A Hidden Markov Model Approach to the Classification of Acoustic Transients

M. K. Shields and C. W. Therrien
Naval Postgraduate School
Monterey, California 93943

Abstract

A system for the detection and classification of acoustic transients based on Hidden Markov Model (HMM) methods is described. The system is tested using two different sets of data. The results from these tests are summarized and compared to some other known algorithms and to the performance of human beings. The effect of observation noise on the training and classification process is also discussed, and results using different noise correction techniques are presented.

1 INTRODUCTION

Acoustic transient signals occurring in various monitoring and surveillance problems are difficult to classify because of their short duration, low energy content, and nonstationary behavior. Like many signals (including most notably speech) transients have most of their information content encoded in the temporal variations of their short-duration power spectra. Several ways have been proposed to use this feature of transients to aid in transient event classification. In this research we investigate the classification of acoustic transient events using hidden Markov modeling techniques. The goal of the research was to develop a complete system for the detection and classification of acoustic transients. In this paper we describe the classification system and summarize the results of the classification.

The remainder of this paper contains three sections. The next section presents an overall description of the system. The third section presents the results of several tests using our method and compares the results to those of other algorithms using the same data. The final section gives a summary and conclusions.

2 SYSTEM DESCRIPTION

In problems involving hidden Markov modeling, the classifier is assumed to observe a discrete-time data sequence which is produced by an underlying Markov chain in the following way. When in a particular state the Markov chain produces output observations according to a probability distribution associated with that particular state. At each time epoch the state may change according to the Markov transition probabilities and so the sequence that the classifier observes can correspond to different probability distributions at different times. The sequence of state transitions are not seen by the classifier and so the Markov chain is said to be "hidden".

In our application of HMM techniques to transients (as in the application to speech) the HMM is not assumed to produce the sampled acoustic signal data directly. Rather, each small segment of the signal data is mapped into a feature vector, and the feature vector is associated with one of M symbols or "codewords" from a finite set of possibilities called the "codebook". This is the vector quantization step. The codewords associated with the short segments of acoustic data form the observation sequence produced by the Markov chain. The performance of the system is thus clearly dependent on the choice of the features.

The system for classification in this work is based upon a five state, left-to-right HMM. The states in this process can be thought of as an initial state (S1) where the signal raises out of the background noise, a final state (S5) where the signal decreases back into the background noise, and three states of unspecified length between. Neither the observation probability distribution functions nor the Markov transition probabilities are known and thus both must be estimated from known data. The system performs two tasks: "training" or estimating the "model" (the observation and transition probability distributions) from sets of known data, and classifying observations.
The classification system is composed of five modules: Initial Processor (IP), Vector Quantizer Code Book Generator (VQCBG), Vector Quantizer (VQ), Model Parameter Estimator (MPE), and Classifier (CLASS) as shown in Fig. 1. Each of these is discussed briefly below.

Figure 1: Block diagram of HMM classification system.

2.1 Initial Processing

In the IP portion of the system sampled acoustic data is mapped into a feature space related to the spectrum. The IP portion of the system also performs detection of the beginning and ending of the transient event and possible correction for background noise. The system is configured to use feature sets consisting of either LPC coefficients, samples of the cepstrum, or the power in one of three different sets of FFT bandpass filters (see 3.1). The latter processing is referred to as FFTFB 1, 2, and 3.

Because transient signals are observed in colored noise which can be considered stationary over periods of time much longer than the duration of the transient, it is possible to measure the noise from before or after the transient event, and correct the feature vector for its effects.

For the FFTFB IP sections we use a simple method where some portion of the estimated noise power is subtracted from each FFT bin. In the classification results we refer to “full correction”, which takes the difference of the signal and estimated noise spectra in FFT bins and sets any resulting negative values to zero; and “limited correction”, which merely subtracts some fraction of the measured noise.

2.2 Code Book Generator and Vector Quantizer

The VQCBG and VQ blocks of the system map the vectors from the IP, which span a continuous feature space, into one of $M(24 \leq M \leq 256)$ discrete codewords. The codebook generation algorithm seeks to find a set of codewords that represents the collection of input training vectors with minimum total distortion. Minimizing “distortion” in this context involves matching the feature vector to a codeword in the feature space according to some distortion or distance measure and assigning it to the codeword to which it is in some sense closest. The algorithm begins with an initial selection of codewords and assignment of training vectors to codewords and iteratively perturbs the codewords and assignments until a minimum distortion configuration is reached.

We use the generalized Lloyd algorithm [1] with three different initialization methods which we call “unseeded”, “seeded”, and “limited seeding”. The first uses a previously chosen initial condition (not related to the data in any way); the second uses the means of feature vectors for each class of training vectors to seed the codebook perturbing process; the third limits the codebook generation process to only feature vectors from the high energy portions of the training events and then uses two of these feature vectors as the initial codebook.

VQ is based on the same distortion measure used in VQCBG, and assigns each feature vector to the codeword that “best” represents it according to the distortion measure. The output of the VQ block is a discrete observation, which we represent with the integers 1 to $M$, for each feature vector input.

