Universidad de Alcalá
BIBLIOTECA

Document downloaded from the institutional repository of the University of Alcala: http://ebuah.uah.es/dspace/

This is a preprint version of the following published document:

Tejedor, J., Macias-Guarasa, J., Martins, H. F., Martín-López, S. & González-Herráez, M. 2019, "A Contextual GMM-HMM Smart Fiber Optic Surveillance System for Pipeline Integrity Threat Detection", JLT, vol. 37, no. 18, pp. 4514-4522

Available at http://dx.doi.org/10.1109/JLT.2019.2908816

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/republishing this material for advertising or promotional purposes, creating new collective Works for resale or redistributrion to servers or list, or reuse of any copyrighted components of this work in other works.

(Article begins on next page)

This work is licensed under a

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

A Contextual GMM-HMM Smart Fiber Optic Surveillance System for Pipeline Integrity Threat Detection

Javier Tejedor, Javier Macias-Guarasa, *Member, IEEE*, Hugo F. Martins, Sonia Martin-Lopez, and Miguel Gonzalez-Herraez

***Abstract*—This paper presents a novel pipeline integrity surveillance system aimed to the detection and classification of threats in the vicinity of a long gas pipeline. The sensing system is based on phase-sensitive optical time domain reflectometry (ϕ-OTDR) technology for signal acquisition and pattern recognition strategies for threat identification. The proposal incorporates contextual information at the feature level in a Gaussian Mixture Model-Hidden Markov Model (GMM-HMM)-based pattern classification system and applies a system combination strategy for acoustic trace decision. System combination relies on majority voting of the decisions given by the individual contextual information sources and the number of states used for HMM modelling. The system runs in two different modes: (1) machine+activity identification, which recognizes the activity being carried out by a certain machine, and (2) threat detection, aimed to detect threats no matter what the real activity being conducted is. In comparison with the previous systems based on the same rigorous experimental setup, the results show that the system combination from the contextual feature information and the GMM-HMM approach improves the results for both machine+activity identification (7.6% of relative improvement with respect to the best published result in the literature on this task) and threat detection (26.6% of relative improvement in the false alarm rate with 2.1% relative reduction in the threat detection rate).**

***Index Terms*—Distributed fiber sensing, Acoustic sensing, Vibration sensing, Pipeline integrity, phase-sensitive OTDR, Pattern recognition**

I. Introduction

Fiber optic distributed acoustic sensing (DAS) with phase-sensitive optical time-domain reflectometer (ϕ-OTDR) technology has been widely used to build systems that aim to continuous monitoring of potential threats to the pipeline integrity. By adding a pattern recognition system (PRS), we can effectively reduce the number of false alarms in the system to an acceptable level and increase the cost-effectiveness of the solution [1]–[6].

In [7], [8], we showed that most of the DAS+PRS presented to that date had significant issues with respect to the pattern classification design and experimental evaluation procedures. Since then, new works have been presented [9]–[26], though they also suffer from similar issues, as there is a lack of rigorous and realistic experimental procedures: no real classification nor results are presented [10], [14], [15], [21]; not enough details on the system description or experimental procedure are provided [11]–[13], [17], [18], [26]; the data were not acquired in a realistic field environment [9], [11]–[13], [16], [17], [19], [20], [22]–[25]; and the lack of testing signals [18], [20], [23].

To address the aforementioned issues, we presented a DAS+PRS strategy that addressed pipeline integrity surveillance under fully realistic conditions, and using a rigorous experimental procedure [7], [27]–[30].

In our previous work [30], we presented a smart fiber optic surveillance system for pipeline integrity threat detection based on a Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) approach, which showed to outperform our baseline GMM-based system [28]. On the other hand, in [7], we also showed that adding contextual feature information into our baseline GMM-based system also improved the system performance.

The proposal presented in this paper extends our previous work [30] by (1) incorporating contextual feature information in the GMM-HMM modelling, and (2) presenting new decision combination techniques based on the contextual feature information and different HMM sets.

The pipeline integrity surveillance system consists of a combination of hardware and software modules. The hardware side refers to the DAS system used to record the data, and the software side refers to the pattern classification system that classifies the acoustic data acquired by the sensing system. Two different operation modes were set up in the system: machine+activity identification, where both the machine and the activity are identified, and threat detection, where just the occurrence of a threat in the pipeline must be detected.

The rest of the paper is organized as follows: Section II presents the pipeline integrity threat detection system. The experimental procedure is presented in Section III. The experiments, results, and discussion are presented in Section IV, and Section V provides some conclusions and ideas for future work.

II. Pipeline Integrity Threat Detection System

The pipeline integrity threat detection system integrates different modules, as shown in Figure 1, being an evolution of the architecture described in [30]. These modules are explained in more detail next.

J. Tejedor is with Fundación Universitaria San Pablo CEU.
H.F. Martins is with FOCUS S.L., Spain.
J. Macias-Guarasa. Martin-Lopez, and M. Gonzalez-Herraez are with the Department of Electronics, University of Alcalá, Spain.

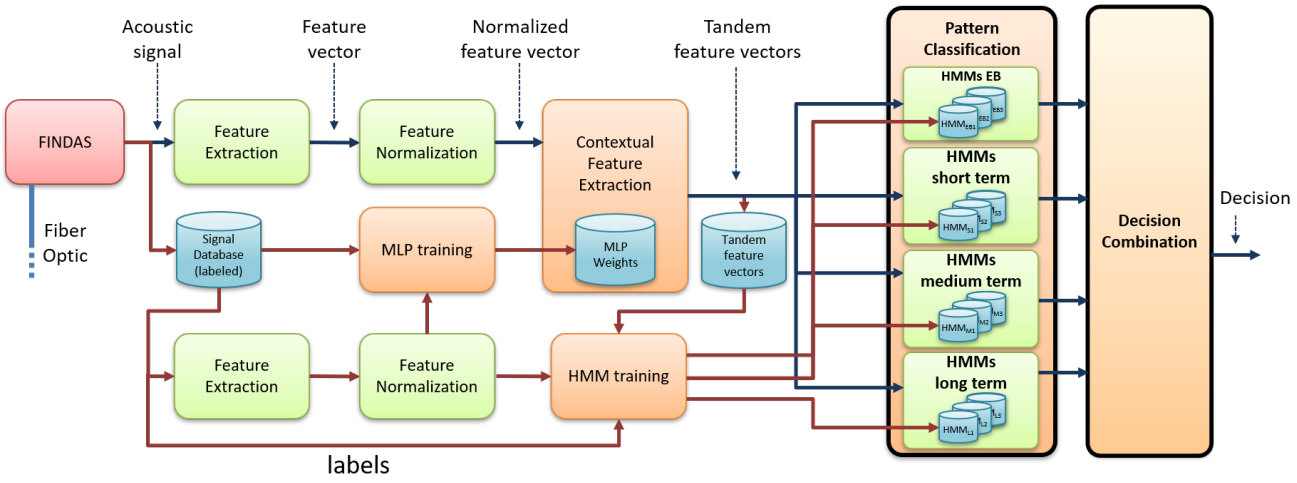

Fig. 1: Pipeline integrity threat detection system architecture. Modules in bold are the new ones with respect to [7], [30].

