

Global Maximum Likelihood Voice Decoding

Richard A. Dean

Dept. of Defense, Ft. Meade, Md. 20755

ABSTRACT

This paper presents research on a new technique for improving speech coder performance in the presence of errors. As speech coding is used in more communications systems to enhance bandwidth and channel performance, the error performance of the coder takes on more significance. This paper shows that significant performance advantages can be achieved by combining conventional maximum likelihood decoding with the context within the speech signal to make global decisions. The Hidden Markov Model is the most popular technique for capturing this underlying structure. In this research the HMM is integrated into a combined channel and speech decoder. A theoretical development of a Global Maximum Likelihood decoder is presented. Results of testing in channel errors is presented which demonstrate enhanced error performance.

INTRODUCTION

Attention to channel performance is appropriate as digital encoding of voice is finding its way into many new applications where error control is important. In the past digital voice was accepted as a necessary inconvenience associated with encryption of speech or associated with digital telephony. In these cases the bandwidth expansion and expense of digital speech was an acceptable price for the associated service. Today's advances in speech coding encourage the further introduction of digital voice into communications systems because of the enhancement possible for the communications grade of service and for voice quality enhancement [1,2].

This new era in voice communication requires a fresh look at the techniques used for the design of the system for error control. Digital speech has unique properties that offer the potential for improved

error performance but good speech quality also requires minimum delay. This makes the design tradeoffs quite different from classical digital communications. Exploring these possibilities requires an encompassing look at voice coding, digital communications, and channel effects. Today these disciplines, as applied to digital voice systems, are disjoint. Current designs apply error control techniques that are proven for data applications with few, if any, modifications.

A broader perspective for the design of digital voice into a communications system can be achieved by viewing voice data compression and error control as part of a continuum. Voice coding and vector quantization can be viewed in terms of Rate Distortion Theory. The Rate Distortion measure $R(\delta)$ is an effective measure for data compression because it provides an estimate of the required data rate R as a function of the entropy of source U and of $U|Y$ and as a function of the allowable distortion $E(d)$. It is expressed as:

$$R(\delta) = \lim_{k \rightarrow \infty} \frac{1}{k} \cdot \{ \text{MIN} [H(U) - H(U|Y)] : E(d) \leq k\delta \} \quad (1)$$

The rate distortion function is produced by a search over source codebook A_u and receive codebook A_v for the best match in the transmit and receive codebook to maximize the entropy $H(U|Y)$, the average information provided about $\{U\}$ from $\{Y\}$, in the region where the distortion $E(d)$ is below the level δ .

A similar expression is available for channel distortion. The case of Phase Shift Keyed modulation will be considered as it is a common scheme for radio and wireline applications. Here the probability of channel error, can be expressed as a function of rate R as

$$P_E = .5e^{-\frac{(V^2/2^{R-1}N_0)}{2}} \quad (2)$$

where the V^2 is the signal power, R is the number of bits per symbol, and N_0 is the usual channel noise parameter. The selection of rate can then be seen as a trade off between channel distortion and speech distortion.

Treating digital voice as a form of data communications has been a convenience for digital communication designers. A rich and powerful inventory of tools from Coding and Information Theory are available to accommodate errors from most channels. Separating the problem into the classical disciplines of source coding and channel coding, designers have solutions for most applications. Shannon's channel coding theorem leads to a typical design for error control on a burst channel as in Figure 1. However when channels have burst errors the delay associated with interleaving coupled with the coding delay can be intolerable for natural voice communications. This dilemma is presented graphically in Figure 2 for an HF application. Note that error performance improves with delay while voice performance degrades resulting in no region of satisfactory performance.

GLOBAL MAXIMUM LIKELIHOOD DECODING

Improvements to speech decoding in the presence of errors come through an extension of Maximum Likelihood Estimation (MLE) techniques. A likelihood function is proposed by extending the likelihood function over a sequence of channel data $\{V\}$ and associated MLE decision sequence $\{Y\}$. In this case Y is a discrete random variable from the VQ codebook set A_y and V is the continuous random variable corresponding to the channel signal and noise. We also introduce the discrete random variable Z_n , also over the set A_y , which corresponds to the element of the set A_y that maximizes the likelihood function $L(\{Y\}, \{V\})$ over the joint region of $\{Y\}$ and $\{V\}$.

