KALMAN FILTER FOR SPEECH ENHANCEMENT IN COCKTAIL PARTY SCENARIOS USING A CODEBOOK-BASED APPROACH

Mathew Shaji Kavalekalam¹, Mads Græsbøll Christensen¹, Fredrik Gran² and Jesper B. Boldt²

¹Audio Analysis Lab, AD:MT, Aalborg University, Denmark {msk,mgc}@create.aau.dk ²GN Resound A/S, DK 2750, Ballerup, Denmark {jboldt}@gnresound.com

ABSTRACT

Enhancement of speech in non-stationary background noise is a challenging task, and conventional single channel speech enhancement algorithms have not been able to improve the speech intelligibility in such scenarios. The work proposed in this paper investigates a single channel Kalman filter based speech enhancement algorithm, whose parameters are estimated using a codebook based approach. The results indicate that the enhancement algorithm is able to improve the speech intelligibility and quality according to objective measures. Moreover, we investigate the effects of utilizing a speaker specific trained codebook over a generic speech codebook in relation to the performance of the speech enhancement system.

Index Terms— speech enhancement, kalman filter, autoregressive models

1. INTRODUCTION

Enhancement of speech degraded by background noise has been a topic of interest in the past decades due to its wide range of applications. Some of the important applications are in digital hearing aids, hands free mobile communications and in speech recognition devices. Speech enhancement algorithms that have been developed can be mainly categorised into spectral subtraction methods [1], statistical model based methods [2, 3] and subspace based methods [4, 5]. The primary objectives of a speech enhancement system are to improve the quality and intelligibility of the degraded speech. Multi-channel speech enhancement algorithms proposed in [6] have been able to show improvements in speech quality and intelligibility [7]. In comparison to multi-channel algorithms, conventional single channel speech enhancement algorithms have not been successful in improving the speech intelligibility, in presence of non-stationary background noise [8, 9]. Babble noise, which is commonly encountered among hearing aid users is considered to be highly non-stationary noise. Thus, an improvement in speech intelligibility in such scenarios is highly desirable.

In this paper we investigate a speech enhancement framework based on Kalman filtering. Kalman filtering for speech enhancement in white background noise was first proposed in [10]. This work was later extended to deal with coloured noise in [11, 12], where the speech and noise short term predictor parameters (STP) required for the functioning of the Kalman filter is estimated using an approximated expectationmaximisation algorithm. The work presented in this paper uses a codebook-based approach [13] for estimating the speech and noise STP parameters. We also investigate the effects of utilizing a speaker specific trained codebook over a generic speech codebook in relation to the performance of the enhancement system, which has not been considered in previous studies. Objective measures such as Short Term Objective Intelligibility (STOI) [14], Perceptual Evaluation of Speech Quality (PESQ) [15] and Segmental Signal to Noise ratio (SegSNR) have been used to evaluate the performance of the enhancement algorithm in presence of babble noise.

The remainder of the paper is structured as follows. Section 2 explains the signal model and the assumptions that will be used in the paper. Section 3 explains the speech enhancement framework in detail. Experiments and results are presented in Section 4 followed by conclusion in Section 5.

2. SIGNAL MODEL

We now introduce the signal model and assumptions that will be used in the remainder of the paper. It is assumed that clean speech signal s(n) is additively interfered with the noise signal w(n) to form the noisy signal z(n) according to

$$z(n) = s(n) + w(n)$$
 $\forall n = 1, 2...$ (1)

It is also assumed that the noise and speech are statistically uncorrelated with each other. The clean speech signal s(n) is modelled as a stochastic autoregressive (AR) process,

$$s(n) = \sum_{i=1}^{P} a_i(n)s(n-i) + u(n) = \mathbf{a}(n)^T \mathbf{s}(n-1) + u(n),$$
(2)

where $\mathbf{a}(n) = [a_1(n), a_2(n), \dots a_P(n)]^T$ is a vector containing the speech Linear Prediction Coefficients (LPC), $\mathbf{s}(n - \mathbf{s})$

