Robustness metric-based tuning of the augmented Kalman filter for the enhancement of speech corrupted with coloured noise

Aidan E.W. George, Stephen So, Ratna Ghosh, Kuldip K. Paliwal

PII: S0167-6393(17)30467-3
DOI: https://doi.org/10.1016/j.specom.2018.10.002
Reference: SPECOM 2596

To appear in: *Speech Communication*

Received date: 16 January 2018
Revised date: 21 August 2018
Accepted date: 10 October 2018

Please cite this article as: Aidan E.W. George, Stephen So, Ratna Ghosh, Kuldip K. Paliwal, Robustness metric-based tuning of the augmented Kalman filter for the enhancement of speech corrupted with coloured noise, *Speech Communication* (2018), doi: https://doi.org/10.1016/j.specom.2018.10.002

Highlights

- Proposed method reduces Kalman filter estimation bias caused by noisy speech input
- The Kalman filter gain is adaptively and dynamically tuned by a robustness metric
- The proposed method is significantly preferred over the other treatment types

Robustness metric-based tuning of the augmented Kalman filter for the enhancement of speech corrupted with coloured noise

Aidan E.W. George[12], Stephen So[1], Ratna Ghosh[3], Kuldip K. Paliwal[4]

Abstract

In this paper, we describe a tuning method based on a robustness metric and extended to work with the augmented Kalman filter for enhancing coloured-noise-corrupted speech. The method proposed within utilises the robustness metric to provide dynamic and adaptive tuning of the Kalman filter gain in order to reduce the residual noise that results from poor speech model estimates. An analysis of the Kalman filter recursion equations is presented that augments the robustness metric equations to include coloured noise model parameters. Objective and blind AB subjective listening tests were performed on the NOIZEUS speech corpus for both white and coloured noises with the results being compared with the MMSE method. In the blind AB subjective testing, the 15 English-speaking listeners showed preference for the proposed method over both the MMSE and oracle Kalman filter methods (where clean speech parameters were used). These results imply that the proposed tuned Kalman filter produces more perceptibly-acceptable enhanced speech than the oracle Kalman filter, which is considered the ideal for this enhancement technique.

Keywords: Robustness metric, Kalman filter, Speech enhancement

1 Introduction

The undesirable presence of background noise in digitally-recorded speech causes unreliability in speech processing applications. For example, the presence of noise in speech that has been coded for mobile telephony is known to result in a loss of quality and intelligibility [1]. Speech enhancement aims to reduce the impact of noise to improve the quality or intelligibility of processed speech. Many speech enhancement methods have been reported in the literature, which include the Wiener filter [2], spectral subtraction [3], Minimum Mean Square Error - Short-Time Spectral Amplitude (MMSE-STSA) estimation [4], subspace methods [5], the Kalman filter [6], Deep Neural Networks [7, 8], and wiener filtering utilising wavelet thresholding on multi-tapered spectra (W-WT) [9]. These methods typically process speech in the frequency (acoustic) or time domain, but the application of speech enhancement algorithms in the subband domain ([10, 11]) and modulation domain have also been recently investigated (e.g. spectral subtraction [12], MMSE [13] and Kalman filtering [14]). The focus of this paper is on single-channel speech enhancement, where only the additive noise corrupted speech is available.

[1]School of Engineering, Gold Coast campus, Griffith University, QLD, 4222, Australia.
[2]Corresponding email: aidan.george@griffithuni.edu.au
[3]Instrumentation and Electronics Engineering, Jadavpur University, Kolkata 700098, India
[4]Signal Processing Laboratory, Griffith University, Nathan campus, QLD, 4111, Australia

The Kalman filter is an unbiased linear MMSE estimator originally developed for control system applications [15] and was first applied to speech enhancement by Paliwal and Basu [6]. The Kalman filter expands on the stationary signal input operating conditions of the Wiener filter to include non-stationary signals by the use of dynamic models. When a noisy speech signal (also known as the observation or measurement) is provided, the Kalman filter estimates the clean speech state vector by combining the noisy measurement with speech model predictions. The prediction component is determined by LPA (Linear Prediction Analysis) to form a speech production model that is representative of the vocal tract for the particular sound. Typically, the autocorrelation method [16] is used to estimate, from a frame of speech, the linear prediction coefficients (or LPCs) and the excitation variance for the speech model. The Kalman filter produces an *a posteriori* [1] state vector estimated through the use of the Kalman filter recursive equations.

The main limitation with the Kalman filter is that the *oracle case*, in which the model parameters of clean speech are available, rarely occurs in practice. As a result, the speech model parameters need to be estimated from the noise corrupted speech. Depending on the level and type of background noise, the speech model parameters will be inevitably biased, which results in the presence of residual noise in the enhanced speech. The effect of the bias was investigated in [17], where an offset in the temporal trajectory of the Kalman filter gain was observed and was particularly evident during regions that were absent of speech. This Kalman filter gain offset resulted in a portion of the measurement noise to be included in the output.

Reduction of the estimation bias has been previously investigated, in particular the use of recursive algorithms. Iterative algorithms typically estimate the speech model parameters from speech that has been pre-processed by another enhancement algorithm. For instance, the iterative Kalman filter [18] recursively applies the Kalman filter algorithm to provide an enhanced speech model source, and this enhanced speech model is used for filtering the original noise corrupted speech signal. Similarly, in the Kalman-PSC filtering algorithm [19], the LPCs are estimated from speech that has been processed by the Phase Spectrum Compensation (PSC) algorithm [20]. An alternate method was reported in [17], where the application of tapered windows prior to LPC estimation in the first iteration was shown to be beneficial in reducing estimation bias in the iterative Kalman filter. However, the bias reduction provided was of a static nature and often led to the introduction of speech distortion that is characteristic of 'over-suppression', particularly in the regions containing the harmonic structure of voiced speech [17].

In this paper, a non-iterative Kalman filtering algorithm will be presented that dynamically tunes the Kalman filter gain in order to reduce the estimation bias and its adverse effects. The basis for the Kalman filter gain modification is a robustness metric that was first proposed in the instrumentation literature [21], which quantifies the level of robustness of the Kalman filter's speech model at each time sample. Also complementing the understanding of the Kalman filter's operation is the shift of focus from estimating state vectors to that of estimating scalar outputs, as detailed in [17]. Adopting this paradigm shift to the Kalman filter enables a more detailed and insightful understanding of its operation. We have previously reported the application of the robustness and sensitivity metrics in improving the Kalman filter for white noise [22, 23, 24]. The proposed algorithm produces a non-iterative Kalman filter that achieves competitive performance with current speech enhancement methods for speech corrupted with coloured noise. A detailed derivation of the metrics for speech corrupted with coloured noise will be presented within this paper. Objective and subjective listening tests were performed using the NOIZEUS speech corpus [1] to evaluate the proposed algorithm and compare its performance against conventional Kalman filtering (both oracle and non-oracle) as well as the MMSE-STSA method [4].

[1] An *a posteriori* estimate is formed after a measurement or observation is made.

The rest of the paper is organised as follows: in Sections 2, and 3, the recursive Kalman filtering equations will be presented to formalise the mathematical notation utilised in this paper. These equations will then be re-written to show the scalar version of the Kalman filter recursive equations, which are more applicable to speech enhancement. The robustness measure will also be derived, and its application to tuning the Kalman filter will be discussed. Section 5 contains the objective and subjective results that compare proposed Kalman filter with other speech enhancement methods. Section 6 will feature the discussion and conclusion as well as final insights for this method and future directions.

2 Coloured noise speech enhancement using the Kalman filter

2.1 The Kalman recursive equations

The additive noise model that is generally assumed in the problem of speech enhancement can be expressed as follows:

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

where $x(n)$ and $v(n)$ are the clean speech and corrupting (or measurement) noise, respectively, and $y(n)$ is the noise-corrupted speech that is the only signal available in practice.

In the Kalman filter that is used for speech enhancement [17], a pth order linear predictor is used to model the speech signal:

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + w(n) \tag{2}$$

where $\{a_k; k = 1, 2, \ldots, p\}$ are the linear prediction coefficients (or LPCs) and $w(n)$ is the excitation (or process) noise. The speech state vector $\boldsymbol{x}(n)$ is given by:

$$\boldsymbol{x}(n) = \begin{bmatrix} x(n) \\ x(n-1) \\ x(n-2) \\ \vdots \\ x(n-p+1) \end{bmatrix} \tag{3}$$

and the state transition matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \begin{bmatrix} -a_1 & -a_2 & \ldots & -a_{p-1} & -a_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix} \tag{4}$$

The speech production model in Eq. (2) is re-written in terms of these state-space variables:

$$\boldsymbol{x}(n) = \boldsymbol{A}\boldsymbol{x}(n-1) + \boldsymbol{d}_x w(n) \tag{5}$$

where $\boldsymbol{d}_x = [1\ 0\ 0\ \ldots\ 0]^T$.

Traditionally, the Kalman filter formulation assumes that the measurement noise, $v(n)$, is a zero-mean, white Gaussian noise that is uncorrelated with $x(n)$. In this study, we will utilise

the augmented-matrix Kalman filter formulation for the handling of speech that is corrupted by coloured noise, as described in [18]. The coloured measurement noise is modelled by a qth order linear prediction:

$$v(n) = -\sum_{k=1}^{q} b_k v(n-k) + u(n) \tag{6}$$

where $\{b_k; k = 1, 2, \ldots, q\}$ are the measurement noise LPCs and $u(n)$ represents the random component of the measurement noise. We define the measurement noise state vector, $\boldsymbol{v}(n)$, and state transition matrix, $\boldsymbol{B}$, as follows:

$$\boldsymbol{v}(n) = \begin{bmatrix} v(n) \\ v(n-1) \\ v(n-2) \\ \vdots \\ v(n-q+1) \end{bmatrix} \tag{7}$$

$$\boldsymbol{B} = \begin{bmatrix} -b_1 & -b_2 & \ldots & -b_{q-1} & -b_q \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix} \tag{8}$$

This allows us to form a state-space equation for the coloured measurement noise:

$$\boldsymbol{v}(n) = \boldsymbol{B}\boldsymbol{v}(n-1) + \boldsymbol{d}_v u(n) \tag{9}$$

where $\boldsymbol{d}_v = [1\ 0\ 0\ \ldots\ 0]^T$. The state-space equations for the speech and measurement noise can then be combined into *augmented form*:

$$\begin{bmatrix} \boldsymbol{x}(n) \\ \boldsymbol{v}(n) \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}(n-1) \\ \boldsymbol{v}(n-1) \end{bmatrix} + \begin{bmatrix} \boldsymbol{d}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{d}_v \end{bmatrix} \begin{bmatrix} w(n) \\ u(n) \end{bmatrix} \tag{10}$$

or

$$\bar{\boldsymbol{x}}(n) = \bar{\boldsymbol{A}}\bar{\boldsymbol{x}}(n-1) + \bar{\boldsymbol{D}}\bar{\boldsymbol{w}}(n) \tag{11}$$

Likewise, the additive noise measurement equation Eq. (1) can be represented in terms of the augmented states:

$$y(n) = \begin{bmatrix} \boldsymbol{c}_x^T & \boldsymbol{c}_v^T \end{bmatrix} \begin{bmatrix} \boldsymbol{x}(n) \\ \boldsymbol{v}(n) \end{bmatrix} \tag{12}$$

or

$$y(n) = \bar{\boldsymbol{c}}^T \bar{\boldsymbol{x}}(n) \tag{13}$$

where $\boldsymbol{c}_x = [1\ 0\ 0\ \ldots\ 0]^T$ and $\boldsymbol{c}_v = [1\ 0\ 0\ \ldots\ 0]^T$ are $p \times 1$ and $q \times 1$ vectors, respectively. Equations (11) and (13) together form the augmented state-space representation of the Kalman filter when the noise is coloured.

The Kalman filter recursively computes an unbiased and linear MMSE estimate $\hat{\bar{\boldsymbol{x}}}(n|n)$ of the augmented state vector $\bar{\boldsymbol{x}}(n)$ at given time n, given the noisy measurement $y(n)$, by using the

following equations:

$$\bar{\boldsymbol{P}}(n|n-1) = \bar{\boldsymbol{A}}\bar{\boldsymbol{P}}(n-1|n-1)\bar{\boldsymbol{A}}^T + \bar{\boldsymbol{D}}\bar{\boldsymbol{Q}}\bar{\boldsymbol{D}}^T \tag{14}$$

$$\bar{\boldsymbol{K}}(n) = \bar{\boldsymbol{P}}(n|n-1)\bar{\boldsymbol{c}}\left[\bar{\boldsymbol{c}}^T\bar{\boldsymbol{P}}(n|n-1)\bar{\boldsymbol{c}}\right]^{-1} \tag{15}$$

$$\hat{\bar{\boldsymbol{x}}}(n|n-1) = \bar{\boldsymbol{A}}\hat{\bar{\boldsymbol{x}}}(n-1|n-1) \tag{16}$$

$$\hat{\bar{\boldsymbol{x}}}(n|n) = \hat{\bar{\boldsymbol{x}}}(n|n-1) + \bar{\boldsymbol{K}}(n)\left[y(n) - \bar{\boldsymbol{c}}^T\hat{\bar{\boldsymbol{x}}}(n|n-1)\right] \tag{17}$$

$$\bar{\boldsymbol{P}}(n|n) = \left[\boldsymbol{I} - \bar{\boldsymbol{K}}(n)\bar{\boldsymbol{c}}^T\right]\bar{\boldsymbol{P}}(n|n-1) \tag{18}$$

In these equations, $\bar{\boldsymbol{Q}}$ is the process noise covariance matrix and is given by:

$$\bar{\boldsymbol{Q}} = \begin{bmatrix} \sigma_w^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix} \tag{19}$$

where σ_w^2 and σ_u^2 are the process noise variances of the speech and measurement noise, respectively.

