Detection and Time-of-Arrival Estimation of Underwater Acoustic Signals

Roee Diamant*, Ryan Kastner†, and Michele Zorzi‡
*Department of Marine Technology, University of Haifa, Israel
†Department of Computer Science and Engineering, University of California, San Diego, USA
‡Department of Information Engineering, University of Padova, Italy
Email: roeed@univ.haifa.ac.il, kastner@ucsd.edu, zorzi@dei.unipd.it

***Abstract*—We focus on detection and time-of-arrival (ToA) estimation of underwater acoustic signals of unknown structure. The common practice to use a detection threshold may fail when the assumed channel model is mismatched or when noise transients exist. We propose to detect and evaluate the ToA by labeling samples of observed data as 'signal' or 'noise'. Then, signal is detected when enough samples are labeled as 'signal', and ToA is estimated according to the position of the first 'signal'-related sample. We take a clustering approach, thereby obviating the need for a detection threshold and training. Our method combines a constrained expectation-maximization (EM) with the Viterbi algorithm, and becomes handy when channel conditions are rough, noise statistics is hard to estimate, and signal-to-noise ratio is low. Numerical and experimental results show that, at the cost of some additional complexity, our proposed algorithm outperforms common benchmark methods in terms of detection and false alarm rates, and in terms of accuracy of ToA estimation.**

***Index Terms*—Expectation-maximization, Viterbi algorithm, Detection in low SNR, Time-of-Arrival estimation, Acoustic detection, Clustering**

I. Introduction

Detection and time-of-arrival (ToA) estimation of underwater acoustic signals of unknown structure are the cornerstone of a multitude of applications. Detection of underwater acoustic signals is used to explore acoustic emissions of marine mammals, to passively identify projected noise of vessels, and to discover events such as underwater volcano eruptions. The ToA of the detected signals is required to estimate the range to source, and to study the source of emissions. Without prior knowledge of the structure of the required signals through e.g., a learning or a training phase, detection of underwater acoustic signals is challenging. The signals are of unknown length and structure, and may change in time. In addition, the distribution of the ambient noise is unknown and may include noise transients.

While simple blind detection techniques such as variants of the energy detector [1] seem the ideal solution for the problem at hand, in the considered harsh conditions these techniques will likely fail. This is because when the distribution of the ambient noise is mismatched with the assumed model, threshold-based detection may be triggered by any small transient. Another approach is cyclostationary analysis (cf. [2]), where detection is based on estimating some cyclic features the signal is assumed to possess (e.g., carrier frequency). However, the performance of this approach dramatically decreases when the desired signal is in the form of impulse-like or pseudo-random noise. Instead, we propose to jointly perform detection and ToA estimation by labeling the samples of observed data as *signal* or *noise*. Then, detection is decided based on, e.g., a minimum number of observations identified as 'signals', and ToA is determined as the position of the first signal-related observation.

To better explain the considered scenario and the problem at hand, consider the reception of an acoustic signal of short duration. The ambient noise is modeled as a Gaussian process $\mathcal{N} \sim (0,1)$ and the emitted signal is $\cos(2\pi t_i F_c)$, $t_i = 1, \dots, \frac{T}{F_s}$, where $F_c = 100$ Hz, $F_s = 500$ Hz, $T = 150$ ms. The signal is passed through a delay line channel with five taps. At the beginning of the observed buffer, the observations are related to the 'noise' state. At time instant 0.2 s, the signal arrives and the observations remain related to the 'signal' state until time instant 0.5355 s, after which the observations are related to the 'noise' state. This example is demonstrated in Figure 1, where we show detection results of a K-means clustering [3] by marking the noise-related observations as 'X' and the signal-related ones as circles, as well as detection results for the constant false alarm rate (CFAR) energy detector with a target false alarm probability of 10^{-4} [1]. The figure shows that using K-means clustering, the the time-of-arrival error is roughly 50 ms with a 60% error in the identification of the signal state. Using the energy detector, we observe a significant error in the estimated ToA as well as two false alarms. As Figure 1 reveals, due to the periodic nature of the signal and the strong ambient noise, there are similarities between the values of some of the observations related to the 'signal' state and the observations related to the 'noise' state. As a result, it may be hard to distinguish between the different states, and naive clustering of the observations assuming an i.i.d. state vector yields frequent clustering errors.

