

An empirical mode decomposition based hidden Markov model approach for detection of Bryde's whale pulse calls

.....
Olayinka O. Ogundile,^{1,2} Ayinde M. Usman,¹ and Daniel J. J. Versfeld¹

¹Department of Electrical and Electronic Engineering, Stellenbosch University,
Stellenbosch, South Africa

²Department of Physics and Telecommunication, Tai Solarin University of Education, Ijebu-Ode, Nigeria
ogundileoo@gmail.com, 23172908@sun.ac.za, djversfeld@sun.ac.za

Abstract: This letter proposes an empirical mode decomposition (EMD) based hidden Markov model (HMM) approach for the detection of *mysticetes*' pulse calls such as the Bryde's whales. The HMM detection capabilities depend on the deployed feature extraction (FE) technique. The EMD is proposed as a performance efficient alternative to the popular Mel-scale frequency cepstral coefficient (MFCC) and linear predictive coefficient (LPC) FE techniques. The amplitude modulation–frequency modulation components derived from the EMD process are modified to form feature vectors for the HMM. Also, the ensemble EMD (EEMD) is adapted in a similar way as the EMD. These proposed EMD-HMM and EEMD-HMM approaches achieved better performance in comparison to the MFCC-HMM and LPC-HMM approaches. © 2020 Acoustical Society of America

[Editor: Charles C. Church]

Pages: EL125–EL131

Received: 27 November 2019 Accepted: 15 January 2020 Published Online: 11 February 2020

1. Introduction

Some *mysticetes* such as the Bryde's whale produce short pulse calls for communications or echo-locations. These characteristic pulse calls are difficult to study with the traditional visual inspection. Therefore, reliable algorithms which can accurately detect these pulse calls with little or no false positive rate, $\mathcal{F}$, is necessary. There have been numerous detection algorithms put forward in recent times to proffer a solution in this respect. Among others, machine learning (ML) tools such as k-means clustering,¹ Gaussian mixture models (GMMs),² support vector machines (SVMs),³ hidden Markov models (HMMs),⁴ and so on, are widely used in the literature because of their ability to model the characteristics of *mysticetes* pulse calls.⁵ Likewise, these ML tools have been used to detect and analyse the movement of other obscure but vocal marine species.⁶

With respect to HMMs, it has been widely used for detection and classification of cetaceans' sounds in general. HMM is an efficient and reliable ML tool that can be easily used to model cetaceans' vocal repertoire.⁷ It can easily detect and classify sounds from a defined series of observation, thereby achieving a high sensitivity, $\mathcal{S}$, and a low sensitivity, $\mathcal{F}$.^{5,6} Regardless, the performance of the HMM depends on the feature vector, which we consider a reliability vector, $\mathcal{R}$, of the HMM. In most cases, the more reliable the $\mathcal{R}$, the better the $\mathcal{S}$ and $\mathcal{F}$ performance of the HMM. The reliability vector, $\mathcal{R}$, fed into the HMM can be derived using different feature extraction (FE) techniques. However, the Mel-scale frequency cepstral coefficients (MFCCs)⁸ and linear predictive coefficients (LPCs)⁹ FE techniques are widely used for cetaceans' vocal signals in general, especially for *mysticetes*' (suborder of cetaceans) pulse calls detection and classification.

In this letter, we introduced the empirical mode decomposition (EMD)¹⁰ based HMM approach for the detection of *mysticetes*' pulse calls. The EMD is a data-driven multi-resolution decomposition tool suitable for the analysis of non-linear and non-stationary dataset. It decomposes a signal over a random duration of time into amplitude-modulation–frequency-modulation components, called intrinsic mode functions (IMFs). These IMFs are modified in this letter to form reliability vectors, $\mathcal{R}$, which are fed into a HMM. Furthermore, as a result of the mode mixing drawback of the EMD, the ensemble EMD (EEMD)¹¹ is adapted in a similar approach as the EMD to form reliability vector which can be combined with the HMM. Although the EMD have been used in sound and speech recognition in recent times,^{12,13} the EMD have not been adapted as a feature vector for the HMM as proposed in this letter. The performance of these proposed EMD-HMM and EEMD-HMM approaches were evaluated on the acoustic dataset of a continuous recording of an inshore Bryde's whale pulse calls sampled at 9600 Hz,

