# Real-time Monitoring and Diagnostics of Anomalous Behavior in Dynamical Systems

Sudeepta Mondal, Chandrachur Bhattacharya, Najah F. Ghalyan and Asok Ray

**Abstract** Real-time condition monitoring of complex dynamical systems is of critical importance for predictive maintenance. This chapter focuses on data-driven techniques of fault diagnostics with an emphasis on real-time detection of anomalous behavior in combustion systems. It presents the applications of well-known statistical learning techniques such as *D*-Markov modeling and hidden Markov modeling (HMM) as possible data-driven solutions for anomaly detection in combustion systems. From the perspective of real-time monitoring and diagnostics, such statistical tools are applicable to stochastic dynamical systems in general. Both *D*-Markov and HMM algorithms have been validated on experimental data from a laboratory apparatus, which is an electrically heated Rijke tube.

## 1 Introduction

Anomaly in a dynamical system is defined as deviation of the system performance from the expected or nominal behavior. Gradual evolution of anomalies is usually a consequence of slow parametric or non-parametric changes within the system, which often leads to degraded performance and eventually premature end of the service life. Therefore, anomaly detection is considered to be essential for sustaining order and normalcy in human-engineered complex systems.

Sudeepta Mondal
Department of Mechanical Engineering, Pennsylvania State University e-mail: sbm5423@psu.edu

Chandrachur Bhattacharya
Department of Mechanical Engineering, Pennsylvania State University e-mail: czb28@psu.edu

Najah F. Ghalyan
Department of Mechanical Engineering, Pennsylvania State University e-mail: nfg103@psu.edu

Asok Ray
Department of Mechanical Engineering, Pennsylvania State University e-mail: axr2@psu.edu

The tasks of anomaly detection and failure prediction in engineering systems can be broadly classified into two categories: model-based and dynamic data-driven. In the model-based category, observer-induced techniques are commonly used, where certain residuals or diagnostic signals are generated for use in (possibly) optimal or adaptive threshold functions to detect the presence of faults. Residuals are generated by estimating the system's measured variables and using a deterministic (e.g., [1]) or a stochastic (e.g., [2]) observer. These observers are often designed based on a linear model or a (Jacobian) linearization of the nonlinear system model at selected operating points. These linear or linearized models are constructed as reasonable approximations of local behavior of the complex system when operating under the nominal condition; however, in case of evolving anomalies, the effects of system nonlinearities may gradually become too large to be ignored or approximated. Furthermore, it may be very difficult to model an anomalous system behavior, because such anomalies (e.g., those due to incipient faults) are usually unknown and may be too complex to model by solely relying on the fundamental laws of physics. Furthermore, a fault might be entering the system in a more complex manner than as an additive disturbance. These issues have motivated the study of anomaly detection in dynamical systems using a dynamic data-driven approach, where the dependence on a physics-based system model is de-emphasized; however, the knowledge of a physics-based model, if available, may supplement the information generated from the measured time-series data. In this context, an application example is presented in this chapter.

The application addresses occurrence of thermoacoustic instabilities (TAI) in combustion systems, which are caused by spontaneous excitation of one or more natural modes of acoustic waves [3]. TAI are typically manifested by large-amplitude self-sustained chaotic pressure oscillations in the combustion chamber [4], which may lead to damage in mechanical structures if the pressure oscillations match one of the 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 the control signals. This mandates accurate detection of instabilities from short-length sensor data. While traditional techniques for growth-rate measurement are reported in combustion literature [5, 6], they are only suitable for online data-intensive computations and the lengthy observation sequences are likely to cause large delays relative to the time scale of TAI.

This chapter focuses on formulation and validation of a decision-making algorithm for sensor-based automation (e.g., real-time monitoring and active control of dynamical systems). These algorithms have been built upon the theory of hidden Markov modeling (HMM) to extract features and classify patterns of the process behavior from short-length sensor time series. The underlying concept of the proposed HMM-based algorithms is validated on an ensemble of experimental data from a laboratory apparatus. The performance in each case is evaluated by comparison with that of symbolic time series analysis (STSA) [7, 8, 9, 10]. The results show consistent improvements for the proposed HMM-based method under varying configurations. From these perspectives, the major contributions of this chapter are delineated below.

1. Development of HMM-based algorithms for feature extraction and pattern classification from short-length time series to facilitate real-time monitoring and active control of dynamical systems.
2. Experimental validation of the anomaly detection algorithms in thermoacoustic instabilities in combustors.

The rest of the chapter is organized as follows. Section 2 presents a brief decription of the experimental apparatus. Section 3 provides an introduction to the theory of hidden Markov modeling (HMM) and its usage for symbolization of time series. Section 4 presents the technical approach of the reported work based on developing HMM algorithms for feature extraction and pattern classification. Section 5 develops a modification of traditional STSA methods for detecting anomalies in a dynamical system. Section 6 presents HMM-based methodologies for early detection of anomalous regimes for monitoring and control of TAI. Section 7 summarizes and concludes the chapter along with recommendations for future research.

## 2 Experimental Apparatus

The data used in this article has been generated from a 1.5 m long electrically heated horizontal Rijke tube with an external cross-section of 4"×4" with a wall thickness of 0.25". A Parker P32E series air filter-regulator eliminates pressure fluctuations and impurities. The flow rate is controlled using a 0 - 1000 SLPM Alicat Mass Flow controller. The heating element is a square weave 40×40 nichrome wire mesh, placed at quarter length in the upstream end of the tube. The heating element is powered by a programmable power supply from TDK Lambda. As a safety measure to prevent accidental physical contact with the hot downstream end of the tube, the latter is insulated. The pressure data is procured by a series of 8 acoustic sensors placed inline and flushed with the inner wall of the tube. The sensor data is sampled at a rate of 8192 Hz and the data acquisition process is automated through NI Labview 2016 in conjunction with DAQ devices from National Instruments. The acquired data is filtered using a high-pass filter of cutoff frequency 40 Hz in order to eliminate low-frequency components due to noise and acoustics in the ambient. More details about the experimental setup and data acquisition can be found in the work by Mondal et al [11].

Knowing the stability map of the system provides an operator with the prior idea of which operating conditions lead to stable and unstable operations respectively. Figure 2 depicts the stability map of the system at different operating conditions. The operating conditions for the different experiments are different values of airflow rate (Q) and primary heater power input ($E_{in}$). The current article is focused on the online detection of instabilities in the transient growth phase of thermoacoustic oscillations. Thermoacoustic instabilities are understood to be an outcome of the subcritical Hopf bifurcations [12] due to which the system transits from a stable operation to an unstable limit cycle behavior. This transient growth period is a rich source of information for developing online detection algorithms. Early detection of

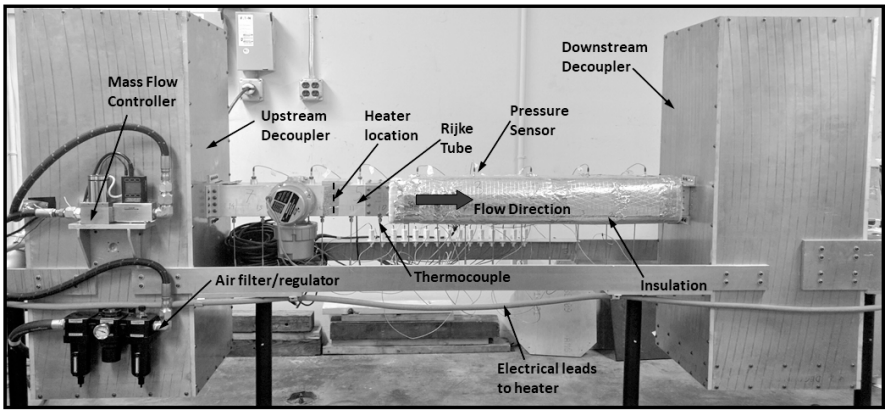


Fig. 1: Experimental setup : Electrically heated Rijke tube (reproduced with permission from Mondal et al. [11])

imminent instabilities from the transient period can provide adequate lead time for the control system to take appropriate actions in order to revert the system to a stable mode of operation. For this purpose, in order to acquire the acoustic data through the transient phase, $E_{in}$ has been increased in such a fashion that the system passes through the Hopf point, in way similar to the one followed by Rigas et al. [6]:

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 ($E_{in}$) of $\approx$200 W.
3. Then the power input has been abruptly increased to a higher value that showed limit cycle behavior.
4. Pressure data has been procured using the acoustic sensors for 30 seconds of operation at a sampling rate of 8192 Hz.

