INNOVATION CODING WITH A CROSS-CORRELATED QUANTIZATION NOISE MODEL

Søren Vang Andersen, Morten Olesen, Søren Holdt Jensen, and Egon Hansen
CPK, Aalborg University
Fredrik Bajers Vej 7
DK-9220 Aalborg Øst, Denmark
e-mail: sva@cpk.auc.dk

ABSTRACT

We present the use of a cross-correlated quantization noise model in the recently proposed Kalman innovation speech coding scheme. Computer simulations and informal listening tests indicate that the incorporation of a cross-correlated noise model yields an improvement in both SNR and perceptual quality when compared to a uncorrelated noise model.

1 INTRODUCTION

The problem of efficiently coding a speech signal at low bit rates has lead to a variety of coding schemes. Many of these schemes achieve high coding efficiency through the use of statistical modeling of the speech signal. At low bit rates quantization noise becomes a non-negligible part of the decoded signal. Therefore statistical modeling of not only the speech signal but also the quantization noise may lead to novel, improved schemes.

The innovation speech coding scheme, proposed by Ramabadran and Sinha [4], is an example of a coding scheme with statistical modeling of quantization noise. To be specific, the decoder uses a Kalman estimator to estimate the speech signal present at the encoder and the estimation is based on innovations which are corrupted by quantization noise. The speech signal is modeled as an outcome of an autoregressive (AR) process, and the quantization noise is modeled as an outcome of a white process that is uncorrelated with other involved processes. This model for the quantization noise is here termed the uncorrelated model (UCM).

In this paper, we propose an extension of the innovation speech coding scheme. Our scheme uses a more comprehensive model for the quantization noise. This model incorporates the cross-correlation, which in fact exists, between the quantization noise process and the process for the input to the quantizer. Such a model is here termed a cross-correlated model (CCM).

2 INNOVATION CODING

The innovation coding scheme will be described briefly in the following. To generalize the description, we here consider vector quantization; this follows as a direct generalization of the scalar scheme described in [4].

Let the AR model for the speech signal be given by the equation[1]

$$X_{t+1} = \boldsymbol{\alpha}_t \begin{bmatrix} X_t & X_{t-1} & \dots & X_{t-p+1} \end{bmatrix}^T + Y_t,$$

where t is the time index, p is the order of the model, $\boldsymbol{\alpha}_t$ is a row vector containing model coefficients, X_t is the process modeling the speech signal, and Y_t is a white process that is uncorrelated with other involved processes. A state space formulation of this AR model can be written as

$$\boldsymbol{S}_{t+j} = \boldsymbol{\Theta}_{t+j,t}\boldsymbol{S}_t + \boldsymbol{R}_{t+j,t}, \tag{1}$$

where j is the number of samples we advance in the signal for each iteration of the state space model. The state transition matrix $\boldsymbol{\Theta}_{t+j,t}$ is a square matrix defined by[2]

$$\boldsymbol{\Theta}_{t+j,t} \equiv \begin{cases} \begin{bmatrix} \boldsymbol{\alpha}_t \\ \boldsymbol{I} \quad \boldsymbol{0} \end{bmatrix}, & j = 1, \\ \boldsymbol{\Theta}_{t+j,t+j-1}\boldsymbol{\Theta}_{t+j-1,t}, & j > 1. \end{cases}$$

The state $\boldsymbol{S}_t$ and state excitation $\boldsymbol{R}_{t+j,t}$ are column vectors defined by

$$\begin{aligned} \boldsymbol{S}_t &\equiv \begin{bmatrix} X_t & X_{t-1} & \dots & X_{t-p+1} \end{bmatrix}^T \\ \boldsymbol{R}_{t+j,t} &\equiv \begin{cases} \begin{bmatrix} Y_t & \boldsymbol{0} \end{bmatrix}^T, & j = 1, \\ \boldsymbol{\Theta}_{t+j,t+j-1}\boldsymbol{R}_{t+j-1,t} + \boldsymbol{R}_{t+j,t+j-1}, & j > 1. \end{cases} \end{aligned}$$

In the following, we will suppress time indexes to avoid an overcrowded notation. Instead we introduce a delay operator $\mathsf{D}[\cdot]$ such that e.g. (1) becomes $\boldsymbol{S} = \mathsf{D}[\boldsymbol{\Theta}\boldsymbol{S} + \boldsymbol{R}]$.