This work was supported by Innovations fund Denmark



Fig. 1. Basic block diagram of the speech enhancement framework

1) = $[s(n-1), \ldots s(n-P)]^T$, P is the order of the AR process corresponding to the speech signal and u(n) is a white Gaussian noise (WGN) with zero mean and excitation variance $\sigma_u^2(n)$. The noise signal is modelled as an AR process,

$$w(n) = \sum_{i=1}^{Q} b_i(n)w(n-i) + v(n) = \mathbf{b}(n)^T \mathbf{w}(n-1) + v(n), \quad (3)$$

where $\mathbf{b}(n) = [b_1(n), b_2(n), \dots b_Q(n)]^T$ is a vector containing noise LPC, $\mathbf{w}(n-1) = [w(n-1), \dots w(n-Q)]^T$, Q is the order of the AR process corresponding to the noise signal and v(n) is a WGN with zero mean and excitation variance $\sigma_v^2(n)$. LPC along with excitation variance generally constitutes the STP parameters.

3. METHOD

This section introduces the enhancement framework investigated in this paper. A single channel speech enhancement technique based on Kalman filtering has been used. A basic block diagram of the speech enhancement framework is shown in Figure 1. It can be seen from the figure that the noisy signal is fed as an input to Kalman smoother, and the speech and noise STP parameters required for the functioning of the Kalman smoother is estimated using a codebook-based approach. The principles of the Kalman filter based speech enhancement is explained in Section 3.1, and the codebook based estimation of the speech and noise STP parameters is explained in Section 3.2.

3.1. Kalman filter for Speech enhancement

The Kalman filter enables us to estimate the state of a process governed by a linear stochastic difference equation in a recursive manner. It is an optimal linear estimator in the sense that it minimises the mean of the squared error. This section explains the principle of a fixed lag Kalman smoother with a smoother delay $d \ge P$. Kalman smoother provides the MMSE estimate of s(n) which can be expressed as

$$\hat{s}(n) = \mathcal{E}(s(n)|z(n+d), \dots, z(1)) \quad \forall n = 1, 2 \dots$$
 (4)

The usage of Kalman filter from a speech enhancement perspective requires the AR signal model in (2) to be written as a state space form as shown below

$$\mathbf{s}(n) = \mathbf{A}(n)\mathbf{s}(n-1) + \mathbf{\Gamma}_1 u(n), \tag{5}$$

where the state vector $\mathbf{s}(n) = [s(n)s(n-1)\dots s(n-d)]^T$ is a $(d+1) \times 1$ vector containing the d+1 recent speech samples, $\mathbf{\Gamma}_1 = [1, 0 \dots 0]^T$ is a $(d+1) \times 1$ vector and $\mathbf{A}(n)$ is the $(d+1) \times (d+1)$ speech state evolution matrix written as

$$\mathbf{A}(n) = \begin{bmatrix} a_1(n) & a_2(n) & \dots & a_P(n) & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 1 & 0 & \vdots & \dots & 0 \\ 0 & \dots & \dots & 1 & 0 & \dots & 0 \\ \vdots & \dots & \dots & 0 & 0 & 1 & 0 \end{bmatrix} .$$
(6)

Analogously, the AR model for the noise signal shown in (3) can be written in the state space form as

$$\mathbf{w}(n) = \mathbf{B}(n)\mathbf{w}(n-1) + \mathbf{\Gamma}_2 v(n), \tag{7}$$

where the state vector $\mathbf{w}(n) = [w(n)w(n-1)\dots w(n-Q+1)]^T$ is a $Q \times 1$ vector containing the Q recent noise samples, $\mathbf{\Gamma}_2 = [1, 0 \dots 0]^T$ is a $Q \times 1$ vector and $\mathbf{B}(n)$ is the $Q \times Q$ noise state evolution matrix

$$\mathbf{B}(n) = \begin{bmatrix} b_1(n) & b_2(n) & \dots & b_Q(n) \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{bmatrix}.$$
 (8)

