# A MULTI-RESOLUTION HIDDEN MARKOV MODEL USING CLASS-SPECIFIC FEATURES

*Paul M. Baggenstoss*

Naval Undersea Warfare Center
Newport RI, 02841, USA
phone: (+001) 401-832-8240
email: p.m.baggenstoss@ieee.org
web: www.npt.nuwc.navy.mil/csf

## ABSTRACT

We address the problem in signal classification applications, such as automatic speech recognition (ASR) systems that employ the hidden Markov model (HMM), that it is necessary to settle for a fixed analysis window size and a fixed feature set. This is despite the fact that complex signals such as human speech typically contain a wide range of signal types and durations. We apply the probability density function (PDF) projection theorem to generalize the hidden Markov model (HMM) to utilize a different features and segment length for each state. We demonstrate the algorithm using speech analysis so that long-duration phonemes such as vowels and short-duration phonemes such as plosives can utilize feature extraction tailored to the their own time scale.

## 1. INTRODUCTION

The Hidden Markov Model (HMM) [1] combined with spectral analysis using cepstral coefficients [2] on fixed-length analysis windows remains at the forefront of automatic speech recognition (ASR) technology. One problem with this architecture is the necessity of using a fixed analysis window size. This constraint is a problem because in speech and other natural processes, the various phenomena that are being tested (such as phonemes in speech) may occur with differing time scale. The window size used on speech analysis is a compromise between phonemes with long time scale such as vowels and phonemes with shorter time scales such as plosives. The need for a fixed-size window arises from the fundamental probabilistic approach that underlies the method and depends on the comparison of likelihood functions formed on a common feature space. One could not directly compare two likelihood functions if they are defined on different feature spaces. Even if pains are taken to normalize the behavior of similar features obtained from differing-size data windows, the fundamental basis for comparison is suspect.

With the introduction of the class-specific feature theorem [3], [4], [5], and later the probability density function (PDF) projection theorem (PPT) [6], the freedom now exists to use a different feature set for each class, even for each state in a HMM [7], and as we now show, different analysis window lengths for each state. Thus, the topic of this paper is to apply the PPT to the problem of using varying-size analysis windows within the framework of a HMM.

## 2. THE HMM AND MULTI-RESOLUTION HMM (MRHMM) ON RAW DATA

We assume familiarity with hidden Markov models (HMMs). A good reference is an article by Rabiner [1] from which we borrow notation. If we ignore the effects of overlapped processing, the underlying assumption when a time-series is segmented for processing is that the data in two different segments are conditionally statistically independent (CSI). In other words, the data in two segments are statistically independent conditioned on the system states in the two segments being known. The CSI property enables the efficient calculation of the joint PDF using the forward procedure. Let there be a raw data time-series, denoted by $\mathbf{X}$, consisting of an integer multiple of $T$ samples, where $T$ is the basic time quantization. The traditional approach, which we describe simply as the HMM, is to divide the data into uniform $T$-sample segments which are to be processed separately. Let $\mathbf{x}_t$ represent the data in time-step $t$ consisting of data samples $1+(t-1)T$ through $tT$. In the HMM, it is assumed that:

1. during any $T$-sample segment, the data is governed by one of $M$ possible states.
2. any two samples, no matter how close together , that are contained in two different segments, are CSI.

For the MRHMM, however, we assume that:

1. during any $T$-sample segment, the data is governed by one of $M$ possible states.
2. for each state $s$ , there is an associated minimum time duration. Once the system transitions to state $s$, it must remain in that state for $nK_sT$ samples, where $K_s$ is the integer minimum duration parameter for state $s$, and $n \geq 1$.
3. Two data samples $x_i$ and $x_j$ are assumed to be CSI and are processed separately if (a) the system has made at least one state transition between times $i$ and $j$, or (b) the system has been continually in the same state $s$ but samples $i$ and $j$ are in different length-$K_sT$ segments. Otherwise, samples $x_i$ and $x_j$ are processed jointly.
4. To allow for the system being in a state for a number of length-$T$ segments not divisible by $K_s$, we define a number of *slave* states, say states $s'$, $s''$, that are slaved to state $s$ in a way to be described, with $K_s > K_{s'} > K_{s''}$.

