Broadcast System Source Codes: A New Paradigm for Data Compression *

Qian Zhao
California Institute of Technology
Department of Electrical Engineering
Pasadena, California. 91125. USA
qianz@z.caltech.edu

Michelle Effros
California Institute of Technology
Department of Electrical Engineering
Pasadena, California. 91125. USA
effros@caltech.edu

Abstract

Broadcast systems play a central role in an enormous variety of network technologies in which one system node must simultaneously send either the same or different information to multiple nodes in the network. Systems incorporating broadcast components include such diverse technologies as wireless communications systems, web servers, distributed computing devices, and video conferencing systems. Currently, the compression algorithms (or source codes) employed in these devices fail to take advantage of the characteristics specific to broadcast systems. Instead, they treat a single node transmitting information to a collection of receivers as a collection of single-transmitter, single-receiver communications problems and employ an independent source code on each. This approach is convenient, since it allows direct application of traditional compression techniques in a wide variety of broadcast system applications. Nonetheless, we here argue that the approach is inherently flawed. Our innovation in this paper is to treat the general broadcast system (with an arbitrary number of receivers and both specific and common information) as an inseparable whole and consider the resulting source coding ramifications. The result is a new paradigm for data compression on general broadcast systems. In this work, we describe this broadcast system source coding paradigm and examine the potential gains achievable by moving away from more conventional methods.

1. Introduction

A broadcast system, in its simplest form, is a network with one transmitter and two receivers. The transmitter broadcasts information, and each receiver is then able to decode some subset of the transmitted information. The transmitter's goal is to simultaneously communicate both "common" information intended for both receivers and "specific" information intended for each receiver independently. Each receiver's goal is to successfully interpret both the common information and the specific information intended for him. In general, we would have M receivers and up to $2^M - 1$ messages intended for the $2^M - 1$ distinct (and nonempty) subsets of channel receivers.

Broadcast systems are very common and important among network systems. The simplest example is a radio or TV station, where all of the information transmitted is common information intended for all users. Another example is a mobile communication system, such as a pager system. In a pager system, common information such as weather reports, TV schedules, and stock market information may be sent to many customers simultaneously. At the same time, each user can also get his private paging messages. Computer networks and video conferencing systems may also be viewed as technologies containing broadcast systems and requiring a mixture of common and specific information.

A key bottleneck common to many broadcast systems is bandwidth. As the use of communication technologies like wireless phones, the world wide web, and teleconferencing systems becomes more prevalent, the need for efficient ways to represent information for transmission through the broadcast components within these systems likewise increases. As a result, achieving good compression in such a system is critical.

Consider a broadcast system with one transmitter sending information to M receivers. Let S denote the set of all groups of users to whom the transmitter sends a distinct message. Then for any $r \in \{1, \ldots, M\}$, $\{r\} \in S$ implies that the transmitter sends private information to receiver r. Likewise, for any $S \subseteq \{1, \ldots, M\}$ such that $|S| > 1$, $S \in S$ implies that the transmitter sends common information to all receivers in set S. For any set $S \in S$, let $\{X_S(i)\}_{i=1}^{\infty}$ denote the source to be described to all users in set S.

Figure 1. A general broadcast communication system.

Note: $\hat{X}_{1,2\_1} = \hat{X}_{1,2\_2}$
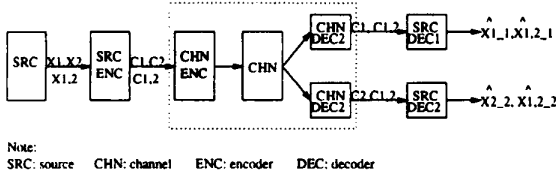
Figure 2. Traditional (single-encoder single-decoder) source codes for broadcast system.

Using existing compression techniques in the given broadcast system requires building an independent source code for each source X_S. Thus for each $S \in \mathcal{S}$, the source encoder for X_S is placed at the transmitter and one copy of the source decoder for X_S is placed at each receiver r such that $r \in S$. While the encoders for all sources $\{X_S\}_{S \in \mathcal{S}}$ are co-located at the transmitter, the encoders operate independently. Likewise while the decoders for all sources X_S such that $r \in S$ are co-located at the receiver indexed by r, the decoders operate independently.