$$Z_n = \text{MAX}[L(\{Y\}, \{V\})] \quad (3) \\ \text{over } A_y$$

Now in the case where the Y_n are correlated and the V_n are independent

a likelihood function for the global decision can be developed as:

$$\text{MAX}[P(Z_n|\{Y\}) \cdot P(\{Y\}|\{V\})] \quad (4) \\ \text{over } A_y$$

This structure enables the incorporation of a probability filter $P(Z_n|\{Y\})$ into the likelihood function which when paired with the channel data $\{V\}$ narrows the uncertainty of the decision. The probability $P(Z_n|\{Y\})$ provides the additional context for the speech vector sequence $\{Y\}$. This filter can be readily developed from the Hidden Markov Model (HMM)[3] which has been shown to be an effective stochastic model for speech. The HMM enables a variety of structures that can be used in the decoding process. One of these structures is presented in Figure 3 where a sequence of speech VQ vectors $\{Y\}$ are converted into speech state decisions $\{X\}$ from which the probability filter $P(Y=y_i|X=x_j)$, directly available from the HMM, can be used in the likelihood function. The state is a phoneme like event with very low entropy relative to the channel data. Correct state decisions can be expected even in the presence of large errors. The state decision can then be used reliably to enhance the vector decision $Z_n=y_i$, the global most likely decision. In this case the probability filter $P(Z=y_i|\{Y\})$ can be developed by the introduction of the probability of state X_n as:

$$P(Z=y_i|\{Y\}) = P(Z=y_i|X=x_k) \cdot P(x_k|\{Y\}) \quad (5)$$

The first term is obtained directly from the HMM observation probability matrix. The second component is provided directly from the HMM state computation [3]. Then combining the speech and the channel components into the Global MLE for Z results in:

$$\text{MAX}[P(Z=y_i|X=x_k) \cdot P(X=x_k|\{Y\}) \cdot P(V_n|y_i)] \quad (6) \\ \text{all } y \text{ in } A$$

EXPERIMENTAL RESULTS

The testbed developed for this research is shown in Figure 4. The testbed incorporates a complete simulation of a very low rate voice

communication system. An LPC Vector Quantizer (LPCVQ) is used for voice coding. The resulting voice spectrum vectors $\{Y\}$ are passed through a channel simulator that introduces the effects of random and burst noise. A special decoding operation is performed which includes likelihood estimates of these decisions for use in the global decoder. A Global MLE decoder then incorporates both channel and speech data into a composite decision on the VQ vector Z . This is followed by synthesis of speech using a VQ decoder and LPC. The testbed also incorporates a Hidden Markov Model (HMM) based on the vectors produced by the LPCVQ. A 64 state, 1024 observation HMM is trained on 11 minutes of speech data.

Preliminary testing has been performed on a global decoder in the testbed shown in Figure 4. Global decoding was performed on received channel vectors with errors rates ranging from .2% to 3%. These results demonstrate that global decoding enhances performance in errors in two ways. The global decoder detects and corrects approximately 30% of the channel errors. In addition the global decoder detects and enhances approximately 60% of the erroneous vectors. This effect is the byproduct of correctly identifying the correct HMM state even when the vector decision is incorrect. While raw error performance is indicative, the most significant measure of performance is the resulting speech distortion. This is computed directly by comparing the distortion of the decoded vectors against the input spectrum. In this case a distortion measure using Line Spectral Pairs was used as below:

$$D(i,j) = \sum_{k=1}^n (W_k \cdot [LSP_i(k) - LSP_j(k)]^2) \quad (7)$$

where W_k is a weighting function

A plot of the cumulative distortion is shown in Figure 5. The cumulative distortion of the global decoder is about 25% of the conventional decoder in channel errors. Comparing distributions of the spectral distortion is also instructive.. The variance of the global decoder spectral distortion is clearly superior.

CONCLUSIONS

A framework for improved speech decoding in the presence of channel errors has been presented. Global MLE decoding can be used to enhance voice coder performance on a variety of channels without the overhead or delay associated with classical encoding techniques. The early results presented here offers encouragement to the potentially significant performance advantages.

- [1] T.E. Tremain et al., "A 4.8 kbps Code Excited Linear Predictive Coder", Proc. Mobile Satellite Conf., May, 1988, pp491-496.
- [2] J. Rothweiler, "Low Rate Voice Coder for HF ECCM Applications", Military Speech Tech 88, San Francisco, Ca, Oct 88, pp213-217.
- [3] L. Rabiner, B. Juang, "An Introduction to Hidden Markov Models", IEEE ASSP Magazine, Jan. 1986, pp4-16.