Let $Q = [s_1, s_2 \ldots s_N]$ be a set of state values, where $1 \leq s_t \leq M,\ 1 \leq t \leq N$. We call $Q$ a *trajectory* because it defines one of the many paths through the state diagram or trellis. Let $p(\mathbf{X}|Q)$ be the likelihood function of the raw data given

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| **2008** | | **00-00-2008 to 00-00-2008** |

| 4. TITLE AND SUBTITLE | 5a. CONTRACT NUMBER |
|---|---|
| **A Multi-Resolution Hidden Markov Model Using Class-Specific Features** | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | 5d. PROJECT NUMBER |
| | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Undersea Warfare Center,Newport,RI,02841** | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S) |
| | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES

14. ABSTRACT
**see report**

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT **unclassified** | b. ABSTRACT **unclassified** | c. THIS PAGE **unclassified** | **Same as Report (SAR)** | **5** | |

the trajectory $Q$. Using the CSI property, we may write

$$\text{HMM}: \quad p(\mathbf{X}|Q) = \prod_{t=1}^{N} p(\mathbf{x}_t|s_t). \tag{1}$$

There are no restrictions on $Q$ except for those restrictions imposed by the initial state probabilities $\pi = \{\pi_m,\ 1 \le m \le M\}$, and the state transition matrix (STM) $\mathbf{A} = \{A_{l,m}\ 1 \le l \le M,\ 1 \le m \le M\}$. For the MRHMM, we can encode all the above restrictions imposed on state transitions by properly structuring $\pi$ and $\mathbf{A}$. For each state $s$, we can define a *partition* of states, which we call *wait states*, of size $K_s$. Let $\mathbf{A}^e$ be the *expanded* MRHMM STM and let $\pi^e$ be the expanded set of prior state probabilities. We structure $\mathbf{A}^e$ so that state transitions *into* the state $s$ partition are only allowed into the first wait state. From the first wait state, the state is forced to increment to the second, third, ... and finally to wait state $K_s$. From wait state $K_s$, the state is allowed to transition to the first wait state of any state partition. Note that although $\mathbf{A}^e$ is dimension $M^e \times M^e$ where $M^e = \sum_{m=1}^{M} K_m$, there are only $M^2$ free parameters in $\mathbf{A}^e$.

At this point, the MRHMM can be seen as nothing more than a HMM with a specially structured $\pi$ and $\mathbf{A}$. But the more important difference, which we will explain below, is in the way that $p(\mathbf{X}|Q)$ is calculated. For the moment, let us talk about our goal. We seek an algorithm to solve the following four problems:

1. Segmentation. Find the most likely trajectory through the trellis subject to the restrictions described above.
2. State probabilities. Determine the *a posteriori* state probabilities $\gamma_{t,m} = p(s_t = m|\mathbf{X})$. This is a more complete description of the trajectories than knowing the single most likely trajectory.
3. Joint PDF. The joint likelihood function of all the data given the model is given by

$$L(\mathbf{x}) = \sum_{Q \in \mathscr{Q}} p(\mathbf{X}|Q)\ p(Q), \tag{2}$$

   where $\mathscr{Q}$ is the set of all possible trajectories and $P(Q)$ is the *a priori* probability of a given trajectory through the trellis. Note that $L(\mathbf{X})$ averages $p(\mathbf{X}|Q)$ over all trajectories through the trellis weighted by the probability of the trajectory. Invalid trajectories have zero contribution.
4. Re-estimation. We would like to estimate the model parameters from the data. Parameters include $\pi$, $\mathbf{A}$, and the parameters $\theta_s$ of the conditional state PDFs $p(\mathbf{x}_t|s,\theta_s)$.

