quality assessment [5,46]. It results from the integration of the perceptual analysis measurement system (PAMS) and an enhanced version of the Perceptual Speech Quality Measure(PSQM), PSQM99. The model begins by level, aligning both signals to a standard listening level [46]. It is obtained by a linear combination of the average symmetric disturbance value D_{ind} and the average asymmetrical disturbance value A_{ind}

$$PESQ = a_0 - a_1 D_{ind} - a_2 A_{ind} \tag{5}$$

where $a_0 = 4.5$, $a_1 = -0.1$, $a_2 = -0.0309$. All the three parameters equation (5) were optimized for speech proceed through networks. The calculation of disturbance values D_{ind} and A_{ind} uses the follows L_p norm formula:

$$L_p = \left(\frac{1}{N}\sum_{m=1}^{N} disturbance[m]^p\right)^{1/p}.$$

The details can be found in [46]. The PESQ score ranges from -0.5 to 4.5, corresponding to low to high speech quality.

Studies have proved that the PESQ measure yielded the highest correlation with the overall quality and signal distortion [46]. Besides, PESQ has also been shown to be consistent in measuring speech intelligibility [36]. Due to the qualities above, it is considered an effective indicator in processed speech.

3.1.2 A Short-Time Objective Intelligibility measure(STOI)

STOI is an objective machine-driven intelligibility measure based on short time segments, which shows a high correlation with the intelligibility of noisy and time-frequency weighted noisy speech. STOI is a function of a time-frequency-dependent intermediate intelligibility measure, which compares the temporal envelopes of the clean and degraded speech in short-time regions utilizing a correlation coefficient. The clean and degraded speech denoted by s and y, the Time-Frequency cell amplitudes of the clean and degraded speech denoted by $S_j(m)$ and $Y_j(m)$, and the short-time temporal envelope denoted by $s_{j,m}$

and $y_{j,m}$, respectively. The intermediate intelligibility is then defined as the sample correlation coefficient between the two vectors, denoted as

$$d_{j,m} = \frac{(s_{j,m} - \mu_{s_{j,m}})^T \tilde{y}_{j,m}}{||s_{j,m} - \mu_{s_{j,m}}|| \, ||\tilde{y}_{j,m} - \mu_{\tilde{y}_{j,m}}||},$$

where $\mu_{(\cdot)}$ is the mean value of the corresponding vector, and $\tilde{y}_{j,m}$ is the corresponding modulation vector of

$$\tilde{Y}_j(m) = \min(Y_j(m), 6.33 \cdot \frac{||y_{j,m}||}{||s_{j,m}||} S_j(m)).$$

The final average of the intermediate intelligibility measure overall bands and frames is calculated as

$$d = \frac{1}{JM} \sum_{j,m} d_{j,m} \tag{6}$$

where M represents the total number of frames and J the number of one-third octave bands. The details of the calculation of the STOI can be found in [50].

Experiments showed that STOI has a better correlation with speech intelligibility compared to other reference objective intelligibility models. Due to this, many people evaluate the improvement in speech intelligibility by comparing the STOI scores [47,17], especially in assessing single-channel algorithms [38].

3.2 Filter Design and Optimal Method

To design the coefficients of the equalizer, the multipliers $w(i)$ is given as a vector

$$\mathbf{w} = (w(0), w(1), \cdots, w(N-1))^T \tag{7}$$

where N is the frame size of STFT. Assume a noise equalizer parameterized by the vector $\mathbf{w}$ is employed, the output signal in the time domain after the equalizer denoted by $y(n)$. Then, the scores of PESQ and STOI could be calculated according to (5) and (6), defined as $P(\mathbf{w})$ and $S(\mathbf{w})$. The equalizer design problem can be written as two maximization problems, denoted by

$$\max \ P(\mathbf{w}) \tag{8a}$$
$$s.t \quad 0 \leq w(i) \leq 1 \quad i = 0, \cdots, N-1 \tag{8b}$$

and

$$\max \ S(\mathbf{w}) \tag{9a}$$
$$s.t \quad 0 \leq w(i) \leq 1 \quad i = 0, \cdots, N-1 \tag{9b}$$

where N is the length of the equalizer.

As each objective measure has its own property in evaluating the speech quality, we attempt to form a combination of the two objective measures which would adjust the speech quality in different aspects. Since a single optimal solution will not minimize/maximize all the objective functions. The problem is formulated as a multi-objective optimal problem which is described as

$$\max \ f(\mathbf{w}) = [f_1(\mathbf{w}), f_2(\mathbf{w})] \tag{10a}$$
$$s.t \quad 0 \leq w(i) \leq 1 \quad i = 0, \cdots, N-1, \tag{10b}$$

where $f_1(\mathbf{w}) = P(\mathbf{w})$ and $f_2(\mathbf{w}) = S(\mathbf{w})$.

It is not possible to mathematically define a single optimal solution minimizing all objective function simultaneously, since vector function $f(w)$ induce on the set of feasible solutions a partial order [40]. In addition, according to the experimental results, the single optimal solution does not exist in this multi-objective problem. However, if we can have a compromise or trade-off between the two objectives (that is guaranteeing one objective can not be improved without degrading the other one), then we can find a set of possible solutions. The solutions for the problem (10) depend on a priori on which objective measure is chosen for treating the problem. To compare the candidate solutions, we can introduce the concepts of Pareto dominance and Pareto optimality [12], originally raised by Francis Ysidro, and then developed by Vilfredo Pareto [15], which allows one preference criterion better off without making use of at least one individual or preference criterion worse off. To understand the concept of Pareto optimal better, we firstly introduce the concept of Pareto-dominate.

