

# Hidden Markov Modeling-Based Decision-Making Using Short-Length Sensor Time Series

**Najah F. Ghalyan**

Department of Mechanical Engineering,  
The Pennsylvania State University,  
University Park, PA 16802  
e-mail: nfg103@psu.edu

**Sudeepta Mondal**

Department of Mechanical Engineering,  
The Pennsylvania State University,  
University Park, PA 16802  
e-mail: sbm5423@psu.edu

**David J. Miller**

Professor  
Department of Electrical Engineering,  
The Pennsylvania State University,  
University Park, PA 16802  
e-mail: djmiller@enr.psu.edu

**Asok Ray<sup>1</sup>**

Distinguished Professor  
Fellow ASME  
Department of Mechanical Engineering and Mathematics,  
The Pennsylvania State University,  
University Park, PA 16802  
e-mail: axr2@psu.edu

*Real-time decision-making (e.g., monitoring and active control of dynamical systems) often requires feature extraction and pattern classification from short-length time series of sensor data. An example is thermoacoustic instabilities (TAI) in combustion systems, caused by spontaneous excitation of one or more natural modes of acoustic waves. The TAI are typically manifested by large-amplitude self-sustained pressure oscillations in time scales of milliseconds, which need to be mitigated by fast actuation of the control signals, requiring early detection of the forthcoming TAI. This issue is addressed in this technical brief by hidden Markov modeling (HMM) and symbolic time series analysis (STSA) for near-real-time recognition of anomalous patterns from short-length time series of sensor data. An STSA technique is first proposed, which utilizes a novel HMM-based partitioning method to symbolize the time series by using the Viterbi algorithm. Given the observed time series and a hidden Markov model, the algorithm generates a symbol string with maximum posterior probability. This symbol string is optimal in the sense of minimizing string error rates in the HMM framework. Then, an HMM likelihood-based detection algorithm is formulated and its performance is evaluated by comparison with the proposed STSA-based algorithm as a benchmark. The algorithms have been validated on a laboratory-scale experimental apparatus. The following conclusions are drawn from the experimental results: (1) superiority of the proposed STSA method over standard methods in STSA for capturing the dynamical behavior of the underlying*

<sup>1</sup>Corresponding author.

Contributed by the Dynamic Systems Division of ASME for publication in the JOURNAL OF DYNAMIC SYSTEMS, MEASUREMENT, AND CONTROL. Manuscript received October 11, 2018; final manuscript received March 28, 2019; published online May 8, 2019. Assoc. Editor: Youngsu Cha.

*process, based on short-length time series and (2) superiority of the proposed HMM likelihood-based algorithm over the proposed STSA method for different lengths of sensor time series.*  
[DOI: 10.1115/1.4043428]

*Keywords: sensor-based automation, hidden Markov modeling, symbolic time series analysis, combustion instability*

## 1 Introduction

Sensor-based automation in cyber-physical systems [1] often requires real-time decision-making for early detection of anomalous behavior in the physical process, whose relevant features are extracted from measurement data for pattern classification. This is also true for dynamic data driven application systems, where the computational and instrumentation aspects of an application are dynamically incorporated to enable reliable and fast modeling of the characteristics and behaviors of the underlying process [2]. An application example is monitoring and near-real-time active control of thermoacoustic instabilities (TAI) in combustion systems, which are usually caused by spontaneous excitation of one or more natural modes of the acoustic waves [3]. The TAI phenomena are typically manifested by large-amplitude self-sustained (possibly chaotic) pressure oscillations in the combustion chamber [4], which may lead to damage in mechanical structures if the pressure oscillations match one of the natural frequencies of the system. The time scales of TAI are on the order of milliseconds, which must be mitigated by fast actuation of control signals. This mandates early detection of instabilities from short-length sensor data. Along this line, Sarkar et al. [5] have used symbolic time series analysis (STSA)-based algorithms [6,7] for online detection of lean blowout in combustion systems. For classification of different operational regimes in combustion systems, Mondal et al. [8] have proposed a hidden Markov modeling (HMM)-based algorithm that relies on short-length time series.

This technical brief develops STSA and HMM-based algorithms for anomaly detection using short-length sensor time series. These algorithms are executable in real time and require training a nominal (null) model only, which can be performed in an online fashion. The developed algorithms have been validated on a laboratory-scale apparatus for early detection of TAI. However, the underlying algorithms are applicable for decision-making in dynamical systems in general, where fast detection/decision-making is highly desirable. In this context, major contributions of the work, reported in this technical brief, are delineated as follows:

- Hidden Markov model likelihood-based algorithm for feature extraction and pattern classification:* This algorithm is built upon short-length time series to facilitate real-time monitoring and control in cyber-physical systems [1].
- Hidden Markov model-based partitioning for STSA:* This partitioning method makes use of the (dynamic programming-based) Viterbi algorithm [9] to convert a time series into a symbol string with the maximum posterior probability.
- Validation with experimental data:* The anomaly detection algorithms are validated on a laboratory-scale experimental apparatus for early detection of TAI in combustion systems.