For the HMM, the above problems are solved by the *forward procedure* and the associated *backward procedure* and the Baum-Welch algorithm [1]. For the MRHMM, we need to adapt these algorithms, not only by structuring the $\pi$ and $\mathbf{A}$, but by changing the way that $p(\mathbf{X}|Q)$ is calculated. We will explain by example. Let the first state partition be length 3 ($K_1 = 3$) and let the partition for state $s = 1$ consist of the wait states $q = 1$, $q = 2$, and $q = 3$. Let

$$Q = [4,6,7,\mathbf{1},\mathbf{2},\mathbf{3},4,5,6,7,10,11,12,5,6,7,\mathbf{1},\mathbf{2},\mathbf{3},10]$$

be a particular valid length-20 state trajectory. Being a valid trajectory, wait states $q = 1$ through $q = 3$ occur *only as part of the sequence* $1,2,3$. Here is the point at which the HMM and MRHMM differ. For the HMM, we have

$$p(\mathbf{X}|Q) = p(\mathbf{x}_1|q_1 = 4) \cdots\ p(\mathbf{x}_{20}|q_{20} = 10). \tag{3}$$

We can gather all the state PDF values into the matrix $P_{t,q} = p(\mathbf{x}_t|q)$. Then, (2) may be computed by the well known *forward procedure* [1] operating on $P_{t,q}$ and using parameters $\pi$ and $\mathbf{A}$. To change the HMM into a MRHMM, we need two steps:

**Step 1**. Partial PDF values. For each valid trajectory $Q$ and each state $s$, collect all terms in $p(\mathbf{X}|Q)$ associated with the wait state sequence for state partition $s$ and replace the terms by the *partial PDF value*. Define $p(\mathbf{x}_t^{K_s}|s) = \prod_{i=1}^{K_s}\ p(\mathbf{x}_{t+i-1}|q_i)$, where $q_1 \ldots q_{K_s}$ is the sequence of wait states in the state $s$ partition. Define $\left[p(\mathbf{x}_t^{K_s}|s)\right]^{1/K_s}$ as the *partial PDF value* (the geometric mean of the PDF terms in the sequence). In the above example, the first occurrence of the wait state sequence 1,2,3 is the sequence of terms

$$p(\mathbf{x}_4|q_4 = 1)\ p(\mathbf{x}_5|q_5 = 2)\ p(\mathbf{x}_6|q_6 = 3),$$

which we denote by $p(\mathbf{x}_4^3|s = 1)$. We replace each of the three PDF factors by the partial PDF value $[p(\mathbf{x}_4^3|s = 1)]^{1/3}$. This substitution **does not change the value of** $p(\mathbf{X}|Q)$.

Note that we can accomplish this by changing $P_{t,q}$. Associated with every possible occurrence of partition $s$ sequence is a diagonal line in matrix $P_{t,q}$ of length $K_s$. The diagonal starts with the first wait state of partition $s$ at any time $t$ and ends with the last wait state of partition $s$ at time $t + K_s - 1$. Each such sequence is replaced with the geometric mean as described. The resulting matrix is called the partial PDF matrix $P_{t,q}^p$. Note that applying the *forward procedure* to $P_{t,q}^p$ gives precisely the same result as $P_{t,q}$ *provided* $\pi$ *and* $\mathbf{A}$ *reflect the restrictions to state transitions that were described above*. Matrix $P_{t,q}^p$ if viewed as an image appears to have diagonal "streaks" of constant value.

