ON GMM KALMAN PREDICTIVE CODING OF LSFS FOR PACKET LOSS

Shaminda Subasingha[†], Manohar N. Murthi[†], and Søren Vang Andersen[‡]

[†] Dept. of Electrical and Computer Engineering University of Miami, USA s.subasingha@umiami.edu, mmurthi@miami.edu

ABSTRACT

Gaussian Mixture Model (GMM)-based Kalman predictive coders have been shown to perform better than baseline GMM Recursive Coders in predictive coding of Line Spectral Frequencies (LSFs) for both clean and packet loss conditions However, these stationary GMM Kalman predictive coders were not specifically designed for operation in packet loss conditions. In this paper, we demonstrate an approach to the the design of GMM-based predictive coding for packet loss channels. In particular, we show how a stationary GMM Kalman predictive coder can be modified to obtain a set of encoding and decoding modes, each with different Kalman gains. This approach leads to more robust performance of predictive coding of LSFs in packet loss conditions, as the coder mismatch between the encoder and decoder are minimized. Simulation results show that this Robust GMM Kalman predictive coder performs better than other baseline GMM predictive coders with no increase in complexity. To the best of our knowledge, no previous work has specifically examined the design of GMM predictive coders for packet loss conditions.

Index Terms— speech coding, Kalman filtering, GMM, vector quantization

1. INTRODUCTION

Vector quantization based on Gaussian Mixture Models (GMMs) for vector data such as Line Spectral Frequencies (LSFs) has been widely investigated in recent years ([1],[2],[3]). In [4, 5], we applied Kalman filtering principles to GMM predictive coding of LSFs to account for the quantization noise. With this GMM-Kalman framework, one obtains an online a-posteriori GMM (online GMM-KF) that adapts its transforms and codebooks at each time instance, leading to better predictive coding. Moreover, we demonstrated how the online GMM-KF can be used to obtain a stationary GMM Kalman predictive coder (GMM-KF-ST) which uses fixed transforms. The GMM-KF-ST provided better performance in both clean and packet loss conditions

[‡] Department of Electronic Systems Aalborg University, Denmark sva@es.aau.dk

than the standard GMM-RC ([2]) at roughly the same complexity. While the GMM-KF-ST in [4, 5] performed well in packet loss conditions, this predictive coder was not explicitly designed to operate in packet loss conditions. In fact, to the best of our knowledge, no previous papers have explicitly attempted to design GMM predictive coding of LSFs for packet loss conditions. This is the subject of this paper.

In particular, we explore how to take the stationary GMM-KF-ST, and convert it into a predictive coder more robust for packet loss conditions. In contrast to the GMM-KF-ST, the robust design presented in this paper consists of: 1) An encoder which utilizes a set of robust encoding Kalman gains; 2) A decoder which switches its Kalman decoding gains based on its previous channel outcomes; 3)An improved Packet Loss Concealment (PLC) strategy by developing a suitable signal probability density function (pdf) at the decoder. As with the GMM-KF-ST, this robust GMM Kalman predictive coder (RB-GMM-KF) is of a low complexity, but provides much better performance than the baseline GMM-KF-ST, and the GMM-KF-RC in packet loss conditions.

The rest of the paper is organized as follows: Section 2 contains a brief review of the GMM-KF-ST. Section 3 presents the design of the Robust GMM Kalman coder; Section 4 presents simulations and Section 5 presents conclusions.

2. REVIEW OF STATIONARY GMM-KF CODER

First we define the state vector x_k based on p source vectors, each of dimension d, as $x_k = [s'_k \dots s'_{k-p+1}1]'$. Now consider the GMM-KF-ST coder [4, 5] based on the a-priori pdf (1)

$$p(x_k|Z_{k-1}^{ab}) = \sum_{j=1}^{ab} P(M_{j,k}|Z_{k-1}^{ab}) p(x_k|M_{j,k}, Z_{k-1}^{ab}), \quad (1)$$

where each Gaussian pdf $p(x_k|M_{j,k}, Z_{k-1}^{ab})$ is defined by stationary covariances $\overline{\Sigma}_i^-$ and conditional means $\hat{x}_{i,k|k-1}$. Note that $Z_{k-1}^{ab} = [z_1^{ab}, \dots, z_{k-1}^{ab}]$ denotes the set of common *absolute* measurements (defined by quantization) available to all coders, and that $M_{j,k}$ denotes the j^{th} model at time k. With each conditional density $p(x_k|M_{j,k}, Z_{k-1}^{ab})$, we associate a linear plant and a measurement model given as,

$$x_{k+1} = F_i x_k + w_{i,k}; \quad z_{i,k} = H_{i,k} x_k + v_{i,k}.$$
(2)

This work was supported in part by the National Science Foundation via Awards CCF-0347229, and CNS-0519933.