A. *Distributed Acoustic Sensing System*

The DAS system is a commercially available ϕ-OTDR-based sensor named FINDAS, manufactured and distributed by FOCUS S.L. A detailed description of the sensing principle and experimental setup used in the FINDAS sensor can be found in [31].

The FINDAS has an (optical) spatial resolution of 5 meters (readout resolution of one meter) and a typical sensing range of up to 45 km, using standard single-mode fiber (SMF). A sampling frequency of $f_s = 1085$ Hz was used for signal acquisition. When the energy of the vibrations monitored by the FINDAS in a certain fiber position is above a predefined threshold, the acoustic samples are recorded to form a 20-second length acoustic trace. This acoustic trace is then sent to the software module (described next) for classification purposes. The simultaneous detection of multiple activities at different positions is also possible, by setting a different threshold for each fiber position.

B. *Feature Extraction + Normalization*

Feature extraction aims to extract the meaningful information from the acoustic traces recorded by the FINDAS. This information is stored in the so-called feature vectors, each of which consists of numerical values that convey that meaningful information. To do so, each 20-second length acoustic trace is first split into segments (frames), which are 1-second long each. For each frame, a feature vector is computed, and each frame overlaps with the previous one 95% of its signal values (i.e., 1030 values for the given sampling frequency $f_s = 1085$ Hz), hence producing a *smooth* change in the frame values (i.e., 95% of the next acoustic segment share the same signal values that the current acoustic segment), so that a *smooth* change also occurs in the feature vector, as shown in [28]. Then, the Fast Fourier Transform (FFT) is applied to each frame to obtain its spectral information (i.e., the acoustic frame is converted from the time-domain to the frequency domain), since we showed in [28] that the machines and activities presented a consistent spectral behavior that made suitable for classification. The FFT size was set to 8192 points so that each frequency bin comprises 0.066 Hz (being 542.5 Hz the maximum frequency in the acoustic signal). Then, as in our previous work [30], the energy values corresponding to 100 frequency bands for a 100 Hz bandwidth are computed from the spectral information. These energy values comprise the numerical values stored in each feature vector. Then, a sensitivity-based normalization, which consists of normalizing each energy value by the energy above the considered bandwidth (from 100 Hz to 542.5 Hz) (see [28] for more details) is applied to the feature vectors to deal with the signal degradation when the distance between the sensed point and the sensor increases. Therefore, N_P-dimensional normalized feature vectors, which will be referred as baseline feature vectors, comprise the output of this module (N_P=100).

C. *Contextual Feature Extraction*

The contextual feature extraction bases on the same multi-layer perceptron (MLP) approach described in our previous work [7]. An MLP, whose typical architecture is shown in Figure 2, is a class of feedforward artificial neural network. This consists of multiple layers of computational units (an input layer, one or more hidden layers, and an output layer), which are interconnected in a feed-forward way. Each neuron in one layer has directed connections to the neurons in the next layer (hence, denoting the *forward* direction). MLPs are widely used in artificial intelligence tasks such as speech recognition [32], [33], image recognition [34], [35], audio segmentation [36], [37], etc., due to their ability of producing a classification output from input features.

The MLP used in this work has three layers, as shown in Figure 2: an input layer that consists of $N_P \cdot W_{size}$ baseline feature vector values, where W_{size} is the number of baseline feature vectors used as contextual information (for an acoustic frame analyzed at time t, the MLP will use the $W_{size}/2$ baseline feature vectors before t and the $W_{size}/2$ baseline feature vectors after t, along with the baseline feature vector generated for time t), a hidden layer, whose number of units is taken from our previous work [7], and an output layer, with

the number of units equal to the number of classes involved in the system modes (eight in the machine+activity identification mode and two in the threat detection mode). Specifically, three MLPs will be used to model the behavior of short, medium and long temporal contexts, using W_{short}, W_{medium} and W_{long} feature temporal window sizes, respectively. As in our previous work [7], the time lengths of each temporal context are 5 s, 12.5 s and 20 s, corresponding to the short, medium and long temporal contexts, respectively. 100 and 3 units were also employed in the hidden layer of the MLP for the machine+activity identification and threat detection modes, respectively [7].

The MLP models required for each temporal context (referred to as MLP_S, MLP_M and MLP_L in Figure 3) are trained by the MLP training module in Figure 1. The standard back-propagation algorithm [38] is employed to learn the MLP weights (i.e., connections between all of the units of the input and hidden layers and connections between all of the units of the hidden and output layers, as shown in Figure 2) so that the classification error on the training data is minimized. Therefore, three different sets of weights are learned (one for each temporal context), which are used next to obtain the posterior probability vectors.

The contextual feature extraction involves two different stages, which are applied to each of the different temporal contexts:

1) Posterior Probability Vector Computation: For each set of normalized feature vectors (i.e., baseline feature vectors) and using the weights computed during MLP training, the MLP is employed to calculate a posterior probability for each class to be identified that is next used to enhance the baseline feature vector. A posterior probability is the probability given to each of the classes involved in the classification when the baseline feature vector is fed in the MLP. Therefore, a set of N_C-dimensional posterior probability vectors per MLP (i.e., per temporal context) is obtained, as shown in Figure 2.