recorded in the endmost southwest of South Africa, as shown in Fig. 1 (image produced using Sonic Visualiser). The contribution and result of this letter is innovative. First, the EMD-HMM and EEMD-HMM produce good sensitivity, $\mathcal{S}$, and negligible false positive rate, $\mathcal{F}$, in comparison with the traditional techniques such as the MFCC-HMM and LPC-HMM. Second, the EMD-HMM and EEMD-HMM reduces the computational load added to HMM as compared to the MFCC-HMM and LPC-HMM. However, the EEMD-HMM approach achieved a better FE characteristic of the pulse call in comparison to the EMD-HMM because of the mixing mode problem associated with the EMD. Though we verified the EMD-HMM and EEMD-HMM results with the Bryde's whale pulse calls in this letter, the technique can be extended to other *mysticetes* that produce characteristic pulse calls.

2. HMM

The HMM is a ML tool popularly used for human and animal sound analysis and recognition as discussed in a tutorial by Rabiner.⁴ HMM is a stochastic state machine, where a change in the hidden states will end with the emission of a symbol. It is pliable and can effortlessly model and classify the features of a set of sound waveforms. The HMM process can be viewed in two steps: training step and detection step. During training, the HMM estimates three major parameters from the $\mathcal{R}$ of a sound waveform: (1) start probability, $\mathbb{P}_s$, (2) transition matrix, $\mathcal{T}$, and (3) emission distribution, $\psi = \{\nu, \Sigma, \phi\}$, where ν is the mean, Σ is the covariance matrix, and ϕ is the mixture weight. Thus, HMM represents the signal as a sequence of states across time as defined by $\mathcal{T}$. The $\mathcal{T}$'s contain probability values, which are used to transit from one state to another. The ψ parameters are computed using the Baum-Welch (BM) algorithm.¹⁴ Although, the parameters ν , Σ , and ϕ assume flat starts or random values at the beginning of the BM process, HMMs are quite sensitive to flat values because it limit the performance of the model. Here, the k-means clustering coupled with the GMM are used to initialise the emission distribution parameters as viewed in Fig. 2. The trained ψ parameters are matched to the sound to be recognised in the detection step using the Viterbi algorithm (V-alg).¹⁵ The V-alg takes the $\mathbb{P}_s$, $\mathcal{T}$, and $\mathcal{R}$ from the sound to be recognised (note that this $\mathcal{R}$ have been processed with the trained ψ parameters as indicated in Fig. 2), and it outputs all the closely matched hidden states. The HMM approach have been shown to be viable in marine mammal sound detection and classification, recent works include.^{5,16}



Fig. 1. (Color online) Time series and spectrogram representation of an inshore Bryde's pulse calls.

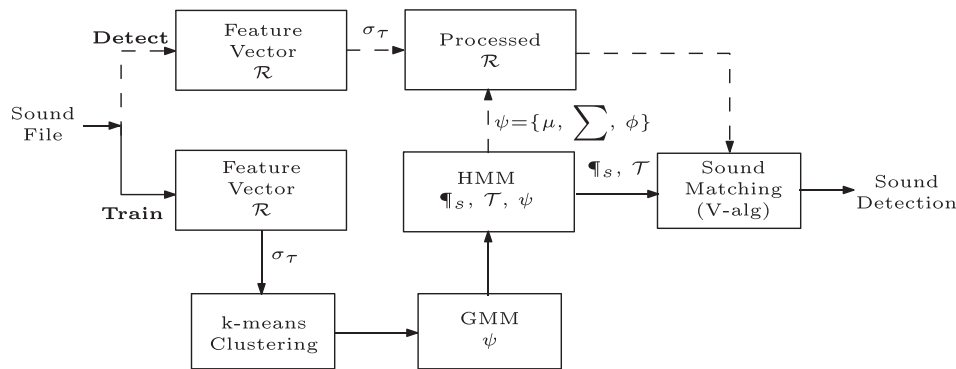


Fig. 2. HMM based detection system.