**Step 2**. Relax CSI assumption. Associated with each streak is a data *analysis window* $\mathbf{x}_t^{K_s}$, which is a segment of data of length $K_s T$ samples ending at sample $(t + K_s - 1)T$. The product of each streak of partial PDF values is a PDF of the data analysis window assuming CSI segments. To relax the CSI assumption within the streak, we replace the partial PDF values with the $K_s$ root of the analysis window PDF processed as a unit. Let $P_{t,q}^f$ represent the "full window" partial PDF values created this way. The value of $L(\mathbf{X})$ calculated by the forward procedure operating on $P_{t,q}^f$ changes, however, it remains a valid joint PDF of $\mathbf{X}$. We know this because all we have done is replace the the conditional PDFs $P(\mathbf{X}|Q)$ assuming all the segments are independent with another PDF that assumes statistical dependence within the wait state sequences associated with a given state.

At this point we have a raw-data based MRHMM model that we can compute efficiently using the forward procedure operating on $P_{t,q}^f$. To create a feature-based MRHMM model, we need only to apply the PPT.

## 3. CLASS-SPECIFIC MULTI-RESOLUTION CLASS-SPECIFIC (CS-MRHMM)

The standard feature-based HMM is the same as the raw-data based HMM with the raw data segments $\mathbf{x}_t$ replaced by the feature vector

$$\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2 \ldots \mathbf{z}_N\},\ \mathbf{z}_t = T(\mathbf{x}_t).$$

With this simple replacement, the forward procedure computes the feature-based likelihood function

$$L(\mathbf{Z}) = \sum_{Q \in \mathcal{Q}} p(\mathbf{Z}|Q)\, P(Q), \tag{4}$$

For the CS-MRHMM, we need to use the PPT to transition to the feature domain. Let

$$\mathbf{x}_t^K = [x_{1+(t-1)T} \ldots x_{(t+K-1)T}],$$

be the length $KT$ sample *analysis window* which starts at sample $1+(t-1)T$. It includes segments $\mathbf{x}_t$ through $\mathbf{x}_{t+K-1}$. The term $p(\mathbf{x}_t^K|s)$ will be calculated using the PDF projection theorem [6]. As we have written several publications on the topic including the tutorial article [8], we describe the method only briefly. Let $\mathbf{x}$ be a general segment of raw time-series data. Let $\mathbf{z}_s = T_s(\mathbf{x})$ be a feature set calculated from $\mathbf{x}$ specifically designed for state $s$. Let $\hat{p}(\mathbf{z}_s|s)$ be a PDF estimate of the feature set $\mathbf{z}_s$ based on training data from state $s$. The feature likelihood function is *projected* from the feature space to the raw data by pre-multiplying by the J-function as follows:

$$\hat{p}_p(\mathbf{x}|s) = J(\mathbf{x}; T_m, H_{0,s})\ \hat{p}(\mathbf{z}_s|s). \tag{5}$$

The function $\hat{p}_p(\mathbf{x}|s)$ can be regarded as a function only of $\mathbf{x}$ by substituting $T_s(\mathbf{x})$ for $\mathbf{z}_s$ and can be shown to integrate to 1 over $\mathbf{x}$ (thus it is a PDF). The J-function is a unique function of $\mathbf{x}$ determined precisely from the feature transformation $T_s$ and the class-dependent reference hypothesis $H_{0,s}$:

$$J(\mathbf{x}; T_s, H_{0,s}) = \frac{p(\mathbf{x}|H_{0,s})}{p(\mathbf{z}_s|H_{0,s})}. \tag{6}$$

Since $J(\mathbf{x}; T_s, H_{0,s})$ is determined *a priori* without regard to training data, it can be considered the *untrained* part of $\hat{p}_p(\mathbf{x}|s)$, while $\hat{p}(\mathbf{z}_s|s)$ is the trained part.