Based on the initial mean temperature of the tube the steady state limit cycle amplitude can vary, since the natural frequency increases with an increase in the velocity of sound. This is depicted in Figure 3. When a lower initial mean temperature of around $27^oC$ was maintained, a lower frequency mode of instability ($\sim$114 Hz) was observed to be excited for a range of flow rates marked with '$\times$' in unshaded boxes in Figure 2. The shaded boxes indicate the conditions when the higher frequency unstable mode ($\sim$131 Hz) and its harmonics were excited with a higher initial mean temperature of about $75^oC$. The cases for which the system remains stable are marked as '$\circ$'.

| 2000 | O | O | O | O | O | O | O | X | X | X | X | X | X | X | X | X | X | X | X | O | O | O | O | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1800 | O | O | O | O | O | O | O | X | X | X | X | X | X | X | X | X | X | X | X | O | O | O | O | O |
| 1600 | O | O | O | O | O | O | O | O | X | X | X | X | X | X | X | X | X | X | O | O | O | O | O | O |
| 1400 | O | O | O | O | O | O | O | O | X | X | X | X | X | X | X | X | O | O | O | O | O | O | O | O |
| 1200 | O | O | O | O | O | O | O | O | X | X | X | X | X | X | X | X | O | O | O | O | O | O | O | O |
| 1000 | O | O | O | O | O | O | O | O | X | X | X | X | X | X | O | O | O | O | O | O | O | O | O | O |
| 800 | O | X | X | X | X | X | X | X | X | X | O | O | O | O | O | O | O | O | O | O | O | O | O | O |
| 600 | O | X | X | X | X | X | X | X | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O |
| 400 | O | X | X | X | X | X | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O |
| 200 | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O | O |
| $E_{in}$ (W) / Q (LPM) | 61 | 72 | 82 | 92 | 102 | 112 | 122 | 132 | 142 | 152 | 162 | 172 | 182 | 191 | 200 | 210 | 219 | 228 | 237 | 246 | 255 | 264 | 273 | 281 | 290 |

Fig. 2: Stability Map of the system as a function of Q and $E_{in}$. Emboldened box margins indicate operating conditions for which the pressure signature is depicted in Figure 3 (reproduced with permission from Mondal et al. [11]).

## 3 Mathematical Background

This section provides the background information on Hidden Markov modeling (HMM) [13] and symbolic time series analysis (STSA) [7, 10] for anomaly detection in dynamical systems[1] [14].

### 3.1 Hidden Markov Modeling

Hidden Markov modeling (HMM) has found its applications in diverse fields [13, 11]. While the theory of HMM is presented in detail in [15], the key concepts are very succinctly outlined below for the sake of completeness.

Consider a time series $X = \{x_1, x_2, \ldots, x_T\}$, $x_n \in \mathbb{R}^N$. In HMM framework, $X$ is considered as a realization of a hidden Markov chain $Z = \{z_1, z_2, \ldots, z_T\}$, where $z_t$ is one of the finitely many states $Q \triangleq \{q_1, \ldots, q_M\}$ 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

[1] 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.

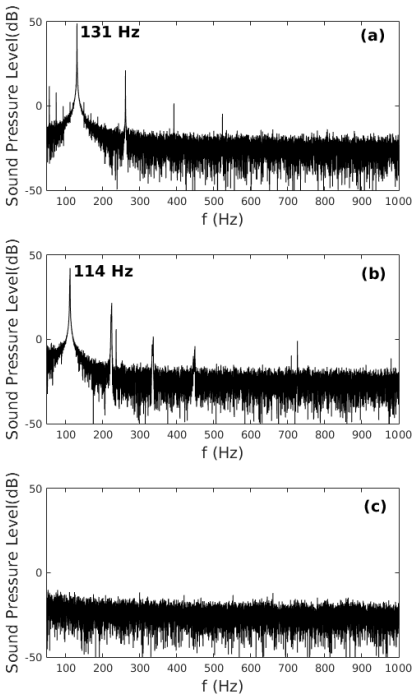


Fig. 3: Representative plots of fluctuating pressure amplitudes (left) and their respective power spectral densities (right) at different operating conditions. (a)$E_{in}$=1400 W, $Q$=162 LPM. (b) $E_{in}$=600 W, $Q$=112 LPM. (c) $E_{in}$=1400 W, $Q$=228 LPM.(reproduced with permission from Mondal et al. [11])

paths. To compute the joint likelihood of $X$ given an HMM $\lambda$, it is necessary to marginalize over all these hidden state paths:

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

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

$$\alpha_n(i) \triangleq p(x_1\ x_2 \ldots x_n, z_n = q_i|\lambda) \tag{2}$$

This variable is recursively computed by using the *forward procedure* [13], and the observation likelihood $p(X|\lambda)$ is evaluated as:

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

The *Baum-Welch* algorithm [13] is a standard (expectation maximization) tool for training HMMs; it finds a triple $\lambda^* = \{A, B, \pi\}$ that locally maximizes the total observation likelihood:

$$\lambda^* = \arg\max_{\lambda}\{p(X|\lambda)\} = \arg\max_{\lambda} \sum_{Z \in \mathcal{Z}} p(X, Z|\lambda) \tag{4}$$

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

$$Z^*(X) = \arg\max_{Z \in \mathcal{Z}}[p(Z|X, \lambda^*)] \tag{5}$$

by using the (dynamic programming-based) *Viterbi algorithm* [15]. Both Eqs. (3) and (5) are used here to define anomaly detection algorithms.

### 3.2 Symbolic Time Series Analysis

The authors now provide the background for symbolic time-series analysis (STSA) and the construction of probabilistic finite state automata (PFSA) [16, 17]. The time series of the measured signal is partitioned (or quantized) and then symbolized as a symbol string. In this process, the signal space is partitioned into a finite number of cells, where the number of cells is identically equal to the cardinality $|\Sigma|$ of the (symbol) alphabet $\Sigma$ and a symbol from the alphabet $\Sigma$ is assigned to each (signal) value corresponding to the cell to which it belongs [18, 19]; details are reported in [17]. Thus, a symbol is associated with that data point when the value of a data point at a given instant of time is located within a particular cell. The following definitions, which are available in standard literature (e.g., [16, 17]), are recalled here for completeness of the chapter.

**Definition 1** A finite state automaton (FSA) $G$, having a deterministic algebraic structure, is a triple $(\Sigma, Q, \delta)$ where:

- $\Sigma$ is a (nonempty) finite alphabet, i.e., its cardinality $|\Sigma|$ is a positive integer.
- $Q$ is a (nonempty) finite set of states, i.e., its cardinality $|Q|$ is a positive integer..
- $\delta : Q \times \Sigma \rightarrow Q$ is a state transition map.

**Definition 2** A symbol block, also called a word, is a finite-length string of symbols belonging to the alphabet $\Sigma$, where the length of a word $w \triangleq s_1 s_2 \cdots s_\ell$ with $s_i \in \Sigma$ is $|w| = \ell$, and the length of the empty word $\epsilon$ is $|\epsilon| = 0$. The parameters of FSA are extended as:

- The set of all words, constructed from symbols in $\Sigma$ and including the empty word $\epsilon$, is denoted as $\Sigma^\star$.
- The set of all words, whose suffix (respectively, prefix) is the word $w$, is denoted as $\Sigma^\star w$ (respectively, $w\Sigma^\star$).

- The set of all words of (finite) length $\ell$, where $\ell$ is a positive integer, is denoted as $\Sigma^{\ell}$.

*Remark 1* A symbol string (or word) is generated from a (finite-length) time series by symbolization.

**Definition 3** A probabilistic finite state automaton (PFSA) $K$ is a pair $(G, \pi)$, where:

- The deterministic FSA $G$ is called the *underlying FSA* of the PFSA $K$.
- The probability map $\pi : Q \times \Sigma \rightarrow [0, 1]$ is called the morph function (also known as symbol generation probability function) that satisfies the condition: $\sum_{\sigma \in \Sigma} \pi(q, \sigma) = 1$ for all $q \in Q$.
- The $\big(|Q| \times |\Sigma|\big)$ morph matrix $\Pi$, which is converted into the $(|Q||\Sigma| \times 1)$ morph vector $\nu$ to serve as an extracted feature in the sequel, is generated by the morph function $\pi$.

Equivalently, a PFSA is a quadruple $K = (\Sigma, Q, \delta, \pi)$.

For anomaly detection using STSA, a time series $X$ of sensor data is first converted into a symbol string. Then, PFSAs are constructed from the symbol strings, which in turn generate low-dimensional feature vectors [9] 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 features of PFSA model is computed by frequency counting [10]. This yeilds the nominal feature vectors.
3. *Select* a new block of the time series up to the current time $t$ and convert it into a symbol string. This yields a new PFSA with a new (quasi-)stationary feature vector o matrix of the system at time $t$.
4. *Compute* the anomaly at time $t$ as the divergence between the nominal feature $(F_{nom})$ and current feature vectors $(F_t)$:

$$\mu(t) = d(F_{nom}, F_t) \tag{6}$$

where $d(\bullet, \bullet)$ is an appropriately chosen distance norm (such as the Kullback-Leibler divergence [20]).

## 4 HMM-based Algorithm for Feature Extraction and Pattern Classification

This section presents novel anomaly detection methods, proposed by [14], used in this work for early detection of TAI onset in combustion systems. The first one amounts to a novel type of STSA technique, in which the time series is converted to a symbol sequence with the maximum posterior probability by using the dynamic

programming method. In the second method, an HMM is first trained using Baum-Welch algorithm, and the likelihood of the observed time series conditioned on the trained HMM is used as anomaly measure. More details are explained in the following subsections.