3. FE methods for HMMs

Executing the HMM with the BM algorithm and the V-alg have been used in several applications as mentioned. Notwithstanding, the $\mathcal{R}$ plays an important role in employing the HMM approach in sound detection and classification. A reliable $\mathcal{R}$ aids the detection capability of the HMM. We emphasise that the more reliable the $\mathcal{R}$, the better is the performance of the HMM both in terms of $\mathcal{S}$ and $\mathcal{F}$. Therefore, this letter focuses on improving the reliability of the feature vector. A couple of FE methods have been used with the HMM but the MFCC and LPC methods are widely used for marine sounds. The MFCC extracts the feature coefficients from the sound signal by converting it from the time domain into the Mel frequency scale. This process can be achieved in seven sequential steps as discussed in Ref. 8: (1) pre-emphasis, (2) framing (3), windowing (4), fast Fourier transform, (5) mel-scale filter bank, (6) logarithm operation, and (7) discrete cosine transform. On the other hand, the LPC calculates the feature coefficients by using a linear combination of the previously regenerated sound signal to predict an approximated value for the current sound signal.⁹ Typically, both the MFCC and LPC use between 10 to 14 coefficients, but we assume 12 coefficients in this letter. That is, the $\mathcal{R}$ is 12-dimensional. Note that the higher the dimension of $\mathcal{R}$, the greater the computational load of the HMM. In this regards, we introduced the EMD and EEMD as a performance efficient alternative to the MFCC and LPC FE methods for *mysticetes* pulse calls such as the Bryde's whale.

4. Empirical mode decomposition

EMD introduced in Ref. 10 has been used by different researchers to solve a number of signal processing problems. Some of the recent signal processing applications of EMD include radar signal emitter recognition,¹¹ speech analysis,¹³ marine mammal sound detection,¹² and so on. EMD is a completely data-driven tool that decomposes a signal $z(t)$ over a random duration of time t into IMFs. The IMFs are the zero-mean functions representing the features of the signal that is decomposed. The IMFs are estimated in an iterative process referred to as sifting, and are arranged from the highest frequency content in Hz to the lowest. The sifting process is used to iteratively remove the large-scale features of the signal to be decomposed until only the fine-scale features of the signal remains. The IMFs must satisfy these two important conditions: (1) the number of extrema of the signal $z(t)$ and the number of zero crossings must correspond or vary by at most one and (2) the mean value of the extrema envelopes must be equal to zero.¹⁰ In summary, the sifting process to decompose a wave signal $z(t)$ is as follows.

- (1) Calculate all the extrema of $z(t)$ over a time t . Thus, the cubic spline function is used to interpolate all the extrema to form the upper and lower envelopes.
- (2) Subtract the mean $\mathcal{U}$ of the extrema envelopes from the original signal $z(t)$ to obtain the first component of $z(t)$. That is, $y(t) = z(t) - \mathcal{U}$.
- (3) The original signal $z(t)$ is replaced with $y(t)$, and further decomposition is performed on $y(t)$ by iteratively repeating steps (1) and (2) while the two IMFs conditions hold.
- (4) The first IMF ($\mathcal{I}_1$) obtained from steps (1)–(3) is subsequently subtracted from the original signal $z(t)$ to obtain a residual. That is, $r = z(t) - \mathcal{I}_1$.
- (5) The residual r is taken has the original signal $z(t)$, and further decomposition is performed as in steps (1)–(3) until the specified number of IMFs are obtained.

Note that the obtained IMFs can be combined to reconstruct the original signal as defined as

$$\bar{z}(t) = \sum_{n=1}^j \mathcal{I}_n + r_j, \quad (1)$$

where $\bar{z}(t)$ is an approximation of $z(t)$ and r_j is the residual signal at the j number of IMFs.

5. Ensemble EMD

The EEMD is an improvement on the mode mixing issue associated with the EMD. An IMF in EEMD is the mean of the IMF obtained over an ensemble of trials, $\mathcal{X}$, using the EMD. Unlike the EMD, additive white Gaussian noise (AWGN) with unique amplitude that is randomly varied for each $\mathcal{X}$ is added to $z(t)$ before decomposition.^{11,17} The EEMD steps is summarised as follows.

- (1) Specify the number of trial, $\mathcal{X}$, and set the trial index, $x = 1$.
- (2) Add AWGN, $N_x(t)$ [the amplitude of $N(t)$ should be unique from the previous amplitudes], to the $z(t)$ to obtained a modified signal. That is, $w_x(t) = z(t) + N_x(t)$.
- (3) Decompose $w_x(t)$ using the EMD steps in Sec. 4 to obtain j IMFs.
- (4) Redo steps (2) and (3) $\mathcal{X}$ times, and increment x , until $x = \mathcal{X} + 1$.
- (5) Compute the mean of the IMFs over the $\mathcal{X}$ trial for each j IMFs to obtain the ensemble IMFs, $e\mathcal{I}$. Note that the $e\mathcal{I}$ can also be combined to reconstructed an approximation of $z(t)$ as in the case of the EMD.