## 2 Background Information

This section provides the background information on HMM [9] and STSA [6,7] for anomaly detection in dynamical systems.<sup>2</sup>

**2.1 Hidden Markov Modeling.** Hidden Markov modeling has found its applications in diverse fields [8,10]. While the theory of HMM is presented in detail in Refs. [9] and [10], the key

<sup>2</sup>An anomaly is a deviation of the behavior of a physical process from its nominal (i.e., healthy) condition; it often evolves as a result of gradual degradation in the system characteristics (e.g., structural durability). Early detection of anomalies is essential for enhancing the system performance as well as for mitigation of potential catastrophic failures.

concepts are very succinctly outlined below for the sake of completeness.

Consider a time series  $X = \{x_1, x_2, \dots, x_T\}$ ,  $x_n \in \mathbb{R}^N$ . In HMM framework,  $X$  is considered as a realization of a hidden Markov chain  $Z = \{z_1, z_2, \dots, z_T\}$ , where  $z_t$  is one of the finitely many states  $\mathcal{Q} \triangleq \{q_1, \dots, q_M\}$  at time  $t$ . In this setting, the HMM is a triple  $\lambda = \{A, B, \pi\}$ , where  $A = [a_{ij}] \in \mathbb{R}^{M \times M}$  with  $a_{ij} = p(z_{t+1} = q_j | z_t = q_i)$ ;  $B = [b_j(x)]$  with  $b_j(x) = p(x | z = q_j)$ ; and  $\pi = [\pi_i] \in \mathbb{R}^{1 \times M}$  with  $\pi_i = p(z_1 = q_i)$  at the initial time  $t = 1$ . It is noted that each row of  $A$  and the row vector  $\pi$  are probability mass functions (i.e., all elements of  $\pi$  are non-negative and sum to 1 and each row in  $A$  has non-negative elements and sums to 1), while each element of  $B$  is a state-conditional probability density function defined for any observation  $x$ .

Given an observed data string  $X$  of length  $T$ , there are  $M^T$  possible hidden state paths (of length  $T$ ) that could generate  $X$ ; let  $\mathcal{Z}$  be the set of all such hidden state paths. To compute the joint likelihood of  $X$  given an HMM  $\lambda$ , it is necessary to marginalize over all these hidden state paths

$$p(X|\lambda) = \sum_{Z \in \mathcal{Z}} p(X, Z|\lambda) \quad (1)$$

To mitigate this computational complexity, the forward variable  $\alpha$  is introduced as

$$\alpha_n(i) \triangleq p(x_1 x_2 \dots x_n, z_n = q_i | \lambda) \quad (2)$$

This variable is recursively computed by using the *forward procedure* [10], and the observation likelihood  $p(X|\lambda)$  is evaluated as

$$p(X|\lambda) = \sum_{i=1}^M \alpha_T(i) \quad (3)$$

The *Baum–Welch* algorithm [10] is a standard (expectation maximization) tool for training HMMs; it finds a triple  $\lambda^* = \{A, B, \pi\}$  that locally maximizes the total observation likelihood

$$\lambda^* = \underset{\lambda}{\operatorname{argmax}} \{p(X|\lambda)\} = \underset{\lambda}{\operatorname{argmax}} \sum_{Z \in \mathcal{Z}} p(X, Z|\lambda) \quad (4)$$

Also, the most probable hidden state path for a given data string  $X$  is computed as

$$Z^*(X) = \underset{Z \in \mathcal{Z}}{\operatorname{argmax}} [p(Z|X, \lambda^*)] \quad (5)$$

by using the (dynamic programming-based) *Viterbi algorithm* [10]. Both Eqs. (3) and (5) are used here to define anomaly detection algorithms.

**2.2 Symbolic Time Series Analysis.** Symbolic time series analysis [11,12] has found its applications in anomaly detection and pattern recognition [6,7,13]. A crucial part in STSA is to convert a time series  $X = \{x_1, \dots, x_T\}$ ,  $x_t \in \mathbb{R}^N$ , into a symbol string  $S = \{s_1, \dots, s_T\}$ ,  $s_i \in \mathcal{A}$ ,  $\mathcal{A}$  a (finite cardinality) alphabet of symbols.<sup>3</sup> This conversion is commonly performed by partitioning the state space  $\mathbb{R}^N$  into  $|\mathcal{A}|$  disjoint regions, with each region assigned a distinct symbol from  $\mathcal{A}$ . Then each point  $x_t$  in the time series is converted into the symbol associated with the region that contains  $x_t$ . Examples of symbolization using state space partitioning are uniform partitioning [15], maximum entropy partitioning (MEP) [15,16], and  $K$ -means [17]. In uniform partitioning, the state space is partitioned into regions with equal Lebesgue measures, while in MEP the state space is partitioned into regions such that the entropy is maximized. In  $K$ -means, the state space is partitioned into  $K$  Voronoi cells such that a distortion measure-based

<sup>3</sup>It is noted that the alphabet size  $|\mathcal{A}|$  should be much smaller than the time series length  $T$  [14].

objective function is minimized [17]. In state space partitioning, the symbol at each time  $t$  is chosen based on the point  $x_t$  only. This may be suboptimal for (possibly nonautonomous) dynamical systems, where the same point in the state space may recur multiple times; yet, for each such occurrence the point may be part of a quite distinct temporal pattern. Using state space partitioning, the time series must be assigned the same symbol at these “recurrence” times, with the symbol sequence at time  $t$ ,  $s_t$ , only providing information about  $x_t$ . Therefore, despite the convenience of state space partitioning for symbolizing time series, it could lead to a symbol string that may not accurately represent the dynamical properties of the underlying process. In this regard, Ghalyan et al. [18] reported a symbolization algorithm that minimizes a clustering objective function to *jointly* convert the time series into a symbol sequence without partitioning the state space. Although this approach is optimal in the sense of estimating a generating partition for the observed time series [18], it is generally suboptimal in the HMM framework [10].