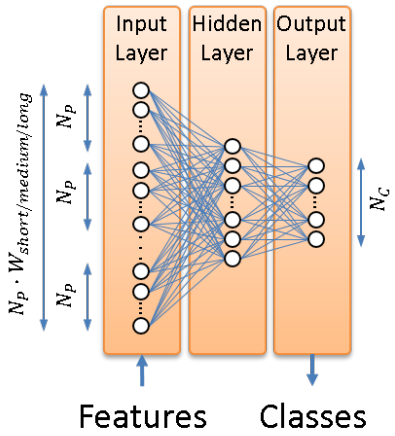

Fig. 2: Architecture of the three-layer MLP employed in the contextual feature extraction module.

Figure 3 shows the detailed architecture of the contextual feature extraction module (that contains the three MLP models MLP_S, MLP_M and MLP_L trained from the MLP training module in Figure 1), and its connection to the HMM-based pattern classification modules.

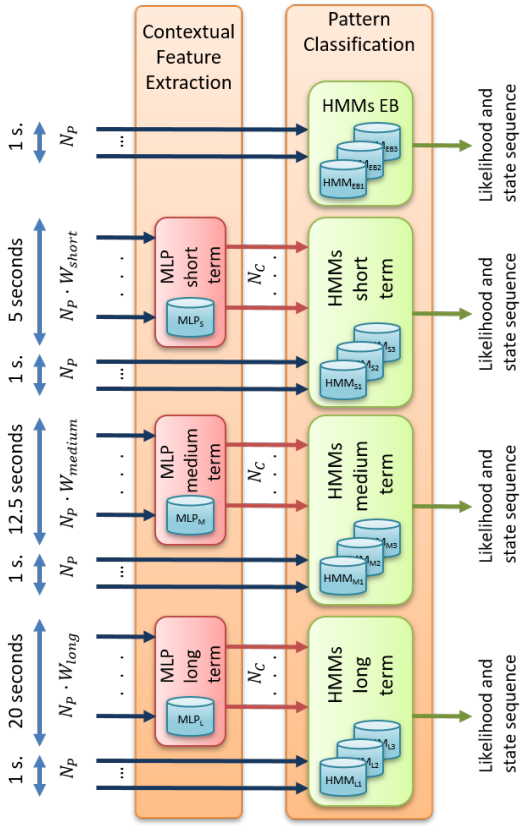

Fig. 3: Detailed architecture of the contextual feature extraction module and its connection to the HMM-based pattern classification modules.

2) Tandem Feature Vector Building: This stage concatenates the baseline N_P-dimensional normalized feature vectors (those generated by the Feature extraction+normalization module) and the N_C-dimensional posterior probability vectors computed by the MLPs. Therefore, $(N_P + N_C)$-dimensional tandem feature vectors are built (in our implementation, $N_P + N_C = 108$ for the machine+activity identification mode and $N_P + N_C = 102$ for the threat detection mode).

For MLP training, posterior probability vector computation and tandem feature vector building, the ICSI QuickNet toolkit [39] has been employed.

D. Pattern classification

Pattern classification is carried out using a GMM-HMM approach, for which training and recognition stages are necessary. HMMs have been widely used in machine learning related tasks such as speech recognition [40], [41], image recognition [42], [43], etc.

1) HMM basis: A Markov model is a statistical model in which each state corresponds to an observable (physical) event. An extension of it produces an HMM, which is a statistical model in which the observation is a probabilistic function of the state (and hence can be modelled from a GMM generating a GMM-HMM). Therefore, the resulting HMM is a doubly

embedded process with an underlying statistical process that is not observable (i.e., hidden), but can only be observed through another set of statistical processes that produce the sequence of observation. An example of an HMM is shown in Figure 4, where each state is modelled from a single Gaussian.

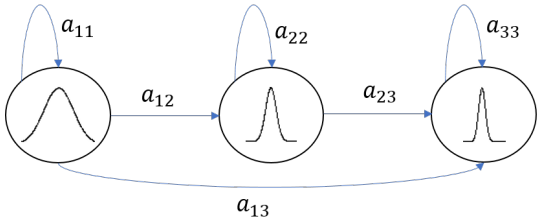

Fig. 4: HMM example with 3 states. Symbols a_{ij} denote the state transition probabilities.

2) HMM training: Training is carried out with data acquired in real field recordings, and just needs to run once. Data employed from HMM training come from (1) the contextual feature extraction and (2) the feature extraction+normalization stages, so that the tandem feature vectors and the baseline feature vectors are both used for HMM training. From data in (1), which are the new ones with respect to the data used in [30], an HMM with a single Gaussian component for each of the activities in the machine+activity identification mode, and two different HMMs (one representing the threat class and the other representing the non-threat class) with a single Gaussian component each, in the threat detection mode[1] are trained. Therefore, given the short, medium, and long temporal contexts, three different sets of HMMs, denoted as **HMMs short term**, **HMMs medium term**, and **HMMs long term**, respectively in Figure 1, are separately built for the machine+activity identification and threat detection modes.

From data in (2), which are the same as those presented in [30], an HMM training that only employs the baseline feature vectors (i.e., the N_p-dimensional feature vectors obtained in the Feature extraction+normalization stage) has also been carried out. This aims to take advantage of the complementary information that *may* exist between the baseline and the *enhanced* feature vectors. This HMM set, referred to as **HMMs EB** in Figure 1, also employs a single Gaussian component.

For each HMM set, three different HMMs that only differ in the number of states have been built. To do so, 1, 2, and 3-state HMMs have been built for each HMM set, which are next used in the recognition stage.

The HMM training consists of the estimation of the mean and the full covariance matrix of the Gaussian component, and the transition matrix probabilities for each state of the HMM from the Baum-Welch algorithm [44]. This Baum-Welch algorithm is a special case of the Expectation-Maximization algorithm applied to HMMs [45].

3) HMM recognition: HMM recognition employs the three sets of tandem feature vectors (i.e., **HMMs short term**, **HMMs medium term**, and **HMMs long term** in Figure 1) along with the baseline feature vector (i.e., **HMMs EB** in Figure 1) for signal classification.

[1]A single Gaussian component was employed for each mode due to its best performance on the GMM-based system presented in [28].

The Viterbi algorithm [44] is used to classify each test acoustic frame as the class (machine+activity or threat/non-threat) with the highest probability. The Viterbi algorithm finds the "best" path between the test acoustic frames and the previously trained HMMs. As shown in Figure 1, four sets of recognition processes (each using the set of HMM with 1, 2, and 3 states) are run from the different temporal contexts (tandem feature vectors) and the baseline feature vector to compute the individual frame-level decisions (likelihood and state sequence) for each test acoustic frame.