The i^{th} coder predicts the a-priori state as $\hat{x}_{i,k|k-1} = F_i \hat{x}_{i,k-1|k-1}$ where the predicted state $\hat{x}_{i,k|k-1}$ becomes the conditional mean. The prediction error $x_k - \hat{x}_{i,k|k-1}$ is Karhunen-Loeve transformed as $\psi_{i,k} = \bar{H}_i(x_k - \hat{x}_{i,k|k-1})$ (with \bar{H}_i obtained from an eigendecomposition of $\bar{\Sigma}_i^-$), and quantized to form the innovation $\hat{\psi}_{i,k}$. This is modeled by the measurement model in equation (2).

The GMM-KF-ST follows the competition and update cycle using the Kalman gains \bar{K}_i and \bar{K}_i respectvely. In the competition cycle each sub-coder forms its candidate quantization by, $\hat{x}_{i,k|k} = \hat{x}_{i,k|k-1} + \bar{K}_i(z_{i,k} - \bar{H}_i\hat{x}_{i,k|k-1})$. The coder i^* which minimizes $||x_k - \hat{x}_{i,k|k}||^2$ is chosen and only the quantization information of the *selected* coder i^* is transmitted. The *non-selected* coders $j \neq i^*$, update their aposteriori state by $\hat{x}_{j,k|k} = \hat{x}_{j,k|k-1} + \bar{K}_j(\tilde{z}_{j,k} - \bar{H}_i\hat{x}_{j,k|k-1})$, where $\tilde{z}_{j,k}$ is created based on $z_{i^*,k}$, which is the measurement of the winning coder and defines the common absolute measurement. This GMM-KF-ST provides better performance than a baseline GMM-RC predictive coder in [2] for both clean and packet loss conditions. Now we explore how to take the GMM-KF-ST, and build a more robust GMM Kalman predictive coder explicitly designed for packet loss.

3. DESIGN OF A ROBUST GMM KALMAN PREDICTIVE CODER FOR PACKET LOSS

In predictive coding based on GMMs, the codebooks are defined by the Gaussian pdfs. As the conditional means change, the location of the codebooks move in the vector space, and as the conditional probabilities change, the number of codepoints assigned to each Gaussian also changes. When packets are lost, the codebooks at the encoder and decoder are inherently mismatched, particularly the conditional means $\hat{x}_{i,k|k-1}$, and conditional coder probabilities $P(M_{j,k}|Z_{k-1}^{ab})$. To achieve better performance at the decoder, the mismatch between the encoder and decoder must be taken into account.

To obtain a more robust GMM Kalman predictive coder, we propose the following steps. First, we determine special Kalman decoding gains that are based on whether the previous packet was lost or received. This is due to the fact that the error covariance at the decoder is contingent upon whether the previous packet was received or lost. Consequently to minimize the mismatch between the conditional means at the encoder and decoder, the decoder must handle the received innovations $\psi_{i,k}$, based on these contingencies. Second, we determine special Kalman encoding gains that are based on the fact that the encoder never knows whether the transmitted innovations are received or lost. That is, we find average expected error covariances for the sub-coders at the encoder (calculated using the decoder error covariances). These average expected error covariances are utilized to obtain the special Kalman encoding gains, once again to minimize the mismatch between the encoder and decoder conditional means. Utilizing the average expected error covariances at the encoder, one can also obtain better bit allocations than in the GMM-KF-ST. Finally, a modified PLC scheme in the decoder builds a signal pdf based on the past decoded source vectors to provide enhanced PLC. To obtain the complete robust GMM-KF predictive coder, we must follow an iterative approach in which one first fixes the encoder, and finds the decoder parameters, and then fixes the decoder, and finds the encoder parameters. Through iteration, one obtains a final coder. Now let us describe these steps.