Without training, classifying the data samples is difficult. Instead, we adopt a Hidden Markov Model (HMM) and perform clustering. Each hidden node is connected to either a signal-related observation or a noise-related observation. Consequently, the connections between the nodes represent the transitions between states. This way, the objectives of signal detection, ToA estimation, and parameter evaluation are transformed into evaluating the value of the hidden nodes.

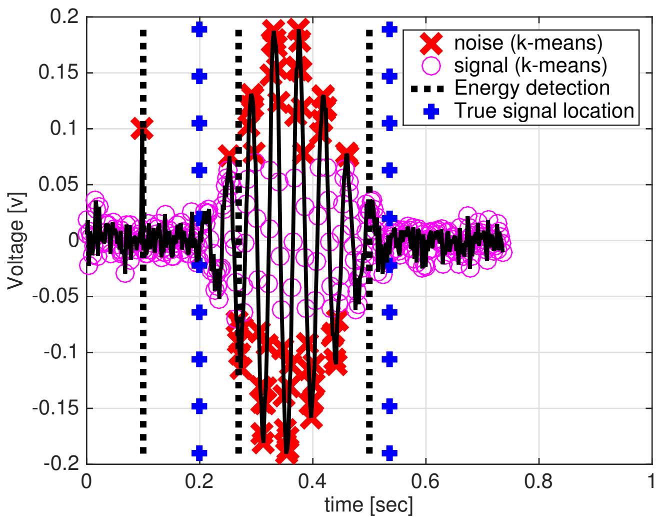

Fig. 1: Example of a signal to detect.

We propose a two step algorithm. In the first step, we use a constrained EM algorithm to calculate the statistical parameters of each state. These parameters are used to evaluate the prior probabilities of the observations and the posterior probabilities of the states. In the second step, we use the estimated statistical information to evaluate the state transition matrix and the observation emission probabilities, and to perform clustering using the Viterbi Algorithm (VA) for HMM. Considering the unknown distribution of both the signal and noise states, we allow model flexibility by using a general Gaussian mixture model. The result is a high-complexity detection scheme that does not assume knowledge of the signal, noise, or channel, and therefore does not use a detection threshold. Results from numerical simulations and a field experiment show that our algorithm outperforms benchmark techniques.

II. System Model

Our setting includes a vector of N observation samples, $\boldsymbol{x}$. We assume the received observation samples are a mixture of two i.i.d. random processes. Specifically, each observation, $x_i \in \boldsymbol{x}$, is associated with a state m where $m = 1$ represents the 'signal' type and $m = 2$ reflects the 'noise' type. The two possible states are represented by hidden nodes to form a $2 \times N$ trellis. The relations between the hidden nodes are expressed by a 2×2 transition matrix $\boldsymbol{T}$ whose $t_{m,n}$ element is the transition probability from state m to state n. The emission probabilities of the observations are denoted as $\Pr(x_i|\omega_j)$, where ω_j is the set of distribution parameters of the jth state.

We consider a binary hypothesis model, where the received sample vector consists of 1) either noise and signal, or 2) noise only. This model renders the formation of consecutive observation sequences of unknown length, and sparse states are unlikely. Specifically, in case of hypothesis 1) there are a maximum of three consecutive groups in the order: 'noise', 'signal', and 'noise', while in case of hypothesis 2) there is only one group of 'noise'. Since the transition between states is rare, we expect $t_{i,i}$ to be close to 1. However, since we do not assume to know the length of the signal, $\boldsymbol{T}$ remains unknown. Furthermore, the properties of the two states are also assumed unknown and hence $\Pr(x_i|\omega_j)$ is not given.