Definition 1 Suppose $\mathbf{w}^1$ and $\mathbf{w}^2$ are two decision vectors. We said $\mathbf{w}^1$ is Pareto-dominate to the vector $\mathbf{w}^2$, if and only if

$$\begin{aligned} \forall i = 1,2,\ f_i(\mathbf{w}^2) \leq\ f_i(\mathbf{w}^1) \\ and\ \exists j = 1,2:\ f_j(\mathbf{w}^2) <\ f_j(\mathbf{w}^1). \end{aligned} \tag{11}$$

This conception compares the two solution: $f(\mathbf{w}_1)$ is better than $f(\mathbf{w}_2)$ for all objectives, and there is at least one objective function for $f(\mathbf{w}_1)$ is strictly bigger than $f(\mathbf{w}_2)$. Then, the Pareto optimal solutions could be described as follows.

Definition 2 A solution $\mathbf{w}^*$ is Pareto optimal, if and only if there does not exist another feasible solution that dominates it. The set of Pareto optimal solutions is called the Pareto set, and the corresponding objective vectors are known as Pareto front.

It means that $\mathbf{w}^*$ cannot be larger in one of the objectives without affecting the other objective.

Many methods have been proposed to solve the problem above. One class of techniques is converting the multi-objective problems into single-objective problem with additional variables, parameters or additional constraints, also known as scarlarization, after which the problem can be solved by classical optimization techniques. The main idea of traditional scarlarization techniques is either aggregating the objective functions or optimizing one and treating the other as constraints. In this problem, as the objectives are not that complex, we consider two traditional methods, which are simple to implement, including weighted sum method and $\varepsilon-$constraint method [39,40]. Weighted sum method involves the aggregating the objective functions by a convex combination, while $\varepsilon-$constraint method considers objectives into constraints. Both of these two methods can give a single solution according to the preferential information about the PESQ and STOI given by decision makers. They can also approach the Pareto front by essentially repeating the solution process after modifying the additional parameters.

The $\varepsilon-$constraint is based on minimization of one preferred or primary objective function, and the other is turned into constraints bound by some allowable levels ε. Then, the a single-objective optimal problem is formulated for the most relevant objective function subjective to additional constraints. By this method, Pareto optimal set can be found by appropriately specifying the values of ε, and problem (10) can be presented in

$$\max\ S(\mathbf{w}) \tag{12a}$$

$$s.t \qquad 0 \leq w(i) \leq 1 \quad i = 0, \cdots, N-1 \tag{12b}$$

$$P(\mathbf{w})/5 \leq \varepsilon \tag{12c}$$

In problem (12), we consider the value of STOI as the objective function and put the value of $PESQ$ in the constraint bound by a considerable range ε. This problem is repeatedly solved under different value of ε to get the whole Pareto set. Also, we can change the positions of $P(\mathbf{w})$ and $S(\mathbf{w})$ in (12) according to decision maker's preference.

In solving (12), we can convert the constraint into the objective by introducing it as a penalty function. By choosing a suitable parameter, the feasible set is unchanged. One simple and popular way is to form a convex combination of the two objectives, also known as weighted sum method. It can be presented as

$$\max\ F(\mathbf{w}) = A * S(\mathbf{w}) + (1-A) * P(\mathbf{w})/5 \tag{13a}$$

$$s.t \quad 0 \leq w(i) \leq 1 \quad i = 0, \cdots, N-1 \tag{13b}$$

where A is a fixed real number from 0 to 1. The weighting coefficients means the the relative importance of the two objective measures, which consists of a trade-off between PESQ and STOI. Users can weight the objectives according to to the importance, and the preference information can be taken into consideration by changing the value of A. This weight must be given in advance. After solving the optimal problem, a single solution, that best satisfies the additional information, is provided.

Above all, the three problems are essentially constrained nonlinear optimization problems, and many optimal methods can be applied to reality in solving the problems. Sequential quadratic programming methods is an effective method for solving nonlinear programming methods. Based on the works of Biggs [8], Han [25] and Powell [44], an SQP method mimics Newton's method for constraint optimization. However, the SQP method requires the gradient information. As the computation of the objective function is complex, the gradient information is supplied analytically calculated gradients [41].

It is an iterative method of starting from some initial point and converging to a constrained local minimum. At each iteration, one obtains search directions from a quadratic program (QP) that is a quadratic model of a certain Lagrangian function subject to the constraints. The Lagrangian function associated with problem (13) is defined as

$$L(\mathbf{w}, \lambda) \triangleq F(\mathbf{w}) + \lambda_1 \mathbf{w} + \lambda_2 (\mathbf{w} - 1) \tag{14}$$

where $\lambda = [\lambda_1\ \lambda_2]^T$ is a vector of nonnegative Lagrange multiplier estimates. Then, the SQP search direction can be calculated by solving a sub-problem described as

$$\max_{\mathbf{w}}\ F(\mathbf{w}) + g(\mathbf{w})^T d + \frac{1}{2} d^T H d \tag{15a}$$

$$s.t\ 0 \leq w + d \leq 1, \tag{15b}$$

where $g(\mathbf{w})$ is the function of analytical gradients, and H is a symmetric matrix. In this problem, we use a Broyden-Fletcher-Goldfarb-Shanno (BFGS) approximation B [20] to represent H. The related works have been proposed in [9] [20], and research shows that a simple BFGS method is efficient in practice.

The algorithm could be described as follows:

(1) Choose an initial guess $\mathbf{w}^0$ that satisfies the constraints and set $k = 0$.
(2) Compute $F(\mathbf{w})$.
(3) Compute the optimum update d^k by solving QP problem. The subproblem is given as

$$\max_{w}\ F(\mathbf{w}^k) + g(\mathbf{w}^k)^T d + \frac{1}{2} d^T B^k d \tag{16a}$$

$$s.t\ 0 \leq \mathbf{w}^k + d \leq 1. \tag{16b}$$

(4) Set $\mathbf{w}^{k+1} = \mathbf{w}^k + \alpha d^k$, with $\alpha \in (0,1)$, which is a suitable steplength parameter.
(5) If $\| \mathbf{w}^k \|_2 / \| \mathbf{w}^{k+1} \|_2 < c$, stop, where $\|\mathbf{w}^k\|_2$ is the Euclidean norm. Otherwise, set $k = k+1$ and then go to (2).