[1] In this paper uppercase Arabic letters are used to denote processes, while corresponding lowercase letters denote their outcomes. Moreover $\mathsf{E}[\cdot]$, $\mathsf{Var}[\cdot]$, and $\mathsf{Corr}[\cdot]$ will be used to denote expectation, variance, and correlation respectively.

[2] We use $\boldsymbol{I}$ to denote an identity matrix and $\boldsymbol{0}$ to denote a zero vector or zero matrix. The dimensions of $\boldsymbol{I}$ and $\boldsymbol{0}$ will follow from the context.

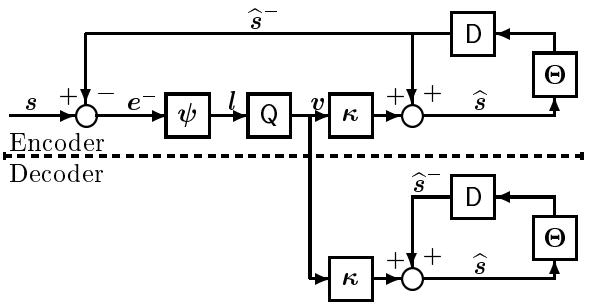

Figure 1: Block diagram of the innovation coding scheme.

With the state space formulation the innovation coding scheme can be illustrated as in Figure 1. The Kalman estimator provides an *a priori* estimate $\widehat{\boldsymbol{s}}^-$ for $\boldsymbol{s}$. From the *a priori* error $\boldsymbol{e}^-$ a vector of linear combinations $\boldsymbol{l} = \boldsymbol{\psi}\boldsymbol{e}^-$ is obtained. This vector is quantized by the quantizer $\mathsf{Q}[\cdot]$ to form the innovation $\boldsymbol{v}$. The innovation is transmitted to the decoder and forms the basis for the estimation in both encoder and decoder. Premultiplying $\boldsymbol{v}$ with the Kalman gain matrix $\boldsymbol{\kappa}$ gives the term to update the *a priori* estimate $\widehat{\boldsymbol{s}}^-$ into the *a posteriori* estimate $\widehat{\boldsymbol{s}}$. From this *a posteriori* estimate the next *a priori* estimate is obtained by pre-multiplication with $\boldsymbol{\Theta}$. The decoded speech signal is j consecutive elements of $\widehat{\boldsymbol{s}}$.

The influence of the quantizer is modeled as a zero-mean, additive process $\boldsymbol{N}$, that is

$$\boldsymbol{V} = \mathsf{Q}[\boldsymbol{L}] = \boldsymbol{L} + \boldsymbol{N}. \tag{2}$$

To determine the correlation matrix for $\boldsymbol{N}$ we restrict the treatment to schemes in which $\boldsymbol{\psi}$ is chosen such that $\mathsf{Corr}[\boldsymbol{L}]$ becomes positive definite and diagonal, and we assume the quantizer to be scaled using the diagonal elements of $\mathsf{Corr}[\boldsymbol{L}]$. It is then reasonable to assume the correlation matrix for $\boldsymbol{N}$ to be

$$\mathsf{Corr}[\boldsymbol{N}] \;=\; \Delta\mathsf{Corr}[\boldsymbol{L}]\,. \tag{3}$$

Here Δ is a diagonal matrix expressing the relative quantization noise variances. At this point we assume the UCM to hold. We can then use the standard Kalman estimator, see e.g. [3], and the encoding algorithm can be stated as a generalization of the algorithm given in [4].

UCM Encoding Algorithm:

$$\begin{aligned}
\widehat{\boldsymbol{s}}^- &= \mathsf{D}[\boldsymbol{\Theta}\widehat{\boldsymbol{s}}] && (4)\\
\boldsymbol{e}^- &= \boldsymbol{s} - \widehat{\boldsymbol{s}}^- && (5)\\
\mathsf{Corr}\left[\boldsymbol{E}^-\right] &= \mathsf{D}\left[\boldsymbol{\Theta}\mathsf{Corr}[\boldsymbol{E}]\,\boldsymbol{\Theta}^T + \mathsf{Corr}[\boldsymbol{R}]\right] && (6)\\
\boldsymbol{l} &= \boldsymbol{\psi}\boldsymbol{e}^- && (7)\\
\mathsf{Corr}[\boldsymbol{L}] &= \boldsymbol{\psi}\mathsf{Corr}\left[\boldsymbol{E}^-\right]\boldsymbol{\psi}^T && (8)\\
\boldsymbol{v} &= \mathsf{Q}[\boldsymbol{l}] && (9)\\
\mathsf{Corr}[\boldsymbol{V}] &= (\boldsymbol{I} + \Delta)\,\mathsf{Corr}[\boldsymbol{L}] && (10)\\
\boldsymbol{\kappa} &= \mathsf{Corr}\left[\boldsymbol{E}^-\right]\boldsymbol{\psi}^T\mathsf{Corr}[\boldsymbol{V}]^{-1} && (11)
\end{aligned}$$