E. Decision combination

The HMM-based recognition stage produces different outputs depending on the number of states and the temporal context/baseline feature vector sets. Therefore, a strategy to combine all these outputs in a single decision for each acoustic trace is necessary. Let O_{EB_i}, O_{S_i}, O_{M_i}, and O_{L_i}, where $i \in 1, 2, 3$, be the output of the recognition processes that employ the baseline feature vector, and the short, medium, and long temporal context information, respectively.

Two different strategies are proposed in this work, namely *State-based decision combination* and *Temporal context-based decision combination*. In the former, a contextual information is selected first, and then we test different state-wise combinations; while in the latter, the state-wise information is selected first, and then we test different temporal context combinations. These are explained in more detail next.

1) State-based decision combination: Given temporal contexts in isolation from the corresponding context information (CI) ($CI = EB$, $CI = S$, $CI = M$, or $CI = L$) or temporal context-wise combination, all individual, pair, and trio HMM state-wise combinations are merged by majority voting. For example for $CI = M$, and the HMM state combination (SC) $SC = 1-2$, the combination employs the recognition outputs O_{M_1} and O_{M_2}. The majority voting scheme means that the class assigned to each acoustic trace is the one for which more frames are classified as that such class.

It must be noted that this decision combination method is an improvement over that presented in our previous work [30], since this combination now employs additional state combinations (in [30], only the state combination $SC = 1-2-3$ was used) and includes context information.

2) Temporal context-based decision combination: From the classification outputs obtained with the 1-state, 2-state, and 3-state HMMs, state combination is employed for a given context information/baseline feature vector. Therefore, given states in isolation ($SC = 1$, $SC = 2$, or $SC = 3$), or state-wise combination, all individual, pair, and trio context information-wise combinations are merged by the majority voting approach explained above. For example, given the state combination $SC = 1-2$, and the context informations $CI = S$ and $CI = M$, the combination employs the recognition outputs O_{S_1}, O_{S_2}, O_{M_1}, and O_{M_2}.

III. Experimental Procedure

A. Database Description

For comparison, the database employed in the experiments is the same as that used in our previous works [7], [28], [30].

TABLE I: Experimental database. 'Big excavator' is a 5 ton Kubota KX161-3. 'Small excavator' is a 1.5 ton Kubota KX41-3V.

Machine	**Activity**	**Duration in each location (in seconds)**							**Threat**
		LOC1	LOC2	LOC3	LOC4	LOC5	LOC6	**Total**	**Non-threat**
Big excavator	Moving along the ground	1100	1100	3540	1740	1620	4160	13260	Non-threat
	Hitting the ground	120	140	240	220	80	260	1060	Threat
	Scrapping the ground	460	460	920	620	200	580	3240	Threat
Small excavator	Moving along the ground	600	500	1700	820	820	1660	6100	Non-threat
	Hitting the ground	200	180	220	220	80	240	1140	Threat
	Scrapping the ground	420	340	780	360	180	520	2600	Threat
Pneumatic hammer	Compacting ground	660	0	580	1320	0	1320	3880	Non-threat
Plate compactor	Compacting ground	740	0	740	1240	0	1680	4400	Non-threat

We provide here the fundamental details, referring the reader to those references for additional information.

An active gas transmission pipeline operated by Fluxys Belgium S.A. was used for the database acquisition, thus operating in a real scenario. The FINDAS sensor is placed at one end of the fiber that runs in parallel to the inspected pipeline. The six different locations (LOC1 through LOC6) cover different pipeline "reference positions" selected at high distances from the sensing equipment (being at 22.24, 22.49, 23.75, 27.43, 27.53 and 34.27 km far from the FINDAS box, respectively) to evaluate the system in conditions close to the actual sensing limits and to ensure feature variabilities in terms of soil characteristics and weather conditions. Table I presents the different machine+activity pairs recorded in each location along with the time duration and the threat/no-threat labelling used in the threat detection mode of the system.

B. Evaluation Metrics

Classification accuracy is the main metric to evaluate the system performance both for the machine+activity identification and threat detection modes. This is computed as the ratio between the number of correctly classified testing machines and activities, and the total number of evaluated activities. A machine+activity pair is considered to be correctly detected in case the machine and the activity output by the system coincides with that of the ground truth and is within the temporal limits of the activity ground-truth time span. In the same way, a threat is correctly detected in case the system generates a threat decision within the ground truth temporal limits.

For the machine+activity identification mode, the full confusion matrix (i.e., a table showing the percentage of testing frames of a given class that have been classified as any of the considered classes) will also be presented.

Additionally, for the threat detection mode, the *Threat Detection Rate* (TDR), which corresponds to the percentage of threat testing activities that are classified as threat, usually referred to as *true positives*, and the *False Alarm Rate* (FAR), which corresponds to the percentage of non-threat testing activities that are classified as threats, usually referred to as *false positives*, were also calculated.

In the result tables, we will compare our new proposal with our previous strategies described in [7] and [30], and to provide a visual cue, we will set the background color of the table cells with the following convention: a green background means that the new proposal is better than [7], blue means that the new proposal is better than [30], and orange means that the new proposal is better than both previous systems.

IV. Experiments and Results

Experiments follow a leave-one-out cross-validation (CV) scheme with 6-folds (each containing the data recorded in each location), where 5 folds are used for any kind of training (GMM-HMMs and MLP weights), and the other location is used for test. The final test results are obtained from averaging the individual test results corresponding to each fold. For both the machine+activity identification and threat detection modes, results from the two different decision combination methods (temporal context-based decision combination and state-based decision combination) are presented in Tables II and Table V, respectively.

