MARYLAND

TECHNICAL
RESEARCH
REPORT

SRC TR 88-107

Switched Scalar Quantizers for Hidden Markov Sources

by

D. Goblirsch and N. Farvardin

SYSTEMS RESEARCH CENTER
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742

Switched Scalar Quantizers for Hidden Markov Sources*

David M. Goblirsch
Aerospace Technology Center
Allied-Signal Aerospace Company
9140 Old Annapolis Road
Columbia, MD 21045
CSNet: dmg@atc.bendix.com

Nariman Farvardin†
Electrical Engineering Department,
Institute for Advanced Computer Studies,
and
Systems Research Center
University of Maryland
College Park, MD 20742
CSNet: farvar@eneevax.umd.edu

December 7, 1988

*Some of the results in this paper were presented at the 1988 IEEE ISIT in Kobe, Japan.
†The work of the second author was supported in part by National Science Foundation grant NSFD MIP-86-57311.

1 Introduction

This paper describes how to design a Switched Scalar Quantizer for a Hidden Markov Source. Our goal in studying Switched Scalar Quantizers is to develop efficient waveform coding systems for nonstationary sources. Unfortunately, the theory for coding nonstationary sources is not as well developed as for the stationary case; most current approaches to coding nonstationary signals are rather ad hoc. Therefore, the design and analysis of waveform coding systems for nonstationary sources remains an important problem.

The class of all nonstationary sources is too broad for analysis. To design and analyze coding systems, we need to make some simplifying assumptions about the time-varying source statistics. One type of nonstationary behavior, observed in speech, is where the source appears to switch between different short-term modes of stationary behavior [1]. However, the order in which different modes become active, the mode transition times, and the actual duration of any particular mode all appear to be random. This leads to the concept of a *composite source* [2,3,4]. A composite source is a (discrete-parameter) stochastic ordered-pair process $(X_t, S_t)_{t=1}^{\infty}$. The process $(X_t)_{t=1}^{\infty}$ is the source *output;* it is this process which is to be transmitted. The process $(S_t)_{t=1}^{\infty}$ is the *switch*. The switch cannot be observed directly, but controls the probability distribution governing the output. The most popular composite source model, particularly in speech recognition, is the *Hidden Markov Model* [5,6], a composite source for which the switch is a Markov chain [7]. The *Hidden Markov Source* (HMS) is an extension of the Hidden Markov Model to sources which produce infinitely long waveforms. The HMS has the advantage of being (if the parameters are properly chosen) a stationary ordered-pair process whose output appears to be nonstationary to an observer who is measuring only local statistics.

Most coding systems intended for nonstationary sources use an *adaptive quantizer* (AQ). An AQ automatically adjusts its parameters in response to perceived changes in

Abstract

This paper describes a new algorithm for designing Switched Scalar Quantizers for Hidden Markov sources. The design problem is cast as a nonlinear optimization problem. The optimization variables are the thresholds and reproduction levels for each quantizer, and the parameters defining the next-quantizer map. The cost function is the average distortion incurred by the system, allowing for a different distortion measure for each subsource. The next-quantizer map is treated as a stochastic map so that all of the optimization variables are continuous-valued, allowing the use of a gradient-based optimization procedure. This approach solves a major problem in the design of switched scalar quantizing systems, that of determining an optimal next-quantizer decision rule. Details are given for computing the cost function and its gradient for weighted-squared-error distortion. Simulation results are presented which compare the new system to current systems, where we see that our system performs better. It is also observed that the optimal system can in fact have a next-quantizer map with stochastic components.

spect to the system parameters. These quantities are needed for a gradient-based descent approach to solving the design problem. Details are worked out for the weighted-squared-error distortion measure, allowing for different weights for each subsource. Section 6 gives a lower bound on the distortion achievable by any switched quantizer when the number of quantization levels is specified. This is followed by Section 7, where the design algorithm is applied to several examples. Comparisons are made against other adaptive and nonadaptive encoding systems. Finally, a summary and ideas for further research are provided in Section 8.