### 4.1 An HMM-Based Partitioning Method for STSA

While most of the partitioning methods in STSA are state space partition, as explained in Subsection 3.2, a novel partitioning method was proposed by [14] 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 $\Sigma$ of hidden states, and the HMM parameters are estimated by using the expectation maximization (Baum-Welch) method. Then, out of $|\Sigma|^T$ possible symbol strings, the algorithm identifies the one with maximum posterior probability, i.e.,

$$S^* = \arg\max_S P(S|X, \lambda) \tag{7}$$

This is efficiently done by using the dynamic programming method (Viterbi algorithm). This string is optimal in the sense of minimizing the string error rate in the HMM framework [15]. 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 (FPR)[2].

### 4.2 A Likelihood-Based HMM Method for Anomaly Detection

As we will experimentally demonstrate in the next subection, the $\text{HMM}_D$ method outperforms other STSA methods. However, $\text{HMM}_D$ makes use of the most prob-

[2] The *Receiver Operating Characteristic* (ROC) curves have been used in the current chapter for assessing the detection performance by varying the parameter $\tau$ from $-\infty$ to $\infty$ [21], 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 FPR, 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.

**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:
$$\lambda^* = \arg\max_{\lambda}\{p(x^*_{t_0:t_0+L}|\lambda)\}$$
2: Use Viterbi algorithm to find the hidden state path:
$$z^*_{t_0:t_0+L} = \arg\max_{z_{t_0:t_0+L}}\{p(z_{t_0:t_0+L}|x^*_{t_0:t_0+L}, \lambda^*)\}$$
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 & 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**

able 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 was proposed in [14], 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:

$$\mu(t) \triangleq -\log[p(x_{t:t+L}|\lambda^*)] \tag{8}$$

where Eq. (8) 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 an abbreviation for HMM Likelihood. A pseudocode of $\text{HMM}_L$ is presented in Algorithm 2. The intuition behind Eq. (8) 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:

$$\lambda^* = \arg\max_{\lambda}\{p(x^*_{t_0:t_0+L}|\lambda)\}$$

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 2* The method $\text{HMM}_L$ effectively considers all possible hidden state paths if the Forward algorithm is used [15]. As demonstrated experimentally in Section 4.3, 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 $X$.

### 4.3 Application to Online Detection of TAI in Combustion Systems

This section demonstrates the performance of Algorithms 1 and 2 for early detection of transience onset to unstable combustion based on experiments performed on Rijke tube [14]. Details on apparatus description and the how the experiments are conducted are given in Section 6. Fifteen (15) 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 sec 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 Figure 4. Further details regarding the apparatus and stability maps at different operating conditions can be found in Section 6.

The performance of $\text{HMM}_D$ is evaluated by comparison with other STSA techniques that use MEP [22, 23] and K-means [20, 24] for partitioning [14]. Then the performance of $\text{HMM}_L$ is compared with that of the best STSA method[3]. 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:

[3] 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 [14].

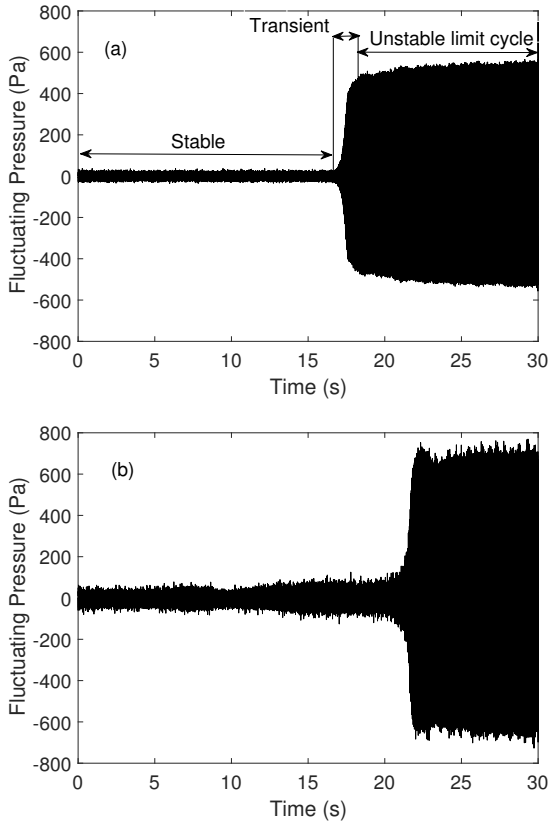


Fig. 4: Unsteady pressure signals showing the transience from stable (nominal) to unstable limit cycle (anomalous) behavior. (a) $E_{in}$ abruptly increased to 1800 W with $Q = 210$ LPM (b) $E_{in}$ abruptly increased to 2000 W with $Q = 250$ LPM.(reproduced with permission from Mondal et al. [11])

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

For each algorithm, an area under the curve (AUC) of the respective receiver operating characteristic (ROC) [21] is obtained by summing results over all experiments, which reflects the respective (anomaly detection) performance. For HMM methods, hidden state set of cardinality 2 (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 (BIC) [20]. 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.

Table 1: Execution time (msec) for TAI Detection ($L = 50$)

| | K-means | | MEP | | $HMM_D$ | | $HMM_L$ | |
|---|---|---|---|---|---|---|---|---|
| D | 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 |

Figure 5 shows[4] the area under the curve (AUC) performance and the ROC curves of these methods using Markov depth $D = 2$, alphabet size $|\Sigma| = 2$ 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 Figures 6, 7, and 8 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 Figures 6, 7, and 8 and Table 1 that $HMM_L$ (which does not depend on the parameter $D$) is generally *faster*[5] and has a much better performance compared to the other methods[6] for all of these values of $D$, with AUC=0.9805. Furthermore, it is seen in Figures 6, 7, and 8 that the performance of $HMM_L$ is significantly superior to that of $HMM_D$, which has the best detection performance amongst STSA techniques. This is consistent with what has been explained earlier in Subsection 4.2, 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 PFSA-based Algorithm for Feature Extraction and Pattern Classification

In this section, the authors initially present a probabilistic finite state automaton (PFSA) approach for online feature extraction from a time-series and the subsequent pattern classification and decision making. In this section, the advantages and disadvantages of PFSA over HMM are also discussed.

[4] The codes for implementing the HMM and STSA-based algorithms and the combustion dataset used in this chapter are available at https://github.com/nfjasim/HMM-codes-for-anomaly-detection.

[5] The algorithms in this chapter were executed on a DELL PRECISION T3400, with an Intel(R) Core(TM)2 Quad CPU Q9550 at 2.83 GHz, with 8 GB RAM, and running under Windows 7.

[6] 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. Some of these states may not be even visited with this small number of data points [14].

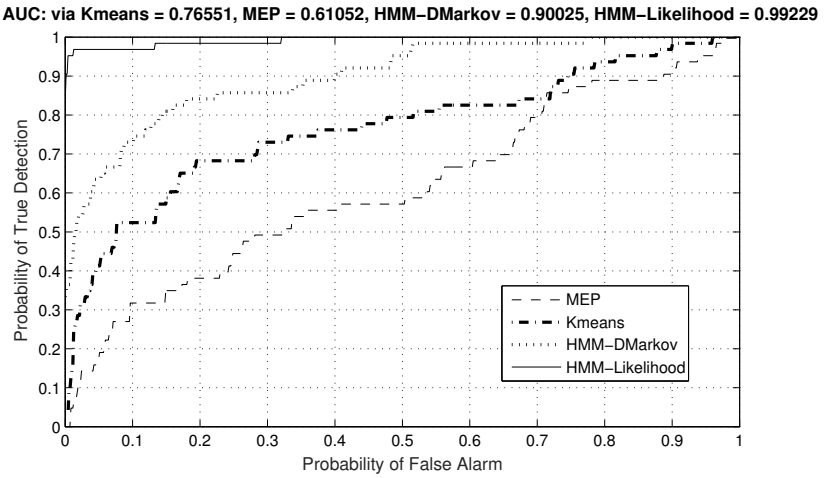


Fig. 5: ROC curves for detection of TAI onset, with $|\Sigma| = 2$, $D = 2$, and $L = 200$.