A. Machine+activity identification mode

Experiments rely on the combination of medium and long temporal contexts due to their best performance in [7]. These temporal contexts are further combined with the results obtained from the baseline feature vector to take advantage of the complementary errors both sets of feature vectors are expected to produce. The HMM state combination relies on a top-down architecture in which the combinations of all the three, two and three states, and just three states have been analyzed. The results presented in Table II show that for each temporal context combination, combining all the recognition outputs from the three states performs the best (see rows with 1-2-3 in the *HMM states* column). This is due to the fact that the models trained with 1, 2, and 3 states, are able to model the feature space in a versatile way so that different activities get complementary behavior, hence improving the overall system performance. The best performance is obtained by combining the medium and long temporal contexts (row **EB + M-L** and **1-2-3** in Table II), which is consistent with the results obtained in our previous work [7]. Adding the short temporal context does not provide any additional gain, probably due to the HMM modelling in medium and long temporal contexts is able to cope with the variability introduced in the short temporal context. This best performance is consistent along the different classes compared with [7] (except for the small excavator+scrapping, for which the performance is roughly similar). When comparing these results with the work presented in [30], the *hitting* and *scrapping* activities obtain a worse performance. This could be due to the fact that these

TABLE II: Decision combination results for the machine+activity identification mode with the best result in bold font. 'EB' denotes the *baseline* feature vector, 'S', 'M' and 'L' denote small, medium and long window sizes, in the contextual feature extraction. 'Acc.' stands for accuracy, 'Mov.' for moving, 'Hit.' for hitting, 'Scrap.' for scrapping, and 'Compact.' for compacting. Combination columns integrate both decision combination methods (Temporal context refers to the temporal context-based decision combination and HMM states refers to the state-based decision combination).

		Machine+activity identification								
Combination		Big excavator			Small excavator			Pneumatic Hammer	Plate Compactor	Acc.
Temporal context	HMM states	Mov.	Hit.	Scrap.	Mov.	Hit.	Scrap.	Compact.	Compact.	
EB (best in [30])	1-2-3	44.5%	35.9%	31.5%	46.6%	31.6%	53.1%	78.4%	31.4%	45.7%
M-L (best in [7])	N/A	66.1%	22.2%	33.7%	57.9%	14.3%	36.6%	78.4%	41.3%	54.9%
EB + M	1-2-3	61.7%	18.9%	37.0%	60.0%	12.3%	42.3%	79.9%	44.1%	54.7%
	2-3	56.1%	37.7%	34.0%	59.0%	22.8%	33.1%	79.4%	36.8%	51.5%
	3	63.0%	26.4%	32.1%	58.0%	42.1%	17.7%	78.5%	34.5%	52.5%
EB + L	1-2-3	70.7%	26.4%	30.9%	62.6%	15.8%	46.9%	83.0%	42.3%	58.7%
	2-3	60.3%	34.0%	30.2%	62.3%	26.3%	34.6%	82.0%	38.2%	53.8%
	3	69.4%	35.8%	29.6%	61.3%	40.4%	23.8%	80.4%	38.2%	56.5%
EB + M-L	**1-2-3**	**69.8%**	**22.6%**	**34.6%**	**63.9%**	**14.0%**	**45.4%**	**83.0%**	**45.5%**	**59.1%**
	2-3	63.3%	32.1%	35.8%	64.3%	17.5%	40.0%	80.9%	41.8%	56.2%
	3	70.6%	26.4%	36.4%	62.6%	19.3%	32.3%	79.4%	43.6%	58.0%
EB + S-M-L	1-2-3	69.7%	24.5%	36.4%	62.3%	8.8%	44.6%	83.0%	48.2%	59.1%
	2-3	63.7%	34.0%	37.0%	63.3%	8.8%	40.0%	82.0%	46.8%	56.7%
	3	70.1%	30.2%	35.8%	61.6%	12.3%	36.9%	81.4%	48.2%	58.6%

activities have the lowest amount of training data, which makes the MLP training obtain a less discriminative set of contextual features that drops the HMM performance.

Table III shows the confusion matrix corresponding to the best combination (row **EB + M-L** and **1-2-3** in Table II), where the values below chance (1/8 = 12.5%) have been removed to ease the visualization and analysis, and where color information as a visual aid has been added. In general, it is clearly seen that the diagonal contains the highest figures for each class (with at least 11.8% absolute better accuracy compared to the second most recognized one, i.e., 34.6%−22.8% = 11.8% in the big excavator+scrapping class), except for the *hitting* activity, which is confused with *moving* and *scrapping*. *Hitting* and *scrapping* are the classes with the lowest amount of training data, which derives in a less robust HMM modelling, and they present different acoustic behaviors (moving up the shovel, moving it down, hitting, scrapping, moving, etc.), which may be more difficult to model.

It is also worthy the significant improvements in the identification rates with respect to the baseline system [30] and also with the proposal in [7], as shown in Table IV. The relative performance improvement between the baseline [30] and our new proposal has an average value of 29.3%. When compared with [7] (which achieves the best results published so far in the literature on this task), we still get an average improvement of 7.6%, which is especially important for the small excavator+scrapping class (24.1% relative improvement) while it achieves smaller improvements for the other *hitting* and *scrapping* activities, with a marginal degradation from 14.3% to 14.0% in the small excavator+hitting class.

B. Threat detection mode

Experiments rely on the combination of all the three temporal contexts due to its best performance in [7]. Different HMM state combinations are presented in our results below: (1) The combination that employs all the three states; (2) the combination that merges the recognition outputs from 1 and 3 state-HMMs. This second combination aims to take advantage of a short-duration model (that may fit better the *stable* activities such as *moving* and *compacting*) and a long-duration model (that may fit better the non-stable activities such as *hitting* and *scrapping*); (3) the combination with a single state to let the reader know the actual limits of the system. Results in Table V show the following findings: (1) The best threat detection rate is obtained by combining the results from the baseline feature vector and the short temporal context information with 1 and 3 state-HMMs. On the one hand, the use of short and long duration models is giving complementary errors so that the combination obtains a better result. On the other hand, the short temporal context is able to cope with the intra-class variability that exists in the threat detection mode of the system, in which different activities are modelled in a single model; (2) the lowest false alarm rate is obtained with the fusion of the recognition outputs obtained from the baseline feature vector, and the medium and long temporal contexts with a single state for HMM modelling. The non-threat classes present a more *stable* behavior than the threat classes and those non-threat classes largely benefit from a single state and longer temporal contexts, since just a *single* behavior needs to be modelled. Since the amount of non-threat test data is larger than that of the threat classes, the FAR will decrease as long as these non-threat classes are correctly identified. An optimal trade-off between both rates can be found when combining the results obtained from 1-state HMM-based recognition processes, each employing the baseline feature vector, and the short and medium temporal contexts (TDR=89.1%, FAR=39.4% as shown in row **EB + S-M** and **1** in Table V). This represents a 26.6% relative improvement in the FAR with 2.1% relative reduction in the TDR compared with the baseline GMM-HMM system [30],

TABLE III: Confusion matrix of the EB + M-L temporal context-based decision combination and 1-2-3 HMM state-based decision combination in Table II for the machine+activity identification mode. Classification accuracy is shown in each cell. The values between brackets represent the number of frames that are classified as the recognized class or that belong to the real class.

