Types-Based Data Embedding for ITU G.711[⚲]

Mark G. Kokes and Jerry D. Gibson
Multimedia Communications and Networking Laboratory
Department of Electrical Engineering
Southern Methodist University
Dallas, Texas, USA
E-mail: {kokes, gibson}@seas.smu.edu

Abstract **In this paper, we expand on a new technique for embedding digital information into G.711 encoded speech signals. Using the *method of types* as a tool to analyze the statistical nature of a digital speech signal, we demonstrate that our embedding scheme is capable of embedding up to 1.6 Kbps of additional information at an average embedded error rate of 10^{-4}. We are able to achieve these embedded rates by not requiring the data to be either *hidden* or decoded *error free*. This additional bandwidth can be used for various low data rate applications. We offer such a scheme as a possibility for use in other existing telecommunications links, both wired and wireless, for the purpose of rate enhancement without changing allocated bandwidths or source compression methods.**

I. Introduction

The field of information hiding contains several subfields [1], including steganography, where a message is concealed in another data stream, and watermarking, where ownership data is included in digital objects to be protected. A third subfield is the area of data embedding, wherein additional information is incorporated in a transmitted data stream by using a key and distorting the original object slightly. The embedded information cannot be reconstructed without the use of the key.

We propose an approach to data embedding that is based upon the *method of types* [2] and *universal classification*. In this approach, a second stream is embedded within a primary host stream without an increase in overall transmitted data rate. The embedded data is extracted using a type-based universal receiver [3, 4, 5], without the use of a key. The choice of type and rate for the embedded stream is based upon analysis of portions of the host stream. The universal receiver learns the embedded type from the received data alone, and hence, there is no side information as in previous data embedding techniques. The embedding process and the receiver are both data adaptive, so the host stream can be reconstructed without error.

Two important differences between previous work in data hiding and the proposed research are that we do not seek to hide the embedded data from other users and that we require the host stream to be decoded *error free*. The overall goal is to increase the effective received data rate without increasing the transmitted data rate. At the outset, we do not preclude the case where the embedded stream may be decoded with errors. However, it is envisioned that in many applications, by suitable choice of the encoded types with respect to each frame, the embedded stream can be decoded essentially error free. Using G.711 [6] as a test case, one goal of this research is to investigate the tradeoffs between maximizing the embedded data rate and keeping the error rate in the reconstructed data stream at an arbitrarily small level.

The conceptual steps of the approach are as follows. The host stream to be transmitted is analyzed to determine possible inherent data types. Modifications to these types are established which can be used to transmit the embedded data. These modifications must be accurately detectable by a type-based receiver. For each frame of host data, the data type is modified in such a way to represent the embedded content. A universal receiver operating on the received data extracts the type representing the embedded symbol and both streams are processed and sent to the user.

We investigate this universal approach to data embedding by identifying the intrinsic characteristics of a G.711 stream that facilitate data embedding. Because of the time-domain waveform following nature of G.711, certain traits are guaranteed. We study techniques for selecting the embedded data types that will allow the highest rates for the embedded stream. We examine the tradeoffs between embedded rate and errors in the embedded content. We illustrate this approach to data embedding for G.711 but offer the possibility that this approach can be used to expand the delivered data rate of other existing telecommunications links, both wired and wireless, without changing allocated bandwidths or source compression methods.

II. Data Embedding and Universal Classification

Our data embedding approach consists of the following conceptual steps [7, 8] (see Fig. 1). The host data stream to be transmitted is analyzed to determine the data types that commonly occur in the stream. Modifications to these types are then determined that can be used to send the embedded data and that can be accurately detected by type-based universal receivers. Then, for each individual frame of host data to be transmitted, the data type is modified in such a way to represent the embedded data. The universal receiver operates on the received data stream and extracts the data type that represents the embedded data symbols. The embedded data stream is then decoded and sent to the user. After removing the modifications to the received data sequence due to the embedded data, the host data can be decoded. Primarily, we expect applications where the host

[⚲] This research was supported, in part, by NSF Grants NCR-9796255 and CCR-0093859.

data streams are generated by lossy data compression schemes.