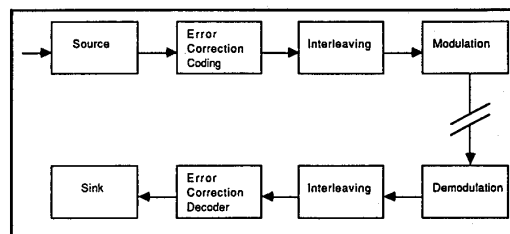


Figure 1: Classical Digital Communication System

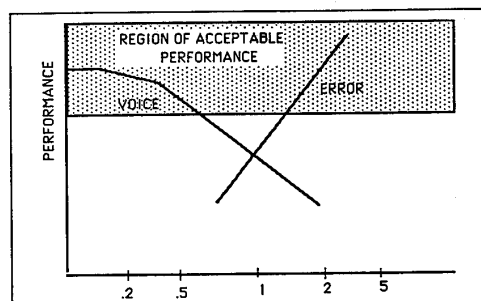


Figure 2: Tradeoff of Voice and Error Performance Against Delay

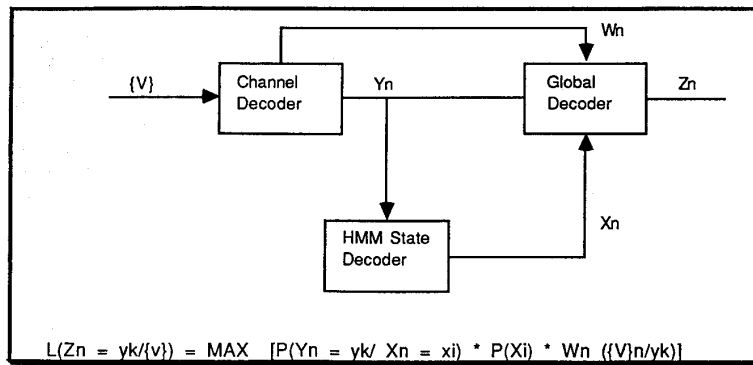


Figure 3: State Based Global Decoder

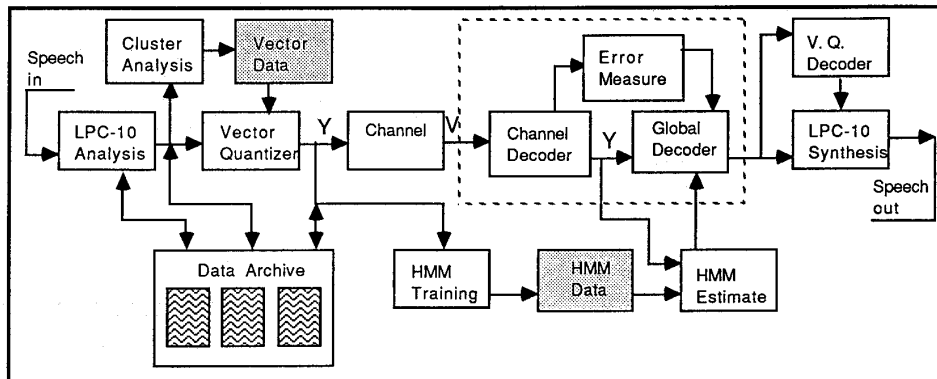


Figure 4: Global Decoder - Testbed Block Diagram

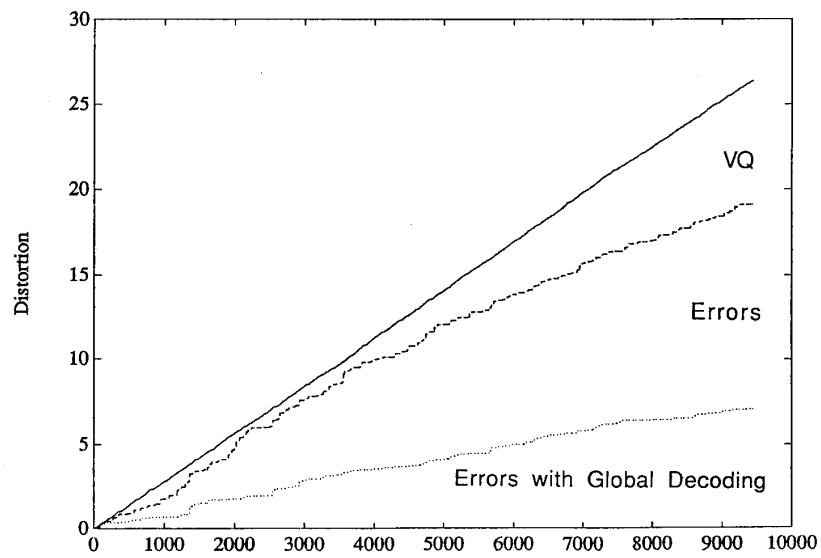


Figure 5.: Cumulative Distortion Comparison