$$\begin{aligned}
\widehat{\boldsymbol{s}} &= \widehat{\boldsymbol{s}}^- + \boldsymbol{\kappa}\boldsymbol{v} && (12)\\
\mathsf{Corr}[\boldsymbol{E}] &= (\boldsymbol{I} - \boldsymbol{\kappa}\boldsymbol{\psi})\,\mathsf{Corr}\left[\boldsymbol{E}^-\right]. && (13)
\end{aligned}$$

The decoding algorithm is identical except for the absence of equations (5), (7), and (9).

3 THE CROSS-CORRELATED QUANTIZATION NOISE MODEL

In this section we derive the modified encoding algorithm which incorporates the CCM. The basic assumption made is that $\left[\begin{array}{cc}\widehat{\boldsymbol{S}}^{-T} & \boldsymbol{E}^{-T}\end{array}\right]^T$ is a zero-mean Gaussian process. Hence, for any two processes that are linear functions of $\left[\begin{array}{cc}\widehat{\boldsymbol{S}}^{-T} & \boldsymbol{E}^{-T}\end{array}\right]^T$, uncorrelatedness will imply independence [5, p.51]. In various contexts we need this assumption in arguments of the following form:

> Let $\boldsymbol{A}$ be a stochastic process that is uncorrelated with $\boldsymbol{E}^-$, and for which $\left[\begin{array}{cc}\boldsymbol{A}^T & \boldsymbol{E}^{-T}\end{array}\right]^T$ is a zero-mean Gaussian process. Then $\boldsymbol{A}$ and $\boldsymbol{E}^-$ are independent. The quantization noise process $\boldsymbol{N}$ is a zero-mean process and a deterministic function of $\boldsymbol{E}^-$. Thus independence of $\boldsymbol{A}$ and $\boldsymbol{E}^-$ implies independence of $\boldsymbol{A}$ and $\boldsymbol{N}$. Because $\boldsymbol{A}$ and $\boldsymbol{N}$ are zero-mean and independent, they are uncorrelated. The argument above also holds for $\boldsymbol{E}^-$ replaced with $\boldsymbol{L}$.

We refer to this argument as the *independence argument.*

The starting point of the derivation is the following important statement:

> If the quantizer satisfies the centroid condition for a squared error distance measure, then the following relation holds:
>
> $$\mathsf{E}\left[\boldsymbol{L}\boldsymbol{N}^T\right] = -\mathsf{Corr}[\boldsymbol{N}]\,. \tag{14}$$

This follows using the same method of proof as in [2, p.357]. Because $\mathsf{Corr}[\boldsymbol{L}]$ is diagonal, the *independence argument* can be used to generalize the above statement to quantizers constructed as the product of lower dimensional quantizers that all satisfy the centroid condition. We are now ready to modify the Kalman estimator to incorporate relation (14) into the noise model.

The Kalman gain for the purpose of updating the *a priori* estimate to the *a posteriori* estimate is given in [3, p.311] as

$$\boldsymbol{\kappa} \;=\; \mathsf{E}\left[\boldsymbol{S}\boldsymbol{V}^T\right]\mathsf{Corr}[\boldsymbol{V}]^{-1}. \tag{15}$$

If we use the general property that $\widehat{\boldsymbol{S}}^-$ and $\boldsymbol{E}^-$ are uncorrelated, and we use the *independence argument*, then (15) may be written as

$$\begin{aligned}
\boldsymbol{\kappa} \;=\;& \mathsf{Corr}\left[\boldsymbol{E}^-\right]\boldsymbol{\psi}^T\mathsf{Corr}[\boldsymbol{V}]^{-1}\\
&+\mathsf{E}\left[\boldsymbol{E}^-\boldsymbol{N}^T\right]\mathsf{Corr}[\boldsymbol{V}]^{-1}. \qquad (16)
\end{aligned}$$