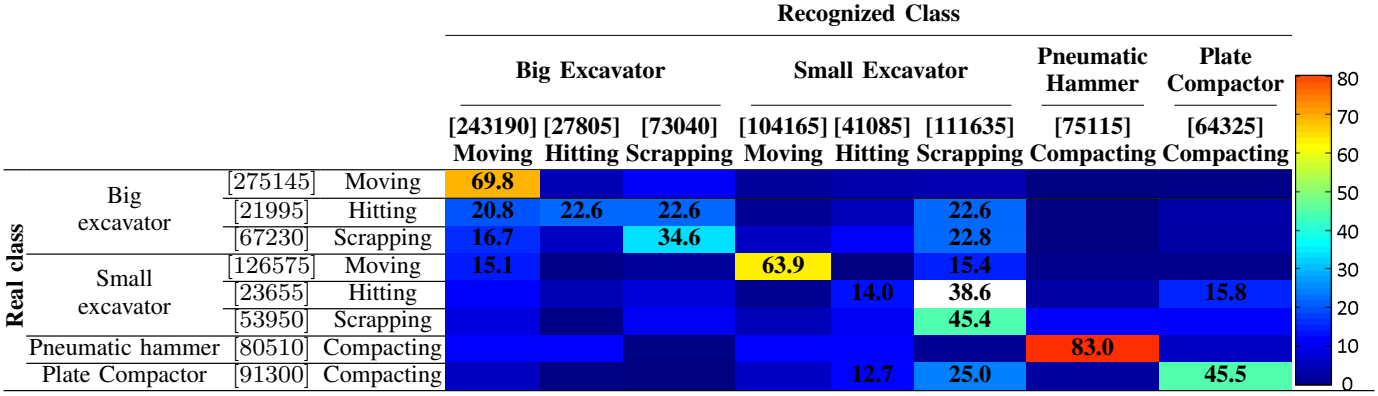

TABLE IV: Machine+activity identification mode rate comparison with previous works [7], [30]. Relative improvement is calculated as $100 \cdot \frac{(novel_{accuracy} - baseline_{accuracy})}{baseline_{accuracy}}$. The new proposal refers to the results obtained from the EB + M-L temporal context-based decision combination and the 1-2-3 HMM state-based decision combination in Table II.

	Big excavator			Small excavator			Pneumatic Hammer	Plate Compactor	Accuracy
	Moving	Hitting	Scrapping	Moving	Hitting	Scrapping	Compacting	Compacting	
Baseline [30]	44.5%	35.9%	31.5%	46.6%	31.6%	53.1%	78.4%	31.4%	45.7%
Context [7]	66.1%	22.2%	33.7%	57.9%	14.3%	36.6%	78.4%	41.3%	54.9%
New proposal	69.8%	22.6%	34.6%	63.9%	14.0%	45.4%	83.0%	45.5%	59.1%
Relative improv. new proposal vs. [30]	**56.9%**	**-36.9%**	**9.8%**	**37.2%**	**-55.6%**	**-14.5%**	**5.9%**	**44.8%**	**29.3%**
Relative improv. new proposal vs. [7]	**5.7%**	**2.2%**	**2.5%**	**10.4%**	**-2.0%**	**24.1%**	**5.9%**	**10.1%**	**7.6%**

and 9.9% relative improvement in the TDR with a relative increase of 11.3% in the FAR compared with the baseline contextual feature information GMM system [7].

V. Conclusions and Future Work

This paper has presented an evolution of the system presented in [30] combined with the strategy described in [7], and is able to continuously monitor potential threats to the pipeline integrity in real field conditions. An augmented system that combines contextual feature information with HMM modelling is able to outperform the baseline HMM system and the contextual feature system by merging both approaches. The proposal has been tested on a machine+activity identification task, where both the machine and the activity that is carried out are detected, and in a threat detection task, where just the occurrence of a threat is detected.

In the machine+activity identification mode, the new system combination proposal achieves an accuracy improvement of 29.3% compared with the baseline system described in [30], and is still better than the best results obtained so far in this task and described in [7], with an average improvement of 7.6% in this case.

In the threat detection mode, the new proposal achieves a TDR that is very close to the best one published so far in [30] (89.1% versus 91%), and obtains very relevant improvements in the FAR (39.4% versus 53.7%) and the overall accuracy (67.0% versus 56.4%). Although our previous system in [7] gets slightly better rates in false alarms and overall accuracy, its threat detection rate is much lower (81.1%).

Future work will analyze new decision combination methods that aim to take advantage of the complementary errors different classification systems may convey.

Acknowledgments

Some authors were supported by funding from the European Research Council through Starting Grant UFINE (grant number #307441), Water JPI, the WaterWorks2014 Cofunded Call, the European Commission (Horizon 2020) through project H2020-MSCA-ITN-2016/722509 - FINESSE, the Spanish Ministry of Economy and Competivity, the Spanish "Plan Nacional de I+D+i" through projects TEC2013-45265-R, TEC2015-71127-C2-2-R, and TIN2016-75982-C2-1-R, and the regional program SINFOTONCM: S2013/MIT-2790 funded by the "Comunidad de Madrid". HFM acknowledges funding through the FP7 ITN ICONE program, grant number #608099 funded by the European Commission. SML acknowledges funding from the Spanish Ministry of Science and Innovation through a "Ramón y Cajal" contract.

References

[1] A. Papp, C. Wiesmeyr, M. Litzenberger, H. Garn, and W. Kropatsch, "A real-time algorithm for train position monitoring using optical time-domain reflectometry," in *Proc. of Intelligent Rail Transportation*, 2016, pp. 1–5.

[2] Q. Sun, H. Feng, X. Yan, and Z. Zeng, "Recognition of a phase-sensitivity OTDR sensing system based on morphologic feature extraction," *Sensors*, vol. 15, pp. 15 179–15 197, 2015.

TABLE V: Decision combination results for the threat detection mode with the best result in bold font. 'TDR' stands for threat detection rate and 'FAR' for false alarm rate. 'EB', 'S', 'M', 'L', and 'Acc.', denote the same as in Table II. Combination columns integrate both decision combination methods (Temporal context refers to the temporal context-based decision combination and HMM states refers to the state-based decision combination).