The step length parameter α is required to enforce global convergence of the optimization algorithm, and it is usually chosen to satisfy a certain Armijo condition [3]. It guarantees a reduced direction at each step of a procedure. Besides, c is used to estimate whether the current solution is convergence. The algorithm stops if sufficient descent is not observed after a certain number of iterations. If the tested stepsize falls below machine precision or the accuracy by which model function values are computed, the merit function cannot decrease further.

4 Experimental Settings

In the experimental part, we evaluate the fixed equalizer on the speech enhancement task. The single indicator and the multi-indicator algorithm are considered, respectively. For the training set, a clean speech data set from the TIMIT database were prepared, and the noisy signal is given by adding the noise signal. For each utterance, we can get a set of parameters by solving an optimal problem. The average parameters of the training set are considered as the parameter of the equalizer. We firstly analyse the impact of the training data set size and the frame size of STFT on the equalizer performance. The noise data included three types of noises (babble, subway, white noise signals), and noisy data was created by adding noises to the clean data. Four levels of signal to noise ratio (SNR) were made as 0, 5, 10, and 15 dB. To quantify the performance of the proposed method, a new set of test data which consists of another 100 utterances were used. Note that all the data utterances were randomly chosen to avoid the problem of data overfitting. To better understand the impact of training mismatch, we also investigate the results of situations where the noise type and SNR levels of the training and test sets were matched and mismatched. This is to ascertain if the proposed system is only tuned to data set which is trained upon. We also compared our results with a Wiener filter. The optimal problems in the training procedure were solved in the MATLAB environment, and nonlinear programming function "fmincon" in the optimization toolbox was used in our experiments. As the algorithm is not that sensitive to the initial guess, we start form $\mathbf{w}^0 = \mathbf{1}$. For solving a particular single optimization problem, it takes approximately 80 minutes using an Intel(R) Core(TM) i7-4790 CPU.

4.1 Results of single-objective optimization algorithm

In this section, we consider single indicator STOI or PESQ, individually, by solving the problem (8) or (9). For every equalizer, two objective quality measures, PESQ and STOI, were used for evaluating the performance of equalizers. In addition, a comparasion with a speech enhancement system based on classical Wiener filter is given.

4.1.1 Training data set size

First of all, we discuss how the quality of the enhanced signals changes as the training data size increases. To ascertain the impact of training data size on the optimization, we increased the sizes of the training data set gradually. We choose STOI as the objective function. The noisy signal is given by adding the noise signal to the clean signal, denoted as

$$x(n) = s(n) + \alpha v(n),$$

where $s(n)$ is clean signal, $v(n)$ is the noise signal, and α is a scalar value. To evaluate the power of noise signal, we introduce SNR in decibels(dB), defined as

$$SNR_{dB} = 10\log_{10}(\frac{p_s}{p_n}),$$

where p_s is the power of the clean signal, and p_n is the power of the noise signal. According to the definition of SNR, the value of α can then be obtained by

$$\alpha = \sqrt{\frac{\|s(n)\|_2^2}{\|SNR \cdot v(n)\|_2^2}},$$

where $\|\cdot\|_2$ represents the Euclidean norm. In this part, a babble noise with SNR 0dB is added, and the number of STFT is 1024. The average results with a various number of the training set were shown in Tab. 1. S_0 and P_0 means the STOI and PESQ value of the noisy signal and S^* and P^* stand for the STOI and PESQ value of the enhanced signals, and ΔP and ΔS means the average percentage improvement of PESQ and STOI. The average percentage of improvement is defined as

$$\Delta P = \frac{P^* - P_0}{P_0} \times 100\% \tag{17}$$

$$\Delta S = \frac{S^* - S_0}{S_0} \times 100\% \tag{18}$$

where P_0 and S_0 are the original PESQ and STOI scores and P^* and S^* are the enhanced scores.

From our table, we can observe that increasing sizes of training data set improved the equalizer's performances but with the cost of increasing training time. In addition, there is no significant improvement based on STOI and even deterioration based on PESQ. Interestingly we observe that as the training set size increases, the performance plateaus. Thus for the subsequent experiments, we fix the training sample size to be 50 utterances. In our following experiments, 50 utterances were used in the training data set.

Table 1 Effect of training data set size (STOI)

Training Set Size	S_0	S^*	$\Delta S(\%)$	P_0	P^*	$\Delta P(\%)$
50	0.67	0.69	3.02	1.69	1.83	8.17
100		0.69	3.02		1.81	7.12
150		0.69	3.03		1.81	7.20
200		0.69	3.04		1.81	7.36
250		0.69	3.10		1.82	7.62
300		0.69	3.12		1.83	8.12

4.1.2 Effect of Frame Size of STFT

The frame size of STFT is directly related to the number of parameters of the equalizer, which is the same as the size of FFT that we use to transform the time domain data to the frequency domain. To test the performance for speech enhancement based on the different frame sizes of STFT, we trained various equalizers with different frame sizes. In this instant, the babble noise with SNR 0dB is considered.

The performance of equalizers trained by STOI-objective algorithm is measured in Table 2. By comparing the fourth and fifth columns of the table, we can observe that increasing the frame size helps increases STOI, and a window of about 1024 produce the best result. The last two columns show the PESQ results, and we obtain the optimal results with 256. All the results of our paper show average values. Table 3 shows the performance of the PESQ-objective algorithm. Comparing the PESQ-objective and the STOI-objective method, the former had a more powerful capacity to enhance PESQ, but the latter was superior at improving STOI. Also, we can see that, in both two algorithms, the average value of PESQ increased, but there was a decrease of STOI in the PESQ-objective method. From the results, it shows that STOI is a more strict evaluation criterion comparing with PESQ. The STOI scores can help increase

PESQ scores, but PESQ cannot help improve STOI. Table 4 gives the results of classical Winer filter. Clearly, STOI-objective algorithm outperforms the Winer filter in both criteria. For the PESQ-objective algorithm, the average PESQ score of the test set has a quite large increase comparing with the Winer filter.