2.3 Model Parameter Estimation

The most difficult task in Hidden Markov Modeling is the estimation of the parameters of the model. The mathematical theory for estimating the parameters of a Markov process where the underlying process cannot be observed is due to Baum [2] and others. The procedure we implement is based on Baum-Welch Reestimation and the Forward Backward algorithm [3].

2.4 Classification of Test Data

One advantage of using Hidden Markov Models is that given the model parameters for a set of $V$ possible models, and a sequence of observations $X$, it is possible to explicitly compute the likelihood function $Pr[v^t_{\text{model}}|X]$ for each model. Then, if the prior
probabilities of each model are the same, the optimum decision in the Bayes sense is the model which has the maximum likelihood. The forward-backward equation provides a simple way to calculate the log likelihood function given the model [5]. This is used in the classification.

3 CLASSIFICATION RESULTS

The system was tested on two data sets representative of a variety of types of transients. We will refer to these data sets as Data Set I and Data Set II.

3.1 Data Set I Tests

Our initial tests used Data Set I and were conducted as follows. The data set was manually examined to determine the beginning, ending, and classification of the various events. This produced a set of 42 usable transient events. In these tests we use the FFTFB IP section with the three different band pass filter banks, FB1, FB2, and FB3. In FB1 and FB2 the 256 FFT bins were evenly divided into 19 and 30 bands respectively. In FB3 the frequency range was divided into 19 one-third octave pass bands. We used the two different types of noise correction discussed above in conjunction with each FFTFB giving nine different feature sets. Four codebooks (of sizes 32, 64, 128, and 256 for the unseeded and limited cases, and 24, 48, 98, and 192 for the seeded case) were generated for each of these feature sets using all three types of VQCBG algorithms. Thus a total of 108 codebooks were generated and the MPE algorithm was applied to the observations using each of these 108 codebooks. Each of the sets of models was then used to classify the 42 transients. The percentage of correctly classified transients for the mid-size codebooks (128 codewords for the unseeded and limited and 98 for the seeded) are shown in Table 1. Observe that for all three types of codebook generation, the specific set of spectral bands used had little effect on the results. Thus we could conclude that the simplest method, FB1, was adequate. Table 1 also indicates that for all three types of codebook generation, the use and type of noise correction had relatively little effect on this data. This will not always be the case.

Figure 2 shows the the results of all code book generation procedures as a function of the code book size. Observe that for the small codebook sizes, the limited codebook produces results which are better than the others, and as the codebook size increases the results improve. Because the limited codebook uses many less frames to determine the codeword (about 1 tenth in our tests) it is much faster. Note also that larger codebook sizes produced less variance in the results.

For comparison, the AR Spectral Distance Classifier (ARSDC) algorithm of Lashkuri [4] was used to classify the same data. While the HMM algorithms produced percentages of correct classification between 83 (worst case) and 100 (in several cases) the ARSDC using the same data gave 45 percent correct classification.

3.2 Data Set II Tests with Added Noise

Our second set of tests used Data Set II with added white Gaussian noise. The purpose of these tests was to determine the effect of observation noise on the classification of transients. Tests were performed with a variety of different combinations of IP and noise correction. Best results were obtained with the FFTFB
features. The LPC and Cepstrum features gave significantly poorer results, probably because the relative energy from successive frames was not accounted for in these features.

These tests performed were:

**TEST 1** - Train the HMM on events with no observation noise and classify the same events with observation noise SNR from 5 dB to $\infty$.

**TEST 2** - Train the HMM on events with observation noise and classify the same events with the same noise levels.

Additionally we repeated the tests to determine the effect of a limited training set. In these cases the models were trained on only five of the events, and were conducted using FFTFB2 without noise correction. Table 2 gives a summary of the results for comparison. The first column under each test shows the results of training and testing on all of the data. The second column shows results when only five events were used for training and testing was performed using all of the data. The LPC and Cepstrum features gave results ranging from a low of 23% (correct) at the 5 dB SNR to 97.4% with no noise. The data were also classified using the ARSDC. This algorithm gave correct classifications of 30% or less for the various cases.

Table 3 compares the results of our algorithm on the TEST 2 data to those of another algorithm [5] based on dynamic time warping and the mel-based cepstrum, and to results of classification of the sounds by three engineers. It is interesting to note that especially at the 5dB signal-to-noise ratio, the humans' classification performance was worse than that of the best automatic classification algorithms.

## 4 SUMMARY AND CONCLUSIONS

An application of Hidden Markov Model methods to the classification of acoustic transients was described. The classification system was based on a HMM using spectral features. Simple noise correction methods were tested and three methods of codebook initialization using the Lloyd algorithm were explored. The procedure called *limiting* slightly improves the classification results, while giving a significant reduction in the time to produce a codebook.

Results were reported on testing with two different data sets. The HMM system with features representing the output of a bank of bandpass filters produced high percentages of correct classification over a wide variety of conditions. Noise correction was found to slightly improve performance.

### References