The  $D$ -Markov machine [6,7] provides an efficient tool to model the stationary dynamics of the resulting symbol string for pattern recognition, where the Markov depth  $D$  is a positive integer. The concept of  $D$ -Markov machine relies on an algebraic structure, called the (irreducible) probabilistic finite state automaton (PFSA)  $\mathcal{K} = (\mathcal{A}, \mathcal{Q}, \delta, \mathcal{M})$ , where  $\mathcal{A}$  is a finite cardinality alphabet as defined earlier,  $\mathcal{Q}$  is a finite set of states,  $\delta : \mathcal{Q} \times \mathcal{A} \rightarrow \mathcal{Q}$  is a state transition map, and  $\mathcal{M} : \mathcal{Q} \times \mathcal{A} \rightarrow [0, 1]$  generates the individual entries of the emission (also called morph) matrix [7]. The parameters  $\delta$  and  $\mathcal{M}$  are used to construct the  $(|\mathcal{Q}| \times |\mathcal{Q}|)$  state transition probability matrix which, in turn, generates the state probability vector  $P$  as the (sum-normalized) left eigenvector corresponding to the unique eigenvalue 1 [7]. The  $D$ -Markov machine is a PFSA corresponding to a stochastic symbolic stationary process for which the probability of the current symbol depends only on the previous (at most)  $D$  consecutive symbols.

In summary, for anomaly detection using STSA [13], a time series  $X$  of sensor data is first converted into a symbol string. Then, PFSA's are constructed from the symbol strings, which in turn generate low-dimensional feature vectors [15] that are used for detection of anomalous patterns. The procedure is executed in the following steps:

- (1) *Select* a block of a time series, called the nominal block, for which the system is in a healthy condition.
- (2) *Construct* a partition for the nominal block and convert it into a symbol string to construct the nominal PFSA model. The emission matrix of the PFSA model is computed by frequency counting [7]. This *learned* nominal model generates a (quasi-)stationary probability vector  $P_{\text{nom}}$  that represents the nominal pattern.
- (3) *Select* a new block of the time series up to the current time  $t$  and convert it into a symbol string using the learned nominal partition. This yields a new PFSA with a new (quasi-)stationary probability vector  $P_t$  that represents the feature vector of the system at time  $t$ .
- (4) *Compute* the anomaly at time  $t$  as the divergence between the nominal feature and current feature vectors

$$\mu(t) = d(P_{\text{nom}}, P_t) \quad (6)$$

where  $d(\cdot, \cdot)$  is the Kullback–Leibler divergence [17].

### 3 Technical Approach

This section first presents a novel partitioning method that symbolizes the time series in an optimal fashion without partitioning the state space. In particular, the time series  $X$  is considered as a realization of a stochastic process that is represented by an HMM,  $\lambda$ , with the alphabet  $\mathcal{A}$  of hidden states, and the HMM

parameters are estimated by using the expectation maximization (Baum–Welch) method. Then, out of  $|\mathcal{A}|^T$  possible symbol strings, the algorithm identifies the one with maximum posterior probability, i.e.,  $S^* = \operatorname{argmax}_S P(S|X, \lambda)$ , by using the Viterbi algorithm. This string is optimal in the sense of minimizing the string error rate in the HMM framework [10]. The string is also expected to extract more information from  $X$  about the underlying dynamical system and to have more power to capture sequential patterns in  $X$  than MEP and  $K$ -means. Unlike standard state space partitions, the proposed partitioning *jointly* symbolizes the entire time series, with  $s_t$  at each time  $t$  providing some information about the entire time series.

The proposed partitioning has been used to develop an STSA technique for anomaly detection. This method is denoted as  $\text{HMM}_D$ , as an abbreviation for HMM  $D$ -Markov machine, whose pseudocode is given in Algorithm 1, where the parameter  $\tau$  is a user-specified threshold, set to achieve a specified false positive rate.<sup>4</sup>

---

#### Algorithm 1 HMM $D$ -Markov ( $\text{HMM}_D$ ) method

---

INPUT: Threshold  $\tau$  and a data block  $x_{t:t+L}$ .

OUTPUT: The decision on whether the system is nominal or anomalous.