For the prior probabilities, we consider a mixture of distributions:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{N} \sum_{m=1}^{2} k_m p(x_i|\omega_m) \ , \tag{1}$$

where $\boldsymbol{\theta} = [\omega_1, k_1, \omega_2, k_2]$, ω_m are the parameters of the mth distribution, and k_m is the probability of the mth distribution, with $k_1 + k_2 = 1$. We model the probability density function $p(x_i|\omega_m)$ using the generalized Gaussian PDF [4],

$$p(x_i|\omega_m) = \frac{\beta_m}{2\sigma_m \Gamma\left(\frac{1}{\beta_m}\right)} e^{-\left(\frac{|x_i - \upsilon_m|}{\sigma_m}\right)^{\beta_m}} \tag{2}$$

with parameters $\omega_m = [\beta_m, \upsilon_m, \sigma_m]$. While common practice uses the normal Gaussian mixture model, (2) provides flexibility, where $\beta_m = 1$, $\beta_m = 2$, and $\beta_m \to \infty$ correspond to the Laplace, Gaussian, and uniform distributions, respectively.

We assume that the desired signal changes rapidly in time, and that its variance can be upper bounded by a *sanity check* bound, T_σ, and lower bounded by the variance of the noise signal. Thus, denoting ς_1, ς_2 as the respective variances of observations related to the two states, we have

$$\varsigma_2 < \varsigma_1 < T_\sigma \ . \tag{3}$$

Since, for the PDF (2),

$$\varsigma_m = (\sigma_m)^2 \frac{\Gamma\left(\frac{3}{\beta_m}\right)}{\Gamma\left(\frac{1}{\beta_m}\right)} \ , \tag{4}$$

and since by (4) ς_m does not change much with β_m, constraint (3) can be modified into

$$\sigma_2 < \sigma_1 < T_\sigma \sqrt{\frac{\Gamma\left(\frac{1}{\beta_1}\right)}{\Gamma\left(\frac{3}{\beta_1}\right)}} \ . \tag{5}$$

III. Labeling of Samples

Our algorithm starts with an initialization phase to find an estimation for the set of distribution parameters, $\boldsymbol{\theta}^0$, and to pre-group observations seemingly related to the same state. Next, a constrained EM is performed to find the prior and posterior probabilities, $\Pr(x_i|\eta_i)$ and $\Pr(\eta_i|x_i)$, respectively. Finally, we estimate the transition matrix and the emission probabilities and use the VA to obtain the clusters $\eta_i, \ i = 1, \ldots, N$.

A. Initialization

To estimate $\boldsymbol{\theta}^0$, we perform initial clustering using the K-means algorithm [3]. The result is an initial clustering $\hat{\eta}_i^0$, followed by grouping of all sample observations for which

$\hat{\eta}_i^0 = m,\ m = 1, 2$ into vectors $\boldsymbol{x}_m$. Then, we statistically find the parameters ω_m^0 of distribution (2) by solving

$$k_m^0 = \frac{|\boldsymbol{x}_m|}{|\boldsymbol{x}|} \tag{6a}$$
$$\upsilon_m^0 = E\left[\boldsymbol{x}_m\right] \tag{6b}$$
$$\varsigma_m^0 = \mathrm{Var}\left[\boldsymbol{x}_m\right] \tag{6c}$$
$$\kappa^0 = \mathrm{K}\left[\boldsymbol{x}_m\right]\ , \tag{6d}$$

where $|\boldsymbol{x}|$ is the size of vector $\boldsymbol{x}$, $\varsigma_m = \frac{\sigma_m^2 \Gamma\left(\frac{3}{\beta_m}\right)}{\Gamma\left(\frac{1}{\beta_m}\right)}$, and $\mathrm{K}\left[\boldsymbol{x}_m\right]$ is the Kurtosis of $\boldsymbol{x}_m$ which for distribution (2) is $\kappa^0 = \frac{\Gamma\left(\frac{5}{\beta_m}\right)\Gamma\left(\frac{1}{\beta_m}\right)}{\Gamma\left(\frac{3}{\beta_m}\right)^2} - 3$.