While it is true that $\hat{p}_p(\mathbf{x}|s)$ is a PDF, it is only an estimate of $p(\mathbf{x}|s)$. The degree to which $\hat{p}_p(\mathbf{x}|s)$ is a good estimate of $p(\mathbf{x}|s)$ depends on (a) the accuracy of $\hat{p}(\mathbf{z}_s|s)$ and (b) the degree to which $\mathbf{z}_s$ is a *sufficient statistic* for the binary hypothesis test between $s$ and $H_{0,s}$. In the rare case that $\mathbf{z}_s$ is in fact a sufficient statistic, the accuracy of $\hat{p}_p(\mathbf{x}|s)$ depends only upon the accuracy of the low-dimensional PDF estimate $\hat{p}(\mathbf{z}_s|s)$. The J-function takes many forms [6], one of which can be used when $\mathbf{z}_s$ are maximum likelihood (ML) estimates of a set of parameters. In this case, $J(\mathbf{x}; T_s, H_{0,s})$ has a simple form based on the Fisher's information matrix [6].

## 4. PRACTICAL IMPLEMENTATION DETAILS

Let us briefly review what we have done so far. We have described how to compute the likelihood function of the CS-MRHMM. To do this, we identify every time-shifted analysis window. For each state $s$ and time step $t$, we identify the analysis window that starts at time step $t$ and is of length $K_s T$ samples. On this window, we extract the state-dependent feature set, then use the PPT to compute the raw-data PDF of the analysis window. We then take the $K_s$ root of the PDF value and insert this value into the length $K_s$ diagonal streak in the matrix $P_{t,q}^f$. Then, we apply the well-known forward procedure using the expanded parameters $\pi^e$, $\mathbf{A}^e$. When $K_s$ is large, this requires a highly overlapped set of windows. The amount of processing required can be mitigated, by recursive processing. For example, the FFT or autocorrelation function (ACF) of a segment can be updated to reflect data that has been shifted out and data that has been shifted in [9]. Applying the MRHMM to real data warrants additional details beyond what has been so far described.

### 4.1 States vs. Signal classes

Let *signal class* refer to a particular signal phenomenon observed in the data. Let *signal state* refer to an instance of a signal class. In the simplest situation, signal class and signal state are synonymous. But, if a signal class is observed to repeat, additional signal states may be used to represent the additional occurrences. These in turn give rise to additional partitions.

### 4.2 Slave Partitions

We have already introduced the notion of slave partitions (slave states). Up to now, this has only meant the necessity of adding additional states with lower $K$. To compute $L(\mathbf{X})$ with the forward procedure, there is nothing else that needs to be done. However, to train the parameters, we will need to discuss the process of *ganging* states.

### 4.3 Training the CS-MRHMM.

In the standard Baum-Welch algorithm for re-estimation of HMM parameters [1], the state feature PDFs for state $s$ are trained by maximizing log-likelihood functions weighted by $\gamma_{s,t}$. Since the standard HMM does not differentiate between wait states, we would need to a separate PDF estimate for each wait state. However, for the CS-MRHMM, there are only PDF estimates associated with the initial wait states, the first wait states of each partition. Logically, the CS-MRHMM produces values of $\gamma_{q,t}$ that are constant in diagonal streaks in a partition. That is, $\gamma_{q,t} = \gamma_{q+1,t+1}$ if wait states $q$ and $q+1$ are in the same partition. Thus, in the CS-MRHMM, each analysis window can be traced to a given constant-valued streak in the $\gamma_{q,t}$ matrix. When training the CS-MRHMM, the features from the associated analysis window are weighted by the corresponding value of $\gamma_{q,t}$ in the streak. Training becomes slightly more complicated, however, once we consider slave partitions and if the number of signal states exceeds the number of signal classes. While each partition is associated with a PDF estimate, we may not want all partition PDF estimates to be independent. To remedy this situation, we "gang together" all partitions that associate with a given signal class. To gang partitions, we first create a compressed version of $\gamma_{q,t}$, denoted by $\gamma_{m,t}^c$ , which sums $\gamma_{q,t}$ over all wait states associated with signal class $m$. Then we then weight an analysis window by the *smallest* value of $\gamma_{m,t}$ in the set of time steps $t$ contained in the analysis window. This works very well in practice but is a clear departure from the Baum-Welch algorithm and may produce an algorithm without guaranteed monotonicity.