In the Kalman estimator for the UCM the second term in (16) is a zero matrix. For the CCM this, however, cannot be true. To see this, use (7) to obtain the following relation:

$$\mathsf{E}\left[\boldsymbol{L}\boldsymbol{N}^T\right] = \boldsymbol{\psi}\mathsf{E}\left[\boldsymbol{E}^-\boldsymbol{N}^T\right]. \tag{17}$$

From (14) we see that the left-hand side of (17) is not a zero matrix which means that $\mathsf{E}\left[\boldsymbol{E}^-\boldsymbol{N}^T\right]$ cannot be a zero matrix. Because $\boldsymbol{\psi}$ is chosen such that $\mathsf{Corr}[\boldsymbol{L}]$ becomes positive definite, $\boldsymbol{\psi}$ cannot have more rows than columns and the rank of $\boldsymbol{\psi}$ will equate its number of rows. Hence, if $\boldsymbol{\psi}$ is square, then it has a unique inverse, and $\mathsf{E}\left[\boldsymbol{E}^-\boldsymbol{N}^T\right]$ can be determined from (17). On the other hand, if $\boldsymbol{\psi}$ is not square, some more care must be taken. One solution for this case is outlined in the following paragraph.

Consider the expansion of $\boldsymbol{\psi}$ into a square full rank matrix $\boldsymbol{\psi}_E \equiv \left[\begin{array}{cc} \boldsymbol{\psi}^T & \boldsymbol{\psi}_A^T \end{array}\right]^T$ such that the expansion of $\boldsymbol{L}$,

$$\boldsymbol{L}_E \equiv \boldsymbol{\psi}_E\boldsymbol{E}^- = \left[\begin{array}{cc} \boldsymbol{L}^T & \boldsymbol{L}_A^T \end{array}\right]^T, \tag{18}$$

will have positive definite and diagonal correlation matrix $\mathsf{Corr}[\boldsymbol{L}_E]$. Note that $\boldsymbol{\psi}_E$ will exist provided that $\mathsf{Corr}\left[\boldsymbol{E}^-\right]$ is positive definite. Now, suppose no information about the extra linear combinations $\boldsymbol{l}_A$ is passed through the quantizer. Then the extra linear combinations have not changed the available information and consequently they have not changed the estimate. Formally this can be expressed by the expanded innovation process $\boldsymbol{V}_E \equiv \left[\begin{array}{cc} \boldsymbol{V}^T & \boldsymbol{0} \end{array}\right]^T$, where the zero information part is assigned the *a priori* expected value for $\boldsymbol{l}_A$, i.e. $\boldsymbol{0}$. The expanded quantization noise process thus becomes $\boldsymbol{N}_E \equiv \boldsymbol{V}_E - \boldsymbol{L}_E = \left[\begin{array}{cc} \boldsymbol{N}^T & -\boldsymbol{L}_A^T \end{array}\right]^T$. We observe that $\boldsymbol{L}$ and $\boldsymbol{L}_A$ are uncorrelated, zero-mean, Gaussian processes and use the *independence argument* to obtain

$$\mathsf{E}\left[\boldsymbol{L}_E\boldsymbol{N}_E^T\right] = -\mathsf{Corr}[\boldsymbol{L}_E]\,\Delta_E, \tag{19}$$

where Δ_E is defined as

$$\Delta_E \equiv \left[\begin{array}{cc} \Delta & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{array}\right].$$

Now (18) and (19) leads to

$$\begin{aligned} \boldsymbol{\psi}_E\mathsf{E}\left[\boldsymbol{E}^-\boldsymbol{N}_E^T\right] &= \mathsf{E}\left[\boldsymbol{L}_E\boldsymbol{N}_E^T\right] \\ &= -\mathsf{Corr}[\boldsymbol{L}_E]\,\Delta_E \\ &= -\boldsymbol{\psi}_E\mathsf{Corr}\left[\boldsymbol{E}^-\right]\boldsymbol{\psi}_E^T\Delta_E. \end{aligned}$$

Since $\boldsymbol{\psi}_E$ has a unique inverse, we have

$$\mathsf{E}\left[\boldsymbol{E}^-\boldsymbol{N}_E^T\right] = -\mathsf{Corr}\left[\boldsymbol{E}^-\right]\boldsymbol{\psi}_E^T\Delta_E.$$

Equating first columns in this matrix expression yields the unknown cross-correlation

$$\mathsf{E}\left[\boldsymbol{E}^-\boldsymbol{N}^T\right] = -\mathsf{Corr}\left[\boldsymbol{E}^-\right]\boldsymbol{\psi}^T\Delta. \tag{20}$$

Note that this is the result which we can obtain directly from (17) in the case that $\boldsymbol{\psi}$ is square.