B. The Constrained EM

The constrained EM algorithm is performed iteratively. In the $(q+1)$th iteration, the expectation of the log-likelihood function is maximized to find the set of parameters $\boldsymbol{\theta}^{q+1}$. This maximization is performed under constraints (5). The procedure is then repeated for a pre-determined number Q of iterations (or until convergence is reached), and the convergence to a local maximum is proven [3]. Let $\boldsymbol{\lambda}$ be the clustering solution vector of the observations in $\boldsymbol{x}$. Given the previous estimate $\boldsymbol{\theta}^q$, the expectation of the log-likelihood function with respect to the conditional distribution of $\boldsymbol{\lambda}$ is

$$\begin{aligned} L(\boldsymbol{\theta}^{q+1}|\boldsymbol{\theta}^q) &= E\left[\ln\left(\Pr(\boldsymbol{x},\boldsymbol{\lambda}|\boldsymbol{\theta}^{q+1})\right)|\boldsymbol{x},\boldsymbol{\theta}^q\right] \\ &= \sum_{m=1}^{2}\left[\sum_{i=1}^{N}\Pr(\lambda_i = m|x_i,\boldsymbol{\theta}^q)\ln p(x_i|\omega_m^{q+1}) \right. \\ &\quad \left. + \sum_{i=1}^{N}\Pr(\lambda_i = m|x_i,\boldsymbol{\theta}^q)\ln k_m^{q+1}\right]\ , \end{aligned} \tag{7}$$

where $\ln x$ is the natural logarithmic function. Since the observations are assumed independent,

$$\Pr(\lambda_i = m|x_i,\boldsymbol{\theta}^q) = \frac{k_m^q p(x_i|\omega_m^q)}{p(x_i|\boldsymbol{\theta}^q)} = \frac{k_m^q p(x_i|\omega_m^q)}{\sum_{j=1}^{2} k_j^q p(x_i|\omega_j^q)}\ . \tag{8}$$

Observing (7), we estimate ω_m^{q+1} by maximizing the first term of (7) and k_m^{q+1} by maximizing the second term. For the latter,

$$k_m^{q+1} = \frac{1}{N}\sum_{i=1}^{N}\Pr(\lambda_i = m|x_i,\boldsymbol{\theta}^q), \quad m = 1, 2\ . \tag{9}$$

For the PDF (2), let us denote the first term of (7) as

$$\begin{aligned} f(\upsilon_m^{q+1},\sigma_m^{q+1},\beta_m^{q+1}) &= \sum_{i=1}^{N}\Pr(\lambda_i = m|x_i,\boldsymbol{\theta}^q) \\ &\cdot\left[\ln\beta_m^{q+1} - \ln(2\sigma_m^{q+1}) - \ln\Gamma(\frac{1}{\beta_m^{q+1}}) \right. \\ &\left. - \left(\frac{|x_i - \upsilon_m^{q+1}|}{\sigma_m^{q+1}}\right)^{\beta_m^{q+1}}\right]\ . \end{aligned} \tag{10}$$

Then, ω_m^{q+1} is found by solving

$$\omega_1^{q+1},\omega_2^{q+1} = \underset{\omega_1,\omega_2}{\mathrm{argmin}}\ -\sum_{m=1}^{2} f(\upsilon_m,\sigma_m,\beta_m) \tag{11a}$$
$$\text{s.t. }: \sigma_2 - \sigma_1 \leq 0\ , \tag{11b}$$
$$\sigma_1 - T_\sigma\sqrt{\frac{\Gamma\left(\frac{1}{\beta_1}\right)}{\Gamma\left(\frac{3}{\beta_1}\right)}} \leq 0\ . \tag{11c}$$

We observe that problem (11) is non-convex. To solve it, we use the alternating optimization approach (cf. [5]), where a multivariate maximization problem is iteratively solved through alternating restricted maximization over individual subsets of the variables.