The state space equations in (5) and (7) are combined together to form a concatenated state space equation as shown in (9)

$$\begin{bmatrix} \mathbf{s}(n) \\ \mathbf{w}(n) \end{bmatrix} = \begin{bmatrix} \mathbf{A}(n) & \mathbf{0} \\ \mathbf{0} & \mathbf{B}(n) \end{bmatrix} \begin{bmatrix} \mathbf{s}(n-1) \\ \mathbf{w}(n-1) \end{bmatrix} + \begin{bmatrix} \mathbf{\Gamma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_2 \end{bmatrix} \begin{bmatrix} u(n) \\ v(n) \end{bmatrix}$$
(9)

which is rewritten as

$$\mathbf{x}(n) = \mathbf{C}(n)\mathbf{x}(n-1) + \mathbf{\Gamma}_3 \mathbf{y}(n), \qquad (10)$$

where $\mathbf{x}(n)$ is the concatenated state space vector, $\mathbf{C}(n)$ is the concatenated state evolution matrix, $\Gamma_3 = \begin{bmatrix} \Gamma_1 & \mathbf{0} \\ \mathbf{0} & \Gamma_2 \end{bmatrix}$ and $\mathbf{y}(n) = \begin{bmatrix} u(n) \\ v(n) \end{bmatrix}$. Consequently, (1) is rewritten as

$$z(n) = \mathbf{\Gamma}^T \mathbf{x}(n), \tag{11}$$

where $\mathbf{\Gamma} = [\mathbf{\Gamma}_1^T \mathbf{\Gamma}_2^T]^T$. The final state space equation and measurement equation denoted by (10) and (11) respectively, is subsequently used for the formulation of the Kalman filter equations (12 - 17). The prediction stage of the Kalman smoother, which computes the a priori estimates of the state

vector $(\hat{\mathbf{x}}(n|n-1))$ and error covariance matrix $(\mathbf{M}(n|n-1))$ is written as

$$\hat{\mathbf{x}}(n|n-1) = \mathbf{C}(n)\hat{\mathbf{x}}(n-1|n-1)$$
(12)

$$\mathbf{M}(n|n-1) = \mathbf{C}(n)\mathbf{M}(n-1|n-1)\mathbf{C}(n)^{T} + \mathbf{\Gamma}_{3} \begin{bmatrix} \sigma_{u}^{2}(n) & 0\\ 0 & \sigma_{v}^{2}(n) \end{bmatrix} \mathbf{\Gamma}_{3}^{T}.$$
(13)

Kalman gain is computed as shown in (14)

$$\mathbf{K}(n) = \mathbf{M}(n|n-1)\mathbf{\Gamma}[\mathbf{\Gamma}^T\mathbf{M}(n|n-1)\mathbf{\Gamma}]^{-1}.$$
 (14)

Correction stage of the Kalman smoother, which computes the a posteriori estimates of the state vector and error covariance matrix is given by

$$\hat{\mathbf{x}}(n|n) = \hat{\mathbf{x}}(n|n-1) + \mathbf{K}(n)[z(n) - \mathbf{\Gamma}^T \hat{\mathbf{x}}(n|n-1)]$$
(15)

$$\mathbf{M}(n|n) = (\mathbf{I} - \mathbf{K}(n)\mathbf{\Gamma}^T)\mathbf{M}(n|n-1).$$
(16)

Finally, the enhanced signal using a Kalman smoother at time index n - d is obtained by taking the $d + 1^{th}$ entry of the a posteriori estimate of the state vector as shown in (17)

$$\hat{s}(n-d) = \hat{\mathbf{x}}_{d+1}(n|n).$$
 (17)

3.2. Codebook based estimation of STP parameters

The usage of Kalman filter from a speech enhancement perspective, as explained in Section 3.1 requires the state evolution matrix C(n) (consisting of the speech LPC and noise LPC), variance of speech excitation signal $\sigma_u^2(n)$ and variance of the noise excitation signal $\sigma_v^2(n)$ to be known. These parameters are assumed to be constant over frames of 25 ms due to the quasi-stationary nature of speech. This section explains the MMSE estimation of these parameters using a codebook based approach. This method uses the a priori information about speech and noise spectral shapes stored in trained codebooks in the form of LPC.