The alternative to the above method of using an independent code for each X_S such that $S \in \mathcal{S}$ is to design a single joint encoder for use at the transmitter and M distinct joint decoders, one for use at each receiver. With this approach, the encoder jointly encodes all messages X_S such that $S \in \mathcal{S}$. Likewise, for any $r \in \{1, \ldots, M\}$, the decoder at receiver r jointly decodes the binary descriptions of all messages X_S such that $r \in S$. The resulting code is called a broadcast system source code.

In [1], Gray and Wyner introduce a related network source coding problem and bound the achievable rate and distortion through that network. This paper focuses on the potential performance gains achievable with broadcast system source codes and the optimal design of broadcast system vector quantizers. The remainder of the paper is organized as follows. In Section 2, we consider the potential benefits of broadcast system source codes. Section 3 contains a brief description of an algorithm for optimal broadcast system source code design. Section 4 contains experimental results. The key contribution of this paper is summarized in Section 5.

2. Broadcast System Source Coding (BSSC)

Figure 1 shows a broadcast communication system. In this paper, we assume that the channel code gives zero probability of error and thereby focus entirely on the source code.

We consider block source codes, and thus the source encoder first blocks each source $\{X_S(i)\}_{i=1}^{\infty}$, for any set $S \in \mathcal{S}$, into vectors of length n. Let X_S^n be one such vector. The encoder maps each collection $(X_S^n)_{S \in \mathcal{S}}$ of n-vector into a corresponding collection $(C_S)_{S \in \mathcal{S}}$ of codewords for
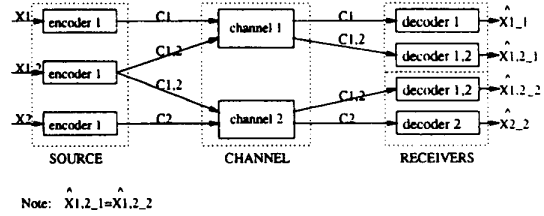
transmission through the broadcast channel. By assumption, receiver r noiselessly receives $(C_S)_{S:r \in S}$ and reconstructs the messages $(X_S^n)_{S:r \in S}$ as $(\hat{X}_{S\_r}^n)_{S:r \in S}$. The notation $\hat{X}_{S\_r}^n$ here distinguishes the reconstruction of X_S at r from the reconstruction of X_S at any other receiver $q \in S$.

To make our description easier to understand, we focus on a simplified system with two receivers. Note, however that the following results apply to arbitrary numbers of receivers. Given the two-receiver assumption, $M = 2$, $r \in \{1, 2\}$, and $\mathcal{S} = \{\{1\}, \{2\}, \{1, 2\}\}$ is the most general possible message set. For notational simplicity, we write X_1, X_2, and $X_{1,2}$ instead of the more correct (but cumbersome) $X_{\{1\}}$, $X_{\{2\}}$, and $X_{\{1,2\}}$. Source X_r is the private information intended for receiver r, and source $X_{1,2}$ is the common information intended for both receivers. Receiver 1's reproductions of X_1 and $X_{1,2}$ are denoted by $\hat{X}_{\_1} = (\hat{X}_{1\_1}, \hat{X}_{1,2\_1})$. Likewise receiver 2's reconstruction is $\hat{X}_{\_2} = (\hat{X}_{2\_2}, \hat{X}_{1,2\_2})$.