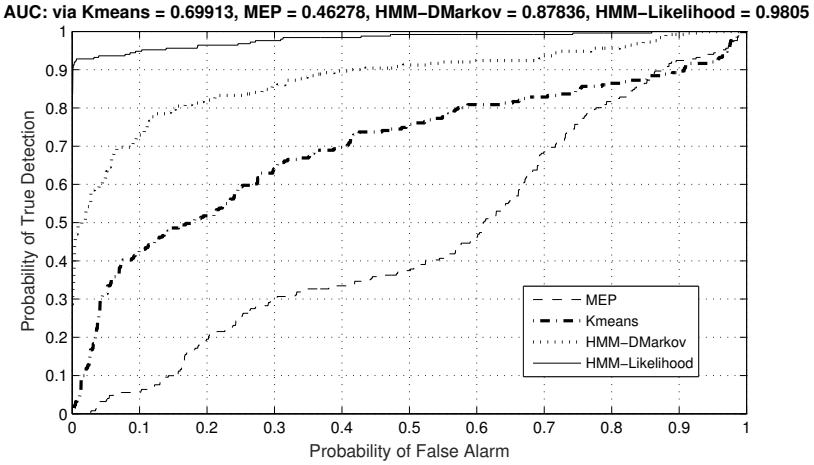


Fig. 6: ROC curves for detection of TAI onset, with $|\Sigma| = 2$, $D = 2$, and $L = 50$.

## 5.1 Reconstruction of time-series from symbolized sequence

As mentioned in Subsection 3.2 using a PFSA involves symbolizing a segment of the time-series using a partitioning method (such as MEP or Uniform as mentioned in Subsection 3.2), and then generating a feature from the generated symbol string,

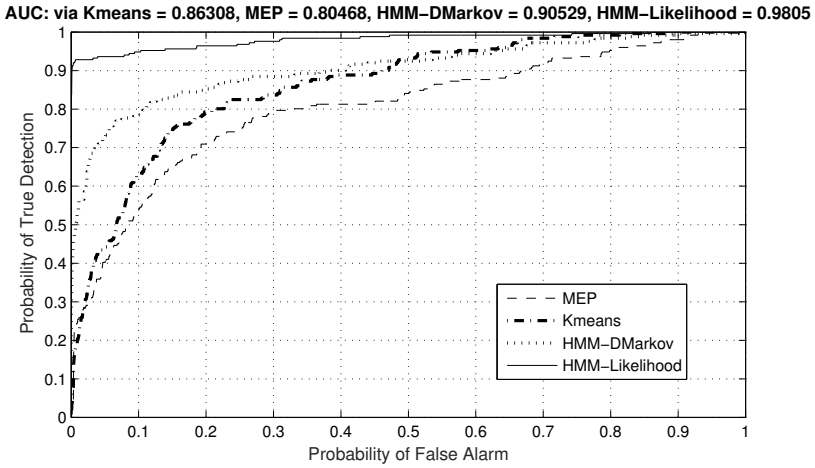


Fig. 7: ROC curves for detection of TAI onset, with $|\Sigma| = 2$, $D = 3$, and $L = 50$.

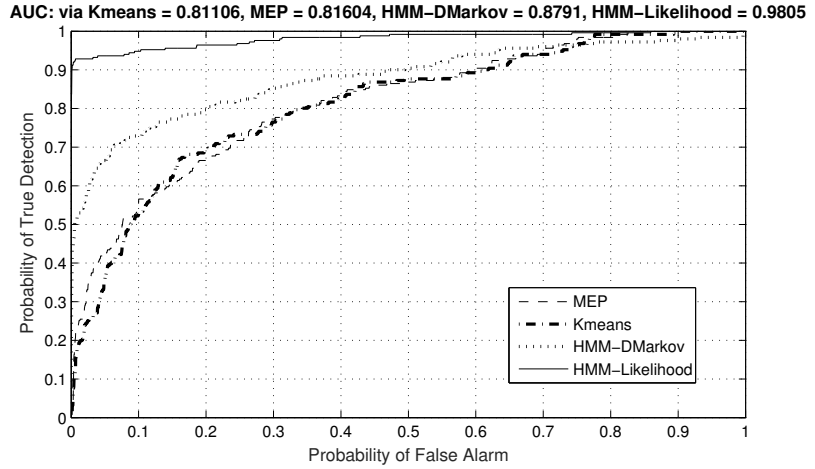


Fig. 8: ROC curves for detection of TAI onset, with $|\Sigma| = 2$, $D = 4$, and $L = 50$.

such as the morph matrix. The naive way to generate the PFSA is using a frequency counting technique for computing the occurrence of each state and the state transition. Selecting appropriate partitioning techniques, depths and alphabet sizes greatly controls how well the PFSA captures the essential dynamics and nature of the signals. A good measure of how well the partitioning method is performing is

the reconstruction error. As the symbol string is supposed to be a representation of the data, it is important to be able to reconstruct the original time series from the symbol string given the following data:

- The symbol string
- The mean value of the time series corresponding to that state (computed by taking the mean of the actual time-series values that are denoted by that state)

The $L_2$ (Euclidian) norm difference between the reconstructed time-series and the original time series is used as a measure of the reconstruction; normalizing the time-series makes it easier to compare the norm differences. The normalized reconstruction errors for 3 different partitioning methods, namely uniform, MEP and K-Means are tabulated in Tab. 2, with the lowest error (best reconstruction) values marked in bold font. In this analysis, the data is symbolized in a windowed fashion and the partition boundaries are recomputed for every window after normalization of the data segment. Recomputing the partitions for each segment greatly enhances the accuracy of the PFSA as compared to those methods described in the previous section. This increases computational time but by a negligible amount since the entire formulation is algebraic and computation is very simple.

This analysis has been done over 2 sets of 40 time-series obtained from the in-house Rijke tube experimental setup from 2 different pressure sensor locations, one having high signal-to-noise ratio (SNR) and the other with significantly more noise (i.e. low SNR). These signals are either initially stable and then become unstable or vice-versa. For each, different alphabet sizes ($|\Sigma|$) have been considered for demonstration purposes, while depth ($D$) has been maintained as $D = 1$ for brevity with only 2 cases having $D = 2, 3$. The reported values are the averaged values across all of the 40 time-series.

Table 2: Reconstruction Error for STSA

| | | | | Low SNR Sensor | | | High SNR Sensor | | |
|---|---|---|---|---|---|---|---|---|---|
| $\|\Sigma\|$ | $D$ | Sampling Frequency | Data Length | Uniform Error ($\times 10^{-3}$) | MEP Error ($\times 10^{-3}$) | K-Means Error ($\times 10^{-3}$) | Uniform Error ($\times 10^{-3}$) | MEP Error ($\times 10^{-3}$) | K-Means Error ($\times 10^{-3}$) |
| 2 | 1 | 100 Hz | 82 points | 0.1366 | 0.1439 | **0.1358** | 0.1564 | 0.1683 | **0.1560** |
| 4 | 1 | 100 Hz | 82 points | 0.0667 | 0.0773 | **0.0658** | 0.0763 | 0.0837 | **0.0748** |
| 6 | 1 | 100 Hz | 82 points | 0.0443 | 0.0504 | **0.0442** | 0.0503 | 0.0574 | **0.0499** |
| 2 | 1 | 50 Hz | 164 points | 0.1384 | 0.1404 | **0.1374** | 0.1580 | 0.1610 | **0.1576** |
| 4 | 1 | 50 Hz | 164 points | 0.0687 | 0.0722 | **0.0671** | 0.0776 | 0.0840 | **0.0755** |
| 6 | 1 | 50 Hz | 164 points | 0.0467 | 0.0503 | **0.0462** | 0.0518 | 0.0572 | **0.0510** |
| 6 | 2 | 50 Hz | 164 points | 0.0406 | 0.0427 | **0.0400** | 0.0436 | 0.0470 | **0.0425** |
| 6 | 2 | 100 Hz | 82 points | **0.0375** | 0.0411 | 0.0381 | 0.0417 | 0.0461 | **0.0406** |

The K-means partitioning method proves to be the most accurate across nearly all combinations of alphabet size, depth and sensor. Also, in the given range, increasing

alphabet size significantly improves performance as does increasing the depth from 1 to 2. However, increasing either of these too much causes degradation as described in the previous section.

## 5.2 Online Anomaly Detection

In this subsection, 3 methods of online anomaly detection using STSA based methods are described and results presented. These methods are:

- Norm difference method
- Entropy method
- ×-entropy method

**Norm difference method:** Using the generated symbol sequence, PFSAs can be generated for each window. Using the features related to the PFSA, such as the morph matrix, we can try capture the dynamics of the system. Using this feature, classification can be done by first training a morph matrix for the nominal state by taking the mean of all the morph matrices generated from the windows corresponding to the nominal state in the training set (the ground truth). During testing, a distance function checks the distance of the morph matrix generated from the unknown data segment. If the distance is above a threshold, the generated morph matrix is said to be 'too distant' from that corresponding to the nominal state and is said to be in an anomalous state.

**Entropy method:** In this method, the morph matrix does not need to be computed, instead a frequency counting technique is used to compute the probability of each state, yielding a probability vector $p$ of length equal to that of the number of states $|Q|$. The Shannon entropy of the data is then computed as $H = \Sigma_{i=1}^{|Q|} p(i) \times log(p(i))$. For the purpose of detection and classification we utilize the fact that the stable (nominal) regime is more noisy than the ordered limit cycle type behavior seen in the off-nominal conditions [25]. Thus, by definition, the entropy should have a higher value during the stable regime and a drop in the same must be an indication of a transition to the anomalous regime [26].