3.1. Decoder Design for the Robust GMM-KF Coder

Let the indicator variable γ_k be defined such that $\gamma_k = 1$ denotes a successful reception and $\gamma_k = 0$ denotes a packet loss at time k. Consider Figure 1 which describes different possibilities at the decoder based on γ_k and γ_{k-1} . We use 'r' to denote the outcome 'received' ($\gamma_k = 1$) and 'e' to denote an erasure or packet loss ($\gamma_k = 0$). Now let us define the possible 'modes' of the decoder based on the γ_k and γ_{k-1} as $\{\Phi_{r|r}, \Phi_{r|e}, \Phi_{e|\{r,e\}}\}$. The actual mode of the decoder at time k is denoted by \mathcal{M}_k^d where $\mathcal{M}_k^d \in \{\Phi_{r|r}, \Phi_{r|e}, \Phi_{e|\{r,e\}}\}$. For example the state $\mathcal{M}_k^d = \Phi_{r|e}$ denotes $\gamma_k = 1$ and $\gamma_{k-1} = 0$. We assume the Bernoulli loss model and loss probability as ρ . Now for each successful reception, each sub-decoder could either have been selected or non-selected at the encoder (i.e., one Gaussian mixture is selected to quantize the LSF vector s_k , the other L-1 mixtures are not selected). We denote these two events by 's' and 'n', respectively. Note that in each subdecoder's domain $r = \{s \cup n\}$. We use $m_{i,k} = s$ to indicate that the i^{th} sub-decoder is selected ('s'), and $m_{i,k} = n$ to indicate non-selected ('n'). If the innovation is lost ('e') at time k, then $m_{i,k} = e'$.

The key is to learn the fixed (stationary) error covariance matrices for the different modes of the decoder given in Figure 1. In particular, for $\Phi_{r|r}$, we need to determine $\bar{\Sigma}_{i,s|s}^$ for all the sub-coders $i = 1, \dots, L$, which corresponds to the case $m_{i,k} = s$ and $m_{i,k-1} = s$. Similarly, we must also determine, $\bar{\Sigma}_{i,n|s}^-, \bar{\Sigma}_{i,s|n}^-$, and $\bar{\Sigma}_{i,n|n}^-$. These four error covariance matrices will be used to determine corresponding Kalman decoding gains $\bar{K}_{i,s|s}, \bar{K}_{i,n|s}, \bar{K}_{i,s|n}$, and $\bar{K}_{i,n|n}$. Then, in operation, when the decoder enters mode $\Phi_{r|r}$, these Kalman gains are utilized both to create the decoded value (e.g.,, $\hat{x}_{i^*,k|k} = \hat{x}_{i^*,k|k-1} + \bar{K}_{i,s|n}\hat{\psi}_{i^*,k}$ if $m_{i^*,k} = s$ and $m_{i^*,k-1} = n$), and to update the non-selected sub-coders. Similarly, for $\Phi_{r|e}$, one must obtain the error covariances $\bar{\Sigma}_{i,s|e}^-, \bar{\Sigma}_{i,n|e}^-$ and corresponding Kalman decoding gains $\bar{K}_{i,s|e}, \bar{K}_{i,n|e}$. How do we obtain these parameters?

We fix the (stationary) encoder, and for simplicity, fix the coder conditional probabilities (and hence the bit allocations), and run the decoder as a stationary GMM decoder over a training set and under packet loss conditions. Whenever the decoder enters the mode $\Phi_{r|r}$ or $\Phi_{r|e}$, for each sub-coder *i*, it computes a set of Kalman recursions related to training the relevant error covariance parameters but completely unrelated

to the actual decoding. That is, for each i, it computes

$$\Sigma_i^s = \Sigma_{i,k|k} = \Sigma_{i,k|k-1} - \bar{K}_i (I - \bar{\Delta}_i) \Sigma_{i,k|k-1} \text{ if } m_{i,k} = s$$

$$\Sigma_i^n = \Sigma_{i,k|k} = \Sigma_{i,k|k-1} - \bar{\tilde{K}}_{i,k} \bar{H}_i \Sigma_{i,k|k-1} \text{ if } m_{i,k} = n$$