Combination		Threat detection		
Temporal context	HMM states	TDR	FAR	Acc.
EB (best in [30])	1-2-3	91.0%	53.7%	56.4%
S-M-L (best in [7])	N/A	81.1%	35.4%	68.3%
EB + S	1-2-3	90.8%	56.5%	54.2%
	1-3	92.8%	58.8%	52.9%
	1	91.5%	48.4%	60.6%
EB + M	1-2-3	88.1%	50.6%	58.1%
	1-3	90.5%	48.7%	60.2%
	1	91.5%	48.8%	60.3%
EB + L	1-2-3	87.1%	47.9%	60.0%
	1-3	89.6%	49.4%	59.4%
	1	91.8%	51.8%	58.0%
EB + S-M	1-2-3	84.8%	47.3%	59.9%
	1-3	89.8%	47.0%	61.3%
	1	**89.1%**	**39.4%**	**67.0%**
EB + S-L	1-2-3	84.6%	44.4%	62.1%
	1-3	89.1%	47.8%	60.5%
	1	89.1%	39.6%	66.9%
EB + M-L	1-2-3	82.8%	41.2%	64.2%
	1-3	85.6%	43.2%	63.3%
	1	84.8%	37.3%	67.7%
EB + S-M-L	1-2-3	84.1%	46.2%	60.7%
	1-3	86.8%	43.6%	63.3%
	1	89.6%	41.1%	65.8%

[3] H. Wu, S. Xiao, X. Li, Z. Wang, J. Xu, and Y. Rao, "Separation and determination of the disturbing signals in phase-sensitive optical time domain reflectometry (ϕ-OTDR)," *Journal of Lightwave Technology*, vol. 33, no. 15, pp. 3156–3162, 2015.

[4] C. Cao, X. Y. Fan, Q. W. Liu, and Z. Y. He, "Practical pattern recognition system for distributed optical fiber intrusion monitoring system based on phase-sensitive coherent OTDR," in *Proc. of Asia Communications and Photonics Conference*, 2015, pp. 145:1–145:3.

[5] H. Wu, Z. Wang, F. Peng, Z. Peng, X. Li, Y. Wu, and Y. Rao, "Field test of a fully distributed fiber optic intrusion detection system for long-distance security monitoring of national borderline," in *Proc. of SPIE*, vol. 91579, 2014, pp. 915 790–1–915 790–4.

[6] C. Conway and M. Mondanos, "An introduction to fibre optic intelligent distributed acoustic sensing (iDAS) technology for power industry applications," in *Proc. of International Conference on Insulated Power Cables*, vol. A3.4, 2015, pp. 1–6.

[7] J. Tejedor, J. Macias-Guarasa, H. Martins, D. Piote, J. Pastor-Graells, S. Martin-Lopez, P. Corredera, and M. Gonzalez-Herraez, "A novel fiber optic based surveillance system for prevention of pipeline integrity threats," *Sensors*, vol. 17, no. 2, pp. E355:1–E355:19, 2017.

[8] J. Tejedor, J. Macias-Guarasa, H. F. Martins, J. Pastor-Graells, S. Martin-Lopez, P. Corredera, G. D. Pauw, F. D. Smet, W. Postvoll, C. H. Ahlen, and M. Gonzalez-Herraez, "Real field deployment of a smart fiber optic surveillance system for pipeline integrity threat detection: Architectural issues and blind field test results," *Journal of Lightwave Technology*, vol. 36, no. 4, pp. 1052–1062, 2017.

[9] F. Bi, T. Zheng, H. Qu, and L. Pang, "A novel fiber optic based surveillance system for prevention of pipeline integrity threats," *Photonic Sensors*, vol. 6, no. 2, pp. 143–152, 2016.

[10] R. Sifta, P. Munster, P. Sysel, T. Horvath, V. Novotny, O. Krajsa, and M. Filka, "Distributed fiber-optic sensor for detection and localization of acoustic vibrations," *Metrology and measurement systems*, vol. XXII, no. 1, pp. 111–118, 2015.

[11] C. Cong, F. Xinyu, L. Qingwen, and H. Zuyuan, "Practical pattern recognition system for distributed optical fiber intrusion monitoring based on ϕ-cotdr," *ZTE communications*, vol. 15, no. 3, pp. 1–4, 2017.

[12] H. Wu, Y. Qian, W. Zhang, and C. Tang, "Feature extraction and identification in distributed optical-fiber vibration sensing system for oil pipeline safety monitoring," *Photonic Sensors*, vol. 7, no. 4, pp. 305–310, 2017.

[13] Q. Yan, Y. Yang, Z. Zhang, S. Jiang, Y. Xu, and C. Zhou, "Distributed vibration sensing system for oil and gas pipelines based on COTDR and BP neural network," in *Proc. of International Conference on Optical Fiber Sensors*, 2018.

[14] Y. Chafiq, M. Tatin, O. Postaspana, J.-B. Paris, D. Maraval, and V. Lamour, "Fiber optic sensing for monitoring structure and health of railway infrastructures," in *Proc. of International Conference on Optical Fiber Sensors*, 2018.

[15] G. Cedilnik, R. Hunt, and G. Lees, "Advances in train and rail monitoring with DAS," in *Proc. of International Conference on Optical Fiber Sensors*, 2018.

[16] L. Shiloh, A. Eyal, and R. Giryes, "Deep learning approach for processing fiber-optic DAS seismic data," in *Proc. of International Conference on Optical Fiber Sensors*, 2018.

[17] F. Jiang, H. Li, Z. Zhang, Y. Zhang, and X. Zhang, "Localization and discrimination of the perturbation signals in fiber distributed acoustic sensing systems using spatial average kurtosis," *Sensors*, vol. 18, no. 9, pp. 2839:1–2839:16, 2017.

[18] S. Liang, X. Sheng, S. Lou, Y. Feng, and K. Zhang, "Combination of phase-sensitive OTDR and michelson interferometer for nuisance alarm rate reducing and event identification," *IEEE Photonics Journal*, vol. 8, no. 2, pp. 6 802 112:1–6 802 112:12, 2016.

[19] W. Zhaoyong, P. Zhengqing, Y. Qing, C. Haiwen, Q. Ronghui, and F. Zujie, "Fast pattern recognition based on frequency spectrum analysis used for intrusion alarming in optical fiber fence," *Chinese Journal of Lasers*, vol. 42, no. 4, pp. 167–172, 2015.