### 4.4 Efficient Implementation

The number of wait states in the expanded HMM problem can be very large. The forward and backward procedures have a complexity of the order of the square of the number of states. Thus, an efficient implementation of the forward and backward procedures and Baum-Welch algorithm may

be needed that takes advantage of the redundancies in the expanded problem. We have obtained a time reduction factor of 42 with a problem that had 7 signal classes and expanded to 274 wait states. The two algorithms were tested to produce the same results within machine precision.

## 5. EXAMPLES

### 5.1 Simulated Data

To illustrate the concepts, we tested the concept of the CS-MRHMM using simulated data. To independent identically distributed (iid) Gaussian noise, we added a low frequency (LF) pulse of autoregressive (AR) process of 128 samples in length with a peak frequency response of 0.4 radians per sample, followed by a random-length gap of at least 256 samples, followed by high frequency (HF) pulse of AR process of 64 samples with a peak frequency response of 1.2 radians per sample. An example of the signal and noise is shown in Figure 1. We implemented the HMM with three signal states,

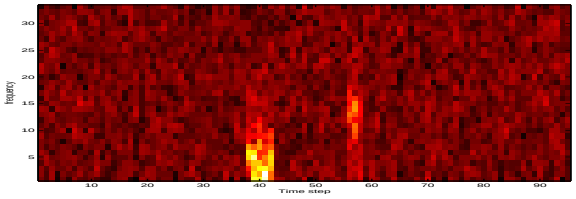


Figure 1: Example of spectrogram of synthetic data. The data consists of three signal classes. Class 1 (noise) occurs first, then a low-frequency pulse of duration 128 samples, then noise, then a high-frequency pulse of duration 64 samples.

each corresponding to a signal class : "noise", "LF pulse", and "HF pulse". We used nine partitions including six slave partitions. The elemental segment length was $T = 32$ samples. There were a total of 25 wait states. Parameters of the nine partitions are listed in table 1. Autoregressive (LPC)

| Partition | Signal class | $KT$ | $K$ | $P$ |
|---|---|---|---|---|
| 1 | Noise | 256 | 8 | 4 |
| 2 | Noise | 128 | 4 | 4 |
| 3 | Noise | 64 | 2 | 4 |
| 4 | Noise | 32 | 1 | 3 |
| 5 | LF Pulse | 128 | 4 | 4 |
| 6 | LF Pulse | 64 | 2 | 4 |
| 7 | LF Pulse | 32 | 1 | 3 |
| 8 | HF Pulse | 64 | 2 | 4 |
| 9 | HF Pulse | 32 | 1 | 3 |

Table 1: Partition parameters for the illustrative example. $K$ is the partition length in elemental segments. $KT$ is the length of the partition in samples. Parameter $P$ is the autoregressive (AR) model order (same as LPC model order).

features of model order $P$ (see table 1) were extracted by overlapped window processing. A separate feature processor was used for each combination of $K$ and $P$. Features were shared between partitions that had the same $K$ and $P$ values. Analysis windows were shifted always by the elemental segment length of 32 samples for each update, so the amount of overlap depended on the length of the analysis window. To handle end effects, data was assumed to wrap around in time.

Features were extracted from each analysis window by first taking the FFT, computing the magnitude squared, then computing the inverse-FFT to produce the autocorrelation function (ACF). The Levinson algorithm was used to produce the reflection coefficients of order $P$. The total power in window is also stored as the $P+1$st feature. The J-function [6] is obtained by use of the saddle-point approximation [10]. Further details of the implementation details of the AR models can be found in [8].