During the operation of the Kalman filter, the noise-corrupted speech, $y(n)$, is windowed into non-overlapped and short (e.g. 20 ms) frames and the LPCs and excitation variance, σ_w^2, are estimated for each frame. These LPCs remain constant during the Kalman filtering of speech samples in the frame, while the Kalman filter gain $\bar{\boldsymbol{K}}(n)$, *a posteriori* error covariance, $\bar{\boldsymbol{P}}(n|n)$, and state vector estimate, $\hat{\bar{\boldsymbol{x}}}(n|n)$, are continually updated on a sample-by-sample basis. The measurement noise model parameters, $\{b_k; k = 1, 2, \ldots, q\}$ and σ_u^2, are estimated during frames where there is no speech.

2.2 Paradigm shift from estimated state vectors to estimated speech samples

The conventional view of the Kalman filter is that it recursively estimates [using Eq. (17)] the augmented *state vector*, $\bar{\boldsymbol{x}}(n|n)$, by correcting the *a priori*[2] state vector estimate, $\hat{\bar{\boldsymbol{x}}}(n|n-1)$, with a weighted innovation signal. However, in the context of speech enhancement, we are ultimately interested in the estimated clean speech sample $\hat{x}(n|n)$ (which is a scalar), rather than the state vector. The remaining elements in the state vector are generally unused in the output, except in special cases such as the delayed Kalman filter [6]. Therefore, as was noted in [17, 21], it is necessary for us to shift the focus of the Kalman filtering formulation from that of estimating state vectors to *estimating speech samples*. The derivation that follows will show that the equations for computing the estimated speech sample can be expressed as *purely scalar quantities*. For the purposes of simplifying the mathematical notation, we will drop the use of the bar for representing augmented vector and matrix variables (e.g. $\bar{\boldsymbol{c}} \equiv \boldsymbol{c}$) except $\bar{\boldsymbol{A}}$, and also drop the use of the hat for estimated vectors [and hence for example, $\hat{\bar{\boldsymbol{x}}}(n|n-1) \equiv \bar{\boldsymbol{x}}(n|n-1) \equiv \boldsymbol{x}(n|n-1)$].

We begin the formulation by defining the innovation signal $q(n)$ and its covariance matrix $\boldsymbol{S}(n)$.

$$q(n) = y(n) - \boldsymbol{c}^T\boldsymbol{x}(n|n-1) \tag{20}$$

The innovation signal is a random process that contains the new information that is provided by the new measurement $y(n)$. The covariance matrix $\boldsymbol{S}(n)$ of the innovation signal is given by:

$$\begin{aligned} \boldsymbol{S}(n) &= \boldsymbol{c}^T\boldsymbol{P}(n|n-1)\boldsymbol{c} \\ &= \boldsymbol{c}^T\bar{\boldsymbol{A}}\boldsymbol{P}(n-1|n-1)\bar{\boldsymbol{A}}^T\boldsymbol{c} + \boldsymbol{c}^T\boldsymbol{D}\boldsymbol{Q}\boldsymbol{D}^T\boldsymbol{c} \end{aligned} \tag{21}$$

where we have substituted Eq. (14). Using some matrix algebra, the second term in the summation

[2] An *a priori* estimate is one that is based on prior knowledge (eg. a model) in the absence of a measurement.

above can be reduced down to:

$$\boldsymbol{c}^T \boldsymbol{D}\boldsymbol{Q}\boldsymbol{D}^T \boldsymbol{c} = \sigma_w^2 + \sigma_u^2 \tag{22}$$

Let us consider the first term of Eq. (21). We first break up the product of terms:

$$\begin{aligned} \boldsymbol{c}^T \bar{\boldsymbol{A}} &= \begin{bmatrix} \boldsymbol{c}_x^T & \boldsymbol{c}_v^T \end{bmatrix} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{c}_x^T \boldsymbol{A} & \boldsymbol{c}_v^T \boldsymbol{B} \end{bmatrix} \end{aligned} \tag{23}$$

The *a posteriori* error covariance matrix, $\boldsymbol{P}(n-1|n-1)$, can be expressed as:

$$\begin{aligned} \boldsymbol{P}(n-1|n-1) &= E\{\boldsymbol{e}(n-1|n-1)\boldsymbol{e}^T(n-1|n-1)\} \\ &= E\left\{ \begin{bmatrix} \boldsymbol{e}_x(n-1|n-1) \\ \boldsymbol{e}_v(n-1|n-1) \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_x^T(n-1|n-1) & \boldsymbol{e}_v^T(n-1|n-1) \end{bmatrix} \right\} \\ &= \begin{bmatrix} \boldsymbol{P}_{xx}(n-1|n-1) & \boldsymbol{P}_{xv}(n-1|n-1) \\ \boldsymbol{P}_{vx}(n-1|n-1) & \boldsymbol{P}_{vv}(n-1|n-1) \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{P}_{xx}(n-1|n-1) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{P}_{vv}(n-1|n-1) \end{bmatrix} \\ &\text{(since } x(n) \text{ and } v(n) \text{ are zero-mean and uncorrelated)} \end{aligned} \tag{24}$$

where,

$$\begin{aligned} \boldsymbol{P}_{xx}(n-1|n-1) &= E\left\{\boldsymbol{e}_x(n-1|n-1)\boldsymbol{e}_x^T(n-1|n-1)\right\} \\ \boldsymbol{P}_{xv}(n-1|n-1) &= E\left\{\boldsymbol{e}_x(n-1|n-1)\boldsymbol{e}_v^T(n-1|n-1)\right\} \\ \boldsymbol{P}_{vv}(n-1|n-1) &= E\left\{\boldsymbol{e}_v(n-1|n-1)\boldsymbol{e}_v^T(n-1|n-1)\right\} \end{aligned}$$

and $E\{\bullet\}$ is the expectation operator.
Substituting this into the first term of Eq. (21), we can express it as a summation:

$$\begin{aligned} \boldsymbol{c}^T \bar{\boldsymbol{A}} \boldsymbol{P}(n-1|n-1) \bar{\boldsymbol{A}}^T \boldsymbol{c} &= \boldsymbol{c}_x^T \boldsymbol{A} \boldsymbol{P}_{xx}(n-1|n-1) \boldsymbol{A}^T \boldsymbol{c}_x + \boldsymbol{c}_v^T \boldsymbol{B} \boldsymbol{P}_{vv}(n-1|n-1) \boldsymbol{B}^T \boldsymbol{c}_v \\ &= \alpha^2(n) + \beta^2(n) \end{aligned} \tag{25}$$

$$\begin{aligned} \text{Where,} \quad \alpha^2(n) &= \boldsymbol{c}_x^T \boldsymbol{A} \boldsymbol{P}_{xx}(n-1|n-1) \boldsymbol{A}^T \boldsymbol{c}_x \\ \beta^2(n) &= \boldsymbol{c}_v^T \boldsymbol{B} \boldsymbol{P}_{vv}(n-1|n-1) \boldsymbol{B}^T \boldsymbol{c}_v \end{aligned}$$

In this equation, $\alpha^2(n)$ and $\beta^2(n)$ represent the transmission of *a posteriori* error variances (of the speech and measurement noise samples, respectively) by the augmented dynamic model from the previous time sample. Therefore, the innovation covariance can be written in terms of scalar quantities only:

$$S(n) = \alpha^2(n) + \beta^2(n) + \sigma_w^2 + \sigma_u^2 \tag{26}$$

We can see that the innovation covariance consists of three error *variance* components: (1) the variance of the transmitted errors from the previous sample due to prediction by the dynamic model $[\alpha^2(n) + \beta^2(n)]$; (2) the variance of the error in the speech production model σ_w^2; and (3) the variance of the error in the measurement noise model σ_u^2.

Next, we will formulate the estimated clean speech sample $\hat{x}(n|n)$, which is the output of the

Kalman filter:

$$\hat{x}(n|n) = \boldsymbol{g}^T \hat{\boldsymbol{x}}(n|n) \tag{27}$$

$$\text{where} \quad \boldsymbol{g} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{28}$$

Substituting Eq. (17) into Eq. (27), we obtain:

$$\begin{aligned} \hat{x}(n|n) &= \boldsymbol{g}^T \hat{\boldsymbol{x}}(n|n-1) + \boldsymbol{g}^T \boldsymbol{K}(n) \left[y(n) - \boldsymbol{c}^T \hat{\boldsymbol{x}}(n|n-1) \right] \\ &= \boldsymbol{g}^T \hat{\boldsymbol{x}}(n|n-1) + \boldsymbol{g}^T \boldsymbol{K}(n) y(n) - \boldsymbol{g}^T \boldsymbol{K}(n) \boldsymbol{c}^T \hat{\boldsymbol{x}}(n|n-1) \\ &= \hat{x}(n|n-1) + K_0(n) y(n) - K_0(n) [\hat{x}(n|n-1) + v(n|n-1)] \\ &= [1 - K_0(n)] \hat{x}(n|n-1) + K_0(n) [y(n) - v(n|n-1)] \end{aligned} \tag{29}$$

where the Kalman gain scalar term, $K_0(n)$, can be expressed as:

$$\begin{aligned} K_0(n) &= \boldsymbol{g}^T K(n) \\ &= \boldsymbol{g}^T \boldsymbol{P}(n|n-1) \boldsymbol{c} S^{-1}(n) \\ &= \left[\boldsymbol{g}^T \boldsymbol{A} \boldsymbol{P}(n-1|n-1) \boldsymbol{A}^T \boldsymbol{c} + \boldsymbol{g}^T \boldsymbol{D} \boldsymbol{Q} \boldsymbol{D}^T \boldsymbol{c} \right] S^{-1}(n) \end{aligned} \tag{30}$$

Let us simplify the bracketed expression of Eq. (30). Using the steps similar to the derivation of Eqs. (22) and (25), it can be shown that:

$$\begin{aligned} \boldsymbol{g}^T \boldsymbol{A} \boldsymbol{P}(n-1|n-1) \boldsymbol{A}^T \boldsymbol{c} = \boldsymbol{g}^T \boldsymbol{A} \boldsymbol{P}(n-1|n-1) \boldsymbol{A}^T \boldsymbol{g} &= \boldsymbol{c}_x^T \boldsymbol{A} \boldsymbol{P}_{xx} \boldsymbol{A}^T \boldsymbol{c}_x \\ &= \alpha^2(n) \end{aligned} \tag{31}$$

$$\boldsymbol{g}^T \boldsymbol{D} \boldsymbol{Q} \boldsymbol{D}^T \boldsymbol{c} = \boldsymbol{g}^T \boldsymbol{D} \boldsymbol{Q} \boldsymbol{D}^T \boldsymbol{g} = \sigma_w^2 \tag{32}$$

and so,

$$\boldsymbol{g}^T \boldsymbol{P}(n|n-1) \boldsymbol{g} = \boldsymbol{g}^T \boldsymbol{P}(n|n-1) \boldsymbol{c} = \boldsymbol{c}^T \boldsymbol{P}(n|n-1) \boldsymbol{g} = \alpha^2(n) + \sigma_w^2 \tag{33}$$

Bringing these results together, the scalar Kalman gain term $K_0(n)$ can be expressed as:

$$K_0(n) = \frac{\alpha^2(n) + \sigma_w^2}{\alpha^2(n) + \beta^2(n) + \sigma_w^2 + \sigma_u^2} \tag{34}$$

Therefore, using Eqs. (29) and (34), we can now re-write the state vector-based Kalman update equation (17) in terms of the estimated speech sample:

$$\hat{x}(n|n) = \frac{\beta^2(n) + \sigma_u^2}{\alpha^2(n) + \beta^2(n) + \sigma_w^2 + \sigma_u^2} \hat{x}(n|n-1) + \frac{\alpha^2(n) + \sigma_w^2}{\alpha^2(n) + \beta^2(n) + \sigma_w^2 + \sigma_u^2} [y(n) - v(n|n-1)] \tag{35}$$

In order to check the validity of this equation, it is useful to compute the output for two extreme cases. The first case is when there is speech but no noise present in the measurement, i.e. $v(n) = 0$. In this case, we can establish that $\beta^2(n) = 0$ and $\sigma_u^2 = 0$, so the output of the Kalman filter