- 1: Initiate the algorithm using a nominal block of data  $x_{t_0:t_0+L}^*$  to find the nominal model:

$$\lambda^* = \operatorname{argmax}_{\lambda} \{p(x_{t_0:t_0+L}^* | \lambda)\}$$

- 2: Use Viterbi algorithm to find the hidden state path:

$$z_{t_0:t_0+L}^* = \operatorname{argmax}_{z_{t_0:t_0+L}} \{p(z_{t_0:t_0+L} | x_{t_0:t_0+L}^*, \lambda^*)\}$$

- 3: Using  $D$ -Markov machine, construct a PFSA based on  $z_{t_0:t_0+L}^*$  as the symbol string to obtain the nominal pattern (i.e., steady-state probability vector)  $P^*$ .
  - 4: Apply Steps 2 and 3, with  $x_{t_0:t_0+L}^*$  replaced by  $x_{t:t+L}$ , to find the pattern (i.e., state probability vector)  $P_t$ .
  - 5: Compute the anomaly statistic  $\mu(t) \leftarrow d(P_t, P^*)$
  - 6: **if**  $\mu(t) > \tau$  **then**
  - 7:     declare the system as anomalous
  - 8: **else**
  - 9:     declare the system as nominal
  - 10: **end if**
- 

The  $\text{HMM}_D$  method outperforms other STSA methods as demonstrated in Sec. 4. However,  $\text{HMM}_D$  makes use of the most probable hidden state path only and ignores all other possible paths. This is a common drawback in STSA which uses a *hard* symbol assignment to convert the time series into a symbol string, rejecting all other possible symbol strings. Some of these rejected symbol strings may involve useful information about the underlying dynamical system, which is not captured by the selected symbol string.

Alternatively, another HMM-based detection algorithm is proposed, which retains all possible hidden state paths. In particular, an HMM null model  $\lambda^*$  is trained by using data from the nominal condition and, for each subsequent block  $x_{t:t+L}$ , the anomaly measure is given by the negative log-likelihood

$$\mu(t) \triangleq -\log [p(x_{t:t+L} | \lambda^*)] \quad (7)$$

where Eq. (7) is obtained by summing over all hidden state paths, as given in Eq. (1). In the sequel, this method is called  $\text{HMM}_L$  as

<sup>4</sup>The receiver operating characteristic (ROC) curves have been used in the current paper for assessing the detection performance by varying the parameter  $\tau$  from  $-\infty$  to  $\infty$  [19], where each point in the ROC curve corresponds to a specific value of  $\tau$ . Therefore, the threshold can be determined from the ROC curve by specifying a maximum allowable false positive rate, which may depend on the application. If the cost of a positive false alarm is low, the maximum FPR could be increased. On the other hand, for applications where the cost for a positive false alarm is high, a small value for the maximum FPR should be selected.

an abbreviation for HMM likelihood. A pseudocode of  $\text{HMM}_L$  is presented in Algorithm 2. The intuition behind Eq. (7) is as follows. Since the HMM is trained using observations generated in the nominal regime, the likelihood of the time series measurements (after occurrence of an anomaly) conditioned on the nominal HMM should decrease. Based on a properly chosen threshold, one can decide whether change has occurred within the block or not using such likelihoods.

---

#### Algorithm 2 HMM likelihood ( $\text{HMM}_L$ ) method

---

INPUT: Threshold  $\tau$  and a data block  $x_{t:t+L}$ .

OUTPUT: Decision on whether the system is nominal or anomalous.

- 1: Initiate the algorithm using a nominal block of data  $x_{t_0:t_0+L}^*$  to find the nominal HMM:

$$\lambda^* = \operatorname{argmax}_{\lambda} \{p(x_{t_0:t_0+L}^* | \lambda)\}$$

- 2:  $\mu(t) \leftarrow -\log [p(x_{t:t+L} | \lambda^*)]$
  - 3: **if**  $\mu(t) > \tau$  **then**
  - 4:     declare the system as anomalous
  - 5: **else**
  - 6:     declare the system as nominal
  - 7: **end if**
- 

*Remark 3.1.* The method  $\text{HMM}_L$  effectively considers all possible hidden state paths if the Forward algorithm is used [10]. As demonstrated experimentally in Sec. 4, this method is considered richer than  $\text{HMM}_D$  in the sense that the information associated with all possible symbol strings is retained and utilized to extract relevant features from the time series  $X$ .

## 4 Experimental Validation

This section validates the proposed algorithms, namely  $\text{HMM}_L$  and  $\text{HMM}_D$ , on data generated from a laboratory-scale experimental apparatus [8]. The first objective of the experimental validation is to evaluate the performance of the proposed STSA method,  $\text{HMM}_D$ , compared to  $K$ -means and MEP-based STSA methods, for early detection of TAI in combustion systems, based on short-length sensor time series and low-dimensional feature vectors.<sup>5</sup> The second objective of the experimental validation is to demonstrate the efficacy of  $\text{HMM}_L$ , compared to the best STSA method, which is in fact  $\text{HMM}_D$ .