2 The Hidden Markov Source

As mentioned in the Introduction, we will restrict attention to the following type of composite source, which is a special case of a Hidden Markov Source (HMS). An HMS is a composite source for which the switch is a Markov chain. Let S be the number of subsources in the HMS. We assume that each subsource is a memoryless source, i.e., an independent and identically distributed sequence of real-valued random variables. Each subsource is assumed to have an absolutely continuous first-order probability distribution function. The probability density function (pdf) of the observation when the switch is pointing to the j-th subsource will be denoted by $g_j(x)$.

We assume that the switch is time-homogenous, irreducible, and aperiodic. Thus, a stationary probability distribution exists for the Markov chain, $(\rho_1, \rho_2, \ldots, \rho_S)$, where

$$\rho_j = P[S_t = j]. \tag{1}$$

The switch transition probability from state j to state k will be denoted by $\lambda_{k|j}$:

$$\lambda_{k|j} = P[S_{t+1} = k | S_t = j]. \tag{2}$$

We also assume that the Markov chain is started in its stationary distribution because then the observation process $(X_t)_{t=1}^{\infty}$ is stationary, although this will not appear to be the source probability distribution in an effort to reduce the effects of source-quantizer mismatch. AQ schemes differ in how and how often these adjustments are made. Many approaches for designing AQ's have been studied [8], mostly for the case where only the variance of the source is changing.

The major drawback of current approaches to AQ is that they do not assume any model for the nonstationary source. Usually, the only assumption is that fluctuations in signal power are completely arbitrary. This is clearly a worst-case approach. A model-based design approach should yield a system which performs significantly better than these worst-case approaches when the nonstationary behavior has some structure, as it does for composite sources. A second drawback is that the design of current AQ systems is rather ad hoc. A third drawback, as we will see later, is that any AQ system with a variable-scale-fixed-shape quantizer is *inherently* suboptimal for the type of source considered in this paper.

AQ systems are specific instances of *Switched Scalar Quantizers.* An optimal design approach, however, should make as few a priori assumptions about the individual quantizers and the switching strategy as possible. The problem that arises is how to select the best parameters and switching strategy. This is the problem addressed in this paper. Our solution is to formulate the design problem as a constrained optimization problem. To do this, we first show how to parameterize the next-quantizer decision rule. We then show how to calculate the average distortion and its gradient.

The paper is organized as follows. The HMS model is described in Section 2. We restrict ourselves to a specific type of HMS, one having memoryless subsources. The mechanics of the Switched Scalar Quantizer are described in Section 3. The design problem is formally stated in Section 4. There, we briefly discuss why finding an optimal next-quantizer decision rule is difficult, and then describe our solution to this problem. In Section 5, we show how to calculate the average distortion and its gradient with re-

In general, the transmitter state can be updated in one of two ways: as a function of the current transmitter state and the input to the transmitter (i.e., the source output), or as a function of the current transmitter state and the output of the transmitter (i.e., the selected channel symbol). When the next transmitter state is a function of the current transmitter state and the transmitter output, a correctly initialized receiver with sufficient memory that receives the channel symbols without error can track the sequence of transmitter states and hence the quantizers used by the transmitter without the need for *side-information*: "overhead" information about which quantizer is currently in use. A transmitter with such a next-transmitter-state map is called *trackable* [9]. If, however, the next transmitter state is a function of the current transmitter state and the input to the transmitter, then side-information is essential. In this paper, we assume that we do not want to transmit side-information, so the next transmitter state will be selected on the basis of the current transmitter state and the observed quantizer cell.

We have outlined the operation of a general finite-state switched quantizer transmitter. The system that will be considered in this paper is simpler, however, in that we identify the transmitter state at time t with the quantizer being used at time t; that is, we assume that there is a one-to-one correspondence between transmitter states and quantizers. Therefore, we will usually refer to the next-transmitter-state map as the *next-quantizer map*. There are two reasons for this simplification. First, the proposed system is more tractable. Second, the examples in Section 7 will show that the simplified system often performs close to theoretically derived bounds that also apply to the more general system, demonstrating that there is usually little loss in performance by making this simplification.