Table 2 Effect of frame size (STOI)

No.FFT	S_0	S^*	$\Delta S(\%)$	P_0	P^*	$\Delta P(\%)$
64	0.67	0.68	1.03	1.69	1.85	9.56
128		0.68	1.43		1.86	10.26
256		0.69	2.28		1.89	12.06
512		0.69	2.56		1.83	8.31
1024		0.69	3.02		1.83	8.17
2048		0.69	2.53		1.76	3.95

Table 3 Effect of frame size (PESQ)

No.FFT	S_0	S^*	$\Delta S(\%)$	P_0	P^*	$\Delta P(\%)$
64	0.67	0.65	-2.88	1.69	2.56	51.98
128		0.65	-3.39		2.52	49.48
256		0.65	-2.41		2.48	47.24
512		0.66	-2.16		2.43	44.09
1024		0.66	-2.05		2.35	39.50
2048		0.67	-0.08		2.03	20.35

Table 4 Effect of frame size (Winer Filter)

No.FFT	S_0	S_w	$\Delta S_w(\%)$	P_0	P_w	$\Delta P_w(\%)$
64	0.67	0.65	-3.25	1.69	1.62	-4.02
128		0.66	-1.38		1.68	-0.39
256		0.67	0.62		1.71	1.60
512		0.68	1.18		1.74	3.13
1024		0.64	-3.98		1.70	0.78
2048		0.61	-8.94		1.58	-6.37

4.1.3 Effect of training with different SNR

To analyze the effect of the training with different SNR conditions, we add the noise with different signal-to-noise (SNR) conditions, 0dB, 5dB, 10dB, 15dB, respectively, and trained four equalizers under these four conditions. To improve the robustness of the trained model, we further merge all the four equalizers and generate a merging equalizer, and the performance of the merging equalizer was tested with different SNR levels signals. To see the impact of SNR, we choose the same babble noise and the same number of FFT, which is 512. By using the objective function as defined in Eq. (8), the optimal STOI and PESQ scores were searched. Table 5 shows the mean value for test samples corrupted under different SNR conditions and enhanced by different equalizers. As observed, both PESQ and STOI improve, except for the 15dB noise processed by 0dB equalizer in STOI measure that acts as a decrease in these results. The merging equalizer is more robust, which has a positive effect on all SNR levels signals. Table 6 shows the results by using objective function in Eq. (9). Similarly, in the PESQ-objective method, the performance of STOI has decreased, but for PESQ, there is an enormous increase. The performance is greater when the training sample matches the SNR of the input signals. However, we see that training the filters with training samples of all SNR levels yields the most consistent performance. The results of Wiener filter are also given in Table 7 for comparison. It can be seen that there are only slightly improvements.

Table 5 Results Based on Different SNR (STOI)

SNR(Test Sample)	S_0	P_0	equalizer	S^*	$\Delta S(\%)$	P^*	$\Delta P(\%)$
0dB	0.67	1.69	0dB	0.69	2.56	1.83	8.31
			5dB	0.69	2.59	1.83	8.41
			10dB	0.68	2.12	1.80	6.43
			15dB	0.68	1.52	1.76	3.88
			merging	0.69	2.30	1.80	6.56
5dB	0.78	2.06	0dB	0.79	1.44	2.18	5.75
			5dB	0.80	2.00	2.17	5.86
			10dB	0.79	1.85	2.15	4.44
			15dB	0.79	1.50	2.11	2.65
			merging	0.79	1.90	2.15	4.55
10dB	0.86	2.41	0dB	0.86	0.23	2.51	4.03
			5dB	0.87	1.24	2.52	4.21
			10dB	0.87	1.40	2.49	3.18
			15dB	0.87	1.30	2.46	1.88
			merging	0.87	1.36	2.49	3.27
15dB	0.92	2.76	0dB	0.91	-0.86	2.84	2.90
			5dB	0.92	0.34	2.85	3.09
			10dB	0.93	0.74	2.83	2.35
			15dB	0.93	0.87	2.80	1.38
			merging	0.93	0.65	2.83	2.41

Table 6 Quality Improvement Based on Different SNR (PESQ)

SNR(Test Sample)	S_0	P_0	equalizer	S^*	$\Delta S(\%)$	P^*	$\Delta P(\%)$
0dB	0.67	1.69	0dB	0.66	-2.16	2.43	44.09
			5dB	0.66	-2.12	2.42	43.80
			10dB	0.66	-1.71	2.40	42.24
			15dB	0.66	-1.21	2.30	36.54
			merging	0.66	-1.44	2.36	40.22
5dB	0.78	2.06	0dB	0.75	-3.72	2.70	31.77
			5dB	0.75	-3.63	2.70	31.62
			10dB	0.75	-3.18	2.68	30.42
			15dB	0.76	-2.66	2.59	26.22
			merging	0.75	-2.90	2.65	29.02
10dB	0.86	2.41	0dB	0.82	-5.08	3.00	24.45
			5dB	0.82	-4.92	3.00	24.41
			10dB	0.82	-4.43	2.98	23.47
			15dB	0.83	-3.91	2.90	20.10
			merging	0.83	-4.12	2.95	22.41
15dB	0.92	2.76	0dB	0.87	-6.13	3.30	19.64
			5dB	0.87	-5.89	3.31	19.75
			10dB	0.87	-5.34	3.29	19.13
			15dB	0.88	-4.79	3.21	16.41
			merging	0.88	-4.97	3.26	18.27

Table 7 Results Based on Different SNR (Wiener filter)

SNR(Test Sample)	S_0	P_0	S^*	$\Delta S(\%)$	P^*	$\Delta P(\%)$
0dB	0.67	1.69	0.68	1.18	1.74	3.13
5dB	0.78	2.06	0.78	0.90	2.08	1.36
10dB	0.86	2.41	0.87	0.41	2.43	0.56
15dB	0.92	2.76	0.92	0.15	2.77	0.27