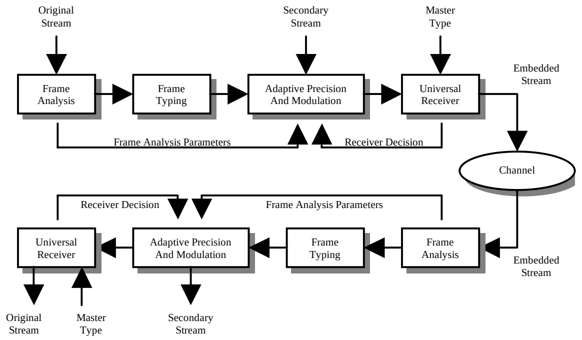

Fig. 1: Data Embedding System Block Diagram

Embedding data into digital signals can be thought of as transmitting information over a communication channel that is corrupted by *strong* interference and channel effects. Such a model for the case of a binary communication system is given as,

$$\begin{aligned} H_0 &: s_0 + \eta(t), \text{ Symbol 0 Transmitted} \\ H_1 &: s_1 + \eta(t), \text{ Symbol 1 Transmitted.} \end{aligned} \tag{1}$$

In this model, a data symbol is hypothesized (i.e. H_x) to be transmitted from one of two sources. The binary data symbol to be transmitted, s_x, corresponds to the data symbol that is to be embedded into the host signal, $\eta(t)$. The strong interference is representative of the host signal.

More generally, consider the following M-ary hypothesis testing problem:

$$\begin{array}{lll} H_1 : & X^n \sim P_1 & \text{Source 1} \\ H_2 : & X^n \sim P_2 & \text{Source 2} \\ & \vdots & \vdots \\ H_M : & X^n \sim P_M & \text{Source M} \\ H_{M+1} : & & \text{Rejection Region} \end{array} \tag{2}$$

where the test vector X^n is of length n. We assume that under hypothesis H_i, the test vector, X^n, is generated by a source with probability measure P_i (unknown to the detector). In addition, due to the absence of an accurate statistical model for the M sources, we assume that there exist training vectors T_i^N, $i = 1,2,...M$ of length N from each of the M possible data sources. Therefore, the classification between source types is made on the basis of the test vector, X^n, and the training vectors, T_i^N, $i = 1,2,...M$.

It has been shown that the asymptotically optimal Generalized Likelihood Ratio Test (GLRT) for determining if a finite alphabet test sequence, X^n, arose from the same source as a finite alphabet training sequence, T_i^N, is:

$$h(X,T_i) = \frac{1}{n}\log\left\{\frac{\sup_{Q_1,Q_2} Q_1(X^n)Q_2(T_i^N)}{\sup_Q Q(X^n,T_i^N)}\right\}, \tag{3}$$

where Q_1, Q_2, and Q denote source densities [9, 10].

From an intuitive point of view, one can see that if the data sequences X^n and T_i^N arise from the same source, then h_i will converge to zero in the limit. Alternatively, if the data originated from different sources, then h_i will converge to some constant greater than zero which will allow for discrimination between the proposed M hypotheses. It was originally shown by Gutman [11] for the classification problem that this test offers asymptotically optimal performance over a very wide range of source statistics.

Unfortunately, due to the requirement of the supremum calculations in (3), the detector is not practical to implement. However, through the use of the method of types, the log-likelihood ratio is reduced to

$$\begin{aligned} h_i(X,T_i,\lambda) = \; & d_{KL}\{Q_{(X^n)}, Q_{(X^n,T_i^N)}\} \\ & + \frac{N}{n} d_{KL}\{Q_{(T_i^N)}, Q_{(X^n,T_i^N)}\} - \lambda. \end{aligned} \tag{4}$$