6. Proposed feature vectors, $\mathcal{R}$

In this letter, the IMFs obtained from the EMD ($\mathcal{I}$) and EEMD ($e\mathcal{I}$) processes are modified to compute $\mathcal{R}$. Given a sound waveform $z(t)$, the estimated $\mathcal{I}$ or $e\mathcal{I}$, are usually of this matrix form

$$\mathcal{I}_{i,j} = \begin{bmatrix} \mathcal{I}_{1,1} & \mathcal{I}_{1,2} & \dots & \mathcal{I}_{1,j} \\ \mathcal{I}_{2,1} & \mathcal{I}_{2,2} & \dots & \mathcal{I}_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{I}_{i,1} & \mathcal{I}_{i,2} & \dots & \mathcal{I}_{i,j} \end{bmatrix}, \quad (2)$$

where i is the number of points the signal is sampled at time t and j is the number of specified IMFs. If the $\mathcal{I}_{i,j}$ has τ number of windows, where each window has approximately l number of sampling points. Thus, each number of sampling point l can be represented as

$$\mathcal{I}_{l,j} = \begin{bmatrix} \mathcal{I}_{1,1} & \mathcal{I}_{1,2} & \dots & \mathcal{I}_{1,j} \\ \mathcal{I}_{2,1} & \mathcal{I}_{2,2} & \dots & \mathcal{I}_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{I}_{l,1} & \mathcal{I}_{l,2} & \dots & \mathcal{I}_{l,j} \end{bmatrix}. \quad (3)$$

Hence, we propose that $\mathcal{R}$ can be calculated for each τ with l sampling points as follows.

Step 1: Calculate the mean of each column of the $\mathcal{I}_{l,j}$ to form a $1 \times j$ vector as

$$\mu_{\mathcal{I}_{l,j}} = \left[\left(\frac{1}{l} \sum_{a=1}^l \mathcal{I}_a \right)_1, \left(\frac{1}{l} \sum_{a=1}^l \mathcal{I}_a \right)_2, \dots, \left(\frac{1}{l} \sum_{a=1}^l \mathcal{I}_a \right)_j \right]. \quad (4)$$

Step 2: Subtract the value in each column of $\mu_{\mathcal{I}_{l,j}}$ from each element in the corresponding column of $\mathcal{I}_{l,j}$ to obtain a filtered $\mathcal{I}_{l,j}$ as

$$\mathcal{I}_{f,l,j} = \mathcal{I}_{l,j} - \mu_{\mathcal{I}_{l,j}} = \begin{bmatrix} \mathcal{I}_{f,1,1} & \mathcal{I}_{f,1,2} & \dots & \mathcal{I}_{f,1,j} \\ \mathcal{I}_{f,2,1} & \mathcal{I}_{f,2,2} & \dots & \mathcal{I}_{f,2,j} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{I}_{f,l,1} & \mathcal{I}_{f,l,2} & \dots & \mathcal{I}_{f,l,j} \end{bmatrix}. \quad (5)$$

Step 3: Compute the root mean square (*rms*) of each row of $\mathcal{I}_{f,l,j}$ to obtain a $l \times 1$ vector

as

Table 1. FE-HMM performance comparison: $l = 500$.

s	$\mathcal{S}$ (%)				$\mathcal{F}_p$ (%)			
	EEMD	EMD	MFCC	LPC	EEMD	EMD	MFCC	LPC
18	99.69	99.44	96.60	91.22	0.06	0.08	0.14	1.29
12	98.90	98.77	95.66	87.60	0.06	0.08	0.16	1.44
6	98.08	97.32	94.54	82.71	0.09	0.10	0.23	1.56

$$\mathcal{I}_{rms_{1,l}} = \begin{bmatrix} \sqrt{\frac{1}{j} \left(\sum_{u=1}^j (\mathcal{I}_{f_{1j}})^2 \right)} \\ \sqrt{\frac{1}{j} \left(\sum_{u=1}^j (\mathcal{I}_{f_{2j}})^2 \right)} \\ \vdots \\ \sqrt{\frac{1}{j} \left(\sum_{u=1}^j (\mathcal{I}_{f_{lj}})^2 \right)} \end{bmatrix}. \quad (6)$$

Step 4: Find the upper (**U**) and lower (**L**) envelopes of $\mathcal{I}_{l,j}$ as

$$\mathbf{U} = \mu_{\mathcal{I}_{l,j}} + \mathcal{I}_{rms_{1,l}} \quad \text{and} \quad \mathbf{L} = \mu_{\mathcal{I}_{l,j}} - \mathcal{I}_{rms_{1,l}}. \quad (7)$$