If each of the T quantizers has C cells, then the transmitter needs to store $T(C-1)$ parameters, namely, the *decision thresholds*

$$\xi_l^i, \quad i = 1,2,\ldots,T, \;\; l = 1,2,\ldots,C-1,$$

the case to an observer who doesn't know about the HMS structure in the sense that the *locally* measured statistics will change with time.

3 Switched Scalar Quantizers

3.1 Finite-State Switched Quantizer Transmitters

In a *switched quantizer transmitter,* a scalar quantizer is used to encode each source sample. The quantizer to be used at a particular time instant is selected from a predetermined set of quantizers on the basis of the transmitter's history. In a *finite-state switched quantizer transmitter,* this set of quantizers is finite.

For a set of T quantizers each having C cells, storing the full history of the transmitter's operation for the past K time instants requires $K \log_2(TC)$ bits. Obviously, this quantity grows with K. In a finite-state switched quantizer transmitter, the history of the transmitter's operation is condensed to one of a finite (and preferably small) number of possibilities. Therefore, the transmitter's history is summarized by a finite-valued variable called the *transmitter state.* The transmitter state at time t uniquely determines which quantizer will be used on the source sample X_t. In general, it is not necessary that there be a one-to-one pairing between transmitter states and quantizers; different states may use the same quantizer. However, the number of transmitter states is not less than the number of quantizers.

Denote the transmitter state at time t by T_t. Let the number of allowable states be T. Without loss of generality, we take $T_t \in \{1, 2, \ldots, T\}$ $\forall t$. Applying the quantizer assigned to T_t to the source output X_t results in a channel symbol C_t, which is the index of the observed cell. After transmitting the selected channel symbol, the system updates the transmitter state, which in turn causes the selection of a quantizer for the next time instant.

where the superscript indexes the reproduction set (i.e., the receiver state) and the subscript indexes the received channel symbol (i.e., the index of the cell observed by the transmitter's quantizer). Observe that the number of distinct reproduction values is not greater than RC. Thus, a system with a finite-state switched-reproduction-set receiver cannot outperform an optimally designed fixed scalar quantizer having RC levels [9, Theorem 2].

For finite-state switched-reproduction-set receivers, the number of receiver states R is finite, but need not be the same as the number of states T in the transmitter. For example, the two numbers may differ when the receiver implements some additional filtering [9]. However, if the transmitter is trackable, one (not necessarily optimal) option is to take the next-receiver-state map to be the same as the next-transmitter-state map. In this case, the receiver only uses its memory to track the the sequence of transmitter states, and the receiver is then called a *tracking receiver* [9]. In the absence of channel errors, we have $T_t = R_t \ \forall t$ for a tracking receiver. For this paper, we assume a tracking receiver. Since the transmitter is a finite-state switched quantizer transmitter, the reproduction map η simply implements the inverse quantizer characteristic. That is, $Y_t = \eta(R_t, C_t) = \eta(T_t, C_t)$ is just the reproduction level assigned to cell C_t of the quantizer associated with transmitter state T_t.

A finite-state switched reproduction set receiver is *time-invariant* if the next-receiver-state map and the reproduction levels do not depend on t. We assume that our receiver is time-invariant.

In summary, our receiver will be a tracking, time-invariant, finite-state switched reproduction set receiver. Since the receiver is a tracking receiver, the number of receiver states is equal to the number of transmitter states, and the next-receiver-state map is identical to the next-transmitter-state map.

where the superscript indexes the quantizer and the subscript indexes the threshold.

A finite-state switched quantizer transmitter is *time-invariant* if the next-quantizer map and the decision thresholds do not depend on t. We assume that our transmitter is time-invariant.

In summary, our transmitter will be a trackable, time-invariant, finite-state switched quantizer. The quantizers are in one-to-one correspondence with the transmitter states. The number of quantizers will be denoted by T. For convenience, it is assumed that all of the quantizers have the same number of cells, this number being denoted by C. We will use τ to denote the next-quantizer map: $T_{t+1} = \tau(T_t, C_t)$.

3.2 Finite-State Switched Reproduction Set Receivers

In general, the receiver implements a *reproduction map* η which depends on the received channel symbol C_t and the current *receiver state* R_t to obtain the *receiver output* Y_t, which is the reproduction of the source output X_t. If Y_t is intended to be a reproduction of the source output $X_{t-\Delta}$, then the receiver has a *delay* Δ. In this paper, we assume that there is to be no receiver delay ($\Delta = 0$).