In order to understand the potential benefits of broadcast system source codes, consider the following lossless coding example, illustrated by Figure 2. To achieve lossless source coding in a broadcast system using only traditional (single-encoder, single-decoder) source codes, we must encode the sources X_1, X_2, and $X_{1,2}$ independently using 3 separate encoders placed at the transmitter. That is, we independently map each blocked vector X_S^n into a codeword C_S from some rate-R_S lossless source code. Receiver 1 noiselessly receives C_1 and $C_{1,2}$, and independently decodes C_1 to give $\hat{X}_{1\_1}$, and $C_{1,2}$ to give $\hat{X}_{1,2\_1}$ using two separate decoders co-located at receiver 1. Likewise, receiver 2 receives C_2 and $C_{1,2}$, and uses independent decoders to reconstruct $\hat{X}_{2\_2}$ and $\hat{X}_{1,2\_2}$ in the same way. The decoder for the common information $C_{1,2}$ located at receiver 1 is the same as the decoder for $C_{1,2}$ at receiver 2. And thus $\hat{X}_{1,2\_1} = \hat{X}_{1,2\_2}$. Further, for this example all codes are lossless, and thus $\hat{X}_S = X_S$ for each $S \in \mathcal{S}$.

Note that the above method decomposes the broadcast system into three separate encoder-decoder subsystems, each of which works on a single information source X_S independently. Thus, the rate R_S required by the decoder to losslessly reconstruct a random variable X_S must be
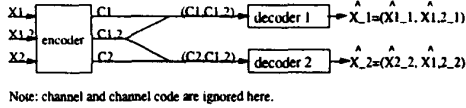
338

Figure 3. Broadcast System Source Coding.

greater than or equal to the entropy $H(X_S)$. As a result, the rate $R_{-1} = R_1 + R_{1,2}$ received at receiver 1 and the rate $R_{-2} = R_2 + R_{1,2}$ received at receiver 2 must be bounded as $R_{-1} \geq H(X_1) + H(X_{1,2})$, and $R_{-2} \geq H(X_2) + H(X_{1,2})$ to achieve lossless compression with the system described above.

Now consider replacing the collection of encoders at the transmitter with a single encoder at the transmitter and replacing the pair of decoders at each receiver by a single decoder at each receiver, as shown in Figure 3. We call the resulting code a *broadcast system source code*. Now, we jointly encode $(X_1^n, X_2^n, X_{1,2}^n)$ into $(C_1, C_2, C_{1,2})$ with a single encoder. Receiver 1 noiselessly receives and jointly decodes $(C_1, C_{1,2})$, while receiver 2 noiselessly receives and jointly decodes $(C_2, C_{1,2})$. In this case, the encoding of X_1^n, X_2^n, and $X_{1,2}^n$ need not be independent, and likewise the decoders at each receiver work cooperatively. The following algorithm can thus be employed. Losslessly describe source $X_{1,2}$ with codeword $C_{1,2}$ from a rate-$H(X_{1,2})$ entropy code. Losslessly describe sources X_1 and X_2 with codewords C_1 and C_2 from rate-$H(X_1|X_{1,2})$ and rate-$H(X_2|X_{1,2})$ conditional entropy code respectively. Receiver 1 receives both C_1 and $C_{1,2}$, while receiver 2 receives both C_2 and $C_{1,2}$. Each decoder can then decode the common information (described by $C_{1,2}$) first and then use that common information to decode the received private information (described by C_1 or C_2). The total rate for receiver 1 and 2 respectively are $R_{-1} = H(X_{1,2}) + H(X_1|X_{1,2}) = H(X_1, X_{1,2})$ and $R_{-2} = H(X_{1,2}) + H(X_2|X_{1,2}) = H(X_2, X_{1,2})$, which are the minimal rates possible for the given transmissions. When X_1 and $X_{1,2}$ are correlated, this rate is smaller than the rate required by the previous method.

The above simple example demonstrates both the problems inherent in using traditional source codes in broadcast systems and the potential benefits of BSSC. In particular, the use of existing (single-encoder, single-decoder) source codes fails to take advantage of the specific characteristics of the broadcast system. Our most basic innovation in the proposal of BSSC is the assertion that source codes should be designed specially for broadcast systems. Dependent encoders at the transmitter and dependent decoders at each receiver are necessary in order to achieve optimal performance. Note, however, that the method given in the earlier example does not generalize to lossless codes with more than two receivers or to lossy source coding, where performance gains are likewise expected. Thus while it serves its

intended purpose of demonstrating the inefficiencies of using independent codes, the method given does not solve the BSSC design problem. In the next section, we describe a practical algorithm for optimal BSSC design. Experimental results follow in Section 4. Before beginning those developments, we pause for several important observations.