Finally, the Kalman gain $\boldsymbol{\kappa}$ can be obtained by inserting (20) into (16) which yields

$$\boldsymbol{\kappa} = \mathsf{Corr}\left[\boldsymbol{E}^-\right]\boldsymbol{\psi}^T\left(\boldsymbol{I}-\Delta\right)\mathsf{Corr}[\boldsymbol{V}]^{-1}.$$

The rest of the derivation follows steps parallel to those for a Kalman estimator for uncorrelated noise, see e.g. [3].

CCM Encoding Algorithm:

The CCM algorithm is obtained from the UCM algorithm by deleting (10), and replacing (11) and (13) with the following:

$$\boldsymbol{\kappa} = \mathsf{Corr}\left[\boldsymbol{E}^-\right]\boldsymbol{\psi}^T\mathsf{Corr}[\boldsymbol{L}]^{-1} \tag{21}$$

$$\mathsf{Corr}[\boldsymbol{E}] = \left(\boldsymbol{I}-\boldsymbol{\kappa}\left(\boldsymbol{I}-\Delta\right)\boldsymbol{\psi}\right)\mathsf{Corr}\left[\boldsymbol{E}^-\right]. \tag{22}$$

4 SIMULATION RESULTS

To investigate the impact of the CCM on the coder performance, simulations on speech signals were carried out. These simulations are described in the following.

The original speech signal was sampled at 8 kHz and linearly quantized using 16 bits per sample. The AR model for the speech signal was a two step filter model consisting of a long delay filter followed by a short delay filter. The short delay filter was a 10'th order all-pole filter. Coefficients for the short delay filter were determined for 10 ms frames using the autocorrelation method with a 30 ms Hamming window covering the last frame, the current frame, and the next frame. The long delay filter was of the form described in [1] using 3 nonzero coefficients. Coefficients for the long delay filter as well as $\mathsf{Var}[Y]$ were estimated for each 10 ms frame. The coefficient vector $\boldsymbol{\alpha}$ was obtained through convolution of the long and short delay inverse filters as described in [4]. The order of this filter was fixed at $p = 120$. Innovations were quantized using the product of 1-bit scalar quantizers optimized for a Gaussian process. The simulations were focused on the issue of quantizing innovations and therefore no quantization of the AR model was included.

Four simulations were made. Results from these simulations are listed in Table 1. In simulations A and B, $\boldsymbol{\psi}$

Table 1: Simulation results

			SEG-SNR [dB]				
ID.	Model	j	F1	F2	M1	M2	Comb.
A	UCM	1	17.4	15.8	16.8	10.1	14.9
B	CCM	1	19.3	20.0	19.5	17.9	19.2
C	UCM	2	18.4	18.2	17.7	14.3	17.2
D	CCM	2	19.1	19.8	19.3	17.8	19.0

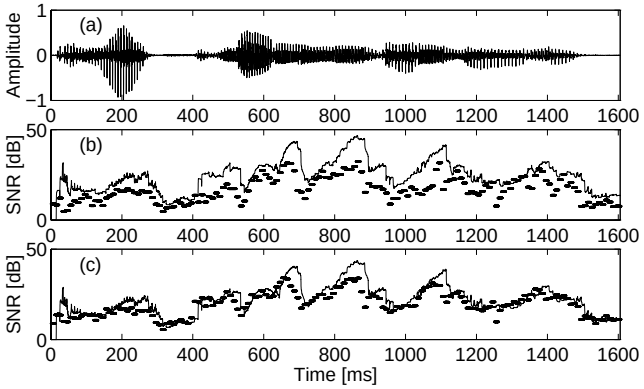

Figure 2: (a) The original speech signal (F2). (b) Time averaged SNR estimate (points) and SNR estimate obtained from the signal model and the *a posteriori* error correlation matrix (solid line), from simulation A with the UCM. (c) Like (b) but from simulation B with the CCM.