The quantities $Q_{(X^n)}$, $Q_{(T_i^N)}$, $Q_{(X^n,T_i^N)}$ and represent the types of the data vectors, T_i^N, X^n, and the concatenated vectors (X^n, T_i^N). These types represent the empirical (*histogram*) estimates of the statistics and joint statistics of the data vectors. The distance metric is the functional d_{KL}, the well known divergence or relative entropy between the probability mass functions in its argument. λ is a positive constant chosen to satisfy some design criterion (i.e. rejection region). In addition to the above, we offer an alternative interpretation for $h_i(X,T_i,\lambda)$ in terms of the entropies of the types,

$$\begin{aligned} h_i(X,T_i,\lambda) = \; & \frac{N+n}{n} H\{Q_{(X^n,T_i^N)}\} - H\{Q_{(X^n)}\} \\ & - \frac{N}{n} H\{Q_{(T_i^N)}\} - \lambda. \end{aligned} \tag{5}$$

The above expression for the discriminant function in terms of the entropies is computationally preferable for online processing as the entropies of the training sequences can be pre-computed. Note that the joint type of X^n and T_i^N in terms of the marginals is defined as

$$Q(X^n,T_i^N) = \frac{nQ_{(X^n)} + NQ_{(T_i^N)}}{n+N}. \tag{6}$$

III. Analysis

Since our approach is data-adaptive, we wish to analyze each sequence of host data to determine if an embedded data stream can be accommodated without substantially compromising the host data. Thus, our classification problem is the M-ary hypothesis problem with rejection [11], where the rejection zone is used for the "no embedded data" case. The number of bits embedded per host data sequence is $log_2\{M\}$.

We do not wish to send any side information, so the first issue to be addressed is under what conditions can an embedded data stream be successfully decoded from the received data stream. More specifically, if we embed $log_2\{M\}$ bits in the host sequence such that the probabilities of falsely decoding embedded hypothesis H_i as one of the other hypotheses, H_j $(j = 1,2,...,M, j \neq i)$, exponentially decreases in n (the host data sequence length) with parameter λ, what can we say about the probability of correctly decoding under the M hypotheses? From [11], we know that if the training sequence N is of insufficient length with respect to n, then there exists an hypothesis H_j such that the probability of choosing rejection given H_j (decoding no embedded data given H_j) approaches 1 as $n \rightarrow \infty$. However, for a sufficiently long training sequence (length N) with respect to n as $n \rightarrow \infty$, the probability of choosing the rejection region under H_j is bounded away from 1.

Since our approach is host data sequence adaptive, these results imply that by adaptively varying the number of bits embedded ($log_2\{M\}$) per host sequence, the receiver will be able to track the data embedding process at the encoder with high probability, and without the transmission of side information.

A second issue concerns how to modify the data type of the host data such that data can be embedded and in such a way that the receiver can determine the modified data type from the received data stream only; that is, without side information. For a given host data sequence to be transmitted, we consider the case where the minimum entropy data type is determined and this minimum entropy data type is modified by shifting within the region of support of the class of data types. We defer justification of the minimum entropy data type as the type to be modified and analyze the process of data embedding via simple shifts of this type. Note that the number of shifts corresponds to the number of hypotheses that must be detected with the universal receiver. We will only consider symmetrized, unimodal type classes in this development.

We know that the optimal likelihood ratio test from the Neyman-Pearson lemma can be written as the difference between two relative entropies [2]. Thus, if we embed data by shifting the symmetrized data type, the number of hypotheses that can be distinguished will be dependent upon the spread of the type class and on the region of support. For $M = 2$, there are three different errors that can occur: (i) Given that H_1 or H_2 has been sent, the detector may decide "no embedded data" and reject both; (ii) Given that H_1 is sent, the detector decides H_2; and (iii) Given that H_2 is sent, the detector decides H_1. Stein's lemma says that we can fix one of these error probabilities at some suitably small value and the others can be made to approach zero exponentially with respect to the relative entropy between hypotheses [2]. However, in our situation, all of these errors may be of equal significance. What is needed is to select the shifts of the minimum entropy data type to obtain equal probabilities of making an error given that any hypothesis or the "no data" case is sent. Thus, we can use a Bayesian approach with specified *a priori* probabilities on the hypotheses, say π_1, π_2, and π_R, and use Sanov's Theorem to bound the error probabilities with respect to the nearest neighbor regions [2].