In Figure 2, we see the partial PDF matrix $P^f_{i,m}$ for a typical sample. Wait states $q = 1$ through $q = 15$ are associated with the "Noise" signal class. wait states $q = 16$ through $q = 22$ are associated with the "LF Pulse" signal class, and wait states $q = 23$ through $q = 25$ are associated with the "HF Pulse" signal class. The gamma probabilities are a by-

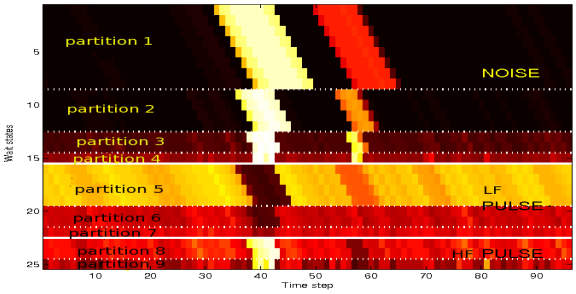


Figure 2: Partial PDF matrix $P^f_{i,m}$ showing devisions between signal classes (solid horizontal lines) and between wait state partitions (dotted lines). Higher probability is darker.

product of the Baum Welch algorithm [1] and indicate the relative probability of each wait state given the data. The gamma probabilities corresponding to figure 2 are shown in Figure 3. This figure can be interpreted as the trajectory through figure 3 that pick up the highest probabilities while meeting the restrictions set by the state transition matrix.

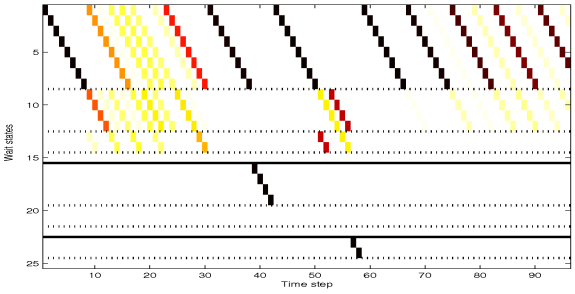


Figure 3: Wait state probabilities for illustrative example.

Note that the wait states for "LF pulse" ($q = 16$ through $q = 19$) are clearly seen where the pulse occurs. The same is true of the "HF pulse" event ($q = 23$ through $q = 24$) . It is possible to see various competing trajectories through the trellis. Note for example in time steps 43 through 56, the noise gap between the two pulses, the HMM is in the noise signal class. In steps 43-50, it is in partition 1, (wait states $q = 1$ through $q = 8$). Then after exiting wait state

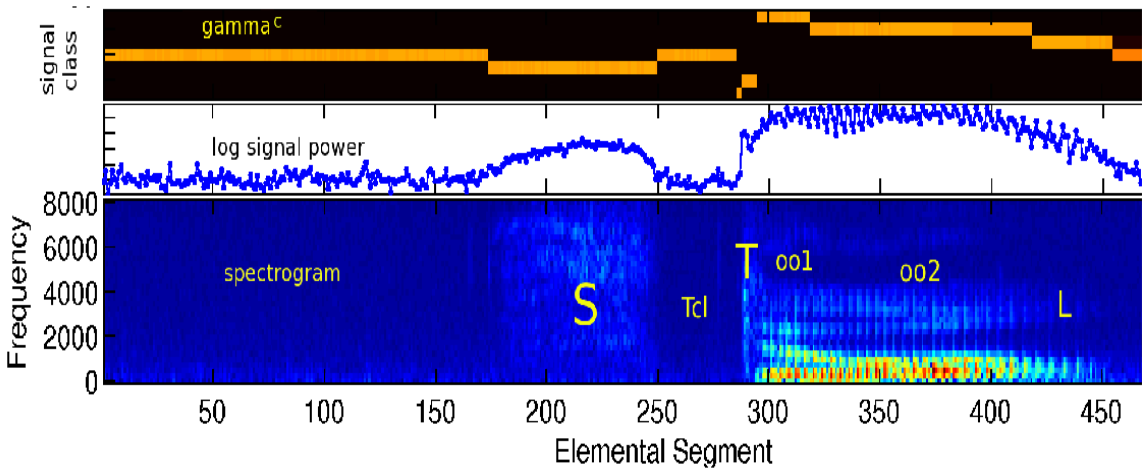


Figure 4: Example of CS-MRHMM operating on the word "stool". From top to bottom: compressed gamma probabilities $\gamma^c_{m,t}$, log signal power, and spectrogram. Short analysis windows have been employed for the "T", while longer processing has been used for background noise and the sounds "S", "oo" and "L". The three components of the "T" can be clearly identified.

