Expressibility Coefficient: An Alternative Approach to Compare Complexity of Two Discrete-Time Signals

Anand Muraleedharan Geetha [1], Vishnu Viswanathan [2]

Abstract- *The work attempts to propose an alternative approach to currently existing methodologies in comparing complexity of any two discrete-time signals. The measure Expressibility Coefficient (EC) is introduced to express the relative easiness of representing one signal as a set of polynomials over the other. The results obtained proves the practical applicability of this methodology when used to analyze the dynamics of signals from complexity perspective.*

Keywords- discrete-time series; complexity; polynomial regression;

I. INTRODUCTION

Approaches adopted to data processing aspects of inputs to a system, such as in case of computer integrated manufacturing or onboard intelligence, can critically influence its performance and scalability. A reliable measure to express the complexity of input signals to a system enables us to design systems that are more adaptable to the needs with respect to dynamics of their inputs. Throughout this work the term complexity would mean the complexity of a time series. The potential applications of a having a concrete concept of complexity would be numerous like standardizing definitions of "complexity to test a given system", "Predictability of the inputs to the system" etc.

Though we cannot, at present, point towards an existing ubiquitously agreed notion on complexity, what could be more agreeable is that a constant signal would be the one with least complexity. The approach followed by [1] is to measure the complexity in terms of the effort to compress the given sequence by a compression algorithm. In the specific case of NSRPS discussed in their work. The famous approach of entropy based measurements have been discussed in many works [2] [3] [4] [5].

However, the current work, attempts to provide an alternative way to compare the complexity of two signals of same length. This approach however won't involve the effects due to repeating patterns within the signal and have not been researched against binary sequences.

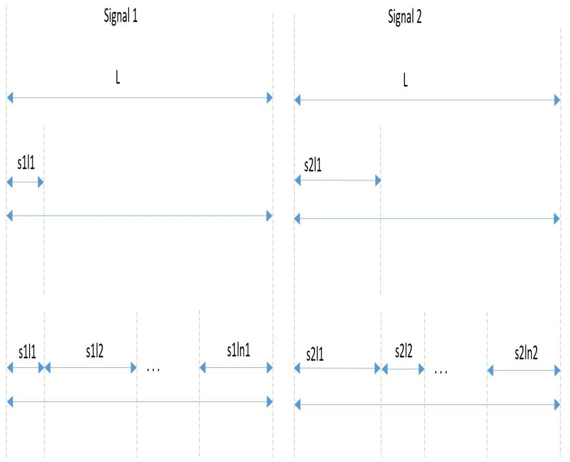
II. PROPOSED ALGORITHM

The approach is to divide these two signals to minimum number of sub-signals such that each sub-signal is the longest time duration for which the amplitudes could be represented as a polynomial of order O without loss of necessary precision. The objective of the algorithm would be to estimate Expressibility Coefficient EC, the relative easiness of fitting these signals to a set of polynomials. EC would essentially be 'the ratio of average length per segment of two signals', such that the signals are divided into least number of sub-signals each represented as a polynomial of order O.

The proposed method to derive EC for a pair of signals is as below

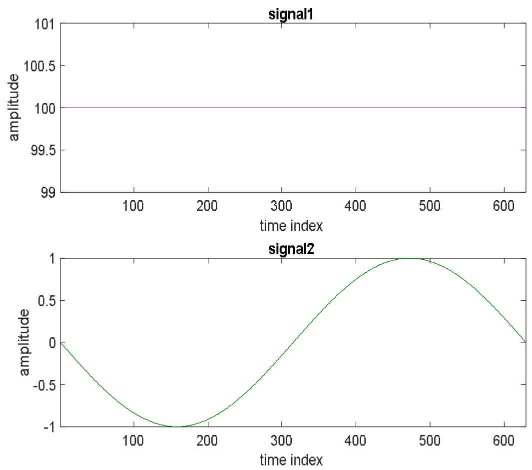
1. Both the signals, S1 and S2, being considered for the comparison must possess same length L
2. Length s1l1, for S1, shall be the longest sub signal starting from the 1^{st} sample which could be represented as a polynomial of order O within acceptable tolerance 'Res'. Polynomial regression is used in the following examples to do the fitting.
3. S1l2 shall be the longest sub-signal starting after s1l1 which could be represented and so on. S1ln1 is the nth sub-signal of S1.
4. Similarly, S2li is the i^{th} sub signal for S2
5. Average length of these sub-signals is an estimate to how complex the representation of these signals could be. If average length of signals is more when compared to other signal, it means it is easier to represent this signal and thus less complex.
6. Expressibility coefficient is given by
$$EC = \frac{n2\ \sum_{i=1}^{n1} length(s1li)^2}{n1\ \sum_{j=1}^{n2} length(s2lj)^2}$$
Where n1 and n2 are respectively the number of sub-signals in S1 and S2
7. EC>1➔ signal 2 is more complex, EC<1 ➔ signal1 is more complex, EC==1➔ same complexity