Once the minimum entropy type has been determined, data can be embedded by constructing hypotheses other than shifts of this data type, such as in [12]. These cases are also being investigated.

IV. Results

Of fundamental importance to type-based data embedding is the fact that this is a lossy approach. By removing the many constraints (i.e. perceptual) in the typical data embedding problem, we plan to embed information in a host signal in such a way that the throughput of the channel is increased without also increasing the transmitted data rate. In order to achieve this additional rate, we are willing to accept a small number of errors in both the original and embedded streams as long as these errors do not significantly affect the quality of the original data stream. We stress the fact that this approach to data embedding is not concerned with attacks or secret key information. This approach focuses on rate enhancement.

In this section, we provide detail regarding the asymptotic analysis in the previous sections. We discuss results regarding the relations between the lengths of the training sequences (N), the lengths of the host sequence (n), and the number of bits embedded in a particular host frame ($log_2\{M\}$). We discuss the amount of distortion associated with making errors in detecting the correct embedded precision and symbols. We also suggest ways to compensate for such errors.

To begin the data embedding process, one must have an understanding of the master type inherent within the original data stream. The master type for G.711 is shown in Fig. 2. This type can be ascertained by observation of a typical G.711 codeword sequence over a reasonable amount of time. The resulting data type often requires some sort of 1:1 mapping in order to obtain the uni-modal characteristic that is conducive to minimal error detection using a shift-based modulation/embedding scheme. This information is key to the detection process for it is shifted versions of the master type that are used to comprise the training data types. So what is a reasonable amount of time over which

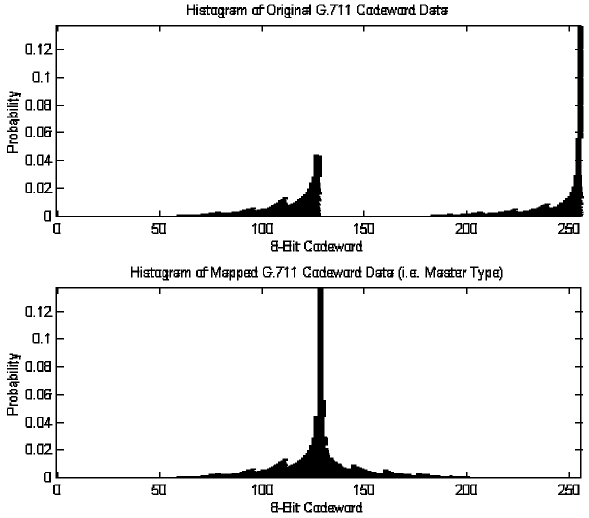

Fig. 2: G.711 Master Type (i.e. Inherent Data Type)

to formulate a master type? In [3], Stolpman suggests that β range from 10^3 to 10^4 (i.e. $\beta = N/n$, the ratio of the length of the training data sequence to the length of the data sequence). Outside of this range there is typically no additional gain in performance for a particular master type. Recall that n is directly proportional to the embedded data rate for a given sequence. In this work, we follow these predetermined guidelines for master type construction. Experimenting with these results, we verified that, on average, increasing β did not significantly affect the detection performance of the embedding system.

Of vital importance to the embedded rates achieved and for that matter, the embedded error rates as well, is the selection of the frame processing length. In the trials described in this paper, n ranges from 4 to 30 depending on the target embedded data rate. For higher target rates, the value of n should decrease. In these trials, the value of n is held constant over the particular speech segments being processed. Suggestions for future work include altering n on a frame-by-frame basis via voice activity detection (VAD) or using a simple spreading measure on the current data type to determine an appropriate value of n for best probability of detection. In any case, the value of n will depend on the variance of the master type produced from the source compression algorithm from which the original data stream in being generated.