First, note that in a system with no common information or independent messages, there is nothing to be gained by BSSC. Independent codes achieve the optimal performance in these special cases (assuming constant complexity).

Second, note that multiple description source coding (MDSC) may be viewed as a special case of BSSC [2]. MDSC addresses the problem where some subset of a binary source description may get lost during transmission. For example, in a simple system with only two "packets," the first or second packet may be lost in transmission (cases 1 and 2 respectively) or all packets may be received (case 3). For this example, we then need three distinct decoders to decode these three cases respectively. This MDSC is a broadcast system with three receivers, one to be used for each of the above cases. In particular, here $M = 3$, $r \in \{1, 2, 3\}$, $\mathcal{S} = \{\{1, 3\}, \{2, 3\}\}$ and $X_{1,3} = X_{2,3} = X$. The BSSC described in Section 3 generalizes the MDSC of [3].

Finally, a secondary advantage of BSSC is the "buy two and get one free" effect. Since the sources $(X_1, X_2, X_{1,2})$ are jointly encoded into $(C_1, C_2, C_{1,2})$, each binary description contains some information about all sources. Thus $(C_1, C_{1,2})$ contains information about X_2 in addition to the intended information about X_1 and $X_{1,2}$. As a result, if desired, we can get some reconstruction of X_2 at receiver 1 without compromising the performance on X_1 and $X_{1,2}$. The same is true for receiver 2. No such benefit is possible using independent codes.

3. Algorithm

The performance of a BSSC is judged in terms of the distortion of the reproductions and the rates used to describe them. In the most general case, each receiver can reproduce all the sources X_S, $S \in \mathcal{S}$. For any receiver r and any set $S \in \mathcal{S}$, let $D_{S_{-r}}$ denote the average distortion for source X_S reproduced at receiver r, where $D_{S_{-r}} = E_S(1/n)d(X_S^n, \hat{X}_{S_{-r}}^n)$. Likewise let R_S denote the average rate used to transmit X_S, where $R_S = E_S(1/n)R_S(X_S^n)$ and $R_S(X_S^n)$ describes the rate used to describe source vector X_S^n.

Let $\mathbf{D}$ be a vector of $M \times |\mathcal{S}|$ distortions ($D_{S_{-r}}$ such that $S \in \mathcal{S}$ and $r \in \{1, \ldots, M\}$) and $\mathbf{R}$ be a vector of $|\mathcal{S}|$ rates (R_S such that $S \in \mathcal{S}$). The set of distortion-rate pairs $(\mathbf{D}, \mathbf{R})$ achievable by BSSC defines a convex region in the $(M + 1) \times |\mathcal{S}|$-dimensional distortion-rate space. As in [3], this distortion-rate region is entirely characterized

by its support functional, $J = \sum_{S,r} \mu_{S\_r}(D_{S\_r} + \lambda_S R_S)$, where μ_{S_r} and λ_S are the associated Lagrangian constants. (Usually we choose $\mu_{S\_r} = 0$ if $r \notin S$.) The functional J may be interpreted as a Lagrangian for minimizing any combination of rates, distortions, or Lagrangians subject to constraints on the remaining quantities.

Like its vector quantization (VQ) predecessors, the BSSC design algorithm proposed here uses an iterative descent approach. The algorithm generates a sequence of codes for which J is non-increasing.

Each iteration proceeds as follows.

1. Optimize the encoder for the decoder. Jointly encode blocked sources $(X_S^n)_{S \in \mathcal{S}}$ into a collection of codewords $(C_S)_{S \in \mathcal{S}}$ such that the reproduced data vectors at each receiver will result in the smallest J.