Figure 1 depicts the experimental apparatus that simulates thermoacoustic instabilities in a laboratory environment. It is an electrically heated Rijke tube, an apparatus that has been commonly used by researchers for studying TAI, because it is easy to operate in the laboratory environment and can simulate the salient features of TAI in real-life combustors [20,21]. It consists of an 1.5 m long horizontal tube with an external cross section of  $4 \times 4''$  with a wall thickness of 0.25''. It is equipped with an air-flow controller that regulates the flow of air ( $Q$ ) at atmospheric pressure through the tube. It has a heating element placed at quarter length of the tube from the air-input end. A programmable direct current power supply controls the power input ( $E_{\text{in}}$ ) to the heater. The experiments have been conducted in the following manner:

- (1) For every run, the air flow rate ( $Q$ ) has been set at a constant value. Different runs have been performed with flow rates ranging from 130 LPM to 250 LPM at intervals of 20 LPM.
- (2) First the system has been heated to a steady-state with a primary heater power input ( $E_{\text{in}}$ ) of  $\approx 200$  W.
- (3) Then the power input has been abruptly increased to a higher value that showed limit cycle behavior as depicted in the stability chart in Mondal et al. [8].

<sup>5</sup>The feature vector is selected to be the state probability vector  $P$ . Given an alphabet  $\mathcal{A}$  and a Markov depth  $D$ , the cardinality  $|\mathcal{Q}|$  of the state set is bounded between  $|\mathcal{A}|$  and  $|\mathcal{A}|^D$  [7]. Therefore, for a given  $D$ , the dimension of the feature vector can be reduced by choosing a small alphabet size (e.g.,  $|\mathcal{A}| = 2$ ).



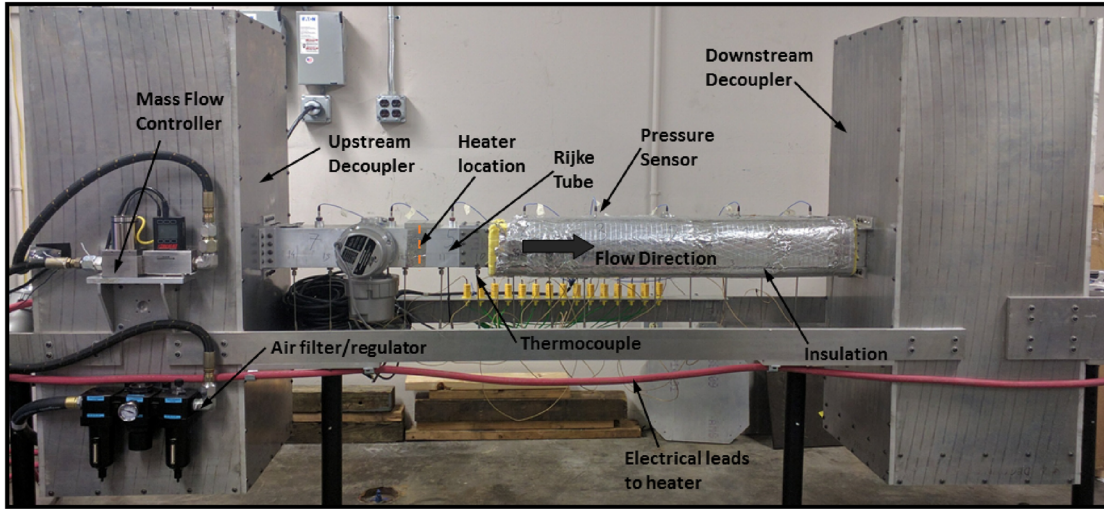


Fig. 1 The electrically heated Rijke tube apparatus

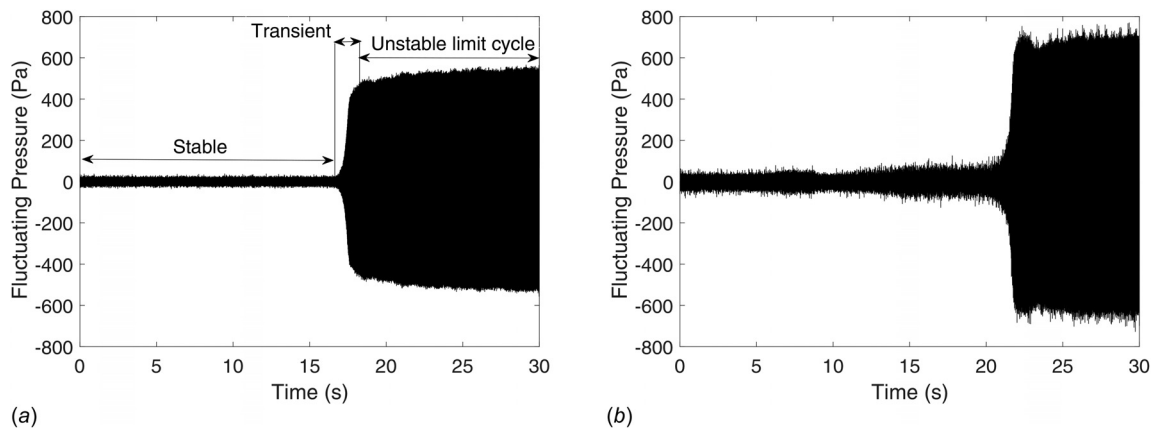


Fig. 2 Unsteady pressure signals showing the transience from stable (nominal) to unstable limit cycle (anomalous) behavior: (a)  $E_{in}$  abruptly increased to 1800 W with  $Q = 210$  LPM and (b)  $E_{in}$  abruptly increased to 2000 W with  $Q = 250$  LPM

At different air-flow rates and heat inputs, bifurcating transition from stable to unstable behavior in the acoustic response of the chamber occurs. Fifteen experiments have been conducted on the Rijke tube apparatus, where the process starts with the nominal or stable behavior, and gradually becomes anomalous (or unstable). A time series of pressure oscillations has been collected over 30 s for each experiment, sampled at 8192 Hz, and filtered to attenuate the effects of low-frequency environmental acoustics; typical profiles of the pressure time series are presented in Fig. 2. Further details regarding the apparatus and stability maps at different operating conditions can be found in Mondal et al. [8].