The parameters to be estimated are concatenated to form a single vector $\theta = [\mathbf{a}; \mathbf{b}; \sigma_u^2; \sigma_v^2]$. The MMSE estimate of the parameter θ is written as

$$\hat{\theta} = \mathcal{E}(\theta | \mathbf{z}),$$
 (18)

where z denotes a frame of noisy samples. Using Bayes' theorem, (18) can be rewritten as

$$\hat{\theta} = \int_{\Theta} \theta p(\theta | \mathbf{z}) d\theta = \int_{\Theta} \theta \frac{p(\mathbf{z} | \theta) p(\theta)}{p(\mathbf{z})} d\theta, \qquad (19)$$

where Θ denotes the support space of the parameters to be estimated. Let us define $\theta_{ij} = [\mathbf{a}_i; \mathbf{b}_j; \sigma_{u,ij}^{2,ML}; \sigma_{v,ij}^{2,ML}]$ where \mathbf{a}_i is the *i*th entry of speech codebook (of size N_s), \mathbf{b}_j is the *j*th entry of the noise codebook (of size N_w) and $\sigma_{u,ij}^{2,ML}, \sigma_{v,ij}^{2,ML}$ represents the maximum likelihood (ML) estimates [16] of speech and noise excitation variances which depends on $\mathbf{a}_i, \mathbf{b}_j$ and \mathbf{z} . ML estimates of speech and noise excitation variances are estimated according to the following equation,

$$\mathbf{E}\begin{bmatrix}\sigma_{u,ij}^{2,ML}\\\sigma_{v,ij}^{2,ML}\end{bmatrix} = \mathbf{D},$$
(20)

where

$$\mathbf{E} = \begin{bmatrix} \left\| \frac{1}{P_{z}^{2}(\omega)|A_{s}^{i}(\omega)|^{4}} \right\| & \left\| \frac{1}{P_{z}^{2}(\omega)|A_{s}^{i}(\omega)|^{2}|A_{w}^{j}(\omega)|^{2}} \right\| \\ \frac{1}{P_{z}^{2}(\omega)|A_{s}^{i}(\omega)|^{2}|A_{w}^{j}(\omega)|^{2}} \\ \mathbf{D} = \begin{bmatrix} \left\| \frac{1}{P_{z}(\omega)|A_{s}^{i}(\omega)|^{2}} \\ \frac{1}{P_{z}(\omega)|A_{w}^{j}(\omega)|^{2}} \\ \frac{1}{P_{z}(\omega)|A_{w}^{j}(\omega)|^{2}} \\ \end{bmatrix} \end{bmatrix},$$
(22)

and $\frac{1}{|A_s^i(\omega)|^2}$ is the spectral envelope corresponding to the i^{th} entry of the speech codebook, $\frac{1}{|A_w^i(\omega)|^2}$ is the spectral envelope corresponding to the j^{th} entry of the noise codebook and $P_z(\omega)$ is the spectral envelope corresponding to the noisy signal. Consequently, a discrete counterpart to (19) can be written as

$$\hat{\theta} = \frac{1}{N_s N_w} \sum_{i=1}^{N_s} \sum_{j=1}^{N_w} \theta_{ij} \frac{p(\mathbf{z}|\theta_{ij}) p(\sigma_{u,ij}^{2,ML}) p(\sigma_{v,ij}^{2,ML})}{p(\mathbf{z})}, \quad (23)$$

where the MMSE estimate is expressed as a weighted linear combination of θ_{ij} with weights proportional to $p(\mathbf{z}|\theta_{ij})$, which is computed according to the following equations