4.1.4 Effect of training with different noise type

To validate the influence of noise type, we trained three equalizers under three different noise types (with different level of stationarity), including babble (most non-stationary), subway (the semi-stationary), and white noise (stationary). The generality of these three types of noise would give a good indication as to how good a system responds in such settings in most real life situations. The samples are corrupted at the SNR level of 0dB, and the frame size of STFT is 512. We tested the equalizers under match and mismatch environments, and three other noise signals are adopted in the test sets, including airport, street restaurant noise. Also, a merging equalizer trained by all these three types of noise is given. Tables 8 and 9 give the results measured by objective evaluation measures, based on (8) and (9), respectively. For the matched-noise environments, the training was conducted with the same noise type as the same type of noise, which contaminates the input speech signal, average PESQ results between PESQ-objective and STOI-objective methods on the test set improves. In contrast, the STOI based on the STOI-objective method increases but decreases in the PESQ-objective method. In the mismatch environments, for the optimal values of STOI based on the STOI-objective method, there is little increase, even decrease. Besides, we found that babble noise is the most sensitive. In the STOI-objective method, the equalizer trained by babble noise can only work on the babble, street and airport noise samples (with little improvements), while the subway and white equalizers can have a positive effect on the other types of noise, significantly the white noise equalizer increases 1.08%, 1.46% and 1.31% STOI on babble noise, subway noise and restaurant noise. Comparing with the STOI, the PESQ-objective method is less sensitive to noise type. These four equalizers can increase the PESQ of the utterances polluted by the six different types of noise. The merging equalizer outperforms the mismatched equalizer, but has a worse performance than the matched equalizer. Table 10 shows the results of Wiener filter with all these six types of noises. SOTI-based algorithm has better results in STOI and PESQ, while the PESQ-based algorithm has much better PESQ scores. In addition, the results indicate that the type of noise is a more sensitive factor than SNR. From the experimental results above, we can see that when the equalizers are tested in the mismatched-SNR and mismatched-noise environments, the optimal values of STOI based on the STOI-objective method have little increase, even decrease. Although the PESQ-objective method is less sensitive to noise type and SNR, there is still slight decrease in the mismatched situations. As the proposed method has no prediction section, when there is a big mismatch between the actual noise and equalizer, the performance would drop.

4.2 Results of multi-objective optimization algorithm

From the results, we can see that there is a natural trade-off between the STOI and PESQ measures. To investigate this trade-off further, we consider a linear combination of PESQ and STOI in the objective function as

$$max \quad A*S+(1-A)*P/5$$

where S represents the value of STOI, P means the value of PESQ, and A denotes the trade-off between the two objective measures. The range of PESQ is from -0.5 to 4.5, while STOI ranges from 0 to 1. Thus, we add a normalization parameter as 5. In this part, the database contains 50 sentences as a training

Table 8 Quality Improvement Based on Different Kinds of Noise (STOI)

Noise(Test Sample)	S_0	P_0	equalizer	S^*	$\Delta S(\%)$	P^*	$\Delta P(\%)$
babble	0.67	1.69	babble	0.69	2.56	1.83	8.31
			subway	0.67	0.31	1.88	11.17
			white	0.68	1.08	1.85	9.46
			merging	0.68	1.91	1.85	9.46
subway	0.65	1.51	babble	0.64	-0.64	1.56	3.09
			subway	0.67	3.49	1.70	12.68
			white	0.66	1.46	1.66	10.44
			merging	0.66	2.25	1.62	7.58
white	0.66	1.42	babble	0.65	-0.83	1.46	2.69
			subway	0.66	0.27	1.59	12.70
			white	0.67	1.91	1.58	11.55
			merging	0.67	1.46	1.53	7.74
street	0.73	1.74	babble	0.73	0.62	1.93	11.01
			subway	0.73	0.22	1.96	12.53
			white	0.74	0.95	1.90	9.23
			merging	0.74	1.14	1.92	10.49
restaurant	0.67	1.70	babble	0.66	-0.52	1.75	3.32
			merging	0.68	1.45	1.83	8.35
			white	0.68	1.31	1.81	6.89
			merging	0.68	1.56	1.80	6.10
airport	0.72	1.81	babble	0.72	0.08	1.90	5.50
			subway	0.72	0.01	1.97	9.42
			white	0.72	0.59	1.94	7.45
			merging	0.72	0.87	1.94	7.34

Table 9 Quality Improvements Based on Different Kinds of Noise (PESQ)

Noise(Test Sample)	S_0	P_0	equalizer	S^*	$\Delta S(\%)$	P^*	$\Delta P(\%)$
babble	0.67	1.69	babble	0.66	-2.16	2.43	44.09
			subway	0.64	-4.58	2.36	40.07
			white	0.65	-2.55	2.34	39.11
			merging	0.65	-2.71	2.38	41.04
subway	0.65	1.51	babble	0.61	-5.52	2.21	47.18
			subway	0.62	-4.23	2.25	50.10
			white	0.61	-5.34	2.16	44.38
			merging	0.62	-4.54	2.21	47.23
white	0.66	1.42	babble	0.61	-6.94	2.13	51.94
			subway	0.61	-7.23	2.10	49.38
			white	0.62	-5.00	2.20	56.91
			merging	0.62	-5.68	2.14	52.74
street	0.73	1.74	babble	0.71	-3.08	2.51	44.81
			subway	0.70	-4.12	2.44	40.71
			white	0.71	-2.55	2.40	38.60
			merging	0.71	-2.74	2.45	41.40
restaurant	0.67	1.70	babble	0.64	-4.36	2.31	36.97
			subway	0.63	-4.84	2.28	34.94
			white	0.64	-3.66	2.25	33.52
			merging	0.64	-3.87	2.28	34.99
airport	0.72	1.81	babble	0.68	-4.60	2.43	35.21
			subway	0.68	-5.76	2.38	32.31
			white	0.69	-3.33	2.38	31.95
			merging	0.69	-4.03	2.40	33.17

Table 10 Quality Improvement Based on Different Kinds of Noise (Winer filter)

Noise Type	S_0	P_0	S^*	$\Delta S(\%)$	P^*	$\Delta P(\%)$
babble	0.67	1.69	0.68	1.18	1.74	3.13
subway	0.65	1.51	0.65	0.35	1.52	0.44
white	0.66	1.42	0.66	0.70	1.44	1.47
street	0.73	1.74	0.73	0.37	1.82	4.68
restaurant	0.67	1.70	0.67	0.81	1.72	1.50
airport	0.72	1.81	0.72	0.07	1.84	2.03

database and 100 sentences as a test database, corrupted by babble noise at global SNR levels of 0dB. We trained a set of equalizers with a different value of A, and the performances are shown in Table 11 and Fig.1.