in which $\bar{\Delta}_i$ is the stationary noise factor for the i^{th} coder. Note that these Kalman recursions are based on the same principles guiding the error covariance updates given in [4, 5]. Similarly, when the decoder enters the mode $\Phi_{e|\{r,e\}}$, it computes $\Sigma_i^e = \Sigma_{k|k-1}$. By running the decoder over a representative packet loss condition, these recursions are run in each mode until they achieve roughly fixed values (i.e., depending upon the mode, the error covariance for each sub-decoder ijumps between fixed Σ_i^s, Σ_i^n , and Σ_i^e). Now let us consider how to obtain the error covariances of $\Phi_{r|r}$ (i.e., $\bar{\Sigma}_{i,s|s}^{-}, \bar{\Sigma}_{i,n|s}^{-}$, $\bar{\Sigma}_{i,s|n}^{-}$, and $\bar{\Sigma}_{i,n|n}^{-}$). In general, one follows Kalman filtering principles as in [4, 5] to obtain predictive error covariances from filtered error covariances, i.e., $\bar{\Sigma}_{i,s|s}^{-} = F_i \Sigma_i^s F_i' + Q_i$ where Q_i is the plant noise covariance matrix. In this manner, for a fixed encoder, one is able to obtain different error covariance matrices for the modes $\Phi_{r|r}$, and $\Phi_{r|e}$, and these error covariances are utilized to determine corresponding Kalman decoding gains as in [4, 5]. This procedure must be repeated for each fixed encoder. Now a natural question is given the decoder defined as in Figure 1, how does one design an appropriate encoder?



Fig. 1. State Transition Diagram at the Decoder

3.2. Encoder Design for the Robust GMM-KF Coder

Suppose we have a fixed decoder based on the error covariances and Kalman decoding gains as described in Figure 1 and in the previous subsection. To design an encoder for this decoder, one must take into account the fact that the encoder never knows which mode the decoder is in, as the encoder never knows whether a transmitted innovation is received or lost. To minimize the mismatch between the encoder conditional means, and to achieve more robust encoding, the encoder modifies its GMM coding error covariances by setting them to expected error covariances, calculated using the decoder error covariances.

In particular, for the fixed decoder, the encoder, for each sub-coder *i*, can calculate expected decoder error covariances $\bar{\Sigma}_{i,s}^{E}$ and $\bar{\Sigma}_{i,n}^{E}$, corresponding to whether the sub-coder *i* is selected or non-selected, i.e., $\bar{\Sigma}_{i,s}^{E} = (1-\rho)\Sigma_{i}^{s} + \rho\Sigma_{i}^{e}$ and $\bar{\Sigma}_{i,n}^{E} = (1-\rho)\Sigma_{i}^{n} + \rho\Sigma_{i}^{e}$. These expected decoder error covariance matrices are based on the uncertainty regarding packet reception or loss, and are used to determine Kalman

encoding gains. Then for each sub-coder i, we obtain new unconditional error covariances $\bar{\Sigma}_i^E$. We find these unconditional error covariances by probabilistic weighting of $\bar{\Sigma}_{i,s}^E$, and $\bar{\Sigma}_{i,n}^E$, in which the probabilities are obtained based on the sub-coder probabilities (i.e., how often the sub-coders are used in practice over training data). The new fixed $\bar{\Sigma}_i^E$, are used to define the new bit allocations which now take into account the uncertainty regarding the reception or loss of a transmitted innovation.

In general, for each re-design of the encoder, the decoder must be re-designed, and vice-versa, until the encoding and decoding parameters achieve roughly fixed values. Now let us examine how the packet loss concealment is designed.

3.3. Packet Loss Concealment

Consider the decoder at mode $\Phi_{e|\{r,e\}}$. As the innovation is lost in this mode we need to employ an effective PLC scheme. In [5], in the event of packet loss the decoder probabilistically combined the sub-coder state predictions to create a synthetic value. That is, to conceal a lost s_k , the decoder computed $\sum_{i}^{L} P(M_{i,k}|Z_{k-1}^{ab})\hat{x}_{i,k|k-1}$ from which a synthetic s_k substitution is obtained. However, this PLC combines all the subcoder state vectors, even when some of the state-vectors are not particularly accurate with regards to the actual previously decoded value. We improve PLC in the GMM-KF system by only using the actual decoded values. Let g_k be the actual decoder value (i.e., the one that is used in the speech decoder in signal synthesis) at time k. That is, $g_k = [\hat{x}_{i^*,k|k}]_d^{\uparrow}$, (top d elements of the vector) in which sub-coder i^* is the sub-coder used at time k.