$$p(\mathbf{z}|\theta_{ij}) = \exp(-d_{\mathrm{IS}}(P_z(\omega), \hat{P}_z^{ij}(\omega)))$$
(24)

$$\hat{P}_{z}^{ij}(\omega) = \frac{\sigma_{u,ij}^{2,ML}}{|A_{s}^{i}(\omega)|^{2}} + \frac{\sigma_{v,ij}^{2,ML}}{|A_{w}^{i}(\omega)|^{2}}$$
(25)

$$p(\mathbf{z}) = \frac{1}{N_s N_w} \sum_{i=1}^{N_s} \sum_{j=1}^{N_w} p(\mathbf{z}|\theta_{ij}) p(\sigma_{u,ij}^{2,ML}) p(\sigma_{v,ij}^{2,ML})$$
(26)

where $d_{\rm IS}(P_z(\omega), \hat{P}_z^{ij}(\omega))$ is the Itakura Saito distortion [17] between the noisy spectrum and the modelled noisy spectrum. More details on the derivation of this method can be found in [13] and the references therein. It should be noted that the weighted summation of AR parameters in (23) should be performed in the line spectral frequency (LSF) domain rather than in the LPC domain. Weighted summation in LSF domain is guaranteed to result in stable inverse filters, which is not always the case in LPC domain [18].

4. EXPERIMENTS

This section describes the experiments performed to evaluate the speech enhancement framework explained in Section 3. Objective measures, that have been used for evaluation are STOI, PESQ and SegSNR. The test set for this experiment consisted of speech from 4 different speakers: 2 male and 2 female speakers from the CHiME database [19] resampled to 8 KHz. The noise signal used for simulations is multi-talker babble from the NOIZEUS database [20]. The speech and noise STP parameters required for the enhancement procedure is estimated every 25 ms as explained in Section 3.2. Speech codebook used for the estimation of STP parameters is generated using the Generalised Lloyd algorithm (GLA) [21] on a training sample of 10 minutes of speech from the TIMIT database [22]. The noise codebook is generated using two minutes of babble. The order of the speech and noise AR model is chosen to be 14. The parameters that have been used for the experiments are summarised in Table 1.

fs	Frame Size	N_s	N_w	P	Q
8 Khz	200 (25ms)	256	12	14	14

Table 1. Experimental setup

The estimated STP parameters are subsequently used for enhancement by a fixed lag Kalman smoother (with d = 40). In this paper, we have also investigated the effects of having a speaker specific codebook instead of a generic speech codebook. The speaker specific codebook is generated by GLA using a training sample of five minutes of speech from the specific speaker of interest. The speech samples used for testing was not included in the training set. A speaker codebook size of 64 entries was empirically noted to be sufficient. The system of Kalman smoother, utilising a speech codebook and speaker codebook for the estimation of STP parameters is denoted as KS-speech model and KS-speaker model respectively. The results are compared with Ephraim-Malah (EM) method [3] and state of the art MMSE estimator based on generalised gamma priors (MMSE-GGP) [23]. Figures 2, 3 and 4 shows the comparison of STOI, SegSNR and PESQ scores respectively, for the above mentioned methods. It can be seen from Figure 2 that the enhanced signals obtained using EM and MMSE-GGP have lower intelligibility scores than the noisy signal, according to STOI. The enhanced signals obtained using KS-speech model and KS-speaker model show a higher intelligibility score in comparison to the noisy signal. It can be seen, that using a speaker specific codebook instead of a generic speech codebook is beneficial, as the STOI scores shows an increase of upto 6%. The SegSNR and PESQ results shown in Figures 3 and 4 also indicate that KS-speaker model and KS-speech model performs better than the other methods. Informal listening tests were also conducted to evaluate the performance of the algorithm.