After producing Y_t, the receiver updates its state according to the *next-receiver-state map*.

In a *finite-state switched reproduction set receiver*, there are a finite number of sets of reproduction levels, each having C elements. The receiver state at time t, denoted by R_t, determines from which of these sets the receiver output Y_t will come. The received channel symbol C_t determines which element of the selected set will be the receiver output. If we denote the number of receiver states by R, the receiver needs to store RC parameters, namely, the *reproduction levels*

$$\eta_l^i, \quad i = 1, 2, \ldots, R, \ \ l = 1, 2, \ldots, C,$$

By convention, $\xi_0^i = -\infty$ and $\xi_C^i = +\infty$ for each quantizer. The *average distortion* D for the SSQ is given by

$$D = \sum_{j=1}^{S} \sum_{i=1}^{T} \pi_{ji} D_{ji}, \tag{4}$$

where

$$\pi_{ji} = P[S_t = j, T_t = i] \tag{5}$$

is the probability of using the i-th quantizer to encode an output generated by the j-th subsource at time t.[1]

The *SSQ design problem* is to select the quantizer thresholds, the reproduction levels, and the next-quantizer map so as to minimize D. The reproduction levels are not truly independent variables, however.[2] For a given set of distortion measures, thresholds, and next-quantizer map, the optimal reproduction levels are completely determined. Therefore, of the quantizer thresholds and reproduction levels, only the thresholds will be adjusted directly by the design algorithm.

The major problem in designing a SSQ is the determination of an *optimal* next-quantizer map. To see this, let $Z_{ki} \subset J_C$ be the set of cell indices for the k-th quantizer that cause the transmitter to select the i-th quantizer for use at the next time instant. To characterize the optimal set of Z_{ki}'s, we can proceed in a manner similar to the derivation of the Bayes decision rule for multiple hypotheses [14]. It turns out that to minimize D for a fixed set of quantizers, the l-th cell of the k-th quantizer should be assigned to Z_{ki}

[1]The use of t in all of the equations that follow is only to help distinguish between the values of variables at the current time instant and the next or past time instants. Because the coding system is time-invariant and the source is stationary, the terms in these equations do not really depend on the actual value of t.

[2]Alternatively, the reproduction levels could be taken as the independent variables and the thresholds could be dependent. This is, in fact, the approach we used in [13] where we showed how to design a single quantiser system. However, now the distortion depends on the next-quantiser decision rule, which depends on the thresholds but *not* on the reproduction levels. Hence, it seems more logical to choose the thresholds as the independent variables.

3.3 Switched Scalar Quantizers

The combination of a trackable, time-invariant, finite-state switched quantizer transmitter and a tracking, time-invariant, finite-state reproduction set receiver will be called a *switched scalar quantizer* (SSQ). This paper is concerned with the design of SSQ's for HMS's having memoryless subsources.

The idea of switching between a finite number of quantizers is not new [10,11,12]. However, our design approach is new and offers the possibility for significant improvement in the performance of such systems because we make use of an HMS model for the nonstationary behavior. Our approach also solves a problem not solved by other switched-quantizer design algorithms: the design of an *optimal* next-quantizer decision rule.

In what follows, we assume that the channel is noiseless, for otherwise the receiver could lose track of the sequence of transmitter states and hence the quantizer sequence being used by the transmitter. The effects of channel noise on SSQ performance will be studied in a later paper.

4 The Design Problem

Let $\mathbf{R}$ denote the real numbers and $\mathbf{R}_+$ the nonnegative real numbers. Define $J_k \triangleq \{1, 2, \ldots, k\}$. Let $d_j : \mathbf{R} \times \mathbf{R} \to \mathbf{R}_+$, $j \in J_S$, be S scalar distortion measures, one for each subsource. These distortion measures need not be different. The first argument of d_j is the source output, while the second argument is the receiver output.