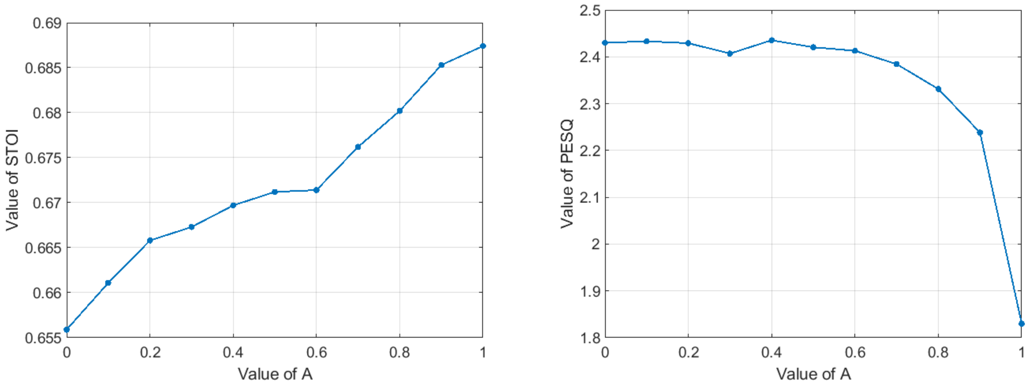
From Fig.1(a), we can observe that when A becomes larger, the score of STOI increases. Conversely, the smaller A obtained the larger PESQ score. Besides, Fig.2 shows the results of a set of Pareto optimal solutions. As observed, when the value of STOI increases, the value of PESQ decreases, which is in accord with Pareto fashion. Comparing with the single objective optimization, the two-objective optimization allows a trade-off between the two criteria. For a STOI-based pattern, the PESQ has been improved slightly. However, for a PESQ-based pattern, although the PESQ value increase greatly, the STOI values decrease. If two criteria are comprised and a suitable A is chosen, we can obtain an equalizer with both significant PESQ improvement and increasing STOI, such as when $A = 0.9$. To analyse the differences in frequency domain between equalizers, we give the spectrograms of one speech from test set as an example. The spectrograms of the clean signal, noisy signal and denoised signals with different A are given in Fig. 3. It can be seen that the SOTI-based equalizer (A=1) mainly has a positive effect on the low frequency part, while the PESQ-based (A=0) equalizer effect more on the high frequency part. The mixed algorithm also has a trade-off between these two single objective algorithm.

Table 11 Quality Improvements of Different Value of A

Value of A	S_0	S^*	$\Delta S(\%)$	P_0	P^*	$\Delta P(\%)$
0		0.66	-2.16		2.43	44.09
0.1		0.66	-1.38		2.43	44.44
0.2		0.67	-0.67		2.43	44.19
0.3		0.67	-0.44		2.41	42.87
0.4	0.67	0.67	-0.07	1.69	2.44	44.58
0.5		0.67	0.15		2.42	43.69
0.6		0.67	0.18		2.41	43.25
0.7		0.68	0.91		2.38	41.54
0.8		0.68	1.50		2.33	38.37
0.9		0.69	2.25		2.24	32.80
1		0.69	2.56		1.83	8.31

5 Conclusion

This paper presents a method to design a fixed single-channel equalizer by optimizing objective measures STOI and PESQ. The proposed method does not need for any noise estimation or a VAD. The equalizer consists of only a single fixed gain in each frequency channel. The simplicity of the proposed structure makes it ideal in many speech applications. The design of the weighting requires a training process, which is posed as a constrained non-linear optimization problem and solved by using the sequential programming method. Experimental results show that there is natural limit as to how big a training sample is



(a) STOI Improvements (b) PESQ Improvements

Fig. 1 Quality Improvements

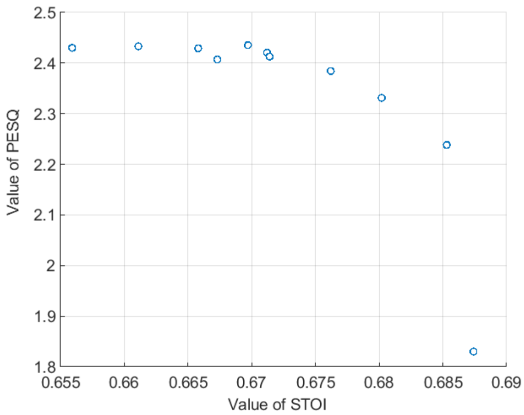


Fig. 2 Results of Pareto optimal solutions

needed and there is greater performance achieved for PESQ improvement compared to STOI. The trade-off between PESQ and STOI is also expressed in a set of Pareto optimal solutions, and the proposed method allows for flexible adjustment of the weighting depending on the requirements. However, as the system is based on a training system, any deviation will result in performance degradation. Future work includes making the system more robust to modelling errors and uncertainties through different learning algorithms.

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Reference

1. J. S. Abel and D. P. Berners. Filter design using second-order peaking and shelving sections. In *ICMC*, 2004.
2. S. An, C. C. Bao, and B. Y. Xia. An adaptive β-order MMSE estimator for speech enhancement using super-Gaussian speech model. In *2013 IEEE China Summit and International Conference on Signal and Information Processing*, pages 327–331. IEEE, 2013.