The performance of  $HMM_D$  is evaluated by comparison with other STSA techniques that use MEP [16,18] and K-means [17] for partitioning. Then the performance of  $HMM_L$  is compared with that of the best STSA method.<sup>6</sup> In each experiment, the entire time series of sensor data is segmented into small disjoint blocks. Based on the ground truth, each block in the experiment is labeled as:

- Class 0 if it belongs to the nominal state;
- Class 1 if it belongs to an anomalous state.

<sup>6</sup>The procedure in Algorithm 1 is followed in MEP and K-means-based algorithms with the symbolization step being replaced by MEP or K-means.

For each algorithm, an area under the curve (AUC) of the respective ROC [19] is obtained by summing results over all experiments, which reflects the respective (anomaly detection) performance. For HMM methods, hidden state set of cardinality

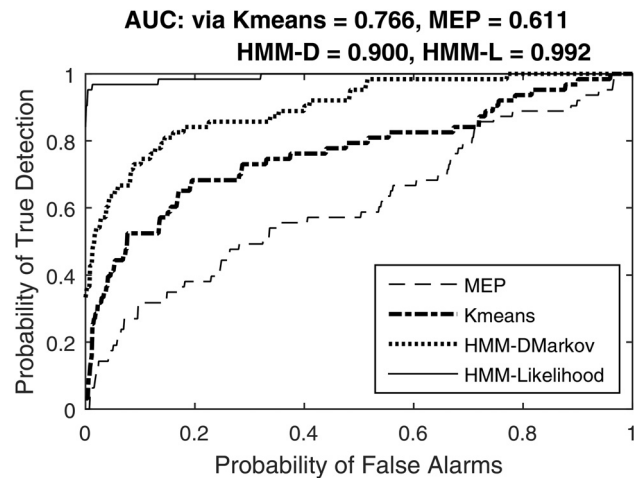


Fig. 3 ROC curves for combustion instability detection having  $|\mathcal{A}| = 2$ ,  $D = 2$ , and  $L = 200$

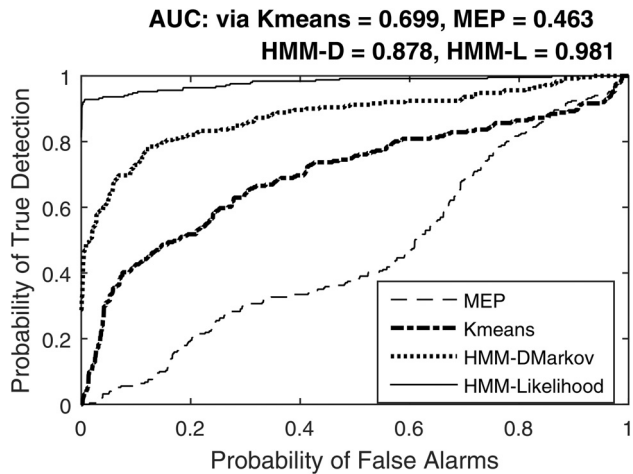


Fig. 4 ROC curves for combustion instability detection having  $|\mathcal{A}| = 2$ ,  $D = 2$ , and  $L = 50$

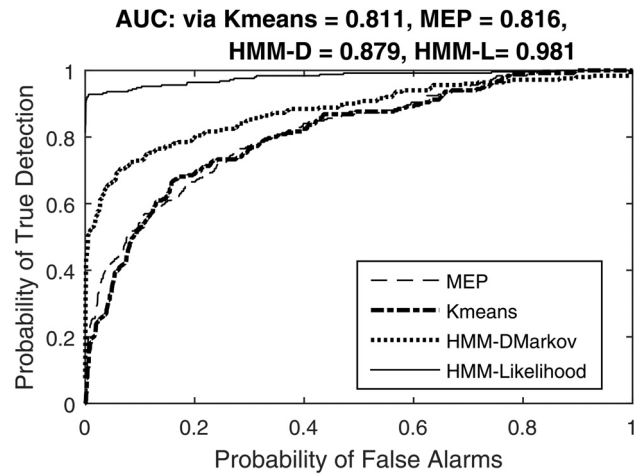


Fig. 6 ROC curves for combustion instability detection having  $|\mathcal{A}| = 2$ ,  $D = 4$ , and  $L = 50$