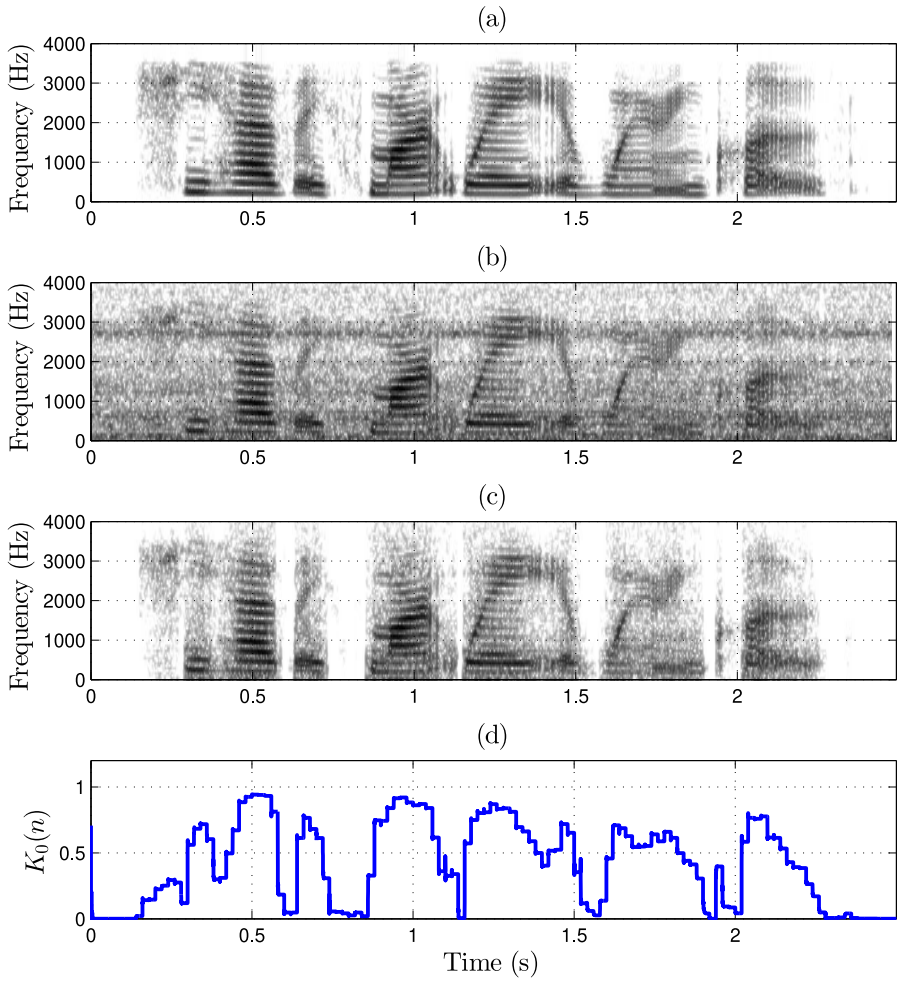

Figure 1: Spectrograms and plots of the Kalman filter gain for speech file sp26.wav (*"She had a smart way of wearing clothes"*): (a) clean speech; (b) speech corrupted with *F16* noise at 5 dB (PESQ = 2.11); (c) output of the Kalman filter (oracle case) when input is noise-corrupted speech (PESQ = 2.38); (d) plot of scalar Kalman filter gain $K_0(n)$ with clean speech input.

becomes:

$$\begin{aligned}\hat{x}(n|n) &= 0 + \frac{\alpha^2(n) + \sigma_w^2}{\alpha^2(n) + \sigma_w^2}[y(n) - 0] \\ &= y(n)\end{aligned}$$

This is a valid result because the measurement signal is clean, so there is no requirement to use the dynamic model. In other words, the scalar Kalman filter gain, $K_0(n)$, is equal to one. In the second extreme case, we assume that there is noise present but no speech in the measurement, i.e. $x(n) = 0$. This corresponds to the silent regions of speech, where $\alpha^2(n) = 0$ and $\sigma_w^2 = 0$:

$$\begin{aligned}\hat{x}(n|n) &= \frac{\beta^2(n) + \sigma_u^2}{\beta^2(n) + \sigma_u^2}\hat{x}(n|n-1) + 0 \\ &= 0 \quad \text{since } a_k = 0 \text{ for } k = 1, 2, \ldots, p\end{aligned} \tag{36}$$

Therefore, in the silent regions, the Kalman filter gain, $K_0(n)$, is equal to zero in order to remove the noisy measurement from the output. Figure 1 shows spectrograms of the output of the Kalman filter in the oracle case for speech that has been corrupted with *F16* noise from the NOISEX-92 database. In Figure 1(d), the scalar Kalman filter gain, $K_0(n)$, has been plotted for the clean speech input to the Kalman filter. When Kalman filtering speech with no noise present, we can see that the scalar Kalman filter gain approaches one for the duration of the speech regions, as noted before. For the opposing case where only noise is present, we can see $K_0(n)$ is approximately zero in these regions of silence (i.e. where there is only noise but no speech present).

2.3 The effect of biased estimates of speech production model parameters

The derivations in the previous section are based on the availability of accurate estimates of the speech production model parameters, as in the case of clean speech. Such parameter estimates enable us to express the Kalman filter gain and the output in terms of purely scalar quantities. However, since clean speech is not available in practice, the parameters for the speech production model need to be estimated from the noise-corrupted speech, $y(n)$. The presence of noise causes the estimates of the LPCs, $\{a_k\}$, and process noise variance, σ_w^2, to be biased. Figure 2(c) shows the spectrogram of the output of the Kalman filter for the non-oracle case. It can be seen that there is a degree of residual noise in the enhanced output, even in the silent regions where there is no speech present. When compared with that for the oracle case (Figure 1(d)), the scalar Kalman filter gain in Figure 2(d) appears to be offset in the positive direction, which increases the overall contribution of the noisy measurement signal in the output.

In order to analyze this effect, let us consider the biased process noise variance obtained when using noise corrupted speech as the input, $\tilde{\sigma}_w^2$, for the linear prediction model of speech. Assuming that $x(n)$ and $v(n)$ are zero mean and not correlated, $\tilde{\sigma}_w^2$ can be estimated from the noise-corrupted speech $y(n)$ as [17]:

$$\begin{aligned}\tilde{\sigma}_w^2 &= R_{yy}(0) + \sum_{k=1}^{p} \tilde{a}_k R_{yy}(k) \\ \text{Where,} \quad R_{yy}(k) &= R_{xx}(k) + R_{vv}(k) \\ &= R_{xx}(0) + \sum_{k=1}^{p} \tilde{a}_k R_{xx}(k) + R_{vv}(0) + \sum_{k=1}^{p} \tilde{a}_k R_{vv}(k)\end{aligned} \tag{37}$$

where $R_{xx}(k), R_{vv}(k), R_{yy}(k)$ are the autocorrelation coefficients of $x(n)$, $v(n)$, and $y(n)$, respectively, and $\tilde{a}_k$ are the biased LPCs.

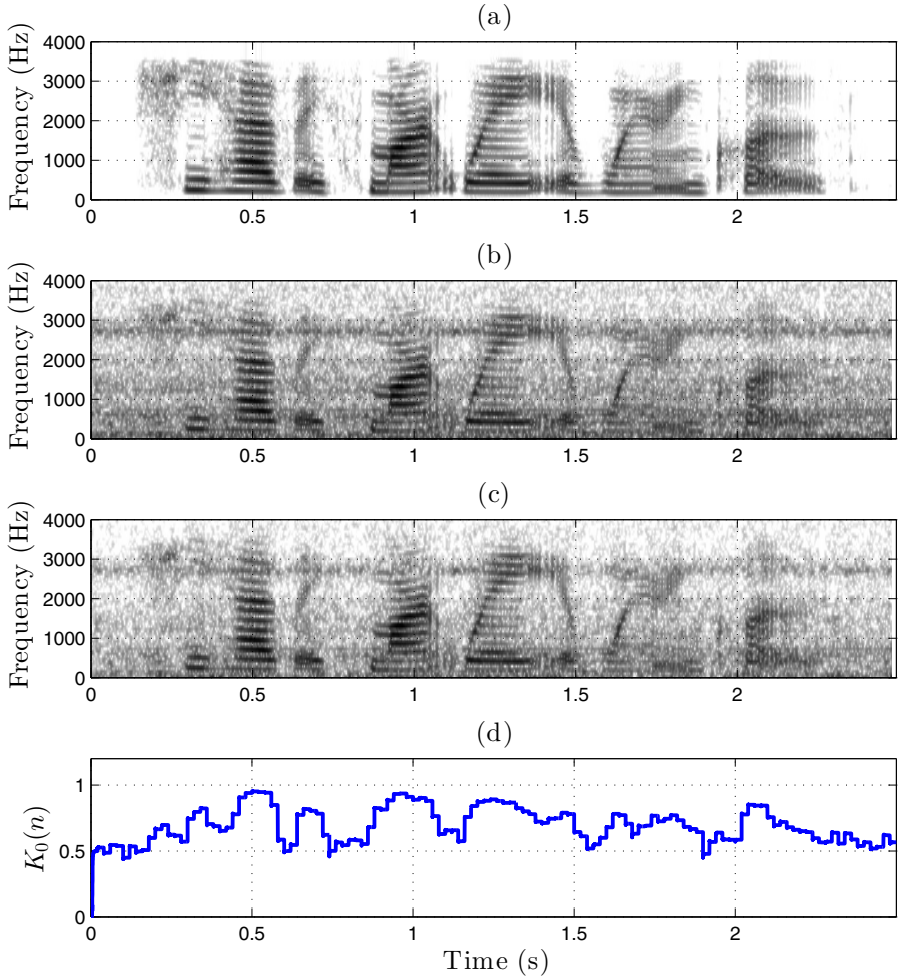

Figure 2: Spectrograms and plots of the Kalman filter gain for speech file sp26.wav (*"She had a smart way of wearing clothes"*): (a) clean speech; (b) speech corrupted with *F16* noise at 5 dB (PESQ = 2.11); (c) output of the Kalman filter (non-oracle case) (PESQ = 2.24); (d) plot of scalar Kalman filter gain $K_0(n)$.

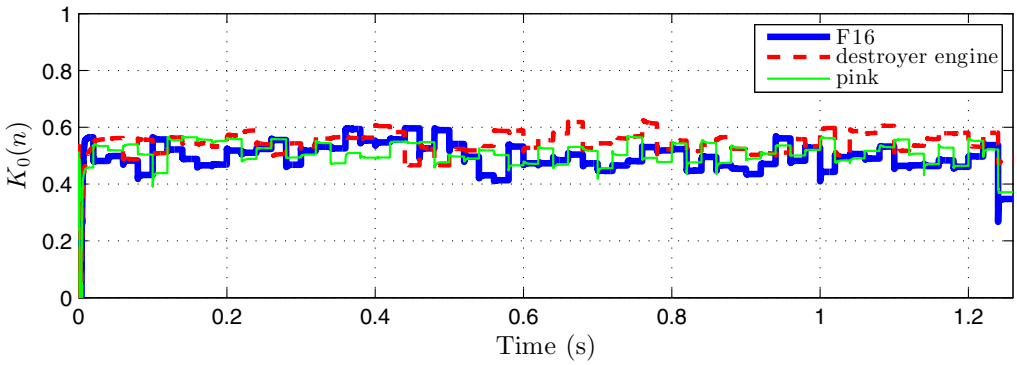

Figure 3: Plot of scalar Kalman filter gains for different coloured noises when no speech is present

In this case, the Kalman update equation (35) becomes:

$$\hat{x}(n|n) = \frac{\beta^2(n) + \sigma_u^2}{\tilde{\alpha}^2(n) + \beta^2(n) + \tilde{\sigma}_w^2 + \sigma_u^2}\hat{x}(n|n-1) + \frac{\tilde{\alpha}^2(n) + \tilde{\sigma}_w^2}{\tilde{\alpha}^2(n) + \beta^2(n) + \tilde{\sigma}_w^2 + \sigma_u^2}[y(n) - v(n|n-1)] \tag{38}$$

Let us consider the case where there is measurement noise but no speech is present [i.e. $x(n) = 0$, hence $R_{xx}(k) = 0$ for all k and $\tilde{\sigma}_w^2 = \sigma_u^2$ (since we are now modelling the measurement noise)]:

$$\hat{x}(n|n) = \frac{\beta^2(n) + \sigma_u^2}{\tilde{\alpha}^2(n) + \beta^2(n) + 2\sigma_u^2}\hat{x}(n|n-1) + \frac{\tilde{\alpha}^2(n) + \sigma_u^2}{\tilde{\alpha}^2(n) + \beta^2(n) + 2\sigma_u^2}[y(n) - v(n|n-1)] \tag{39}$$

When compared with the same equation for the oracle case [Eq. (36)], we can see the effect of using the biased estimates of the speech model parameters. The output sample comprises some fraction [represented by the scalar Kalman filter gain $K_0(n)$] of the measurement noise innovation signal. This case was simulated by Kalman filtering the noise only, where it was observed that the values of $\tilde{\alpha}^2(n)$, $\beta^2(n)$, and σ_u^2 tended to be quite similar. If we assume that $\tilde{\alpha}^2(n) \approx \beta^2(n) \approx \sigma_u^2$, then $K_0(n) \approx 0.5$. We can see in Figure 3, which shows the scalar Kalman filter gain plots for three coloured noises (F16, destroyer engine, pink) that $K_0(n)$ fluctuates around 0.5 when there is no speech present. This result is similar to that reported in [17] for the white noise case, where $b_k = 0$ for all k, hence $\beta^2(n) = 0$, $v(n|n-1) = 0$, and $\tilde{\alpha}^2(n) = 0$.