was a single row vector chosen to be an eigenvector corresponding to the largest eigenvalue of $\mathsf{Corr}[\boldsymbol{E}^-]$. This choice was proposed in [4]. In simulations C and D, $j = 2$ and $\boldsymbol{\psi}$ was chosen with two rows equating eigenvectors corresponding to the two largest eigenvalues of $\mathsf{Corr}[\boldsymbol{E}^-]$. Simulation A and C used the UCM, whereas simulations B and D used the CCM. Simulations were carried out on speech uttered by two female (F) and two male (M) speakers. The combined length of the speech from all four speakers was 10 seconds. As an objective performance indicator segmental SNR (SEG-SNR) was used. The SEG-SNR was calculated using 10 ms frames. Frames with signal power more than 40 dB below the average signal power were left out of the calculation. From Table 1, we see that the CCM for the combined data improved the SEG-SNR with 4.3 dB for the case $j = 1$ and 1.8 dB for the case $j = 2$. This result was supported by an improvement in perceptual quality when judged in informal listening test. In some instants this improvement was substantial.

To describe further the difference between the UCM and the CCM, we investigated the adjustment of the Kalman estimator and quantizer. The signal model provides an estimate of the signal variance and $\mathsf{Corr}[\boldsymbol{E}]$ gives an estimate of the error variance. Using these two estimates, an estimate for the SNR of the decoded speech signal can be made. In Figure 2 this SNR estimate is plotted together with an SNR estimate obtained by time averaging over 10 ms frames. It is seen that these two estimates are more alike when the CCM is used. This observation was the same for all four speakers which we take as an indication that the CCM models the quantization noise more accurately than the UCM. The scaled input to the quantizer is expected to be a unit variance, zero-mean Gaussian process. Based on all 10 seconds of speech in simulations A and B normal probability plots for this process were made. The results are shown

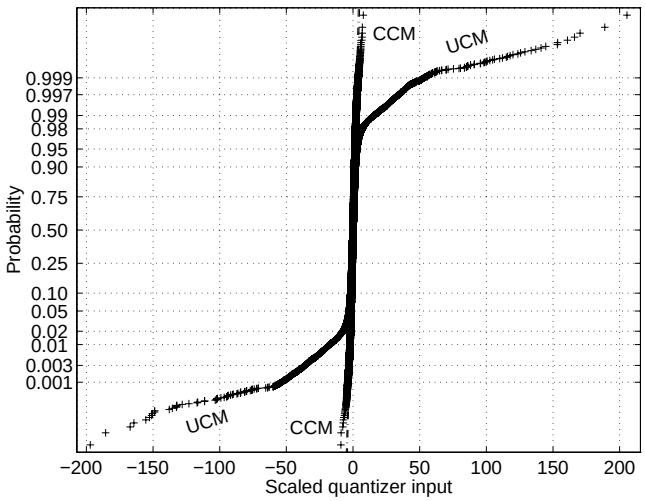

Figure 3: Normal probability plot for the scaled quantizer input. Outcomes of simulation A with the UCM and simulation B with the CCM are plotted. The dash dot line indicates an exact Gaussian process.

in Figure 3. It is clearly seen that when the UCM is used, the scaled quantizer input has a larger amount of outliers than when the CCM is used. Indeed, when the CCM is used the process is close to Gaussian. This result agrees with the assumption made in Section 3.

5 CONCLUSION

The quantization noise modeling in the innovation coding scheme was improved. This improvement was obtained by incorporating the cross-correlation between the quantization noise and the input to the quantizer. In simulations on speech signals the cross-correlated model resulted in a significant increase in the segmental SNR. This increase was supported by an improvement in perceptual quality when judged in informal listening test. In addition to this the adjustment of the estimation statistics and quantizer scaling improved in accuracy.

REFERENCES

[1] B. S. Atal. Predictive Coding of Speech at Low Bit Rates. *IEEE Trans. Comm.*, 30(4):600–614, 1982.

[2] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression.* Kluwer Academic Publishers, 1992.

[3] S. Haykin. *Adaptive Filter Theory.* Printice Hall, 3rd edition, 1996.

[4] T. V. Ramabadran and D. Sinha. Speech Data Compression Through Sparse Coding of Innovations. *IEEE Trans. Speech, Audio.*, 2(2):274–284, 1994.

[5] K. S. Shanmugan and A. M. Breipohl. *Random Signals: Detection, Estimation and Data Analysis.* John Wiley and Sons, Inc., 1988.