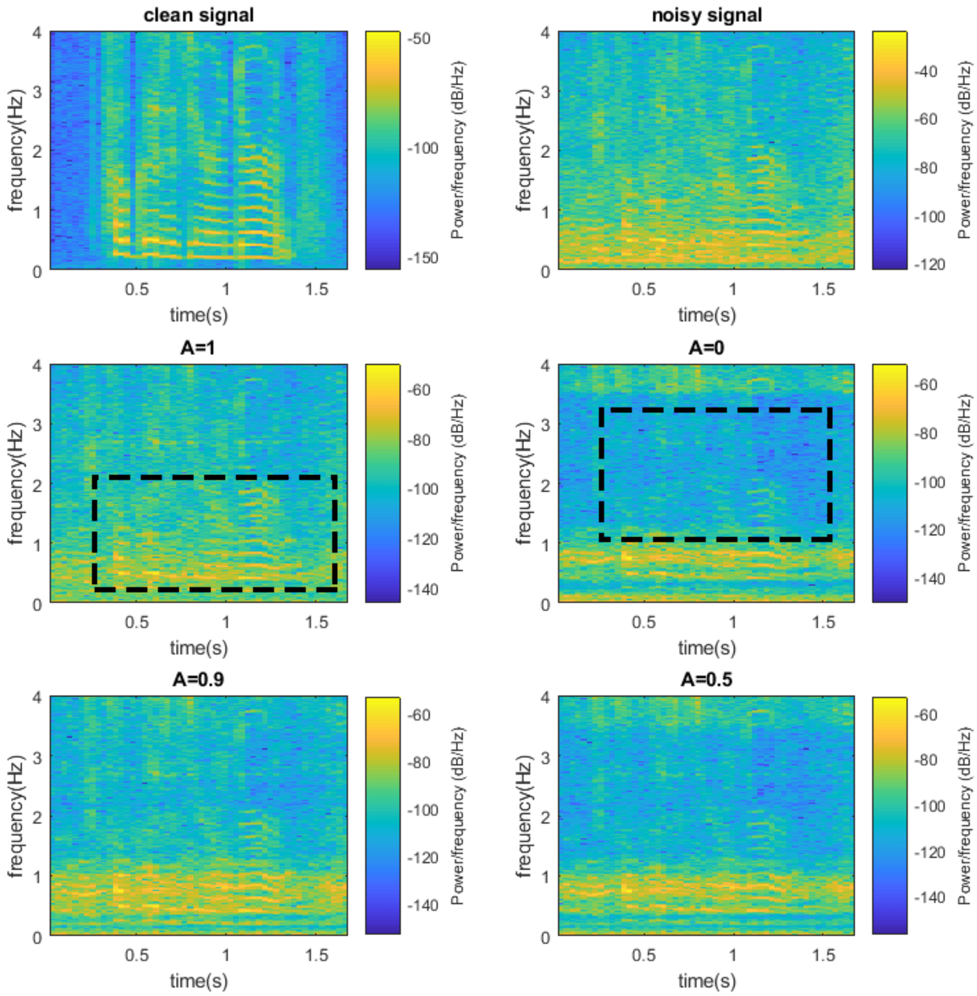


Fig. 3 Spectrogram analysis of clean speech, noisy speech and denoised speech signal with A=1,0,0.9,0.5

3. L. Armijo. Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16(1):1–3, 1966.
4. J. Backman. Digital realisation of phono (RIAA) equalisers. *IEEE Transactions on Consumer Electronics*, 37(3):659–662, 1991.
5. J. G. Beerends, A. P. Hekstra, A. W. Rix, and M. P. Hollier. Perceptual evaluation of speech quality (PESQ) the new itu standard for end-to-end speech quality assessment part ii: psychoacoustic model. *Journal of the Audio Engineering Society*, 50(10):765–778, 2002.
6. J. Benesty and Y. Huang. *Adaptive Signal Processing: Applications to Real-World Problems*. Springer Science & Business Media, 2013.
7. M. Berouti, R. Schwartz, and J. Makhoul. Enhancement of speech corrupted by acoustic noise. In *ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 208–211. IEEE, 1979.
8. M. Biggs. Constrained minimization using recursive equality quadratic programming. *Numerical Methods for Nonlinear Optimization*, pages 411–428, 1972.