We observe from the previous discussion that there are two biased quantities that result from poor estimates of parameters due to the presence of additive noise: (1) biased speech LPC estimates $\{\tilde{a}_k\}$, which are associated with the distortion in the power spectrum of the speech; and (2) biased process noise variance $\tilde{\sigma}_w^2$, which causes an increase in the calculated mean squared error or uncertainty in the model. The biased LPC estimates lead to a higher value of $K_0(n)$, as compared to the oracle case. Hence, a low value of $[1 - K_0(n)]$ multiplies the strong predicted component of speech $\hat{x}(n|n-1)$, leading to speech distortion in the signal regions. In addition, as shown in the case of noise only, the presence of the biased process noise variance, $\tilde{\sigma}_w^2$, results in an offset in the value of $K_0(n)$ during the silent regions. This causes a fraction of the noise in the measurement signal to be included in the output of the Kalman filter. It is therefore desirable to correct the bias in the Kalman filter gain.

3 Kalman filter tuning using the robustness metric

3.1 Robustness in the Kalman filter

The Kalman filter is said to *robust* due to its ability to deal with uncertainties or inaccuracies in the parameters of its dynamic model. These may occur when there is a mismatch between the model-based prediction and the signal to be estimated, either because the dynamic model is not adequate for predicting the signal or since the signal characteristics have somewhat deviated from those used to estimate the model parameters. A robust Kalman filter would mitigate the uncertainty or variance in the model parameters by offsetting their contribution to the *a posteriori* estimate, in favour of the measurement signal. In speech enhancement, the uncertainty or variance of the model relates to the inability of the low-order speech production model to estimate the clean speech. Hence, in the case of voiced speech that has a harmonic structure, such a robust filter would tend to rely less on the low-order speech model, which is based on short-term autocorrelations, since it is unable to capture the long-term correlation information. In contrast, the mean square prediction error, which is also the process noise variance σ_w^2, becomes high for this voiced speech. Hence, the Kalman filter enters into a robust mode of operation with high filter gain $K_0(n)$ and so the measurement $y(n)$ is favoured.

The changes in Kalman filter gain due to the presence of voiced speech can be clearly observed in Figure 1(d) for the oracle case, where the Kalman filter becomes more robust in the voiced speech regions and less in the silent regions, where uncertainty in the speech model is close to zero. In other words, the Kalman filter becomes more sensitive in the silent regions. Generally speaking, *sensitivity* relates to the degree to which a controller changes in accordance to variations in its dynamic model. In the speech enhancement context, a sensitive Kalman filter would favour its dynamic model in response to an unreliable or heavily noise-corrupted (or low signal-to-noise ratio) measurement.

For the practical case where LPCs are estimated from noise-corrupted speech, the Kalman filter is operating in a 'semi'-robust mode during the silent regions due to use of poor LPC estimates, as was shown in Figure 2(d). This leads to residual noise in the output and degrades the enhancement performance of the Kalman filter. What we require is a performance index, or metric, that quantifies the level of robustness in the Kalman filter, so that it can be used to tune its operation. One such set of metrics for the robustness and sensitivity of the Kalman filter was proposed in [21], which we will analyze and adopt for use in the speech enhancement context in the next section.

3.2 Metrics for measuring robustness and sensitivity in the Kalman filter

We begin the derivation by looking at the mean squared error [expressed as $\boldsymbol{g}^T\boldsymbol{P}(n|n)\boldsymbol{g}$] of the output sample $\hat{x}(n|n)$:

$$\begin{aligned}\boldsymbol{g}^T\boldsymbol{P}(n|n)\boldsymbol{g} &= \boldsymbol{g}^T[\boldsymbol{I}-\boldsymbol{K}(n)\boldsymbol{c}^T]\boldsymbol{P}(n|n-1)\boldsymbol{g} \ \text{ [from (18)]}\\ &= \boldsymbol{g}^T\boldsymbol{P}(n|n-1)\boldsymbol{g} - \boldsymbol{g}^T\boldsymbol{K}(n)\boldsymbol{c}^T\boldsymbol{P}(n|n-1)\boldsymbol{g}. \end{aligned} \tag{40}$$

Using (33), $\boldsymbol{g}^T K(n) = K_0(n)$ and (34) in (40), we have

$$\boldsymbol{g}^T\boldsymbol{P}(n|n)\boldsymbol{g} = \alpha^2(n) + \sigma_w^2 - \frac{[\alpha^2(n)+\sigma_w^2]^2}{\alpha^2(n)+\beta^2(n)+\sigma_w^2+\sigma_u^2} \tag{41}$$

Moving the first term on the right hand side to the left and dividing by $\alpha^2(n) + \sigma_w^2$, we get:

$$\begin{aligned}
\frac{\boldsymbol{g}^T \boldsymbol{P}(n|n)\boldsymbol{g} - \alpha^2(n)}{\alpha^2(n) + \sigma_w^2} &= \frac{\sigma_w^2}{\alpha^2(n) + \sigma_w^2} - \frac{\alpha^2(n) + \sigma_w^2}{\alpha^2(n) + \beta^2(n) + \sigma_w^2 + \sigma_u^2} \\
\frac{P_{0,0}(n|n) - \alpha^2(n)}{\alpha^2(n) + \sigma_w^2} &= \frac{\sigma_w^2}{\alpha^2(n) + \sigma_w^2} + \frac{\sigma_u^2 + \beta^2(n)}{\alpha^2(n) + \beta^2(n) + \sigma_w^2 + \sigma_u^2} - 1 \\
\frac{P_{0,0}(n|n) - \alpha^2(n)}{\alpha^2(n) + \sigma_w^2} + 1 &= \frac{\sigma_w^2}{\alpha^2(n) + \sigma_w^2} + \frac{\sigma_u^2 + \beta^2(n)}{\alpha^2(n) + \beta^2(n) + \sigma_w^2 + \sigma_u^2} \\
\Delta P(n|n) + 1 &= J_2(n) + J_1(n)
\end{aligned} \tag{42}$$

Applying the definitions of the sensitivity metric, $J_1(n)$, and robustness metric, $J_2(n)$, to the case of the speech model and coloured noise model considered in this work, we have:

$$J_1(n) = \frac{\sigma_u^2 + \beta^2(n)}{\alpha^2(n) + \beta^2(n) + \sigma_w^2 + \sigma_u^2} \tag{43}$$

$$J_2(n) = \frac{\sigma_w^2}{\alpha^2(n) + \sigma_w^2} \tag{44}$$

We can see that $J_2(n)$ is computed as the proportion of the process noise variance σ_w^2 with respect to the total *a priori* prediction error $[\sigma_w^2 + \alpha^2(n)]$. The higher the process noise variance is, with respect to the transmitted errors of the output speech sample $\alpha^2(n)$, the closer $J_2(n)$ approaches unity. This makes sense because as was mentioned, robustness of the Kalman filter relates to its ability to deal with inaccuracies in the model and use more of the measurement. In this case, a large prediction error, σ_w^2, increases the Kalman filter gain (Eq. 34), hence using more of the measurement $y(n)$, and thereby operating in a more robust mode. Since the process noise variance represents the model-based prediction error, $J_2(n)$ is termed the robustness metric [21].

When interpreting the sensitivity metric, $J_1(n)$ is the proportion of the innovation covariance (in the denominator) that is attributed to the measurement noise. Therefore, when the measurement noise variance σ_u^2 is high or when there is a large transmitted error in the Kalman filter estimate of the measurement noise $\beta^2(n)$, in comparison with the errors related to the speech model ($\alpha^2(n)$ and σ_w^2), the Kalman filter will become more sensitive to variations in its dynamic model, i.e. it will favour the predicted estimate over the measurement.

3.3 Kalman filter gain tuning using the robustness metric $J_2(n)$

Figure 4 shows plots of $J_2(n)$ and $K_0(n)$ for the oracle and non-oracle cases of the Kalman filter. Let us consider the first 0.2 seconds where there is no speech. We can see in Figure 4(b) that the robustness metric is very high for the non-oracle case in this silent region, which indicates that the Kalman filter is operating in its robust mode. Since the LPC estimates are computed from the white noise only, then $\alpha^2(n) \approx 0$, and hence according to Eq. (44), $J_2(n) \approx 1$. The scalar Kalman filter gain $K_0(n)$ in Figure 4(a) is hovering between 0.5 and 1, which is detrimental to the overall quality of the enhanced speech, since residual measurement noise is being passed through to the output.

For the coloured noise case, the behaviour of the robustness metric differs from that seen in the white noise scenario, as represented by the green line in Figure 4(d). The robustness metric does not hover near unity in the coloured noise case, because $\alpha^2(n) = \beta^2(n)$ in the silence regions (since measurement noise is the only signal present), hence according to Eq. (44), the denominator will be larger than the numerator. It is possible to reduce this 'coloured' effect by applying a whitening filter $H_w(z)$ to the speech prior to the estimation of the LPC parameters:

$$H_w(z) = 1 + \sum_{k=1}^{q} b_k z^{-k} \tag{45}$$

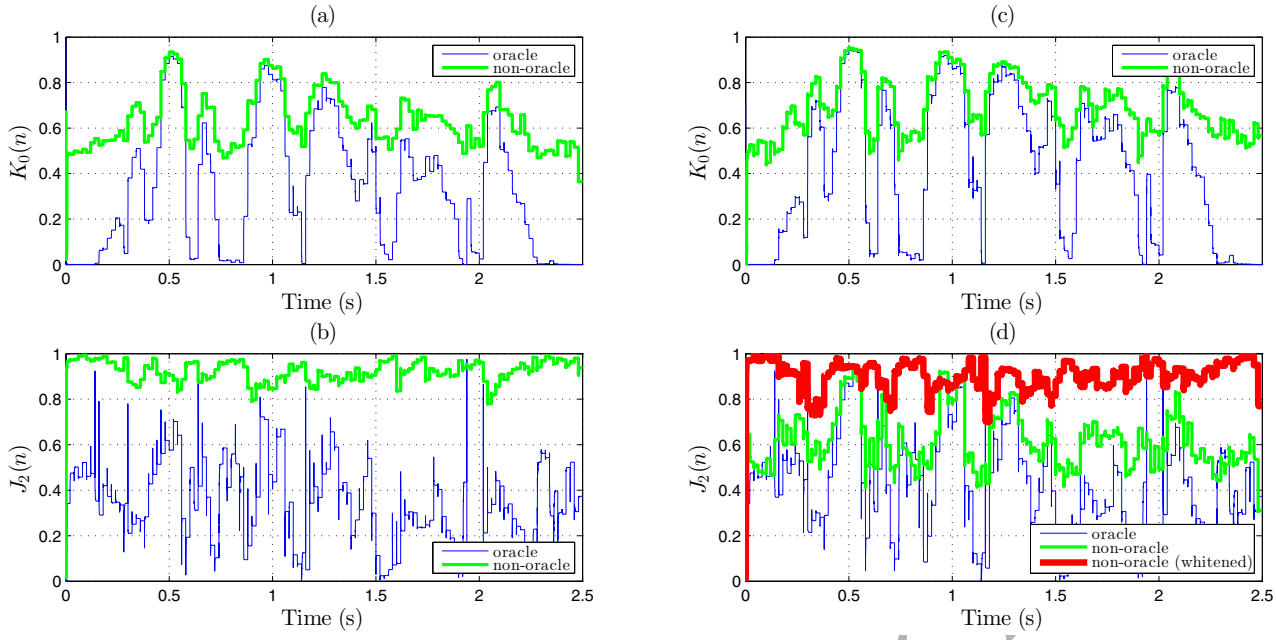

Figure 4: Plot of Kalman filter gain and the robustness metric for the sp26 speech file for the oracle and non-oracle case: (a) scalar Kalman filter gain $K_0(n)$ for *white* noise at 5 dB SNR; (b) robustness metric $J_2(n)$ for *white* noise; (c) scalar Kalman filter gain $K_0(n)$ for *F16* noise at 5 dB SNR; (d) robustness metric $J_2(n)$ for *F16* noise.