C. Clustering

After Q EM iterations, the estimated parameters ω_m are used to estimate the posterior $\Pr(\lambda_i = m|x_i,\boldsymbol{\theta}^Q)$ using (8) and the prior $p(x_i|\lambda_i = m,\boldsymbol{\theta}^Q)$ using (2). The latter is used directly to find the emission probabilities. For the transition matrix, $\boldsymbol{T}$, we find element $t_{m,n}$ by

$$t_{m,n} = \frac{\Pr(\lambda_{i+1} = m,\lambda_i = n)}{\Pr(\lambda_i = n)}, \quad m = 1, 2,\ n = 1, 2\ . \tag{12}$$

While the denominator of (12) can be found by

$$\Pr(\lambda_i) = \frac{\Pr(\lambda_i|x_i)\Pr(x_i)}{\Pr(x_i|\lambda_i)}\ , \tag{13}$$

and

$$\Pr(x_i) = \sum_{m=1}^{2} k_m^Q p(x_i|\omega_m)\ , \tag{14}$$

the numerator of (12) cannot be found analytically. Instead, we find it statistically by setting

$$\lambda_i = \underset{m}{\mathrm{argmax}}\left[\Pr\left(x_i = m|x_i,\boldsymbol{\theta}^Q\right)\right]\ , \tag{15}$$

and counting the percentage of times $\lambda_i = n$ and $\lambda_{i+1} = m$. Once the transition matrix and the emission probabilities are calculated, we use the VA to match observation x_i with state m and to obtain clusters $\eta_i,\ i = 1,\ldots,N$.

IV. Performance Analysis

In this section, we compare the performance of our algorithm for detection and signal characterization (*EM-VA*) with the energy detector (*ED*), the basic unconstrained EM clustering for the Gaussian-mixture model (*EM*), the Baum-Welch algorithm for normal Gaussian-mixture HMM (*Baum-Welch*), and the Baum-Welch algorithm where the EM procedure of the Baum-Welch method is replaced by the simple K-means clustering (*K-means*). For the ED method [1], we use a time window of 10 samples and set the detection threshold according to a target false alarm rate of 10^{-4}. As an upper bound for the variance of the signal we use $T_\sigma = 1$.

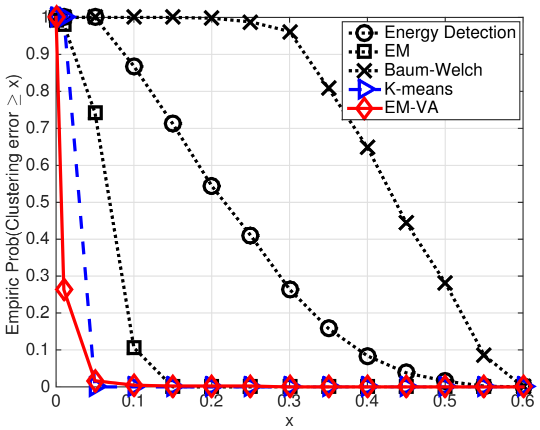

Fig. 2: Empirical C-CDFs of the clustering error for noise-only reception (false alarm) for various detection methods.

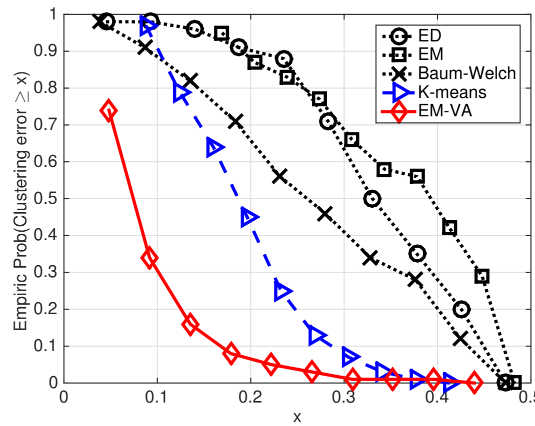

Fig. 3: Empirical C-CDFs of the clustering error for signal+noise reception (detection) for various detection methods.