[20] C. Xu, J. Guan, M. Bao, J. Lu, and W. Ye, "Pattern recognition based on enhanced multifeature parameters for vibration events in ϕ-OTDR distributed optical fiber sensing system," *Microwave and Optical Technology Letters*, vol. 59, no. 12, pp. 3134–3141, 2017.

[21] Y. Wang, B. Jin, Y. Wang, D. Wang, X. Liu, and Q. Bai, "Real-time distributed vibration monitoring system using ϕ-OTDR," *IEEE Sensors*, vol. 17, no. 5, pp. 1333–1341, 2017.

[22] X. Huang, H. Zhang, K. Liu, T. Liu, Y. Wang, and C. Ma, "Hybrid feature extraction-based intrusion discrimination in optical fiber perimeter security system," *IEEE Photonics Journal*, vol. 9, no. 1, pp. 7 800 212:1–7 800 212:12, 2017.

[23] H. Qu, T. Zheng, L. Pang, and X. Li, "A new two-dimensional method to detect harmful intrusion vibrations for optical fiber pre-warning system," *Optik*, vol. 127, no. 10, pp. 4461–4469, 2016.

[24] Z. Yan, L. Shuqin, L. Sheng, and W. Peng, "Study of pattern recognition based on multi-characteristic parameters for ϕ-OTDR distributed optical fiber sensing system," *Chinese Journal of Letters*, vol. 42, no. 11, 2015.

[25] F. Jiang, H. Li, Z. Zhang, and X. Zhang, "An event recognition method for fiber distributed acoustic sensing systems based on the combination of MFCC and CNN," in *Proc. of SPIE*, 2018, pp. 10 618:1–10 618:7.

[26] M. Aktas, T. Akgun, M. U. Demircin, and D. Buyukaydin, "Deep learning based multi-threat classification for phase-OTDR fiber optic distributed acoustic sensing applications," in *Proc. of SPIE*, 2017, pp. 102 080G:1–102 080G:18.

[27] H. Martins, D. Piote, J. Tejedor, J. Macias, J. Pastor, S. Martin, P. Corredera, F. D. Smet, W. Postvoll, C. Ahlen, and M. Gonzalez, "Early detection of Pipeline Integrity Threats using a SmarT fiber-OPtic surveillance system: The PIT-STOP project," in *Proc. of International Conference on Optical Fibre Sensors*, 2015, pp. 96 347X:1–96 347X:4.

[28] J. Tejedor, H. F. Martins, D. Piote, J. Macias-Guarasa, J. Pastor-Graells, S. Martin-Lopez, P. Corredera, F. D. Smet, W. Postvoll, and M. Gonzalez-Herraez, "Towards prevention of pipeline integrity threats using a smart fiber optic surveillance system," *Journal of Lightwave Technology*, vol. 34, no. 19, pp. 4445–4453, 2016.

[29] H. Martins, D. Piote, J. Tejedor, J. Macias, J. Pastor, S. Martin, P. Corredera, F. D. Smet, W. Postvoll, C. Ahlen, and M. Gonzalez,

"Towards detection of Pipeline Integrity Threats using a SmarT fiber-OPtic surveillance system: PIT-STOP project blind field test results," in *Proc. of International Conference on Optical Fibre Sensors*, 2017, pp. 103 231K:4–103 231K:4.

[30] J. Tejedor, J. Macias-Guarasa, H. F. Martins, S. Martin-Lopez, and M. Gonzalez-Herraez, "A Gaussian mixture model-Hidden Markov model (GMM-HMM)-based fiber optic surveillance system for pipeline integrity threat detection," in *Proc. of International Conference on Optical Fiber Sensors*, 2018.

[31] H. F. Martins, S. Martín-López, P. Corredera, M. L. Filograno, O. Frazão, and M. González-Herráez, "Coherent noise reduction in high visibility phase sensitive optical time domain reflectometer for distributed sensing of ultrasonic waves," *Journal of Lightwave Technology*, vol. 31, no. 23, pp. 3631–3637, 2013.

[32] W.-Y. Chen, S.-H. Chen, and C.-J. Lin, "A speech recognition method based on the sequential multi-layer perceptrons," *Neural network*, vol. 9, no. 4, pp. 655–669, 1996.

[33] S. Thomas, P. Nguyen, G. Zweig, and H. Hermansky, "MLP based phoneme detectors for automatic speech recognition," in *Proc. of ICASSP*, 2011, pp. 5024–5027.

[34] S. Bhattacharyya and U. Maulik, "Transformation invariant image recognition using multilayer perceptron," in *Soft Computing for Image and Multimedia Data Processing*, 2013, pp. 73–87.

[35] H. Boughrara, M. Chtourou, C. B. Amar, and L. Chen, "Facial expression recognition based on a mlp neural network using constructive training algorithm," *Multimedia Tools and Applications*, vol. 75, no. 2, pp. 709–731, 2016.

[36] H. Meinedo and J. Neto, "A stream-based audio segmentation, classification and clustering pre-processing system for broadcast news using ANN models," in *Proc. of Interspeech*, 2005, pp. 237–240.

[37] V. Mitra and C.-J. Wang, "Content based audio classification: a neural network approach," *Soft Computing*, vol. 12, no. 7, pp. 639–646, 2008.

[38] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[39] D. Johnson *et al.*, "ICSI Quicknet software package," 2004, http://www.icsi.berkeley.edu/Speech/qn.html, Last access 11/2018.

[40] M. Gales and S. Young, "The application of hidden markov models in speech recognition," *Foundations and Trends in Signal Processing*, vol. 1, no. 3, pp. 195–304, 2007.

[41] R. S. Chavan and G. S. Sable, "An overview of speech recognition using HMM," *International Journal of Computer Science and Information Technology*, vol. 2, no. 6, pp. 233–238, 2013.

[42] G. Vstovsky and A. Vstovskaya, "A class of hidden Markov models for image processing," *Pattern recognition letters*, vol. 14, no. 5, pp. 391–396, 1993.

[43] X. Ma, D. Schonfeld, and A. Khokhar, "Image segmentation and classification based on a 2d distributed hidden markov model," vol. 6822, pp. 51–58, 2008.

[44] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. of the IEEE*, vol. 77, no. 2, pp. 257–286, 1987.

[45] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.