We note that the whitening filter is applied to the speech for LPC estimation only. The whitening filter has two desirable effects on the coloured-noise Kalman filter. Firstly, as represented by the thick red line in Figure 4(d), the behaviour of $J_2(n)$ becomes similar to that seen for the white noise scenario. In other words, $J_2(n)$ hovers close to unity in the silence regions due to the approximate 'whiteness' of the noise ($\alpha^2(n) \approx 0$). Secondly, the whitening filter assists in reducing the bias in the LPC coefficients $\{\tilde{a}_k\}$ and therefore enables the Kalman filter to have a better speech model.

In light of the observations made on the behaviour of the robustness metric for the white noise and coloured-noise Kalman filter with pre-whitened LPC estimation, we propose to modify the scalar Kalman filter gain, so that it is similar to the oracle Kalman filter gain, by using the following equation, which we will justify intuitively as well as mathematically:

$$K_0'(n) = K_0(n)[1 - J_2(n)] \tag{46}$$

3.3.1 The dynamic behaviour of the robustness metric

Intuitively, the robustness metric can be used to scale the Kalman filter gain dynamically over time. Because all terms in Eq. (44) are squared quantities, it is clear that $J_2(n)$ will always be bounded between 0 and 1. In the silence regions, where the non-oracle Kalman filter is operating in an 'artificially' high robust mode, the scaling factor $1 - J_2(n)$ will be approximately zero, hence according to Eq. (46), the Kalman filter gain will be suppressed. However, it is not entirely clear whether this is suitable for the speech-dominant regions, since $J_2(n)$ can be seen in Figure 4(b) to be always larger than 0.8 and subsequently, the Kalman filter gain would be scaled by a factor of 0.2 or less. For voiced speech, a low Kalman filter gain would result in diminished harmonic and voicing characteristics in the enhanced speech output.

We should point out that the $J_2(n)$ plot in Figure 4(b) was obtained from the Kalman filter whose gain was unmodified. Due to the recursive nature of the Kalman filter equations, it is instructive to examine the effects of applying the gain tuning in the subsequent time step. Let us

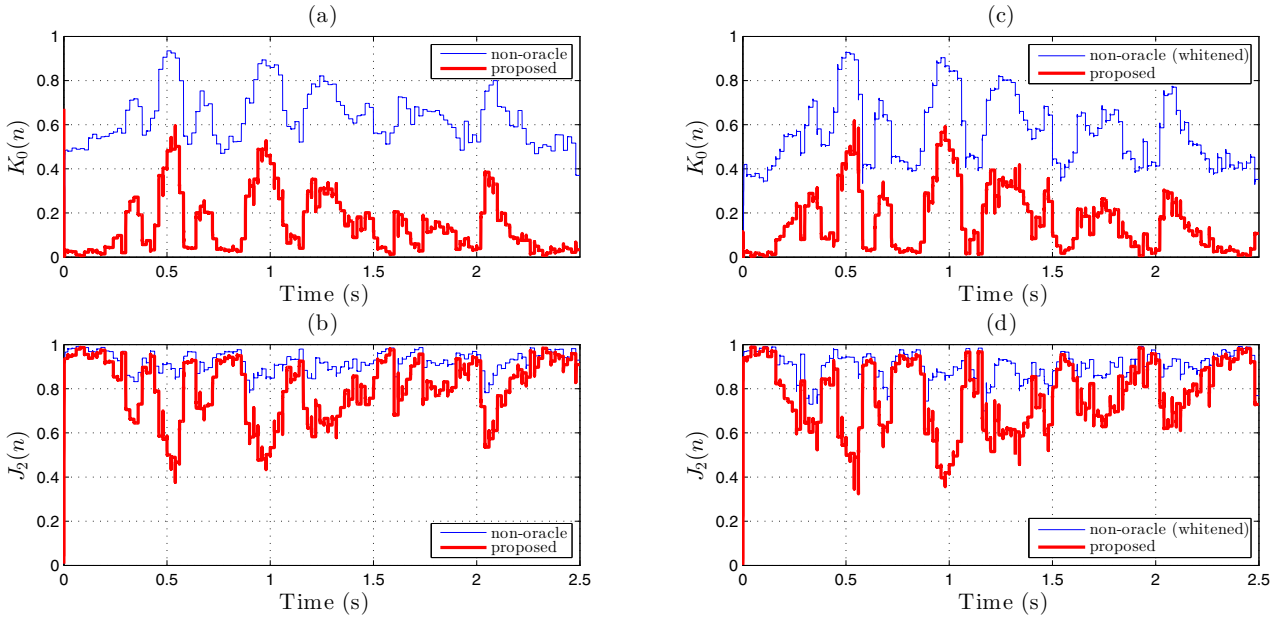

Figure 5: Plot of Kalman filter gain and the robustness metric for the sp26 speech file for the non-oracle and proposed modified case: (a) scalar Kalman filter gain $K_0(n)$ for *white* noise at 5 dB SNR; (b) robustness metric $J_2(n)$ for *white* noise; (c) scalar Kalman filter gain $K_0(n)$ for *F16* noise at 5 dB SNR; (d) robustness metric $J_2(n)$ for *F16* noise.

apply the Kalman filter gain tuning to the *a posteriori* mean squared error at time sample n:

$$\begin{aligned}\boldsymbol{g}^T\boldsymbol{P}'(n|n)\boldsymbol{g} &= \boldsymbol{g}^T\boldsymbol{P}(n|n-1)\boldsymbol{g} - [1-J_2(n)]\boldsymbol{g}^T\boldsymbol{K}(n)\boldsymbol{g}^T\boldsymbol{P}(n|n-1)\boldsymbol{g} \\ &= \alpha^2(n)+\sigma_w^2 - \frac{\alpha^2(n)}{\alpha^2(n)+\sigma_w^2}\frac{[\alpha^2(n)+\sigma_w^2]^2}{\alpha^2(n)+\beta^2(n)+\sigma_w^2+\sigma_u^2} \\ &= \alpha^2(n)+\sigma_w^2 - \frac{\alpha^2(n)[\alpha^2(n)+\sigma_w^2]}{\alpha^2(n)+\beta^2(n)+\sigma_w^2+\sigma_u^2} \end{aligned} \tag{47}$$

Comparing the third term on the right-hand-side of Eqs. (41) and (47), we note that the modified *a posteriori* mean squared error of the latter equation is *larger* than that of the former by the following amount:

$$\boldsymbol{g}^T\boldsymbol{P}'(n|n)\boldsymbol{g} - \boldsymbol{g}^T\boldsymbol{P}(n|n)\boldsymbol{g} = \frac{\sigma_w^2[\alpha^2(n)+\sigma_w^2]}{\alpha^2(n)+\beta^2(n)+\sigma_w^2+\sigma_u^2} \tag{48}$$

This difference is dependent on the values of σ_w^2 and $\alpha^2(n)$. For voiced speech, σ_w^2 is expected to be quite high. Since the computation of $\alpha^2(n+1)$ is dependent on $\boldsymbol{g}^T\boldsymbol{P}'(n|n)\boldsymbol{g}$, therefore according to Eq. (44), we should expect $J_2(n+1)$ to be smaller than what the robustness metric would have been, had no Kalman filter gain modification been made.

The effect of the Kalman filter gain modification can also be seen by substituting Eqs. (34) and (44) into Eq. (46):

$$\begin{aligned} K_0'(n) &= \left[1-\frac{\tilde{\sigma}_w^2}{\tilde{\alpha}^2(n)+\tilde{\sigma}_w^2}\right]\frac{\tilde{\alpha}^2(n)+\tilde{\sigma}_w^2}{\tilde{\alpha}^2(n)+\beta^2(n)+\tilde{\sigma}_w^2+\sigma_u^2} \\ &= \frac{\tilde{\alpha}^2(n)}{\tilde{\alpha}^2(n)+\beta^2(n)+\tilde{\sigma}_w^2+\sigma_u^2} \end{aligned} \tag{49}$$

where we can see that the term $\tilde{\sigma}_w^2$ (and its associated bias) has disappeared. Using $K_0'(n)$ in place

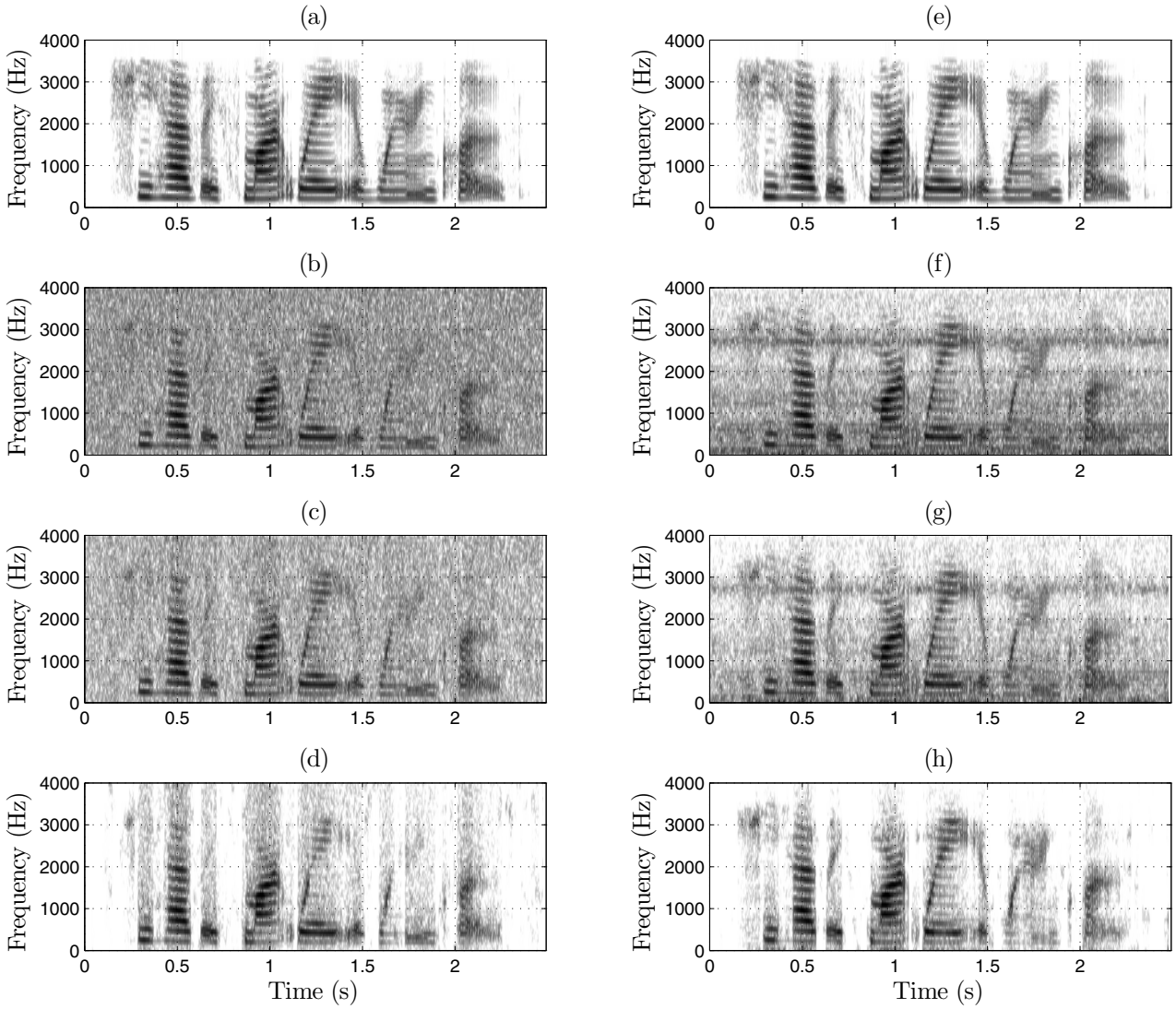

Figure 6: Comparing the spectrograms of the enhanced speech from non-oracle and proposed methods, where input speech was 'She had a smart way of wearing clothes': (a) clean speech; (b) speech corrupted with *white* noise at 5 dB (PESQ = 1.86); (c) output of the Kalman filter (non-oracle case with no gain modification) (PESQ = 2.03); (d) output of the proposed Kalman filter with gain modification (PESQ = 2.26); (e) clean speech; (f) speech corrupted with *F16* noise (PESQ = 2.11); (g) output of the Kalman filter (non-oracle case with no gain modification) (PESQ = 2.24); (h) output of the proposed Kalman filter with gain modification (PESQ = 2.43).

of $K_0(n)$ in (29), the resulting Kalman update equation is thus obtained from (35) as:

$$\hat{x}(n|n) = \frac{\beta^2(n) + \sigma_u^2 + \tilde{\sigma}_w^2}{\tilde{\alpha}^2(n) + \beta^2(n) + \tilde{\sigma}_w^2 + \sigma_u^2}\hat{x}(n|n-1) + \frac{\tilde{\alpha}^2(n)}{\tilde{\alpha}^2(n) + \beta^2(n) + \tilde{\sigma}_w^2 + \sigma_u^2}[y(n) - v(n|n-1)] \tag{50}$$

We have used the tilde notation to indicate the biased estimates in the non-oracle case. If we consider the case where only white (or whitened)[3] noise is present, then $\tilde{\alpha}^2(n) \approx 0$ and $\tilde{\sigma}_w^2 = \sigma_u^2$ (assuming $p = q$), so Eq. (50) becomes:

$$\hat{x}(n|n) \approx \hat{x}(n|n-1) \tag{51}$$

Therefore, the residual measurement noise in the output speech is suppressed. For voiced speech, the Kalman filter gain is expected to be lower due to the loss of the $\tilde{\sigma}_w^2$ term in the numerator as a result of the gain modification. Despite this, the prediction error in the speech model will partly manifest itself in the transmitted *a posteriori* error $\tilde{\alpha}^2(n)$, which will be large for voiced speech.