A. Simulation

Our simulations include sequences of 500 samples of either ambient noise or ambient noise plus signal. In each simulation, the noise is generated as a zero mean general Gaussian random process with parameter β uniformly generated between 1 and 6, and σ is set by the desired SNR. In addition, we include noise transients randomly placed at a rate of 1% across the set of observations. These transients are generated noise samples whose variance is $10\varsigma_2$ (see (4)). For the desired signal, we use a normalized cosine signal passed through a Gaussian window. The length of the signal is chosen uniformly between 10 and 100 samples. The signal is transferred through a tap-delay-line channel impulse response. The number of taps is chosen uniformly between 5 and 50, the tap delay is set uniformly between 10 and 100 samples, and the tap complex amplitude is modeled as a Rayleigh process with variance 0.1. The location of the signal within the observed sequence is randomized uniformly between samples 50 and 400. The SNR is calculated as the power ratio between the sequence of signal-related observations and the ambient noise.

In Figure 2 we show the empirical complementary-cumulative distribution function (C-CDF) of the rate of samples falsely labeled as 'signal' when the received buffer includes only noise (false-alarm). We observe that both the Baum-Welch method, designed mainly for clustering, and the ED produce many false detections. However, the false alarm performance of the EM algorithm is far better. This is because while also in the EM method noise transients are clustered as signal samples, these errors do not propagate to further erroneous decisions as in the cases of the Baum-Welch method and of the energy detector. The dependency of these three benchmark methods in the correct evaluation of the noise and signal distributions is reflected by the results of K-means, which are better than the former. This is because K-means does not assume a given signal or noise distribution function. Yet, the performance gain of our EM-VA method is clearly observed where in most cases no single false alarm is detected.

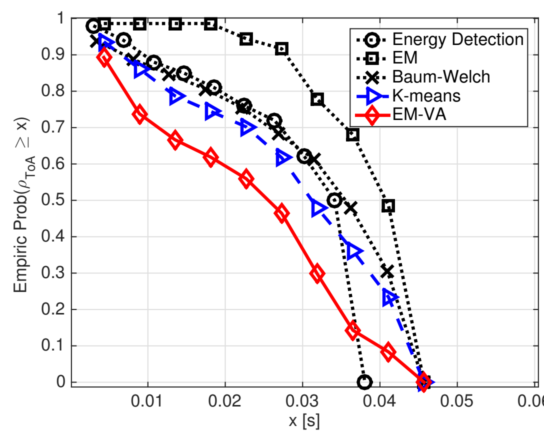

Fig. 4: (Simulation) Empirical C-CDFs of the ToA error. SNR=8.8 dB, $F_s = 1$ kHz.

A similar gain is shown in Figure 3, where we compare the rate of falsely labeled samples when the received buffer contains both noise and signal with an SNR of 8.8 dB.

Figure 4 shows the C-CDF of the error in estimating the ToA. To allow meaningful results, we use a sampling frequency of $F_s = 1$ kHz. For the benchmark clustering methods and for the ED, we observe significant errors in estimating the ToA. This is because, in terms of ToA estimation, even a single clustering error that occurs well before the true ToA due to, e.g., a noise transient, will yield significant error in estimating the ToA. Yet, also in terms of ToA estimation, clearly our EM-VA scheme achieves the best results, with an average error of roughly 15 ms.

B. Sea Experiment

To verify our assumptions and to test the effectiveness of our scheme, we measured its detection performance for 20 min of raw acoustic signals received during a sea experiment, which was conducted off the shores of San Diego. The experiment included a node drifting at depth 10 m and five anchored

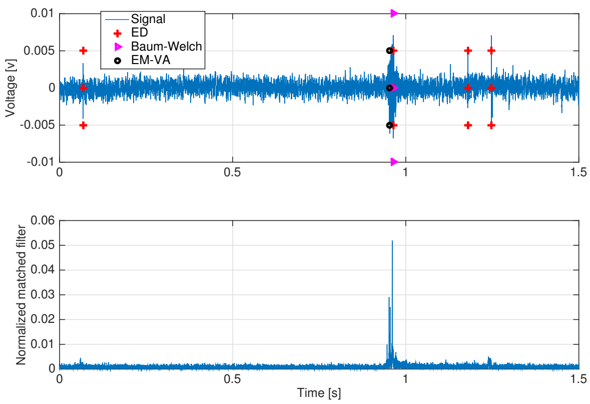

Fig. 5: Example of clustering result from the sea experiment.