**×-entropy method:** Very similar to the above method, the ×-entropy or cross-entropy method uses the probability vectors from not one but 2 synchronous sensors of the same or different type. ×-entropy 'is a measure of the dynamical complexity of the temporal co-dependence from one symbol sequence to another' [26]. As in combustion physics, a slight shift from stable combustion toward thermo-acoustic instability can be captured by an abrupt increase in the computed value of D-Markov entropy rate. The ×-entropy of 2 synchronous signals is computed as $H = \Sigma_{i=1}^{|Q|} p(i) \times log(q(i))$, where $q$ is the probability vector of the states for the second sensor.

The following figures give the ROC curves for various trials using the PFSA norm-difference method. The data used is the same as that used in the previous subsection. Figures 9, 10 and 11 compare different combinations of depths ($D$),

data-window lengths and partitioning methods for the same alphabet size of $|\Sigma| = 6$ which was found to consistently produce better results as seen in Tab. 2. The data was split into a 80:20 train-test split for this analysis.

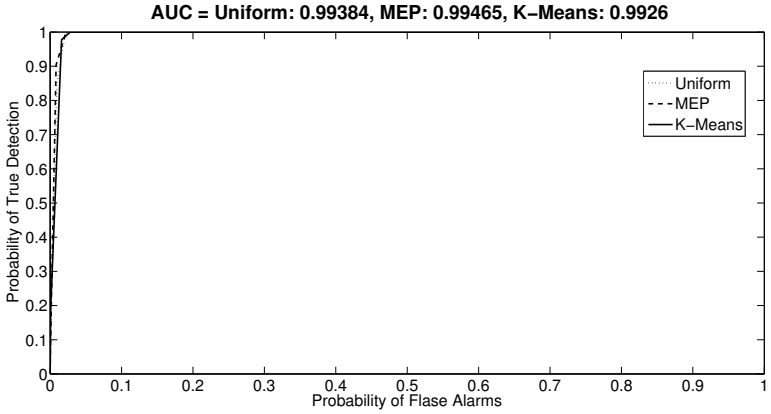


Fig. 9: ROC curves for combustion instability detection using PFSA having $|\Sigma| = 6$, $D = 1$, and $L = 50$ comparing various partitioning methods.

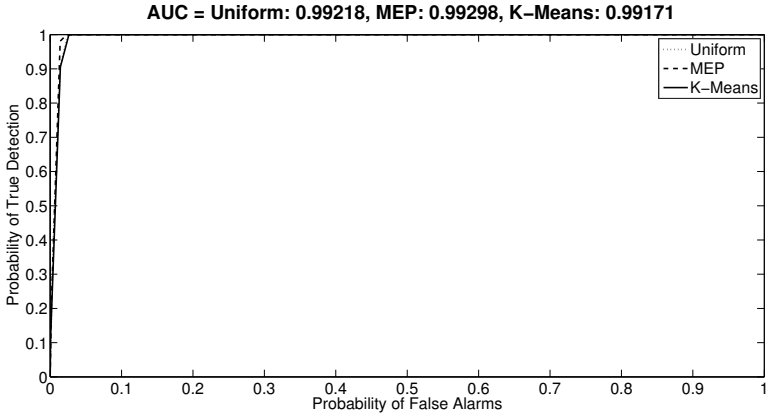


Fig. 10: ROC curves for combustion instability detection using PFSA having $|\Sigma| = 6$, $D = 1$, and $L = 100$ comparing various partitioning methods.

In Fig. 12 the ROC curve for the entropy method is presented for a depth of 2 for various alphabet sizes, while Fig. 13 reports the same for the $\times$-entropy method. The first set of curves uses the other data set used in the previous section consisting of 145 time-series, each starting out stable and then going unstable due to a change in

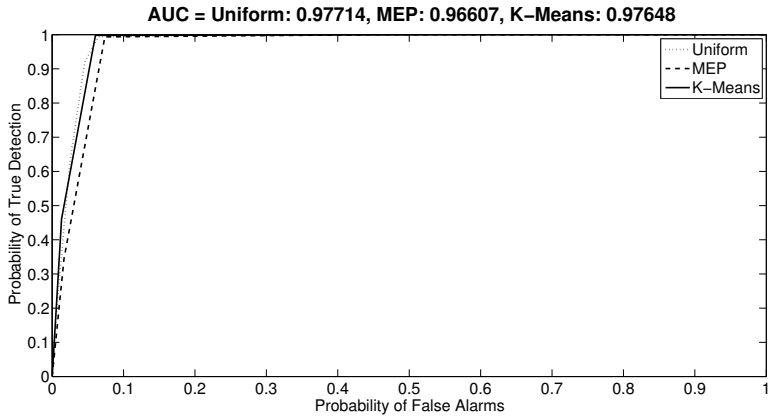


Fig. 11: ROC curves for combustion instability detection using PFSA having $|\Sigma| = 6$, $D = 2$, and $L = 50$ comparing various partitioning methods.

operating characteristics, while the one for the $\times$-entropy method uses the one with the 40 time-series as this data-set has values from 2 spatially-separated synchronous sensors. The uniform partitioning technique is used. The reason for choosing this is that although it does not have the highest reconstruction accuracy, it falls a very close to the K-means technique and gives better AUC values in the ROC curves as compared to the K-means technique. We do not use the MEP partitioning method, as by the nature of its construction, the MEP technique yields uniform probability across all classes and the entropy information is lost.

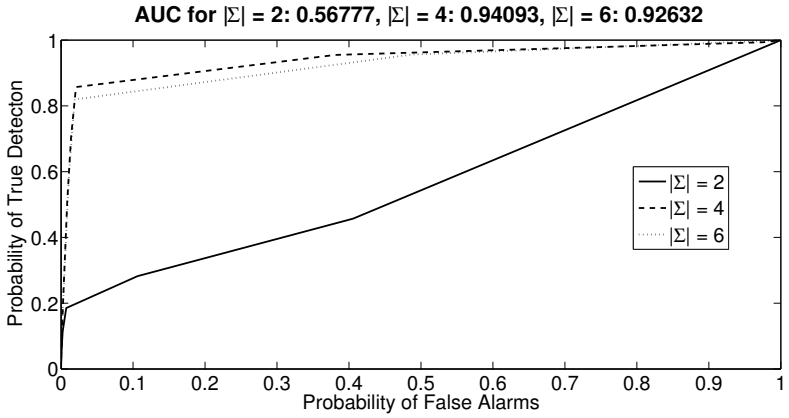


Fig. 12: ROC curves for combustion instability detection using Entropy based method having $D = 1$, and using uniform partitioning technique, comparing various alphabet sizes.

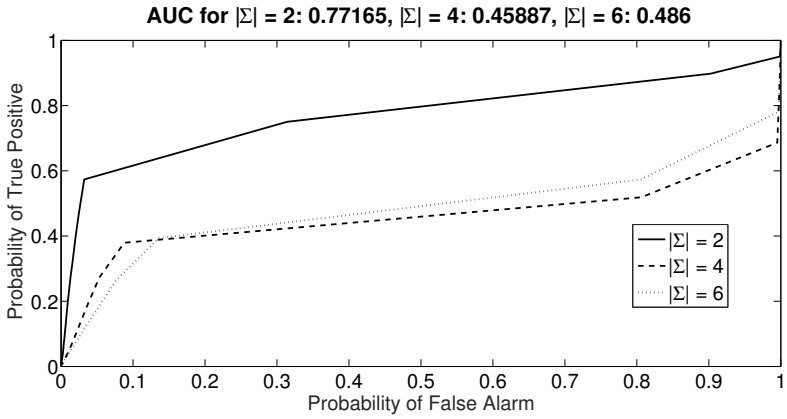


Fig. 13: ROC curves for combustion instability detection using ×-Entropy based method having $D = 1$, and using uniform partitioning technique, comparing various alphabet sizes.

Finally, in Fig. 14 a comparison is presented between the ROC curves computed for various window lengths. The depth is taken as 1, with $|\Sigma| = 6$ using the uniform partitioning technique.

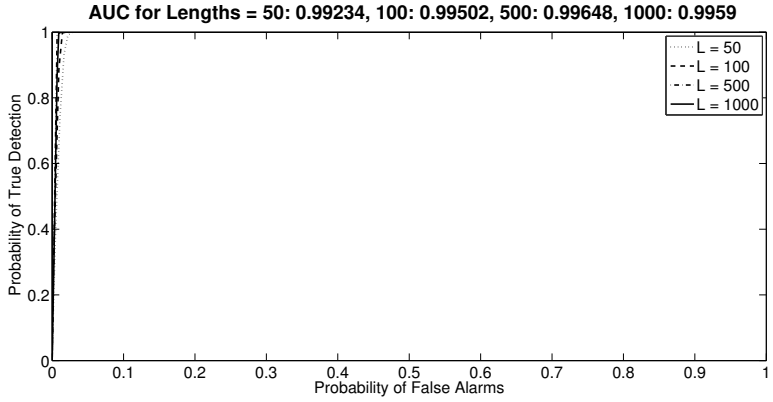


Fig. 14: ROC curves for combustion instability detection using PFSA having $|\Sigma| = 6$, $D = 1$, and using uniform partitioning technique, comparing various windows lengths.