Figure 5 compares the Kalman filter gain and robustness metric between the non-oracle case and the proposed gain modification case. The comparison of $J_2(n)$ shows that once the Kalman filter gain modification is included, the robustness metric exhibits larger variations, especially within the speech frames. This is a desirable property since we do not want the Kalman filter gain to be overly suppressed in the voiced speech regions. As a result of the $J_2(n)$ remaining close to unity in the silence regions, it can be seen that the Kalman filter gain, $K_0(n)$, has been suppressed, thereby reducing the amount of residual measurement noise passing through to the enhanced speech output.

Figure 6 compares the spectrograms for the non-oracle and proposed Kalman filter for speech that has been corrupted by white noise and coloured F16 noise at 5 dB SNR. We can see from Figures 6(d) and (h) that the enhanced speech from the proposed method does not suffer from residual noise in the silence regions, as opposed to the non-oracle method with no gain modification.

4 Experimental Setup

4.1 Corpus

The NOIZEUS speech corpus, which is comprised of 30 phonetically balanced sentences belonging to six speakers (three male, three female) and sampled at 8 kHz [1], was utilised for the experiments presented in this paper. For the objective experiments, a stimuli set was generated that has been corrupted by a range of additive coloured noises at four SNR levels (0, 5, 10, and 15 dB). For the subjective experiments, a stimuli set of two sentences (one for each gender) was corrupted by 5 dB additive coloured (factory2) noise.

4.2 Speech Quality measurement

Objective evaluation of speech quality was performed via three metrics: the PESQ (Perceptual Evaluation of Speech Quality) score [25], and Segmental SNR (SSNR) and Log-Likelihood Ratio (LLR) [1]. For each metric, mean scores for each input SNR, noise type and treatment type were calculated across all 30 sentences of the NOIZEUS speech corpus.

For the subjective listening experiments, blind AB listening tests were used to determine subjective treatment type preference. A total of four speech enhancement treatment types as well as the clean and noisy speech were played as stimuli pairs to a listener. This represented a total of 60 stimuli pairs (30 for each sentence) played in random order to each listener, except for comparisons between

[3]We assume that a whitening filter has been applied before LPC parameter estimation.

the same treatment type. The listener was provided with the following options for each stimuli pair: either the first or second stimuli was perceptually better, or a third response indicating that there was no perceived difference between them. Fifteen english speaking participants completed the listening tests, and subjective results are provided in terms of their average preference score.

4.3 Speech Treatment Types

The proposed algorithm will be compared to the original and oracle Kalman filter as well as the MMSE-STSA method [4]. Stimuli for the following six treatment types were constructed.

1. **Clean**: Original clean speech.
2. **Noisy**: Speech corrupted with specified level of additive coloured noise.
3. **KF-Oracle**: Kalman filter with LPCs estimated from clean speech, 20ms, p = 10, q = 40, no overlap [6].
4. **KF-Normal**: Kalman filter with LPCs estimated from corrupted speech, 20ms, p = 10, q = 40, no overlap [18].
5. **MMSE**: MMSE-STSA method with only corrupted speech, 20ms, p = 10, q = 40, no overlap, no SPU (Speech Presence Uncertainty) [4].
6. **KF-Proposed**: Proposed Kalman filter with robust metric tuning and LPCs estimated from whitened corrupted speech, 20ms, p = 10, q = 40, no overlap.
7. **W-WT**: Wiener filtering based wavelet-thresholding multi-taper spectra with only corrupted speech, 20ms, 16 tapers [9].

5 Results and Discussion

Table 1: Average PESQ results for the comparison of speech enhancement methods over the NOIZEUS speech corpus corrupted with additive *factory2* coloured noise.

Method	Input SNR (dB)			
	0	5	10	15
Noisy	1.98	2.32	2.63	2.96
KF-Oracle	2.51	2.83	3.11	3.43
KF-Normal	2.11	2.45	2.76	3.11
KF-Proposed	2.35	2.70	3.04	3.38
MMSE	2.41	2.72	3.01	3.30
W-WT	1.92	2.36	2.72	3.15

Table 2: Average PESQ results for the comparison of speech enhancement methods over the NOIZEUS speech corpus corrupted with additive *F16* coloured noise.

Method	Input SNR (dB)			
	0	5	10	15
Noisy	1.76	2.08	2.39	2.74
KF-Oracle	2.30	2.58	2.85	3.17
KF-Normal	1.89	2.22	2.53	2.88
KF-Proposed	2.15	2.48	2.82	3.14
MMSE	2.21	2.54	2.84	3.14
W-WT	1.70	2.13	2.50	2.89

Table 3: Average Segmental SNR results for the comparison of speech enhancement methods over the NOIZEUS speech corpus corrupted with additive *factory2* coloured noise.

Method	Input SNR (dB)			
	0	5	10	15
Noisy	-4.53	-1.71	1.48	4.99
KF-Oracle	3.52	5.39	7.44	10.10
KF-Normal	-2.51	0.31	3.33	6.68
KF-Proposed	0.45	1.96	3.26	4.69
MMSE	-0.57	2.19	4.86	7.43
W-WT	0.58	2.68	5.07	7.88

Table 4: Average Segmental SNR results for the comparison of speech enhancement methods over the NOIZEUS speech corpus corrupted with additive *F16* coloured noise.

Method	Input SNR (dB)			
	0	5	10	15
Noisy	-4.69	-1.88	1.24	4.73
KF-Oracle	3.07	4.75	6.75	9.39
KF-Normal	-2.70	-0.07	2.82	6.16
KF-Proposed	-0.04	1.57	3.20	4.84
MMSE	-0.96	1.57	4.01	6.66
W-WT	-0.15	2.02	4.61	7.48

Table 5: Average LLR results for the comparison of speech enhancement methods over the NOIZEUS speech corpus corrupted with additive *factory2* coloured noise.

Method	Input SNR (dB)			
	0	5	10	15
Noisy	0.88	0.68	0.51	0.38
KF-Oracle	0.41	0.32	0.26	0.19
KF-Normal	0.81	0.61	0.46	0.33
KF-Proposed	0.91	0.66	0.49	0.38
MMSE	0.72	0.52	0.40	0.29
W-WT	1.25	1.00	0.81	0.60

Table 6: Average LLR results for the comparison of speech enhancement methods over the NOIZEUS speech corpus corrupted with additive *F16* coloured noise.

Method	Input SNR (dB)			
	0	5	10	15
Noisy	1.09	0.91	0.73	0.54
KF-Oracle	0.52	0.44	0.36	0.27
KF-Normal	1.01	0.83	0.66	0.48
KF-Proposed	1.03	0.81	0.63	0.48
MMSE	0.87	0.68	0.52	0.37
W-WT	1.43	1.23	1.03	0.83

Tables 1 and 2 show the average PESQ scores for all treatment types for stationary coloured noises, demonstrating that all enhancement methods improved upon the speech quality of the

noisy signal (Noisy). When comparing the PESQ results for the Kalman filter treatment types (KF-Normal, KF-Oracle, and KF-Proposed), we can see the negative effect of estimation bias that is introduced by utilising the noise corrupted speech model. This is shown by the PESQ of KF-Proposed improving upon those of KF-Normal and being similar but less than the KF-Oracle results. Comparing KF-proposed to the current methods of MMSE and W-WT, it is seen to achieve consistent improvement over W-WT while it can be considered competitive when comparing it to the MMSE treatment type. In light of these results, it should be noted that PESQ was originally developed for speech coding where the distortions are different to those encountered in speech enhancement.

Tables 3 and 4 show the segmental SNR results for the same two coloured noises. Overall, the KF-Oracle has the highest results. For the lower values of input SNR, the KF-Proposed results are higher than those of the MMSE and KF-Normal treatment types while being competitive with W-WT. As the input SNR increases, which means that the input speech is less noisy, the KF-Proposed SSNRs do not increase as much as the other treatment types. The SSNR metric is sensitive to the energy within the signal frames, such that deviations of energy between the two will have a large effect on the output. The KF-Proposed method alters the Kalman filter gain and it has been mentioned to have an over suppression effect on the speech. This produces an adverse effect on the SSNR as the signal energy will be reduced, and thus the SSNR will be impacted. Tables 5 and 6 display the LLR results for both *factory2* and *F16* coloured noises. It is observable that the KF-Proposed is consistently lower than W-WT while being competitive with the MMSE treatment type. It was found, that both SSNR and LLR objective results do not correlate with the subjective testing.

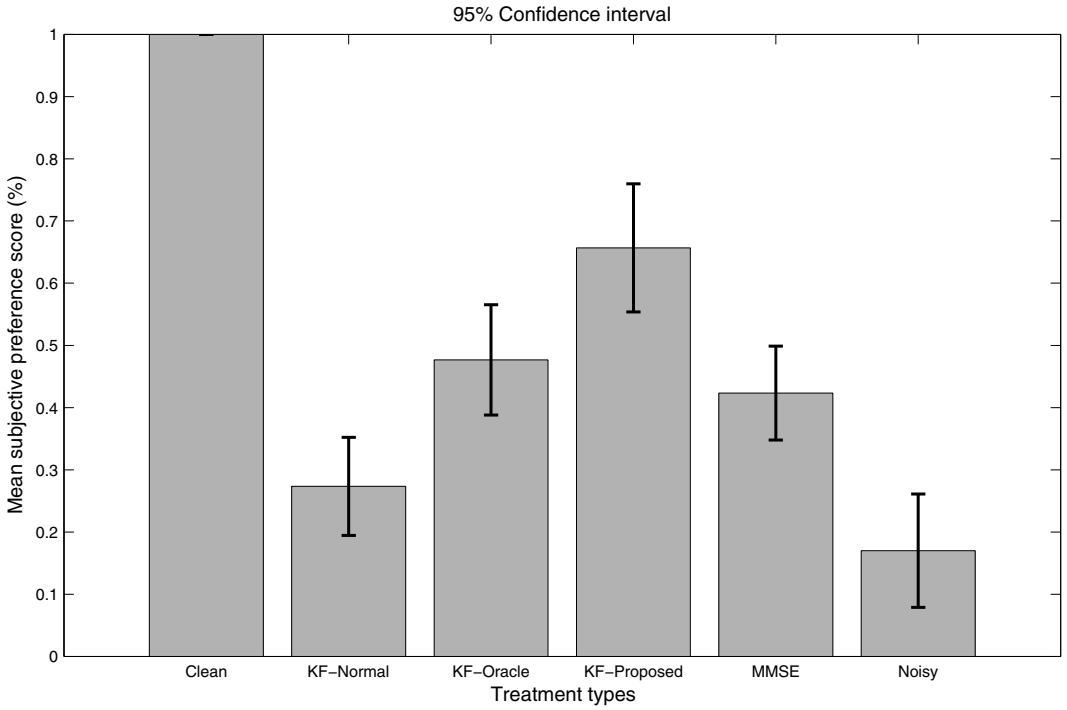

Figure 7: The mean subjective preference score with a 95% confidence interval for the treatment types applied on sp10 (male speaker) from the Noizeus database.

Figures 7 and 8 contain the mean preference scores as well as the 95% confidence interval for blind AB listening tests performed on male and female spoken phrases, respectively. Both figures show KF-Proposed as the most preferred treatment type (excluding clean). The analysis of variance

Table 7: ANOVA table for the male stimuli subjective preference scores ($\alpha = 0.05$). *df* - Degrees of Freedom, SS - Sum of Squares, MS - Mean Squares, $F_{statistic}$ - F statistic, F_α - F critical [26]. The null hypothesis H_0 (means are all equal) is *rejected* ($F_\alpha = 2.324$).

Source	*df*	SS	MS	F-statistic
Treatment	5.000	661.867	132.373	62.952
Error	84.000	176.633	2.103	
Total	89.000	838.500		

Table 8: ANOVA table for the female stimuli subjective preference scores ($\alpha = 0.05$). *df* - Degrees of Freedom, SS - Sum of Squares, MS - Mean Squares, $F_{statistic}$ - F statistic, F_α - F critical [26]. The null hypothesis H_0 (all means are equal) *is rejected* ($F_\alpha = 2.324$).