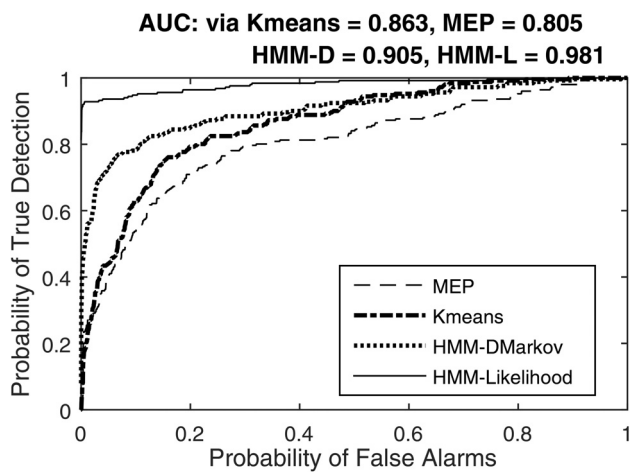


Fig. 5 ROC curves for combustion instability detection having  $|\mathcal{A}| = 2$ ,  $D = 3$ , and  $L = 50$

Table 1 Execution time (ms) for TAI detection ( $L = 50$ )

$D$	$K$ -means		MEP		$HMM_D$		$HMM_L$	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
2	3.3	0.095	1.700	0.037	4.600	0.8755	2.200	0.325
3	4.80	0.101	3.250	0.073	6.1715	0.507	2.200	0.325
4	8.15	0.137	6.659	0.029	9.4160	0.544	2.200	0.325

Note: Std: standard deviation.

on the parameter  $D$ ) is generally *faster*<sup>8</sup> and has a much better performance compared to the other methods<sup>9</sup> for all of these values of  $D$ , with AUC = 0.9805. Furthermore, it is seen in Figs. 4–6 that the performance of  $HMM_L$  is significantly superior to that of  $HMM_D$ , which has the best detection performance among STSA techniques. This is consistent with what has been explained earlier in Sec. 3, because  $HMM_L$  considers all hidden state paths of the learned HMM in contrast to  $HMM_D$ , which considers only the most likely hidden state path by discarding all other possible paths of the learned HMM.

## 5 Summary, Conclusions, and Future Work

This technical brief has proposed a HMM-based pattern recognition tool, called HMM-Likelihood ( $HMM_L$ ), for early detection of anomalous behavior (e.g., TAI in combustion systems). It has also proposed an STSA method, called HMM-based  $D$ -Markov ( $HMM_D$ ), which uses an HMM-based partitioning method that converts a time series into a symbol string with maximum posterior probability using dynamic programming (Viterbi algorithm) [9]. Unlike many other partitioning methods (e.g., uniform,  $K$ -means, and MEP [7,15]), the proposed HMM-based partitioning considers sequential patterns in the time series so that, even with a small alphabet size (e.g.,  $|\mathcal{A}| = 2$ ), the dynamical behavior of the time series is effectively captured. These algorithms have been validated with data collected from a laboratory-scale experimental apparatus [8]. The results show a significant improvement using  $HMM_D$  over  $K$ -means and MEP in the STSA setting. Moreover, the results of the  $HMM_L$  algorithm have been compared with the best STSA technique ( $HMM_D$ ) to show  $HMM_L$ 's consistently

two (i.e.,  $\mathcal{Q} = \{q_1, q_2\}$ ) and Gaussian mixtures for the state-conditional density function  $b_j(x) = p(x|z = q_j)$  have been used. The number of mixture components is found based on the Bayesian information criterion [17]. The detection methods have been tested for quantification of the extent of instability in the setting of transient growth of acoustic oscillations by using short-length blocks of time series.

Figure 3 shows<sup>7</sup> the AUC performance and the ROC curves of these methods using Markov depth  $D = 2$ , alphabet size  $|\mathcal{A}| = 2$ , and data block length  $L = 200$ . As seen in this figure,  $HMM_D$  shows an improvement in the AUC performance over  $K$ -means and MEP, with  $HMM_L$  showing a better performance compared to  $HMM_D$ , scoring an excellent AUC = 0.99229. The same methods are tested again after reduction of the data block length to  $L = 50$ , which can be considered as very short. The corresponding ROC curves are presented in Figs. 4–6 for  $D = 2, 3$ , and 4, respectively. Table 1 shows the mean and standard deviation of the execution time (for both learning and inference) of these methods over the 15 experiments conducted for  $D = 2, 3$ , and 4 with  $L = 50$ . It is seen in Figs. 4–6 and Table 1 that  $HMM_L$  (which does not depend

<sup>7</sup>The codes for implementing the HMM and STSA-based algorithms and the combustion dataset used in this paper are available at <https://github.com/nfjasim/HMM-codes-for-anomaly-detection>

<sup>8</sup>The algorithms in this paper were executed on a DELL PRECISION T3400, with an Intel® Core™2 Quad CPU Q9550 at 2.83 GHz, with 8 GB RAM, and running under Windows 7.

<sup>9</sup>Increasing the depth  $D$  too much will degrade the STSA techniques' performance due to the generated large number of PFSA states for which there is not enough data points (only 50 points in this case) for training [7] (some of these states may not be even visited with this small number of data points).

superior performance for different values of Markov depth  $D$  and time series length  $L$ .