The decoding mode $\Phi_{e|\{r,e\}}$ is continually creating a conditional signal GMM $p(x_k|G_k)$ based on all past decoded source vectors where $G_k = \{g_k, g_{k-1}, ..., g_1\}$. In particular, we create $p(x_k|G_k) = \sum_{i=1}^{L} P(M_{i,k}|G_k)p(x_k|M_{i,k}G_k)$. Once again, utilizing Kalman filtering principles, we consider a linear plant and measurement model associated with conditional pdf $p(x_k|M_{i,k}G_k)$ as,

$$x_{k+1} = F_i x_k + w_{i,k}; \quad g_{i,k} = \prod_i x_k + \nu_{i,k}$$

The terms F_i are identical to the ones given in (2). We see the measurements as noisy observations of the first d elements of the state x_k . We use the notation $\breve{x}_{i,k|k-1}$ and $\breve{x}_{i,k|k}$ to denote state estimates for each sub-coder in the PLC mode $\Phi_{e|\{r,e\}}$ as they are distinct from the states used at the encoder, and in the other decoding modes $\Phi_{r|r}$, and $\Phi_{r|e}$.

Let $\Pi_i = [I_{d \times d} \ \mathbf{0}]$, where I denotes the identity matrix. One can show that $E(\nu_{i,k}\nu'_{i,k}) = E([g_k - [x_k]_d^{\uparrow}][g_k - [x_k]_d^{\uparrow}]') = [\bar{\Sigma}_{i^*s|m_{i,k}}]_{d \times d}^{\uparrow}$ (top left $d \times d$ matrix), where $m_{i,k} \in s, n, e$.

By utilizing the Uncorrelated Noise Model (UNM) assumption ([4]) based Kalman filtering equations ([4]), we update the a-posteriori states, $\check{x}_{i,k|k-1}$ and coder probabilities $P(M_{i,k}|G_{k-1})$. Now we can perform the PLC for time k by calculating $\sum_{i=1}^{L} P(M_{i,k}|G_{k-1})\check{x}_{i,k|k-1}$ given $\gamma_k = 0$.

Now we summarize the operation of the Robust GMM-KF coder as follows. In encoding s_k , the encoder must keep track of $m_{i,k}$ and $m_{i,k-1}$ (i.e., whether the sub-coder i is used to code s_k , and whether it was used to encode s_{k-1} . Based on these $m_{i,k}$, and $m_{i,k-1}$ values, each sub-coder updates its state equation using the special Kalman encoding gains. The quantized innovation is transmitted to the decoder. Depending upon the pattern of receptions and losses, the decoder is in one of 3 modes as illustrated in Fig. 1, and the decoder uses the appropriate Kalman decoding gains if the decoder is in modes $\Phi_{r|r}$, or $\Phi_{r|e}$. If the packet is lost, the decoder is in mode $\Phi_{e|\{r,e\}}$, in which the PLC is performed. In this manner, the encoder and decoder use different Kalman gains in handling the quantized innovations, and as we shall see, this leads to better performance in packet loss conditions. The overall system is still of a low complexity similar to that of the GMM-KF-ST, and GMM-RC.

4. SIMULATIONS

We compare the performance of this Robust GMM-KF predictive coder (RB-GMM-KF) to the baseline stationary GMM-KF-ST as well as a memoryless GMM non-recursive coder (GMM-NRC) [6] in LSF quantization and transmission in packet loss conditions. Training and Testing was based on 100000, and 30000 vectors, respectively.