Source	*df*	SS	MS	F-statistic
Treatment	5.000	618.033	123.607	63.908
Error	84.000	162.467	1.934	
Total	89.000	780.500		

Table 9: Differences between mean subjective scores and significance of pairwise comparisons using Tukey's Honestly Significant Difference test of the male stimuli. Differences larger than 1.092 are significant (marked with *) at the $\alpha = 0.05$ level.

	KF Normal	KF Oracle	**KF Proposed**	MMSE	Noisy
Clean	7.27*	5.23*	3.43*	5.77*	8.30*
KF-Normal		2.03*	3.83*	1.50*	1.03
KF-Oracle			1.80*	0.53	3.07*
KF-Proposed				2.33*	4.87*
MMSE					2.53*

Table 10: Differences between mean subjective scores and significance of pairwise comparisons using Tukey's Honestly Significant Difference test of the female stimuli. Differences larger than 1.047 are significant (marked with *) at the $\alpha = 0.05$ level.

	KF Normal	KF Oracle	**KF Proposed**	MMSE	Noisy
Clean	6.87*	5.73*	3.70*	5.50*	8.20*
KF-Normal		1.13*	3.17*	1.37*	1.33*
KF-Oracle			2.03*	0.23	2.47*
KF-Proposed				1.80*	4.50*
MMSE					2.70*

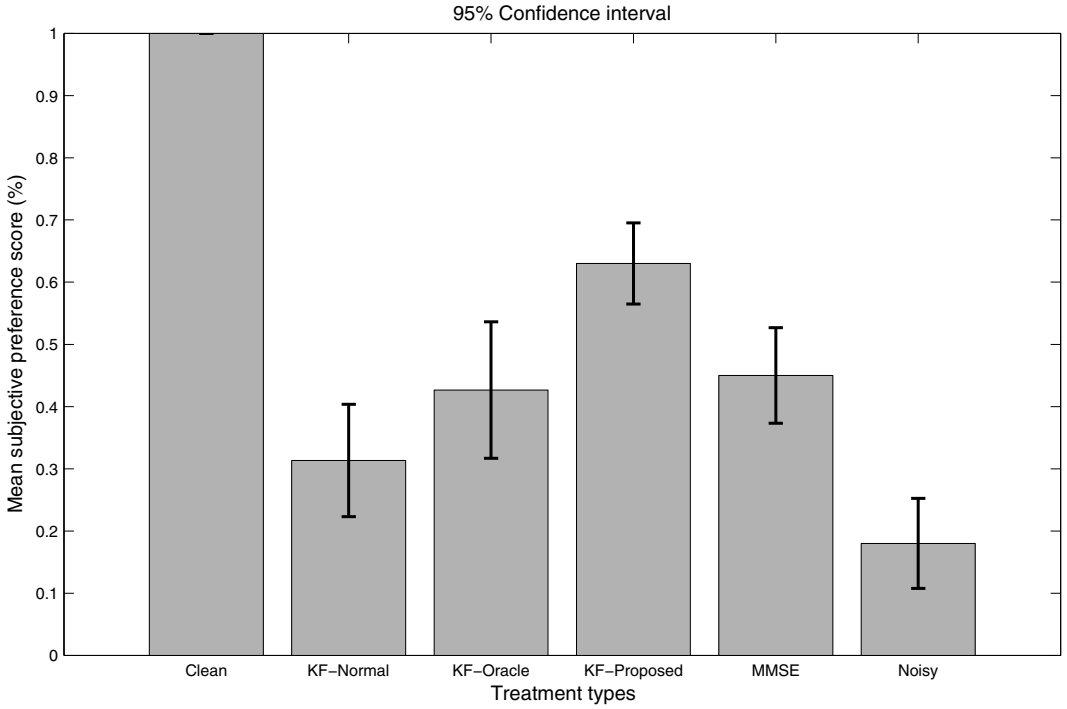

Figure 8: The mean subjective preference score with a 95% confidence interval for the treatment types applied on sp29 (female speaker) from the Noizeus database.

(ANOVA) results are presented in Tables 7 and 8 where the degrees of freedom (df), F-statistic ($F_{statistic}$), sum of squares (SS), and mean squares (MS) [26] are listed. For both tables, the F-statistic is larger than its respective F-critical mean (F_α) at $\alpha = 0.05$, thus the null hypothesis is rejected, which indicates that the different sets of data are not from the same source.

Tables 9 and 10 contains the results of the Tukey Honestly Significant Difference test, in which the KF-Proposed data is seen to be significantly different from all other treatment types for both male and female utterances. Therefore, the KF-Proposed method is shown to be significantly preferred by human listeners in comparison to the other enhancement methods.

Looking again at Figures 7 and 8, it is evident that the KF-Oracle treatment type has a lower preference score. As the speech model of KF-Oracle is derived from the clean speech, the performance is typically preferred, and objectively the KF-Oracle treatment achieves higher PESQ scores. Through informal questioning of the listeners, it was discovered that the abrupt change between noisy speech segments and regions of silence was perceived to be more annoying than the other treatment types, where there are less abrupt changes in the background noise. This is due to the effects of the Kalman filter gain. Since the low order speech model is unable to estimate the long term information of speech, the Kalman filter gain adds a portion of the noisy observation to the *a priori* estimate to best represent that speech segment. In the regions where only noise is present, the oracle Kalman filter functions optimally in suppressing all noise, as seen in Figures 9 (c) and 10 (c), which is dispersed between segments of speech causing the boundaries to perceptually sound disruptive.

When comparing the spectrograms of the Kalman filter treatment types seen in Figure 9, the noise reduction seen when going from KF-Normal to KF-Proposed and KF-Oracle is visibly improved. KF-Oracle has better noise reduction during regions containing noise only and the spectrogram is visibly darker during regions of speech when compared to KF-Proposed. The lessening

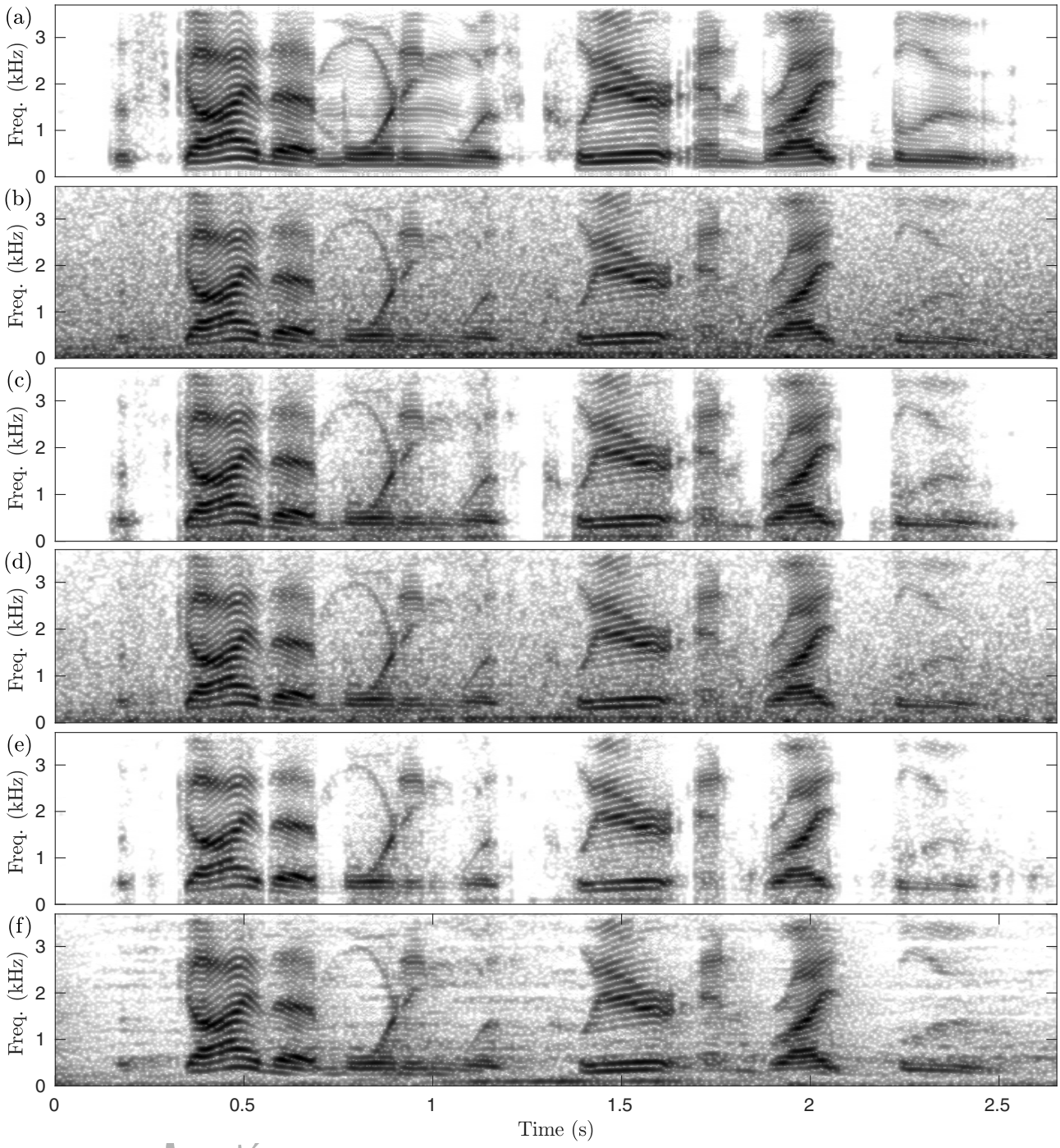

Figure 9: Spectrograms for the treatment types for sp10 corrupted with 5dB *factory2* coloured noise: (a) clean, (b) noise corrupted, (c) oracle Kalman filter (speech model estimated from clean source), (d) Kalman filter (speech model estimated from corrupted source) (e) proposed Kalman filter, (f) MMSE.

of the spectral intensity of the speech sections is due to the bias tuning algorithm, which reduces the noise within the speech regions but also suppresses the speech spectral amplitude. However the subjective testing suggested that this was considered more of an improvement than a distortion since the impact of the noise within the speech region is reduced.

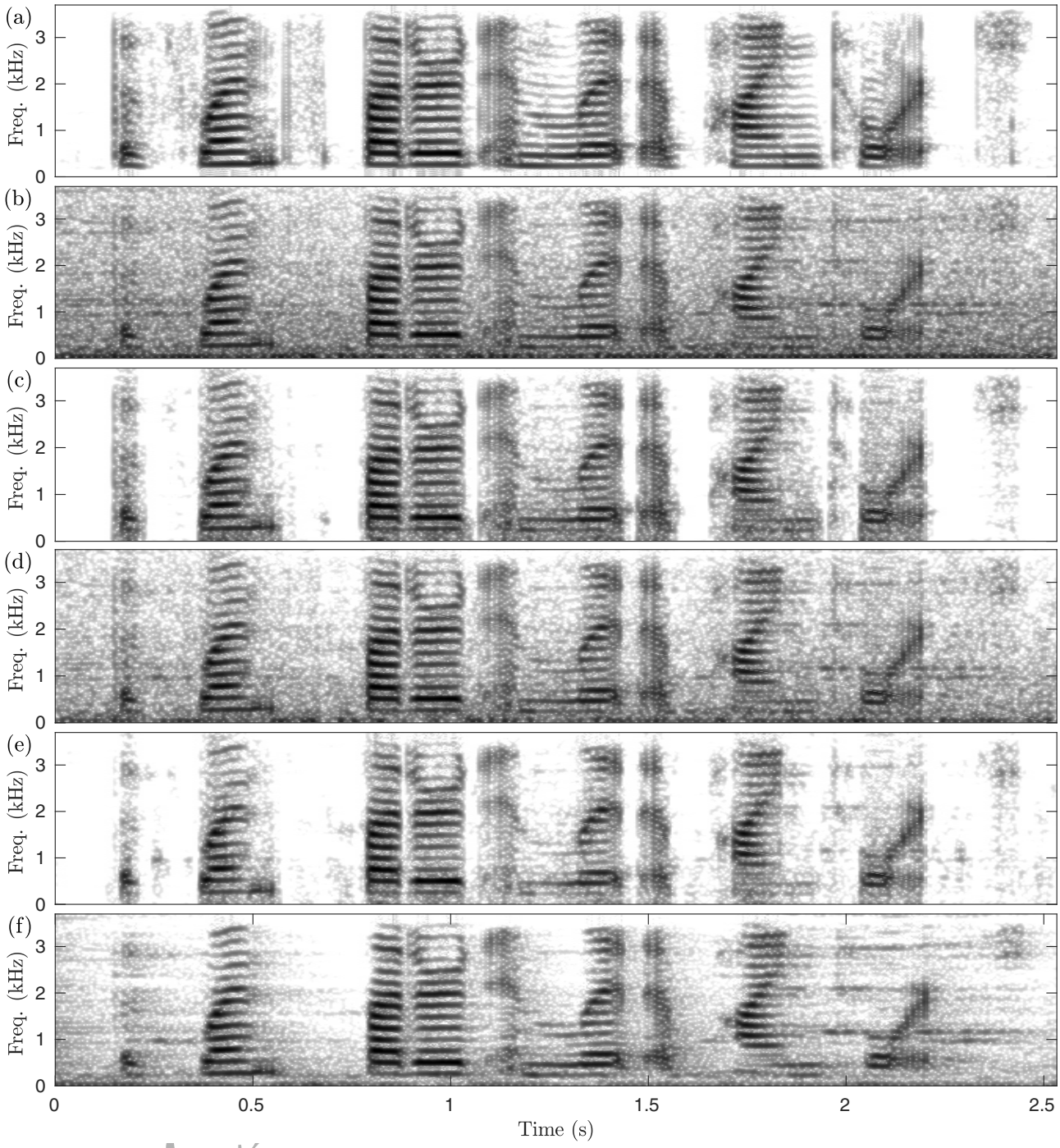

Figure 10: Spectrograms for the treatment types for sp29 corrupted with 5dB *factory2* coloured noise. (a) clean, (b) noise corrupted, (c) oracle Kalman filter (speech model estimated from clean source), (d) Kalman filter (speech model estimated from corrupted source) (e) proposed Kalman filter, (f) MMSE.