This technical brief has considered first-order HMMs that are trained by using the Baum–Welch algorithm. Examples of topics of future research include: (i) higher-order HMMs and (ii) different types of training algorithms (e.g., Gibbs sampling and stochastic variational inference [17]).

## Funding Data

- U.S. Air Force Office of Scientific Research (AFOSR) (Grant Nos. FA9550-15-1-0400 and FA9550-18-1-0135; Funder ID: 10.13039/100000181).

## References

- [1] Rawung, R., and Putrada, A., 2014, “Cyber Physical System: Paper Survey,” International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, Sept. 24–25, pp. 273–278.
- [2] Darema, F., 2004, “Dynamic Data Driven Applications Systems: A New Paradigm for Application, Simulation and Measurements,” Computational Science (ICCS), Kraków, Poland, June, pp. 662–669.
- [3] Lieuwen, T. C., and Yang, V., 2005, *Combustion Instabilities in Gas Turbine Engines: Operational Experience, Fundamental Mechanisms, and Modeling*, American Institute of Aeronautics and Astronautics, Reston, VA, pp. 3–26, Chap. 1.
- [4] Matveev, K., 2003, “Thermoacoustic Instabilities in the Rijke Tube: Experiments and Modeling,” *Ph.D. thesis*, California Institute of Technology, Pasadena, CA.
- [5] Sarkar, S., Chakravarthy, S. R., Ramanan, V., and Ray, A., 2016, “Dynamic Data-Driven Prediction of Instability in a Swirl-Stabilized Combustor,” *Int. J. Spray Combust. Dyn.*, **8**(4), pp. 235–253.
- [6] Ray, A., 2004, “Symbolic Dynamic Analysis of Complex Systems for Anomaly Detection,” *Signal Process.*, **84**(7), pp. 1115–1130.
- [7] Mukherjee, K., and Ray, A., 2014, “State Splitting and Merging in Probabilistic Finite State Automata for Signal Representation and Analysis,” *Signal Process.*, **104**, pp. 105–119.
- [8] Mondal, S., Ghalyan, N. F., Ray, A., and Mukhopadhyay, A., 2018, “Early Detection of Thermoacoustic Instabilities Using Hidden Markov Models,” *Combust. Sci. Technol.* (epub).
- [9] Hajek, B., 2015, *Random Processes for Engineers*, Cambridge University Press, Cambridge, UK.
- [10] Rabiner, L., and Juang, B.-H., 1993, *Fundamentals of Speech Recognition*, Prentice Hall, Upper Saddle River, NJ.
- [11] Daw, C. S., Finney, C. E. A., and Tracy, E. R., 2003, “A Review of Symbolic Analysis of Experimental Data,” *Rev. Sci. Instrum.*, **74**(2), pp. 915–930.
- [12] Graben, P. B., 2001, “Estimating and Improving the Signal-to-Noise Ratio of Time Series by Symbolic Dynamics,” *Phys. Rev. E*, **64**(5 Pt. 1), p. 051104.
- [13] Gupta, S., and Ray, A., 2007, “Symbolic Dynamics Filtering for Data-Driven Pattern Recognition,” *Pattern Recognition: Theory and Applications*, Nova Science Publishers, Hauppauge, NY, pp. 17–71.
- [14] Wen, Y., Mukherjee, K., and Ray, A., 2013, “Adaptive Pattern Classification for Symbolic Dynamic Systems,” *Signal Process.*, **93**(1), pp. 252–260.
- [15] Bahrapour, S., Ray, A., Sarkar, S., Damarla, T., and Nasrabadi, N., 2013, “Performance Comparison of Feature Extraction Algorithms for Target Detection and Classification,” *Pattern Recognit. Lett.*, **34**(16), pp. 2126–2134.
- [16] Chattopadhyay, P., Mondal, S., Bhattacharya, C., Mukhopadhyay, A., and Ray, A., 2017, “Dynamic Data-Driven Design of Lean Premixed Combustors for Thermoacoustically Stable Operations,” *ASME J. Mech. Des.*, **139**(11), p. 111419.
- [17] Murphy, K., 2012, *Machine Learning: A Probabilistic Perspective*, 1st ed., The MIT Press, Cambridge, MA.
- [18] Ghalyan, N. F., Miller, D. J., and Ray, A., 2018, “A Locally Optimal Algorithm for Estimating a Generating Partition From an Observed Time Series and Its Application to Anomaly Detection,” *Neural Comput.*, **30**(9), pp. 2500–2529.
- [19] McDonough, R. N., and Whalen, A. D., 1995, *Detection of Signals in Noise*, 2nd ed., Academic Press, Boston, MA.
- [20] Rigas, G., Jamieson, N., Li, L., and Juniper, M., 2016, “Experimental Sensitivity Analysis and Control of Thermoacoustic Systems,” *J. Fluid Mech.*, **787**, p. R1.
- [21] Jamieson, N. P., Rigas, G., and Juniper, M. P., 2017, “Experimental Sensitivity Analysis Via a Secondary Heat Source in an Oscillating Thermoacoustic System,” *Int. J. Spray Combust. Dyn.*, **9**(4), pp. 230–240.