2. Optimize the decoder for the encoder. Update all reproduction vectors so that the expected distortion between each reproduction vector and the training vectors mapped to it is minimized. In particular, for the squared error distortion measure the updated value of the reproduction vector is the centroid of the training vectors mapped to it.

3. Optimize the entropy code. In variable-rate coding, the description length for each codeword is chosen to match the negative logarithm of the (conditional) probability of this codeword. This length is updated here for each iteration. In fixed-rate coding, this step can be omitted.

A detailed description of this algorithm appears in [4]. The above three steps are iterated until convergence. Convergence to a locally optimal solution is guaranteed.

4. Results

We include results for experiments performed for the two-receiver broadcast system described in Section 2. For these experiments, the training set contains 18 medical brain-scan images, which are evenly divided into 3 groups, each of which acts as the training set for one of the 3 sources. A non-overlapping set of 6 brain-scan images are likewise divided into 3 sources and used as test sets.

We use the algorithm described in Section 3, with the assumption that 1) no receiver cares about the private information of any other, 2) there is a trade-off between two private informations, and 3) the common information has equal importance for both receivers. Thus we choose $\mu_{2\_1} = \mu_{1\_2} = 0$, $\mu_{1,2\_1} = \mu_{1,2\_2} = \alpha/2$, $\mu_{1\_1} = \beta$, $\mu_{2\_2} = 1 - \beta$. Here $\alpha, \beta \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

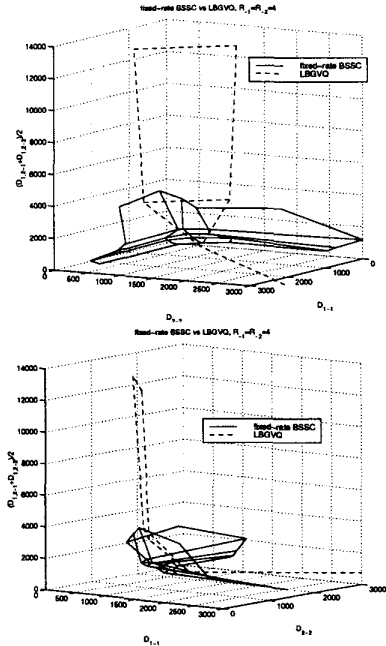All experiments are conducted under the following rate constraint: each receiver is assigned a fixed rate of 4 bits

Figure 4. Fixed-rate BSSC vs LBGVQ.

per pixel (bpp), 2 of which are the common bits for both receivers. Thus in total the system uses a rate of 6 bpp. To meet these rate requirements for fixed-rate coding, we set all $\lambda_S = 0$ and use a depth-6 ternary tree. For variable-rate coding, we use a depth-12 ternary tree, and for each (α, β) pair adjust λ_S to meet the desired rate constraint. (Details on the ternary trees appear in [4].) By fixing the rates in this way, we can focus on the achievable distortions $\mathbf{D} = ((\mathbf{D_{1,2\_1}} + \mathbf{D_{1,2\_2}})/2, \mathbf{D_{1\_1}}, \mathbf{D_{2\_2}})$ (and thereby avoid attempts at visualizing the 6-dimensional rate-distortion space associated with this problem).

For comparison purposes, we consider coding the three sources independently using either fixed-rate VQ (for fixed-rate comparisons), or entropy constrained VQ (ECVQ) (for variable-rate comparisons). We consider a collection of independent codes that meet the above rate constraints, and draw the convex hull of the achieved distortions $\mathbf{D} = (\mathbf{D_{1,2}}, \mathbf{D_1}, \mathbf{D_2})$. Since the rate assigned to each receiver is identical for the BSSC and for the independent codes, the lower the convex hull of distortions in 3-dimension space, the better the performance.