5. CONCLUSION

This paper investigated a speech enhancement method based on Kalman filter, and the parameters required for the function-



Fig. 2. comparison of STOI scores







Fig. 4. comparison of PESQ scores

ing of Kalman filter were estimated using a codebook based approach. Objective measures such as STOI, SegSNR and PESQ were used to evaluate the performance of the algorithm in presence of babble noise. Experimental results indicate that the presented method was able to increase the speech quality and speech intelligibility according to the objective measures. Moreover, it was noted that having a speaker specific trained codebook instead of a generic speech codebook can show upto 6% increase in STOI scores. As future work, it would be interesting to see how a generic speech codebook can be adapted to a speaker specific codebook. Subjective listening tests will also be conducted in the future to validate the results shown here.

6. REFERENCES

- S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 113–120, 1979.
- [2] Y. Ephraim, "Statistical-model-based speech enhancement systems," *Proceedings of the IEEE*, vol. 80, no. 10, pp. 1526–1555, 1992.
- [3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, 1984.
- [4] Y. Ephraim and H. L. V. Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Audio and Speech Process.*, vol. 3, no. 4, pp. 251–266, 1995.
- [5] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Audio and Speech Process.*, vol. 11, no. 4, pp. 334–341, 2003.
- [6] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," *Handbook on Array Processing and Sensor Networks*, pp. 269–302, 2008.
- [7] H. Luts, K. Eneman, J. Wouters, M. Schulte, M. Vormann, M. Büchler, N. Dillier, R. Houben, W. A. Dreschler, M. Froehlich, et al., "Multicenter evaluation of signal enhancement algorithms for hearing aids," *The Journal of the Acoustical Society of America*, vol. 127, no. 3, pp. 1491–1505, 2010.
- [8] R. Bentler, Y. H. Wu, J. Kettel, and R. Hurtig, "Digital noise reduction: Outcomes from laboratory and field studies," *International Journal of Audiology*, vol. 47, no. 8, pp. 447–460, 2008.
- [9] P. C. Loizou, Speech enhancement: theory and practice, CRC press, 2013.
- [10] K. K. Paliwal and A. Basu, "A speech enhancement method based on kalman filtering," *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, 1987.
- [11] J. D. Gibson, B. Koo, and S. D. Gray, "Filtering of colored noise for speech enhancement and coding," *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1732–1742, 1991.
- [12] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and sequential kalman filter-based speech enhancement algorithms," *IEEE Trans. on Speech and Audio Process.*, vol. 6, no. 4, pp. 373–385, 1998.

- [13] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, "Codebook-based bayesian speech enhancement for nonstationary environments," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 2, pp. 441– 452, 2007.
- [14] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of timefrequency weighted noisy speech," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 19, no. 7, pp. 2125–2136, 2011.
- [15] "Perceptual evaluation of speech quality, an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," *ITU-T Recommendation*, p. 862, 2001.
- [16] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, "Codebook driven short-term predictor parameter estimation for speech enhancement," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 14, no. 1, pp. 163–176, 2006.
- [17] K. K. Paliwal and W. B. Kleijn, "Quantization of lpc parameters," *Speech Coding and Synthesis*, pp. 433– 466, 1995.
- [18] Jr. Gray, H. Augustine, and J. D. Markel, "Distance measures for speech processing," *IEEE Trans. Acoust., Speech and Signal Process.*, vol. 24, no. 5, pp. 380–391, 1976.
- [19] J. Barker, R. Marxer, E. Vincent, and S. Watanabe, "The third 'chime' speech separation and recognition challenge: Dataset, task and baselines," *IEEE 2015 Automatic Speech Recognition and Understanding Workshop*, 2015.
- [20] Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech communication*, vol. 49, no. 7, pp. 588–601, 2007.
- [21] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Communications*, vol. 28, no. 1, pp. 84–95, 1980.
- [22] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "Darpa timit acoustic-phonetic continous speech corpus cd-rom. nist speech disc 1-1.1," *NASA STI/Recon Technical Report N*, vol. 93, pp. 27403, 1993.
- [23] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete fourier coefficients with generalized gamma priors," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 6, pp. 1741–1752, 2007.