Table 1 shows the LSD performance of the RB-GMM-KF, GMM-KF-ST and GMM-NRC. We can clearly see the the RB-GMM-KF consistently outperforms the baseline GMM-KF-ST coder in average LSD and outliers (recall that [4, 5] showed that the GMM-KF-ST substantially outperformed a baseline GMM-RC based on [2] in packet loss conditions; we do not show the results of the GMM-RC for packet loss due to space reasons). It is interesting to note that the proposed RB-GMM-KF achieves comparable performance to the GMM-KF-ST at a savings of 2 or more bits. The non-predictive GMM-NRC performs slightly better in outlier performance but the proposed RB-GMM-KF coder achieves slightly better average LSD performance. Table 2 shows an SNR comparison. Since the RB-GMM-KF can be used for vector data other than LSFs, it is interesting to see that the RB-GMM-KF coder outperforms all other coders in SNR.

5. CONCLUSION

In this paper, we have shown how to take a stationary GMM-Kalman predictive coder, and modify it for operation in packet loss conditions. This proposed RB-GMM-KF minimizes the mismatch between the encoder and decoder codebooks by using fixed sets of specially designed encoding and decoding Kalman gains which are obtained offline. The simulation results show RB-GMM-KF outperforms its baseline and shows comparable LSD performance to that of a memoryless GMM-NRC. Therefore, this paper provides one method for taking a GMM predictive coder, and making it more robust to packet loss.

	Packet loss 5%			Packet loss 10%			
Bits	LSD	LSD outlier %		LSD	LSD outlier %		
		2-4dB	>4dB		2-4dB	>4dB	
		RB-GMM-KF					
20	1.24	6.55	2.56	1.55	11.77	5.73	
22	1.13	5.61	2.69	1.50	10.41	5.82	
24	1.06	5.55	2.45	1.40	9.47	5.69	
26	0.99	5.53	2.45	1.29	9.01	5.43	
	GMM-KF-ST						
20	1.39	8.63	5.31	1.72	14.21	7.54	
22	1.22	8.05	4.36	1.63	13.45	7.48	
24	1.17	7.66	4.58	1.50	11.56	7.03	
26	1.14	7.26	5.21	1.43	10.94	7.05	
	GMM-NRC						
20	1.36	6.24	2.53	1.62	10.78	5.34	
22	1.26	5.61	2.48	1.55	9.94	5.27	
24	1.14	4.73	2.51	1.46	8.52	5.27	
26	1.14	4.36	2.37	1.43	8.14	5.19	

Fable 1. LSD performance of	RB-GMM-KF
-----------------------------	-----------

	RB-GMM-KF		GMM-KF-ST		GMM-NRC	
Bits	5%	10%	5%	10%	5%	10%
20	34.79	32.50	33.84	32.01	33.46	31.54
22	35.02	32.61	34.12	32.06	34.08	32.23
24	35.32	32.71	34.46	32.26	34.02	32.01
26	35.57	32.87	34.52	32.45	34.18	32.04

Table 2.	SNR	(in	dB)	performance	comparison
----------	-----	-----	-----	-------------	------------

6. REFERENCES

- J. Samuelsson and P. Hedelin, "Recursive coding of spectrum parameters," *IEEE Trans. on Speech and Audio Proc.*, vol. 9,(5), pp. 492 – 503, 2001.
- [2] A. D. Subramaniam, W. R. Gardner, and B. D. Rao, "Lowcomplexity source coding using gaussian mixture models, lattice vector quantization, and recursive coding with application to speech spectrum quantization," *IEEE Trans. on Audio, Speech & Lang. Proc.*, vol. 14,(2),2006.
- [3] E. R. Duni and B. D. Rao, "High-rate optimized recursive vector quantization structures using hidden markov models," *IEEE Trans. on Audio, Speech, and Language Proc.*, vol. 15(3), pp. 756 – 769, 2007.
- [4] S. Subasingha, M. N. Murthi, and S. V. Andersen, "Gaussian mixture kalman predictive coding of lsfs," in *IEEE Int. Conf.* on Acoustics, Speech, and Signal Proc., April 2008, pp. 4777– 4780.
- [5] S. Subasingha, M. N. Murthi, and S. V. Andersen, "Gaussian mixture kalman predictive coding of line spectral frequencies," *IEEE Trans. on Audio, Speech and Language Proc.*, 2008, Accepted for Publication.
- [6] A. D. Subramaniam and B. D. Rao, "Pdf optimized parametric vector quantization of speech line spectral frequencies," *IEEE Trans. on Speech and Audio Proc.*, vol. 11(2), pp. 130–142, March 2003.