### 5.3 Comparison between PFSA and HMM methods

Also, as we can see, by comparing to results from the previous section, by choosing the partitioning method, alphabet size and depth intelligently high accuracy can be reached, similar to the HMM methods proposed. PFSA methods are typically much faster due to the simplicity and purely algebraic nature of the model. HMMs gain popularity when there are multiple regimes which are very similar to each other. For a binary classification, or for a multi-class classification that are well-separated, a PFSA based classification algorithm is preferable. For very small data windows, a PFSA is usable only if all the states are reached in the frequency counting, which can be achieved by having lower alphabet size or depths, but if the dynamics is too complex, low values of $|\Sigma|$ and $D$ might not be sufficient and that's when HMMs may be more useful. The subsequent section focuses on the HMM methods as a method with wider applicability, specially when the classification is ternary and the transient regime is to be separately identified and classified.

## 6 HMM-based Early Detection of Thermoacoustic Instabilities in the Transient Regime

Early detection of instabilities is a problem that is closely related to the development of active control algorithms to augment the existing passive control systems already present in most commercial combustors. Since instabilities develop in the order of milliseconds, automated detection algorithms must be data efficient in order to detect a divergence from the nominal operation within the transient period before the system goes into the limit cycle operating mode. This is because once the limit cycle instabilities set in, the system goes into a self-driven feedback loop after which it might be too late to implement control actions. Hence, the focus in this work would be the detection of TAI with short windows of time series samples. The dataset used for this purpose comprises 145 pressure time series data samples of 30 seconds duration, each of which starts from a stable operation and ends in a limit cycle instability.

This dataset has been randomly divided into training and test set in the ratio 80:20. The training data provides the prior information for learning 3 HMMs: $\lambda_1$, $\lambda_2$ and $\lambda_3$. This learning phase is offline, and is performed a-priori to the online detection phase. The learnt models are representatives of the different types of data available from the system, albeit they are based on the data that has been procured by prior modeling and/or experiments. The domain knowledge of the presence of stable, transient and the limit cycle modes of operation in the Rijke tube apparatus allows us to train $\lambda_1$ corresponding to the stable regime of operation, $\lambda_2$ corresponding to the transient growth regime and $\lambda_3$ corresponding to the limit cycle unstable regime. This step accomplishes a generative modeling of the system on the basis of the domain knowledge procured by the experiments. Although the experiments have been performed

to encompass a wide range of operating conditions, it is practically unfeasible to perform experiments that cater to all the different conditions of operations that an actual combustor may be subject to. So with a sufficiently dense set of training data procured from different regimes, HMMs can be used to encapsulate the trend in the behavioral characteristics of the system in the different regimes. That is to say that $\lambda_1$, for example, serves as a representative model for all different "types" of stable behavior in the system (since the process is stochastic, no two datasets from the same regime are identical). Such a "context-based" learning makes sense in this respect because the pressure signature of the system in the limit cycle mode of operation, for example, is expected to have a deterministic nature, which is characteristically different from the noisy or chaotic nature generally exhibited in the stable regime. Learning ensemble models for each regime ensures that the modeling is robust to issues like variations in operational conditions and differences in sensor noise levels for even a single mode of operation. This indicates that the method is robust to changes in specific nominal conditions, since the learnt models already take into account a wide range of variations of the conditions based on the range of operation of the experiments.

The test data for instability detection purposes is chosen to be curtailed just before the limit cycle instability sets in, and the performance of the classifier for instability detection is based on successful online detection of the onset of the growth phase. With the pre-trained models, given a window of time series data $\{y_1, ...., y_T\}$, we calculate the following log-likelihood ratio (LLR):

$$\mathcal{L}_{k,1} = log\left[\frac{(p(y_{1:T})|\lambda_k)}{(p(y_{1:T})|\lambda_1)}\right] = log(p(y_{1:T})|\lambda_k) - log(p(y_{1:T})|\lambda_1) \tag{9}$$

where k=2 or 3; and $(p(y_{1:T})|\lambda_k)$ denotes the probability that the observed pressure time series sequence is generated by model $\lambda_k$. This follows from the idea that as the system deviates from a stable operating mode and passes through the bifurcation point towards limit cycle instability, determinism in the form of periodicity sets in the pressure signature, which can be quantified by a greater probability of being generated from either $\lambda_2$ or $\lambda_3$ rather than $\lambda_1$. This is reflected in the log-likelihood ratio as defined in Equation 9 being positive.

We focus on short time-windows of time series data for this analysis, in order to address the issues of real-time detection and control, where early detection of TAI with short data-window could could provide appropriate lead-time for the actuators to implement the control action. The process involved in the analysis is described as follows:

1. First an appropriate window-size is chosen. In this work we have mostly focused on window sizes that correspond to a time scale of ∼10-100 milliseconds.
2. Time series data in subsequent batches of the chosen window length is used to calculate $(p(y_{1:T})|\lambda_k)$, for k= 1, 2 and 3.
3. The log-likelihood ratio as defined by Equation 9 is chosen as the norm for detecting instability onset. The ratio is calculated for each batch of data being

analyzed and the online condition monitoring is performed based on the evolution of the norm over batches of time series samples.

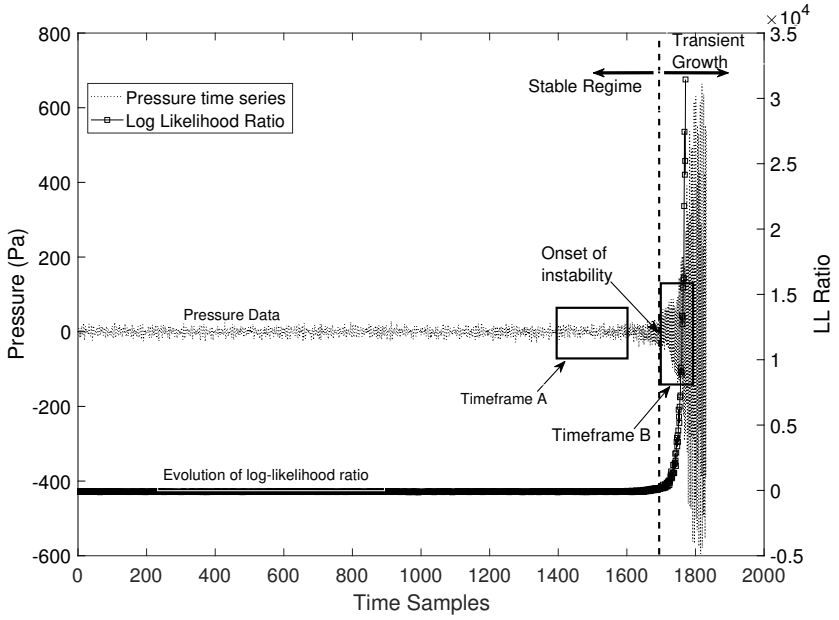


Fig. 15: Evolution of $[L_2 - L_1]$ with the pressure time series data(reproduced with permission from Mondal et al. [11]).

Figures 15, 16 and 17 elucidate the process of online detection of TAI. Figure 15 shows the evolution of $[L_2 - L_1]$ with the analysis being done on batches of a chosen window length of 100 samples. It is clearly seen that there is a sharp rise in the LLR as soon as the instability sets in. The discriminating property of LLR is demonstrated by the fact that the order of magnitude in the change of $[L_2 - L_1]$ is almost $10^4$. Thus, LLR is very sensitive to the rate of growth in the time series data. More insights are obtained by analyzing the variation of the three likelihoods with respect to $\lambda_1, \lambda_2$ and $\lambda_3$ during Timeframes A and B as shown in Figure 15. Timeframe A spans from around 1400-1600 time samples in Figure 15, which is a sample from the stable phase of the data. This is reflected in Figure 16 by the highest relative likelihood of the data corresponding to $\lambda_1$ (indicated by the downfacing shaded triangles), as compared with the likelihoods with respect to $\lambda_2$ (shaded squares) and $\lambda_3$ (circles). The instability sets in around the 1700 s mark, hence Timeframe B spanning from 1600 - 1800 time samples contains the early phase of the transient period. As can be seen in Figure 16, likelihood corresponding to model $\lambda_1$ drops sharply during the early transient period, a property that can be utilized for automated condition monitoring of combustion systems for the onset of TAI.