Let D_{ji} be the average distortion incurred when the i-th quantizer is used to encode a source output generated by the j-th subsource, with respect to the distortion measure d_j. Then

$$D_{ji} = \sum_{l=1}^{C} \int_{\xi_{l-1}^i}^{\xi_l^i} d_j(x, \eta_l^i) g_j(x) dx. \tag{3}$$

and the inequality constraints

$$\tau_{i|kl} \geq 0, \quad i \in J_T,\ k \in J_T,\ l \in J_C. \tag{8}$$

The next-quantizer decision rule is deterministic whenever $\tau_{i|kl}$ is 1 for exactly one $i \in J_T$ and 0 for the remaining i's, for all pairs $(k,l) \in J_T \times J_C$.

It turns out to be more convenient to replace $\tau_{T|kl}$ by $1 - \sum_{i=1}^{T-1} \tau_{i|kl}$, and replace the equality constraints of (7) by the inequality constraints

$$1 - \sum_{i=1}^{T-1} \tau_{i|kl} \geq 0, \quad k \in J_T,\ l \in J_C. \tag{9}$$

The number of design variables is now $T^2C - T$. The number of such inequality constraints is TC. The inequality constraints of (8) are now replaced by $(T-1)TC$ "box" constraints:

$$0 \leq \tau_{i|kl} \leq 1, \quad i \in J_{T-1},\ k \in J_T,\ l \in J_C. \tag{10}$$

The best system with a *deterministic* next-quantizer map can also be found using this approach by solving a similar optimization problem with the $(T-1)TC$ additional equality constraints

$$\tau_{i|kl}(1 - \tau_{i|kl}) = 0, \quad i \in J_{T-1},\ k \in J_T,\ l \in J_C. \tag{11}$$

These new constraints force the τ's to be either one or zero. This approach works best with an optimization algorithm that does not require feasibility during the intermediate stages.

To conclude this section, we note that a stochastic next-quantizer map can be implemented as follows. Suppose that both the transmitter and the receiver have access to a pseudo-random sequence started from the same seed at the same time instant, as might be done in spread-spectrum systems [16]. The elements of this sequence are used whenever a stochastic decision is required, and because the transmitter and receiver sequences are identical, the receiver always knows what decision was made by the transmitter (provided that there are no channel errors).

only if [15]

$$\sum_{j=1}^{S} P[C_t = l | S_t = j, T_t = k] \pi_{jk} \tilde{D}_{ji} \;\leq\; \sum_{j=1}^{S} P[C_t = l | S_t = j, T_t = k] \pi_{jk} \tilde{D}_{jq}$$

for all $q \in J_T$, where $\tilde{D}_{ji}$ is the average distortion incurred when the i-th quantizer is used and the switch at the *previous* time instant was pointing to the j-th subsource. Note that the determination of the optimum Z_{ki}'s, and hence the optimum next-quantizer map, depends on knowing $P[S_t = j, T_t = k]$ for all $k \in J_T$ and $j \in J_S$. However, these quantities themselves depend on the next-quantizer map. This intertwining relationship between the next-quantizer map and the probabilities $P[S_t = j, T_t = k]$ is the source of the difficulty in finding the optimal switching rule. Therefore, this problem will have to be solved using an iterative optimization algorithm. This requires that we parameterize the next-quantizer map.

There are two possible approaches. One is to require that τ be deterministic, resulting in an integer programming problem. The other is to treat τ as a stochastic map, allowing us to work with continuous-valued variables. We take the stochastic approach because it leads to a gradient-based descent algorithm and is more general in that it includes the deterministic case.[3] This idea was first proposed in [9, Section VI] as a suggestion for further research.

Let $\tau_{i|kl}$ be the probability that observing the l-th level of the k-th quantizer causes the transmitter to choose the i-th quantizer for use at the next time instant:

$$\tau_{i|kl} = P[T_{t+1} = i | T_t = k, C_t = l]. \tag{6}$$

For these to be probabilities, they must satisfy the equality constraints

$$\sum_{i=1}^{T} \tau_{i|kl} = 1, \qquad k \in J_T,\; l \in J_C, \tag{7}$$

[3]We will later see that oftentimes the optimum next-quantiser decision rule is in fact a stochastic decision rule.