Figure 4 compares the distortion triple of fixed-rate BSSC with that of fixed-rate independent LBGVQ, showing the 3-dimensional distortion space from two different angles. Figure 5 likewise compares the distortion triple of variable-rate BSSC with that of using independent ECVQs on each source. In both comparisons, the convex hull of BSSC(solid line) is significantly lower than that of independent source code(dashed line) in the 3-dimension space,
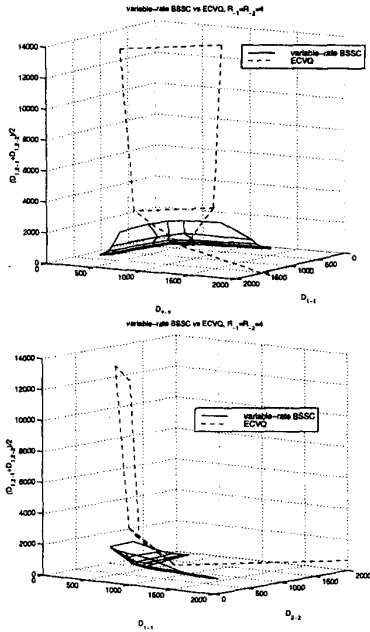
340

Figure 5. Variable-rate BSSC vs ECVQ.

Figure 6. Fixed-rate BSSC vs variable-rate BSSC, in 3 and 2-dimensions.

which shows the performance benefits of BSSC over independent source coding. Finally, Figure 6 compares fixed-rate BSSC with variable-rate BSSC. The solid lines show the variable-rate BSSC, while the dashed lines show the fixed-rate BSSC. Each point corresponds to a single (α, β) pair. As with independent codes, variable-rate codes (of a given dimension) outperform fixed-rate codes (of the same dimension). All vector quantizers given here use vector dimension 4.

When α is fixed, the average distortion of the common information is approximately constant. By varying β, we get a convex curve which shows the trade-off between the distortions D_1 and D_2 on the two private informations. In particular, the greater β is, the lower D_1 but the higher D_2. Thus β in some sense describes the system designer's priorities with respect to the system users' performances. When β remains constant, the greater α is, the lower the average distortion $(D_{1,2\_1} + D_{1,2\_2})/2$ of the common information becomes but the higher the distortions D_1 and D_2 of the private information. Thus α in some sense describes the system designer's priorities with respect to common versus private information. Figure 6 shows this tradeoff in both 2 and 3 dimensions. In 2 dimensions, we label the average distortion of the common information on the curve corresponding to each α and then vary β to trace out the full curve.
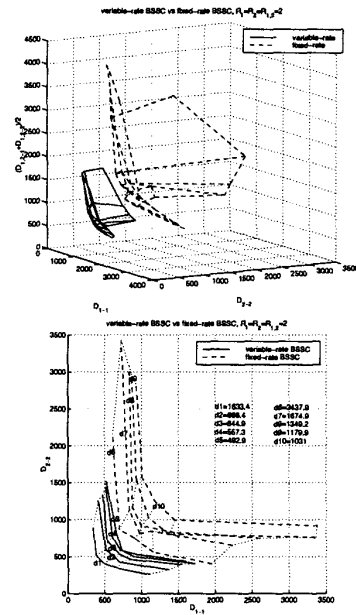
5. Conclusions

We first introduce the concept of BSSC and the assertion that source codes to be used in broadcast systems should be designed specifically for those systems. We then give a simple lossless coding argument to demonstrate that only through BSSC – rather than independent code design – can we hope to achieve optimal source coding performance for broadcast systems. We very briefly describe a practical algorithm for BSSC. (Details appear in [4].) The experimental results for this algorithm demonstrate that BSSC can achieve better performance than independent source coding in two-receiver broadcast systems.

References

[1] R. M. Gray and A. D. Wyner. Source coding for a simple network. *Bell Systems Technical Journal*, 53(9):1681–1721, November 1974.

[2] M. Effros. Network source coding. IEEE. Submitted to ISIT 2000. September 1999.

[3] M. Fleming and M. Effros. Generalized multiple description vector. In *Conference Record, Data Compression Conference*. IEEE, March 1999.

[4] Q. Zhao and M. Effros. Broadcast system source codes. 1999. In preparation.