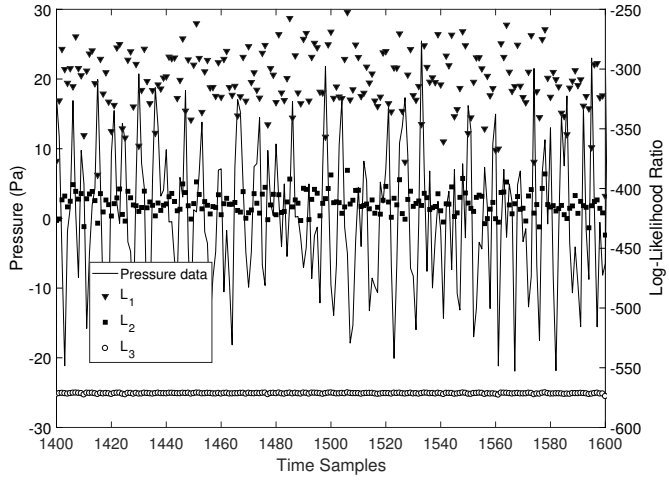


Fig. 16: Evolution of $L_1, L_2$ and $L_3$ with the pressure samples in Timeframe A of Figure 15 (reproduced with permission from Mondal et al. [11]).

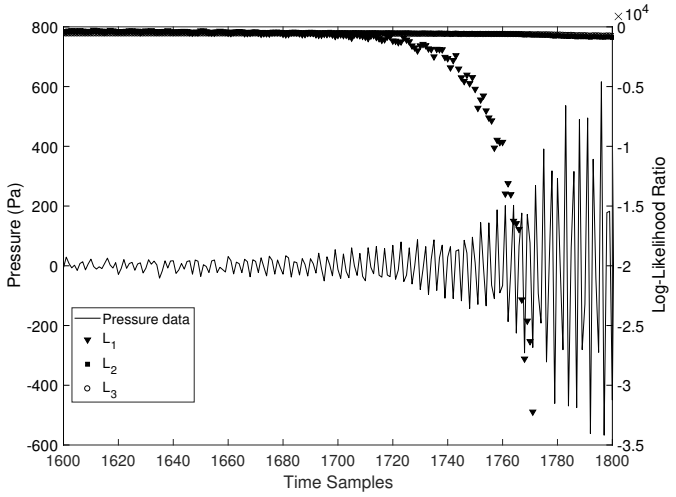


Fig. 17: Evolution of $L_1, L_2$ and $L_3$ with the pressure samples in Timeframe B of Figure 15(reproduced with permission from Mondal et al. [11]).

For analyzing the detection accuracy with respect to different window sizes, we look into a Binary Hypothesis testing problem where our proposed framework classifies each pressure time series window into either *stable* (Class A) or *unstable* (Class B) based on the likelihood-ratio test:

$$log\left[\frac{(p(y_{1:T})|\lambda_{k=2\ or\ 3})}{(p(y_{1:T})|\lambda_1)}\right] \underset{\alpha}{\overset{\beta}{\gtrless}} \tau \tag{10}$$

where $\tau$ is a threshold [27]. A commonly used criterion to choose an appropriate threshold $\tau$ is by using the Receiver Operating Characteristic (ROC) curve. The ROC curve is obtained by varying $\tau$ which provides a trade-off between the probability of successful detection $\left(p_D \triangleq p[Decided\ Class = \beta | True\ Class = \beta]\right)$ and the probability of false alarms $\left(p_F \triangleq p[Decided\ Class = \beta | True\ Class = \alpha]\right)$.

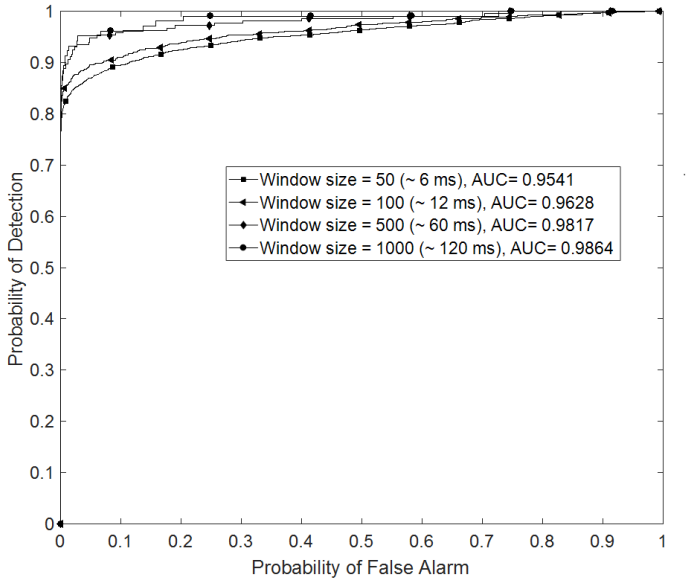


Fig. 18: ROC curves for different window-sizes(reproduced with permission from Mondal et al. [11]).

Figure 18 shows a family of ROC curves for the proposed detection algorithm using different window sizes of data, with the likelihood ratio chosen as $log\left[\frac{(p(y_{1:T})|\lambda_2)}{(p(y_{1:T})|\lambda_1)}\right]$. Window sizes of 50, 100, 500 and 1000 were chosen for this demonstration, which corresponds to time scales in the order of milliseconds considering the sampling frequency of 8192 Hz. A commonly used method for comparing the performance of different classifiers is by comparing the area under the ROC curve (AUC) for each classifier [28]. Higher AUC is generally associated with a better overall performance of a classifier. As depicted in Figure 18, the AUC increases from 0.9541 to 0.9864 as the window size is increased from 50 to 1000, with progressively higher detection rates at specified false alarm rates. This can be accounted to the fact that the classifier is expected to perform better with respect to classifying larger lengths of observation sequences. The high classification accuracy even with

about 6-60 milliseconds of data showcases a prospective application of the HMM based algorithm for detecting instabilities in the early transient phase.

The information from the ROC curves (Figure 18) can be used to choose a particular threshold $\tau$ corresponding to the trade-off between the detection rate required by the user and the allowable false alarm in detection. This can be utilized for regime detection, whereby the classifier has to decide which of the three pre-trained regimes does a short data history belong to. We present the regime detection accuracies in Table 3, where the threshold $\tau$ has been chosen using a ROC of a classifier with $[L_3 - L_1]$ as the log-likelihood ratio, with window size = 100. $\tau$ has been chosen to correspond to about 92% detection rate with 10% allowable false alarm from the ROC curve. The problem of regime detection is particularly challenging in this respect because, with a very short history of data, amplitude-based thresholding or frequency-based detection of instabilities might be inefficient.

Table 3: Classification Accuracy of the Three Regimes(reproduced with permission from Mondal et al. [11])

| Stable Regime | Transient Regime | Limit Cycle Regime |
|---|---|---|
| 90% | 91% | 100% |

This HMM-based algorithm can be naturally extended into problems like growth-rate estimation from acoustic time series data in unstable systems. From an experimental point of view, it is difficult to extract growth rate from dynamic data, as discussed in the work by Moeck et al. [29]. An unstable combustion system generally exhibits a sudden jump from a stable operation regime to a limit cycle mode through a bifurcation, thereby posing serious challenges for data-driven growth rate extraction techniques to detect the growing trend in the data in that short period. Recently, Rigas et al. [6, 30] and Jamieson et al. [30] have reported growth and decay rate measurements using transient acoustic time series data from an electrically heated Rijke tube, similar to the setup used in our current work. Their method is based on extracting regions of linear growth and decay from the Hilbert envelope of the time series signal, and measuring the growth rate by a linear fit in the identified region. A comparison of the HMM-based algorithm with Hilbert envelope-based growth rate extraction can be found in [11], whereby it has been shown that the LLR-based growth rate estimation is significantly efficient in terms of computational expense and data requirements.

From the application point of view, LLR-based regime detection is more versatile as compared to the other data-driven methods reported in literature for prediction of instabilities, e.g. Hurst exponent based characterization by Nair and Sujith [31], minimum permutation entropy based analysis by Gotoda et al. [32] and $\times D$-Markov entropy rate based thresholding by Sarkar et al. [33]. A comparison of time complexities of the different data-driven methods can be found in [11]. In particular, a

comparison with Hurst exponents as shown in Mondal et al. [11] indicated that the computational time for LLR is comparable to the former when the data length is low. But, Hurst exponents are typically used for characterizing instability using stationary datasets, which might require more data to be statistically significant for computing the different Hurst exponents. One major difference in the method proposed in this chapter is that the HMM-based technique can be efficiently applied for transient regimes of operations, whereas most of the other methods have stationarity assumptions in their applicability. Table 4 enlists the processing times of LLR calculation for the different data lengths analyzed in this work, which shows the computational efficiency of this method, making it potentially suitable for regime detection and instability classification.