9. H. G. Bock and K. J. Plitt. A multiple shooting algorithm for direct solution of optimal control problems. *IFAC Proceedings Volumes*, 17(2):1603–1608, 1984.
10. S. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(2):113–120, 1979.
11. M. Brandt and J. Bitzer. Hum removal filters: Overview and analysis. In *Audio Engineering Society Convention 132*. Audio Engineering Society, 2012.
12. Y. Censor. Pareto optimality in multiobjective problems. *Applied Mathematics and Optimization*, 4(1):41–59, 1977.
13. K. Y. Chan, S. Y. Low, S. Nordholm, and K. F. C. Yiu. A decision-directed adaptive gain equalizer for assistive hearing instruments. *IEEE Transactions on Instrumentation and Measurement*, 63(8):1886–1895, 2014.
14. S. R. Chiluveru, S. Chunarkar, M. Tripathy, B. K. Kaushik, et al. Efficient hardware implementation of DNN-based speech enhancement algorithm with precise sigmoid activation function. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2021.
15. C. A. C. Coello, G. B. Lamont, D. A. Van Veldhuizen, et al. *Evolutionary Algorithms for Solving Multi-Objective Problems*, volume 5. Springer, 2007.
16. J. R. Deller Jr. *Discrete-Time Processing of Speech Signals*. Piscataway, NJ: IEEE Press, 2000.
17. V. H. Diaz-Ramirez and V. Kober. Robust speech processing using local adaptive non-linear filtering. *IET Signal Processing*, 7(5):345–359, 2013.
18. Y. Ephraim and D. Malah. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(6):1109–1121, 1984.
19. T. H. Falk, V. Parsa, J. F. Santos, K. Arehart, O. Hazrati, R. Huber, J. M. Kates, and S. Scollie. Objective quality and intelligibility prediction for users of assistive listening devices: Advantages and limitations of existing tools. *IEEE Signal Processing Magazine*, 32(2):114–124, 2015.
20. R. Fletcher. *Practical Methods of Optimization*. John Wiley & Sons, 2013.
21. S. W. Fu, T. W. Wang, Y. Tsao, X. Lu, and H. Kawai. End-to-end waveform utterance enhancement for direct evaluation metrics optimization by fully convolutional neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(9):1570–1584, 2018.
22. T. Gerkmann and M. Krawczyk. MMSE-optimal spectral amplitude estimation given the STFT-phase. *IEEE Signal Processing Letters*, 20(2):129–132, 2012.
23. F. G. Germain, G. J. Mysore, and T. Fujioka. Equalization matching of speech recordings in real-world environments. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 609–613. IEEE, 2016.
24. R. A. Greiner and M. Schoessow. Design aspects of graphic equalizers. *Journal of the Audio Engineering Society*, 31(6):394–407, 1983.
25. S. P. Han. Superlinearly convergent variable metric algorithms for general nonlinear programming problems. *Mathematical Programming*, 11(1):263–282, 1976.
26. S. Haykin. *Adaptive Filter Theory*. Prentice Hall, New Jersey, 1996.
27. D. Hepsiba and J. Justin. Enhancement of single channel speech quality and intelligibility in multiple noise conditions using Wiener filter and deep CNN. *Soft Computing*, pages 1–11, 2021.
28. Y. Hirata. Digitalization of conventional analog filters for recording use. *Journal of the Audio Engineering Society*, 29(5):333–337, 1981.
29. Y. Hu and P. C. Loizou. Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(1):229–238, 2007.
30. L. B. Jackson. Frequency-domain Steiglitz-McBride method for least-squares IIR filter design, ARMA modeling, and periodogram smoothing. *IEEE Signal Processing Letters*, 15:49–52, 2008.
31. S. Kamath, P. Loizou, et al. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. In *ICASSP*, volume 4, pages 44164–44164. Citeseer, 2002.
32. A. Li, M. Yuan, C. Zheng, and X. Li. Speech enhancement using progressive learning-based convolutional recurrent neural network. *Applied Acoustics*, 166:107347, 2020.
33. P. Lockwood and J. Boudy. Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and the projection, for robust speech recognition in cars. *Speech Communication*, 11(2-3):215–228, 1992.
34. P. C. Loizou. Speech quality assessment. In *Multimedia Analysis, Processing and Communications*, pages 623–654. Springer, 2011.
35. P. C. Loizou. *Speech Enhancement: Theory and Practice*. CRC press, 2013.
36. J. Ma, Y. Hu, and P. C. Loizou. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. *The Journal of the Acoustical Society of America*, 125(5):3387–3405, 2009.
37. M. Massberg. Digital low-pass filter design with analog-matched magnitude response. In *Audio Engineering Society Convention 131*. Audio Engineering Society, 2011.
38. P. Mowlaee, R. Saeidi, M. G. Christensen, and R. Martin. Subjective and objective quality assessment of single-channel speech separation algorithms. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 69–72. IEEE, 2012.
39. U. Nangia, N. Jain, and C. Wadhwa. Surrogate worth trade-off technique for multi-objective optimal power flows. *IEE Proceedings-Generation, Transmission and Distribution*, 144(6):547–553, 1997.
40. P. Ngatchou, A. Zarei, and A. El-Sharkawi. Pareto multi objective optimization. In *Proceedings of the 13th International Conference on, Intelligent Systems Application to Power Systems*, pages 84–91. IEEE, 2005.

41. J. Nocedal and S. Wright. *Numerical Optimization*. Springer Science & Business Media, 2006.
42. S. Ogata and T. Shimamura. Reinforced spectral subtraction method to enhance speech signal. In *Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology. TENCON 2001 (Cat. No. 01CH37239)*, volume 1, pages 242–245. IEEE, 2001.
43. E. Plourde and B. Champagne. Generalized bayesian estimators of the spectral amplitude for speech enhancement. *IEEE Signal Processing Letters*, 16(6):485–488, 2009.
44. M. J. Powell. A fast algorithm for nonlinearly constrained optimization calculations. In *Numerical Analysis*, pages 144–157. Springer, 1978.
45. J. Rämö, V. Välimäki, and M. Tikander. Perceptual headphone equalization for mitigation of ambient noise. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 724–728. IEEE, 2013.
46. A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, volume 2, pages 749–752. IEEE, 2001.
47. S. Samui, I. Chakrabarti, and S. K. Ghosh. Improved single channel phase-aware speech enhancement technique for low signal-to-noise ratio signal. *IET Signal Processing*, 10(6):641–650, 2016.
48. N. Sharma, M. K. Singh, S. Y. Low, and A. Kumar. Weighted sigmoid-based frequency-selective noise filtering for speech denoising. *Circuits, Systems, and Signal Processing*, 40(1):276–295, 2021.
49. A. W. Soewito. *Least Square Digital Filter Fesign in the Frequency Domain*. PhD thesis, 1991.
50. C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7):2125–2136, 2011.
51. H. Tao, X. Li, W. Paszke, V. Stojanovic, and H. Yang. Robust PD-type iterative learning control for discrete systems with multiple time-delays subjected to polytopic uncertainty and restricted frequency-domain. *Multidimensional Systems and Signal Processing*, 32(2):671–692, 2021.
52. T. Van Waterschoot and M. Moonen. Fifty years of acoustic feedback control: State of the art and future challenges. *Proceedings of the IEEE*, 99(2):288–327, 2010.
53. Y. Xian, Y. Sun, W. Wang, and S. M. Naqvi. Convolutional fusion network for monaural speech enhancement. *Neural Networks*, 2021.
54. X. Zhang, H. Wang, V. Stojanovic, P. Cheng, S. He, X. Luan, and F. Liu. Asynchronous fault detection for interval type-2 fuzzy nonhomogeneous higher-level Markov jump systems with uncertain transition probabilities. *IEEE Transactions on Fuzzy Systems*, pages 1–1, 2021.
55. N. Zheng, Y. Shi, W. Rong, and Y. Kang. Effects of skip connections in CNN-based architectures for speech enhancement. *Journal of Signal Processing Systems*, 92:875–884, 2020.