5.1 Evaluation on speech corrupted with white noise

In this section, we evaluate the proposed method for speech that has been corrupted with white noise. The coloured noise Kalman filter is expected to appropriately handle the white case as well. Table 11 shows the average PESQ results for the treatment types applied on speech that has been corrupted with white noise. Both the MMSE and KF-Proposed results are of similar values to the KF-Oracle treatment type, and they also have a large objective improvement over the KF-Normal treatment. The KF-Proposed also has a consistently improved result when comparing to W-WT

Table 11: Average PESQ results for the comparison of speech enhancement methods over the NOIZEUS speech corpus corrupted with additive *white* noise.

Method	Input SNR (dB)			
	0	5	10	15
Noisy	1.57	1.83	2.13	2.47
KF-Oracle	2.11	2.34	2.63	2.95
KF-Normal	1.69	2.00	2.32	2.66
KF-Proposed	1.93	2.30	2.63	2.93
MMSE	1.96	2.33	2.64	2.94
W-WT	1.74	2.14	2.46	2.80

Table 12: Average Segmental SNR results for the comparison of speech enhancement methods over the NOIZEUS speech corpus corrupted with additive *white* noise.

Method	Input SNR (dB)			
	0	5	10	15
Noisy	-4.81	-2.00	1.13	4.59
KF-Oracle	2.99	4.63	6.71	9.23
KF-Normal	-2.86	-0.20	2.73	6.05
KF-Proposed	-0.16	1.62	3.80	5.93
MMSE	-0.83	1.54	4.00	6.49
W-WT	0.73	3.12	5.92	8.59

Table 13: Average LLR results for the comparison of speech enhancement methods over the NOIZEUS speech corpus corrupted with additive *white* noise.

Method	Input SNR (dB)			
	0	5	10	15
Noisy	1.47	1.34	1.17	0.97
KF-Oracle	0.84	0.77	0.68	0.57
KF-Normal	1.42	1.27	1.09	0.89
KF-Proposed	1.35	1.18	1.00	0.82
MMSE	1.26	1.08	0.88	0.69
W-WT	1.63	1.54	1.39	1.20

treatment type as well. This again indicates that even though the KF-Proposed does not re-estimate the speech model source, similar results are obtainable by reducing the detrimental estimation bias incurred when utilising the noisy source for the estimation of the speech model. Tables 12 and 13 contain the average SSNR and LLR results for the white noise corrupted speech. It can once again be seen that the KF-Proposed is competitive with MMSE while both are objectively improved when comparing to W-WT.

5.2 The Effect of the Whitening Filter

In this experiment, a whitening filter was applied, as in Figure 11, on the noisy frame during the estimation of the speech model for the proposed Kalman filter. The noise estimate is obtained by recording data during the region of silence where only noise is present. One hypothesis is whether the improvements seen in the results of the KF-Proposed algorithm are due to the effect of the whitening filter only instead of the proposed algorithm. Hence, further investigation was performed

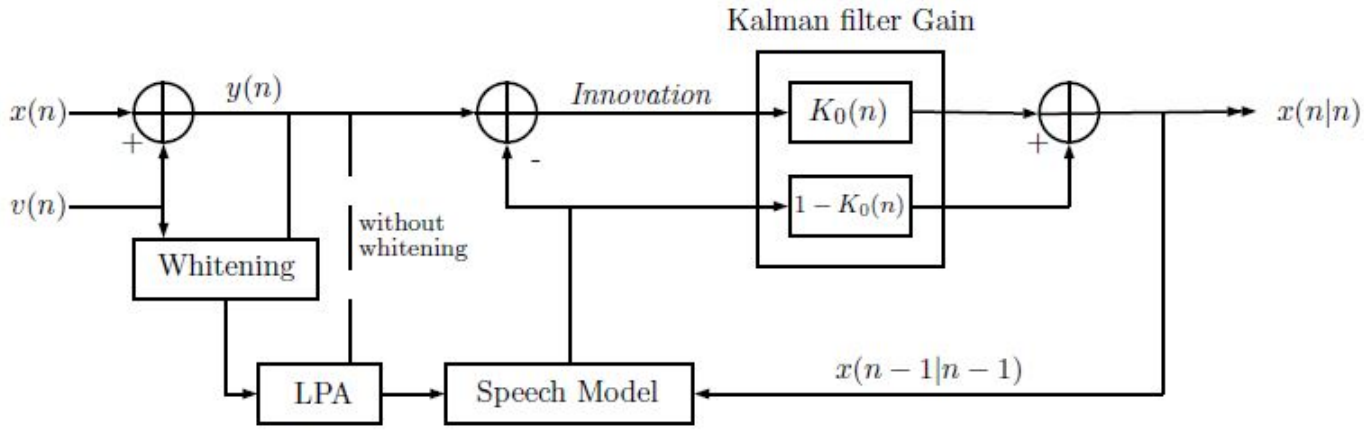

Figure 11: A basic block diagram of the Kalman filter operating with LPCs estimated from whitened speech. The broken line shows normal Kalman filter operation.

Table 14: Average PESQ results for the comparison of speech enhancement methods over the NOIZEUS speech corpus corrupted with additive *factory2* coloured noise.

Method	Input SNR (dB)			
	0	5	10	15
Noisy	1.98	2.33	2.63	2.96
KF-Oracle	2.51	2.83	3.12	3.44
KF-Normal	2.11	2.45	2.76	3.11
KF-Proposed	2.35	2.70	3.04	3.38
KF-Whitened	2.15	2.48	2.82	3.19

to determine whether applying the whitening filter (q = 40) alone can account for the improvements observed in the objective results.

Comparing the average PESQ results within Tables 14 and 15, it is observable that the whitening filter is not the major contributor to the objective improvement of KF-Proposed. Only a small improvement in PESQ score was recorded when comparing KF-Whitened to KF-Normal, while the PESQ score improvement from KF-Whitened to KF-Proposed is larger. This shows that the improvements in the KF-Proposed method are primarily due to the bias reduction algorithm in conjunction with the whitening filter.

Table 15: Average PESQ results for the comparison of speech enhancement methods over the NOIZEUS speech corpus corrupted with additive *F16* coloured noise.

Method	Input SNR (dB)			
	0	5	10	15
Noisy	1.76	2.08	2.39	2.74
KF-Oracle	2.30	2.58	2.85	3.17
KF-Normal	1.89	2.22	2.53	2.88
KF-Proposed	2.15	2.48	2.82	3.14
KF-Whitened	1.94	2.26	2.58	2.95

6 Conclusion

In this paper, a Kalman filter has been proposed that dynamically tunes the Kalman filter gain in order to minimise the effects of estimation bias in the speech model to provide competitive performance with current speech enhancement methods.

The robustness metric, that was earlier reported in the instrumentation literature for quantifying the level of robustness of the Kalman filter, was adapted to the speech enhancement context and applied to dynamically tune the Kalman filter gain. This tuning was shown to reduce the level of residual noise, thus improving the enhancement ability of the Kalman filter.

The PESQ objective results show that the proposed Kalman filter is on par with the speech enhancement performance of the MMSE method and improves upon Wiener WT algorithm. In terms of enhancement performance, the proposed Kalman filter significantly improves upon the non-tuned Kalman filter and is just below the oracle Kalman filter.

The subjective results, from the blind AB listening tests, show a significant preference of the proposed method over all other treatment types, including the oracle Kalman filter, due to the perceived reduction of noise during the regions of speech.

7 Acknowledgements

We would like to thank all the subjects who volunteered to participate in the subjective listening tests.

References

[1] P. Loizou, *Speech Enhancement: Theory and Practice*, 1st ed. CRC Press LLC, 2007.

[2] N. Wiener, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. New York: Wiley, 1949.

[3] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," vol. ASSP-27, no. 2, pp. 113–120, 1979.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," vol. 32, pp. 1109–1121, Dec. 1984.

[5] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," vol. 3, pp. 251–266, Jul. 1995.

[6] K. K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 12, Apr. 1987, pp. 177–180.

[7] Y. Xu, J. Du, L. Dai, and C. Lee, "An experimental study on speech enhancement based on deep neural networks," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 65–68, Jan 2014.

[8] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "A regression approach to speech enhancement based on deep neural networks," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 23, no. 1, pp. 7–19, Jan. 2015. [Online]. Available: http://dx.doi.org/10.1109/TASLP.2014.2364452

[9] Y. Hu and P. C. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 1, pp. 59–67, Jan 2004.

[10] W. R. Wu and P. C. Chen, "Subband Kalman filtering for speech enhancement," vol. 45, no. 8, pp. 1072–1083, Aug. 1998.

[11] C. H. You, S. N. Koh, and S. Rahardja, "Subband Kalman filtering incorporating masking properties for noisy speech signal ," *Speech Communication*,

vol. 49, no. 7–8, pp. 558–573, 2007, speech Enhancement. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639307000325

[12] K. K. Paliwal, K. K. Wojcicki, and B. Schwerin, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain," *Speech Commun.*, vol. 52, no. 5, pp. 450–475, May 2010.

[13] K. K. Paliwal, B. Schwerin, and K. K. Wojcicki, "Single channel speech enhancement using MMSE estimation of short-time modulation magnitude spectrum," in *Proc. Interspeech 2011*, Aug. 2011.

[14] S. So and K. K. Paliwal, "Modulation-domain Kalman filtering for single-channel speech enhancement," *Speech Commun.*, vol. 53, no. 6, pp. 818–829, Jul. 2011.

[15] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng., Trans. ASME*, vol. 82, pp. 35–45, Mar. 1960.

[16] J. Makhoul, "Linear prediction: A tutorial review," vol. 63, no. 4, pp. 561–580, Apr. 1975.

[17] S. So and K. K. Paliwal, "Suppressing the influence of additive noise on the Kalman filter gain for low residual noise speech enhancement," *Speech Commun.*, vol. 53, no. 3, pp. 355–378, Mar. 2011.

[18] J. D. Gibson, B. Koo, and S. D. Gray, "Filtering of colored noise for speech enhancement and coding," vol. 39, no. 8, pp. 1732–1742, Aug. 1991.

[19] S. So, K. K. Wojcicki, J. G. Lyons, A. P. Stark, and K. K. Paliwal, "Kalman filter with phase spectrum compensation algorithm for speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Apr. 2009, pp. 4405–4408.

[20] K. K. Wojcicki, M. Milacic, A. P. Stark, J. G. Lyons, and K. K. Paliwal, "Exploiting conjugate symmetry of the short-time Fourier spectrum for speech enhancement," vol. 15, pp. 461–464, 2008.

[21] M. Saha, R. Ghosh, and B. Goswami, "Robustness and sensitivity metrics for tuning the extended Kalman filter," vol. 63, no. 4, pp. 964–971, Apr. 2014.

[22] S. So, A. E. W. George, R. Ghosh, and K. K. Paliwal, "A non-iterative Kalman filtering algorithm with dynamic gain adjustment for single-channel speech enhancement," in *Proc. International Conference on Signal Processing 2015*, Aug. 2015.

[23] ——, "Kalman Filter with Sensitivity Tuning for Improved Noise Reduction in Speech," *Circuits, Systems, and Signal Processing*, vol. 36, no. 4, pp. 1476–1492, 2017. [Online]. Available: http://dx.doi.org/10.1007/s00034-016-0363-y

[24] A. E. W. George, S. So, R. Ghosh, and K. K. Paliwal, "A Kalman filtering algorithm with joint metrics-based tuning for single-channel speech enhancement," pp. 173–176, Dec. 2016.

[25] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. ITU-T Recommendation P.862." ITU-T, Tech. Rep., 2001.

[26] N. Weiss, "Introductory statistics, 4th edn, addision," 1995.