Determining the number of bits to embed on a frame-by-frame basis contributes significantly to the overall average data rate achievable for a particular speech segment. In the previous section, we suggest that the receiver is able to track the data embedding process at the encoder/decoder (i.e. decoding is present at the encoder) with high probability, and without the transmission of side information. This is made possible by the use of an intermediate type, which we call the *minimum entropy type*. This data type can be formulated at both the encoder and decoder and is shift invariant. The property of shift (i.e. modulation) invariance is fundamental to the calculation of this data type. We use an entropy measure and thresholding procedure on this intermediate type to determine the number of bits to be embedded in the current data frame.

Of particular interest to us is the achievable embedded rates and error rates associated the above mentioned process. Table I demonstrates results from our G.711 trials for 30-second speech samples simulating typical human conversation. We show that we can embed up to an additional 2% (i.e. 1.5 Kbps) of the host stream while maintaining a minimal effect on the original data. Errors in the host stream sound "click"-like in nature and are instantaneous in the sense that they do not linger on in time. This is due to the insignificant delay associated with G.711 speech coding. It is likely that such errors in the host stream can be corrected by the introduction of a slight delay in the data embedding decoder. Such corrections in the host data stream can be accomplished because of the time domain waveform following nature of the G.711 codec. The corrected host stream can then be utilized to adjust for any additional errors detected in the embedded stream as well. This additional processing is currently being explored and could significantly further lower the error rates associated with both the embedded and host data streams and consequently allow us to increase the embedded rates for a desired probability of error.

Table I. Average Embedded Data and Error Rates for G.711

Embedded Data Rate	Embedded Error Rate
1.5 Kbps	10^{-4}
3.2 Kbps	10^{-3}
9.6 Kbps	10^{-2}

References

[1] S. Katzenbeisser and F. Petitcolas, eds., *Information* Hiding: Techniques for Steganography and Digital *Watermarking*, Artech House, Boston, 2000.

[2] T. Cover and J. Thomas, *Elementary Information Theory*, John Wiley & Sons, Inc., New York, 1991.

[3] V. Stolpman and G. Orsak, "Type-Based Receiver For Wideband CDMA," *Proc. of IEEE Wireless Comm. and Networking Conf.*, pp. 1470-1474, Sept., 1999.

[4] S. Paranjpe, V. Stolpman, and G. Orsak, "A Training Free Empirical Receiver For QAM Signals," *Proc. of IEEE Wireless Comm. and Networking Conf.*, pp. 221-225, Sept., 1999.

[5] V. Stolpman, S. Paranjpe, and G. Orsak, "A Blind Information Theoretic Approach To Automatic Signal Classification," Proc. of MILCOM, pp. 447-451, Nov., 1999.

[6] ITU-T G.711, "Pulse Code Modulation (PCM) of Voice Frequencies,", Nov., 1988.

[7] M. Kokes and J. Gibson, "The Method of Types and Lossy Data Embedding," *IEEE DSP Workshop*, Oct., 2000.

[8] M. Kokes, et al, "Embedding Information Into Digital Representations of Signals," *Fifth World Conference on Integrated Design and Process Technology*, June, 2000.

[9] V. Poor, *An Introduction to Signal Detection and Estimation*, New-York: Springer-Verlag, 1988.

[10] O. Zeitouni, J. Ziv, and N. Merhav, "When is the Generalized Liklihood Ratio Test Optimal?," *IEEE Trans. Inform. Theory*, vol. 38, no. 5, pp. 1597-1602, Sept., 1992.

[11] M. Gutman, "Asymptotically Optimal Classification for Multiple Tests With Empirically Observed Statistics," IEEE Trans. Inform. Theory, vol. IT-35, pp.401-408, Mar., 1989.

[12] J. Ziv and N. Merhav, "A Measure of Relative Entropy Between Individual Sequences with Application to Universal Classification," IEEE Trans. Inform. Theory, vol. 39, no. 4, pp. 1270-1279, July, 1993.