Step 5: Amplify the features while suppressing all unwanted sound sources as

$$\mathcal{K} = \mathbf{U} - \mathbf{L} = \begin{bmatrix} \mathcal{K}_{1,1} & \mathcal{K}_{1,2} & \dots & \mathcal{K}_{1,j} \\ \mathcal{K}_{2,1} & \mathcal{K}_{2,2} & \dots & \mathcal{K}_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{K}_{l,1} & \mathcal{K}_{l,2} & \dots & \mathcal{K}_{l,j} \end{bmatrix}. \quad (8)$$

Step 6: Thus, $\mathcal{R}$ is obtained by finding the mean of each column of $\mathcal{K}$, that is

$$\mathcal{R} = \left[\left(\frac{1}{l} \sum_{a=1}^l \mathcal{K}_a \right)_1, \left(\frac{1}{l} \sum_{a=1}^l \mathcal{K}_a \right)_2, \dots, \left(\frac{1}{l} \sum_{a=1}^l \mathcal{K}_a \right)_j \right] = [\sigma_1, \sigma_2, \dots, \sigma_j]. \quad (9)$$

Therefore, for the whole τ window, the extracted features for the sound signal $z(t)$ can be in the form of a matrix as defined as

$$\mathcal{R}_{\tau,j} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,j} \\ \sigma_{2,1} & \sigma_{2,2} & \dots & \sigma_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\tau,1} & \sigma_{\tau,2} & \dots & \sigma_{\tau,j} \end{bmatrix}. \quad (10)$$

Note that the $\mathcal{R}$ extracted using the EMD and EEMD methods is j -dimensional. Thus, the number of IMFs, j determines the dimension of the $\mathcal{R}$. In this letter, the number of IMFs is fixed to six ($j=6$). This is because we observed that increasing the number of IMFs do not necessary

Table 2. FE-HMM performance comparison: $l = 300$.

s	$\mathcal{S}$ (%)				$\mathcal{F}$ (%)			
	EEMD	EMD	MFCC	LPC	EEMD	EMD	MFCC	LPC
18	99.78	99.66	97.72	92.50	0.05	0.08	0.12	1.24
12	99.10	98.98	96.78	88.00	0.05	0.09	0.14	1.39
6	98.24	97.88	95.66	83.88	0.09	0.11	0.21	1.45

Table 3. FE-HMM performance comparison: $l=100$.

s	$\mathcal{S}$ (%)				$\mathcal{F}$ (%)			
	EEMD	EMD	MFCC	LPC	EEMD	EMD	MFCC	LPC
18	99.88	99.72	98.43	97.33	0.04	0.06	0.11	1.10
12	99.45	99.00	97.32	93.09	0.05	0.07	0.11	1.19
6	98.49	98.22	96.52	92.01	0.08	0.09	0.20	1.39

improve the performance of the detection process. Rather, it increases the computational load of the HMM. As a result, with respect to the computational complexity added to the HMM, the EMD and EEMD feature extraction methods are better in comparison to the MFCC and LPC methods.

7. Sampling training, test results, and discussion

The filter sound dataset were annotated and grouped into two: (1) identified Bryde's whale pulse calls and (2) unwanted signals, which includes noise. This implies that there are two trained HMMs assuming four states and two mixture models. One HMM for the unwanted signals ($\mathbb{P}_1, \mathcal{T}_1, \psi_1$) and the other for the pulse calls ($\mathbb{P}_2, \mathcal{T}_2, \psi_2$). The number of test samples, s is varied while training the two models. The test samples are divided into τ windows with approximately l sampling points. So, we varied l for detailed performance comparison. The separately trained HMMs are combined to form an eight state and four mixture HMM to detect the unknown waveform as either a pulse call or an unwanted signal with the aid of the V-alg.