Table 4: Time Complexity of LLR calculation in the online phase(reproduced with permission from Mondal et al. [11])

| Number of samples ($F_s$ = 8192 $Hz$) | 50 | 100 | 500 | 1000 |
|---|---|---|---|---|
| Data length ($milliseconds$) | 6 | 12 | 60 | 120 |
| Processing time ($milliseconds$) | 4 | 6 | 12 | 23 |

## 7 Summary, Conclusions, and Future Research

This chapter has developed and validated a statistical modeling tool for early detection of anomalous behavior leading to thermoacoustic instabilities (TAI) in combustion systems. The underlying algorithms are built upon the concepts of symbolic time series analysis (STSA) and hidden Markov modeling (HMM) to represent the typical behavior of combustion systems at different operational regimes (e.g., steady-state, transient, and unstable). The proposed method focuses on real-time applications with measured short windows of pressure oscillations without compromising the accuracy of TAI prediction and identification of the associated operational regimes. The proposed method has been validated on experimental data from an electrically heated Rijke tube apparatus for predicting the onset of thermoacoustic instabilities.

As shown in Section 4, HMM outperforms STSA for online anomaly detection of TAI onset. In this case, the detection process is restricted to a low-delay tolerance, in which an alphabet size of $|\Sigma| = 2$ has been employed in order to generate a low-dimensional feature vector, given by the stationary state probability vector. This requirement has been relaxed in Section 5, where an alphabet size of $|\Sigma| = 6$ has been adopted, with the morph matrix being used as a feature vector. In this case, detection performance is significantly improved at the expense of a minor increase in the execution time. It is concluded that HMMs are more suited for identifying classes that are very close to each other in terms of the signal texture. For example,

the HMM outperforms STSA techniques for identification of transient regimes that are similar in signal texture to the unsteady regime. For STSA to do this well, a larger window size would be needed, which may not be feasible for online detection. If the detection of the transient regime is of interest (as it is in this chapter), then the HMM approach is preferable. On the other hand, if the major objective is to detect an anomalous operation, then the STSA-based method is recommended. Anomaly detection and regime identification are two important problems that need to be addressed in any online monitoring and control system for mitigating TAI. Since the training period of the STSA-based models are significantly cheaper than that of HMM, they can be used efficiently for detecting anomalies in the system in near real-time. The detected anomalous regimes can then be used as baselines for learning different regime-specific HMMs in a slower time-scale. Since the testing period of HMM is very fast, it can then be used online for regime classification, and the information can be useful for implementation of the required control actions. Moreover, it is worth noting that although the HMM-based detection algorithm has been applied to Rijke tube acoustic instability data, the concept is equally applicable for more complex combustion systems characteristic of turbulent combustors. Since LLR is very sensitive to small variations in texture of the data, intermittent burst to instabilities can be detected online, as they are expected to indicate a sharp drop in the corresponding likelihood to a stable model with respect to an unstable model as soon as the intermittency crops in. While there are many areas of theoretical and experimental research to enhance the work reported in this chapter, the authors suggest the following topics for future research:

1. Development of a unified detection framework by addressing other modes of instabilities (e.g., lean blowouts).
2. Extension of the proposed methods for detection of instabilities in combustion systems operating under different kinds of protocols.
3. Implementation of the proposed methods for (closed-loop) active control of the laboratory-scale apparatus with actuators for controlling instabilities.
4. Extension of the probabilistic approach in the reported work to state estimation for forecasting of future states to predict the temporal behavior.
5. Enhancement of computational efficiency of the proposed methods by using inference-based learning of the probabilistic models.

## References

1. H. Kwakernaak and R. Sivan. *Linear Optimal Control*. Wiley Interscience, New York, USA, 1972.
2. A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, USA, 1970.
3. T.C. Lieuwen and V. Yang. *Combustion Instabilities In Gas Turbine Engines: Operational Experience, Fundamental Mechanisms, and Modeling*, chapter 1, pages 3–26. American Institute of Aeronautics and Astronautics, 2005.

4. K. Matveev. *Thermoacoustic instabilities in the Rijke tube: experiments and modeling*. PhD thesis, California Institute of Technology, 2003.
5. N. Noiray and A. Denisov. A method to identify thermoacoustic growth rates in combustion chambers from dynamic pressure time series. *Proceedings of the Combustion Institute*, 36(3):3843 – 3850, 2017.
6. G. Rigas, N.P. Jamieson, L.K.B. Li, and M.P. Juniper. Experimental sensitivity analysis and control of thermoacoustic systems. *Journal of Fluid Mechanics*, 787, 2016.
7. A. Ray. Symbolic dynamic analysis of complex systems for anomaly detection. *Signal Processing*, 84(7):1115–1130, July 2004.
8. V. Rajagopalan and A. Ray. Symbolic time series analysis via wavelet-based partitioning. *Signal Processing*, 86(11):3309–3320, 2006.
9. S Bahrampour, A Ray, S Sarkar, T Damarla, and N Nasrabadi. Performance comparison of feature extraction algorithms for target detection and classification. *Pattern Recognition Letters*, 34:2126–2134, 2013.
10. K Mukherjee and A Ray. State splitting and merging in probabilistic finite state automata for signal representation and analysis. *Signal Processing*, 104:105 – 119, 2014.
11. S. Mondal, N.F. Ghalyan, A. Ray, and A. Mukhopadhyay. Early detection of thermoacoustic instabilities using hidden markov models. *Combustion Science and Technology*, 0(0):1–28, 2018.
12. E.A. Gopalakrishnan and R.I. Sujith. Influence of system parameters on the hysteresis characteristics of a horizontal rijke tube. *International Journal of Spray and Combustion Dynamics*, 6(3):293–316, 2014.
13. L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, Feb 1989.
14. N.F. Ghalyan, S. Mondal, D.J. Miller, and A.Ray. Hidden Markov modeling-based decision-making using short-length sensor time series. *ASME. J. Dyn. Sys., Meas., Control*, 141(10):104502–104502–6, 2019.
15. L.R. Rabiner and B-H. Juang. *Fundamentals of Speech Recognition*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.
16. Asok Ray. Symbolic dynamic analysis of complex systems for anomaly detection. *Signal Processing*, 84(7):1115 – 1130, 2004.
17. K. Mukherjee and A. Ray. State splitting and merging in probabilistic finite state automata for signal representation and analysis. *Signal Processing*, 104:105 – 119, 2014.
18. V. Rajagopalan and A. Ray. Symbolic time series analysis via wavelet-based partitioning. *Signal Processing*, 86(11):3309–3320, November 2006.
19. A. Subbu and A. Ray. Space partitioning via hilbert transform for symbolic time series analysis. *Applied Physics Letters*, 92(8):084107, February 2008.
20. K. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 1st edition, 2012.
21. R. N. McDonough and A. D. Whalen. *Detection of Signals in Noise*. Academic Press, 2nd edition, 1995.
22. P Chattopadhyay, S Mondal, C Bhattacharya, A Mukhopadhyay, and A Ray. Dynamic data-driven design of lean premixed combustors for thermoacoustically stable operations. *ASME. J. Mech. Des.*, 139(11):111419–111419–10, 2017.
23. N.F. Ghalyan, D.J. Miller, and A. Ray. A locally optimal algorithm for estimating a generating partition from an observed time series and its application to anomaly detection. *Neural Computation*, 30(9):2500–2529, 2018.
24. Y. Chen, Z. Sun, and K. Lam. An effective sub-superpixel-based approach for background subtraction. *IEEE Transactions on Industrial Electronics*, pages 1–1, 2019.
25. Vineeth Nair, Gireehkumaran Thampi, Sulochana Karuppusamy, Saravanan Gopalan, and R. I. Sujith. Loss of chaos in combustion noise as a precursor of impending combustion instability. *International Journal of Spray and Combustion Dynamics*, 5(4):273–290, 2013.
26. S. Sarkar, S.R. Chakravarthy, V. Ramanan, and A. Ray. Dynamic data-driven prediction of instability in a swirl-stabilized combustor. *International Journal of Spray and Combustion Dynamics*, 8(4):235–253, 2016.

27. H.V. Poor. *An introduction to signal detection and estimation*. Springer Science & Business Media, 2013.
28. T. Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861 – 874, 2006. ROC Analysis in Pattern Recognition.
29. J. Moeck, M. Bothien, C. Paschereit, G. Gelbert, and R. King. chapter Two-Parameter Extremum Seeking for Control of Thermoacoustic Instabilities and Characterization of Linear Growth. Aerospace Sciences Meetings. American Institute of Aeronautics and Astronautics, Jan 2007.
30. Nicholas P Jamieson, Georgios Rigas, and Matthew P Juniper. Experimental sensitivity analysis via a secondary heat source in an oscillating thermoacoustic system. *International Journal of Spray and Combustion Dynamics*, 9(4):230–240, 2017.
31. V. Nair and R.I. Sujith. Multifractality in combustion noise: predicting an impending combustion instability. *Journal of Fluid Mechanics*, 747:635âĂŞ655, 2014.
32. H. Gotoda, M. Amano, T. Miyano, T. Ikawa, K. Maki, and S. Tachibana. Characterization of complexities in combustion instability in a lean premixed gas-turbine model combustor. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):043128, 2012.
33. S. Sarkar, S.R. Chakravarthy, V. Ramanan, and A. Ray. Dynamic data-driven prediction of instability in a swirl-stabilized combustor. *International Journal of Spray and Combustion Dynamics*, 8(4):235–253, 2016.