$q = 4$, it has located two possibilities to span the six time steps remaining before HF pulse occurs. It can either go into partition 2 (wait states $q = 9$ through $q = 12$) then partition 3 (wait states $q = 13$ through $q = 14$), or it can choose the reverse, partition 3 then partition 2.

The gamma probabilities can be collapsed to indicate just the signal classes, as shown in Figure 5. The class probabili-

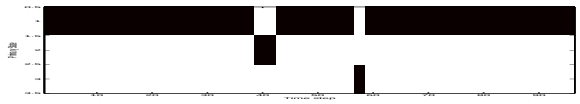

Figure 5: Signal class probabilities calculated by summing figure 3 over the wait states of each class.

ties (figure 5) is an accurate indication of the true content of the data to a time resolution of $T = 32$ samples.

### 5.2 Speech Data

We used the CS-HMM to analyze the spoken word "stool" at 16 kHz sample rate. Space restrictions do not permit a detailed description of the experiment. We identified seven signal classes and assigned values of $K$ and $P$ (LPC order) (1) **Noise** used for both background and the "T" closure: $K = 12$ or 384 samples, $P = 7$, (2) **"S"** : $K = 12$ or 384 samples, $P = 7$, (3) **"T" Burst** : $K = 4$ or 128 samples, $P = 5$, (4) **"T" Aspiration** : $K = 8$ or 256 samples, $P = 6$, (5) **"oo" vowel part 1**: $K = 24$ or 768 samples, $P = 8$, (6) **"oo" vowel part 2**: $K = 24$ or 768 samples, $P = 8$, (7) **"L"** : $K = 24$ or 768 samples, $P = 8$. After adding slave partitions, we had a total of 36 partitions and a total of 258 wait states. The expanded STM was 258 by 258. Using efficient programming, neither the partial probability matrix nor the expanded STM actually need to be created. Figure 4 shows the result of analysis of one example with the CS-MRHMM. Important to note is that the three components of the "T" can be clearly seen by observing $\gamma^c_{m,t}$.

## REFERENCES


[1] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257–286, February 1989.

[2] J. W. Picone, "Signal modeling techniques in speech recognition," *Proceedings of the IEEE*, vol. 81, no. 9, pp. 1215–1247, 1993.

[3] P. M. Baggenstoss, "Class-specific features in classification.," in *IASTED International Conference on Signal and Image Processing*, 1998.

[4] S. Kay, "Sufficiency, classification, and the class-specific feature theorem," *IEEE Trans. Information Theory*, vol. 46, pp. 1654–1658, July 2000.

[5] P. M. Baggenstoss, "Class-specific features in classification.," *IEEE Trans Signal Processing*, pp. 3428–3432, December 1999.

[6] P. M. Baggenstoss, "The PDF projection theorem and the class-specific method," *IEEE Trans Signal Processing*, pp. 672–685, March 2003.

[7] P. M. Baggenstoss, "A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces.," *IEEE Trans. Speech and Audio*, pp. 411–416, May 2001.

[8] P. M. Baggenstoss, "The class-specific classifier: Avoiding the curse of dimensionality (tutorial)," *IEEE Aerospace and Electronic Systems Magazine, special Tutorial addendum*, vol. 19, pp. 37–52, January 2002.

[9] P. Baggenstoss, "Time-series segmentation," *United States Patent 6907367*, June 2005.

[10] S. M. Kay, A. H. Nuttall, and P. M. Baggenstoss, "Multidimensional probability density function approximation for detection, classification and model order selection," *IEEE Trans. Signal Processing*, pp. 2240–2252, Oct 2001.