The results of the HMM detection process are presented in Tables 1–3, in terms of the sensitivity $\mathcal{S}$ and the false positive rate $\mathcal{F}$. The $\mathcal{S}$ is a measure of the detection accuracy of any detector while the $\mathcal{F}$ measures the reliability of the detector. A high value is desired for the $\mathcal{S}$ and the opposite for the $\mathcal{F}$. The tables show the performance of the MFCC, LPC, EMD, and EEMD FE methods with respect to the HMM as a function of l and s . In Table 1, the EMD FE method offered superior $\mathcal{S}$ performance as compared to the MFCC and LPC, respectively. Also, the EMD is the most reliable in comparison to the MFCC and LPC, as it has the lowest $\mathcal{F}$ in Table 1. Similarly, the $\mathcal{S}$ and $\mathcal{F}$ measures of the EMD method outperforms the MFCC and the LPC in Tables 2 and 3. Also, observe that the performance of the all the FE methods improve as s increases and l reduces. Moreover, the EEMD gave a better performance as compared to the EMD method because of the drawback of this EMD method. Yet, both the EMD and EEMD methods can be used as a performance efficient alternative as shown in the results presented. Aside reducing the computational load of the HMM due to their reduced $\mathcal{R}$ dimension, the EMD and EEMD methods offer better $\mathcal{S}$ and $\mathcal{F}$ performances as compared to the MFCC and LPC methods.

8. Conclusion

This letter introduced the EMD and EEMD as a performance efficient alternative FE method for HMM detection of pulse calls. The proposed EMD-HMM and EEMD-HMM approaches were proven to be effective in comparison to the conventional MFCC-HMM and LPC-HMM approaches. The proposed method offered better sensitivity and false positive rate performances, while reducing the computational load of the HMM. Therefore, we suggest that the EMD-HMM and EEMD-HMM approaches can be useful in real-time detection of *mysticetes* pulse calls such as the Bryde's whale.

Acknowledgments

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF (Grant No. 116036).

References and links

- ¹Forgy, E. W. (1965). "Cluster analysis of multivariate data: Efficiency versus interpretability of classification," *Biometrics* **21**(3), 768–769.
- ²Reynolds, D. (2009). *Gaussian Mixture Models* (Springer U.S., Boston, MA), pp. 659–663.
- ³Roch, M., Soldevilla, M., Hoenigman, R., Wiggins, S., and Hildebrand, J. (2008). "Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes," *Can. Acoust.* **36**, 41–47.
- ⁴Rabiner, L. R. (1989). "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE* **77**(2), 257–286.
- ⁵Putland, R., Ranjard, L., Constantine, R., and Radford, C. (2018). "A hidden Markov model approach to indicate Bryde's whale acoustics," *Ecol. Ind.* **84**, 479–487.

- ⁶Yao, R., Johnson, M., Clemins, P., Michael, D., Glaeser, S., Osiejuk, T., and Ebenezer, O.-N. (2009). "A framework for bioacoustic vocalization analysis using hidden Markov models," *Algorithms* **2**, 1410–1428.
- ⁷Jurafsky, D. and James, H. M. (2019). "Speech and language processing," https://web.stanford.edu/~jurafsky/slp3/edbook_oct162019.pdf (Last viewed 15 April, 2019), pp. 1–621.
- ⁸Majeed, S., Husain, H., Abdul Samad, S., and Idbeaa, T. (2015). "Mel frequency cepstral coefficients (MFCC) feature extraction enhancement in the application of speech recognition: A comparison study," *J. Theor. Appl. Inf. Technol.* **79**(1), 38–56.
- ⁹Makhoul, J. (1975). "Linear prediction: A tutorial review," *Proc. IEEE* **63**(4), 561–580.
- ¹⁰Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. London, Ser. A: Math. Phys. Eng. Sci.* **454**(1971), 903–995.
- ¹¹Liu, B., Feng, Y., Yin, Z., and Fan, X. (2019). "Radar signal emitter recognition based on combined ensemble empirical mode decomposition and the generalized S-transform," *Math. Probl. Eng.* **2019**, 1–15.
- ¹²Seger, K., Al-Badrawi, M., Miksis-Olds, J., Kirsch, N., and Lyons, A. (2018). "An empirical mode decomposition-based detection and classification approach for marine mammal vocal signals," *J. Acoust. Soc. Am.* **144**, 3181–3190.
- ¹³Sharma, R., Vignolo, L., Schlotthauer, G., Colominas, M., Rufiner, H. L., and Prasanna, S. (2017). "Empirical mode decomposition for adaptive AM-FM analysis of speech: A review," *Speech Commun.* **88**, 39–64.
- ¹⁴Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.* **41**(1), 164–171.
- ¹⁵Viterbi, A. (1967). "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory* **13**(2), 260–269.
- ¹⁶Brown, J. and Miller, P. J. (2007). "Automatic classification of killer whale vocalizations using dynamic time warping," *J. Acoust. Soc. Am.* **122**, 1201–1207.
- ¹⁷Wu, Z. and Huang, N. (2009). "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.* **1**, 1–41.