beacons. Once every 5 s, each anchor transmitted a signal consisting of three sequential linear chirps, each of duration 10 ms and in the frequency range of 8 kHz to 15 kHz. The transmit power was estimated to be slightly above 190 dB Re 1μPa @1m. The signals were recorded by the drifting node with a sampling frequency of ~65.5 kHz. Since the signals were received at high power, to test the performance at various SNR values, we synthetically added recorded noise to the received signals. To measure the performance in terms of number of corrected and falsely labeled samples, we found the true position of the received signal by implementing the normalized matched filter test as reported in [6].

Detection results for one captured buffer of 1.5 s from the sea experiment are shown in Figure 5. The bottom figure shows the normalized matched filter result with a significant peak at roughly 0.95 s. We observe three false detections identified by the ED method. These detections correspond to noise transients. This shows the sensitivity of the ED method to mismatch in the noise model. No false alarms are observed for the Baum-Welch method. Yet, the ToA estimation is slightly biased compared to the peak location of the normalized matched filter. Similar results were obtained for the K-means. On the contrary, our EM-VA method shows no false alarms and accurately detects the true location of the signal.

Average results for the detection of each of the 3600 linear chirp signals are shown in Table I. In line with the simulation results, we observe that the ED obtains poor performance. Inspecting specific cases, we note that the problem of this detector lies in estimating noise transients as ‘signal”, and as a result, clustering the following noise samples as ‘signal’s’. Due to the smoothing of the VA, the results of the Baum-Welch and the K-means are much better. A surprising result is the good performance of the EM benchmark method. A possible explanation is the noise-like structure of the received acoustic signal. Yet, the best performance is obtained by our EM-VA method. We therefore conclude that also in real system conditions, at the cost of some complexity, the detection performance of our proposed scheme exceeds that of all the benchmark methods.

TABLE I: Average percentage results of correct and false signal labeling from the sea experiment.

SNR [dB]	Measure Type	ED	EM	Baum-Welch	K-means	EM-VA
N/A	False signal labeling	10%	2%	3%	1%	0%
5	Correct signal labeling	90%	92%	93%	91%	97%
10	Correct signal labeling	95%	95%	96%	98%	99%
20	Correct signal labeling	97%	99%	98%	98%	99%

V. Conclusions

In this paper, we proposed a clustering approach for detection and ToA estimation of underwater acoustic signals of unknown structure. We used a combination of a constrained general Gaussian mixture model EM algorithm and VA for HMM clustering. The former allows estimation of the unknown state transition matrix and the observation emission probabilities, while the latter handles the expected low SNR. The result is a robust detection scheme which does not use a detection threshold. Our numerical and experimental results showed that, at the cost of some additional complexity, the performance of our algorithm exceeds that of common benchmark solutions.

Acknowledgments

We would like to thank Jules Jaffe for collecting and providing access to the underwater data set. Paul Roberts and Perry Naughton provided invaluable help in preprocessing the data.

References

[1] D. Torrieri, *Introduction to Radar System*, 1st ed. New York, USA: 2nd ed. McGraw-Hill, 1980.

[2] K. Kyouwoong, I. Akbar, K. Bae, and U. Jung-Sun, “Cyclostationary approaches to signal detection and classification in cognitive radio,” in *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Dublin, Ireland, Apr. 2007, pp. 212–215.

[3] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[4] M. Novey, T. Adali, and A. Roy, “A complex generalized Gaussian distribution: Characterization, generation, and estimation,” *IEEE Trans. Signal Processing*, vol. 58, pp. 1427–1433, Mar. 2010.

[5] J. Bezdek and R. Hathaway, “Some notes on alternating optimization,” in *International Conference on Fuzzy Systems (AFSS)*, London, UK, 2002, pp. 288–300.

[6] R. Diamant, “Probability distribution analysis of a normalized matched filter for detection of underwater acoustic signals,” *Technical report, available online at arXiv:1604.06416 [physics.data-an].*