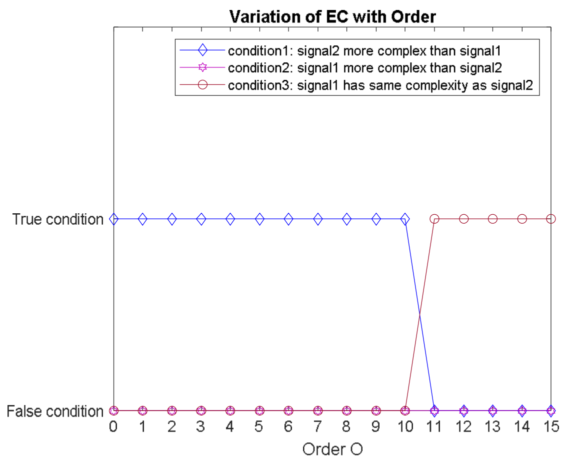
Thus we shall not be interested in the precise numerical value of the EC derived.



III. RESULTS

3.1. constant and sine





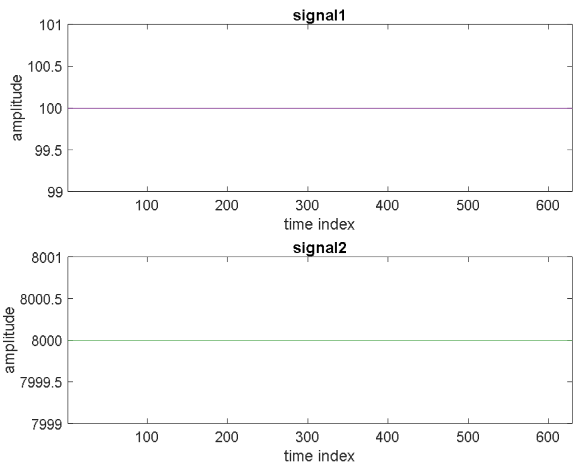
In graphs pertaining to variation of EC with order, ‘True Condition’ represents which among the condition is satisfied according to the derived EC and ‘False Condition’ represents the ones which do not.

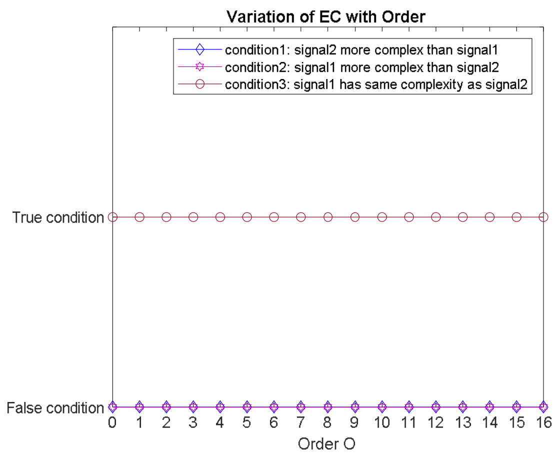
The transition in continuous lines shown in the figure are just for representation and doesn’t imply the particular order at which transition happens.

The result shows that for orders O>=11 with acceptable tolerance ‘Res=10^-6’, the EC comes to 1 and the both sine and constant used have same complexity. For orders O<11, the sine is estimated to be more complex as expected.

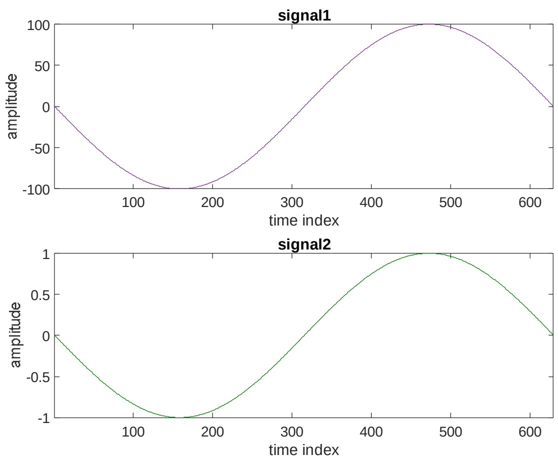
3.2. Effect of amplitude scaling

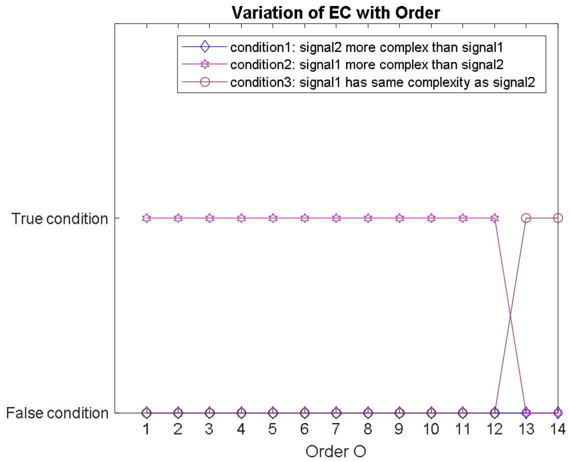
3.2.1. Constant Signals



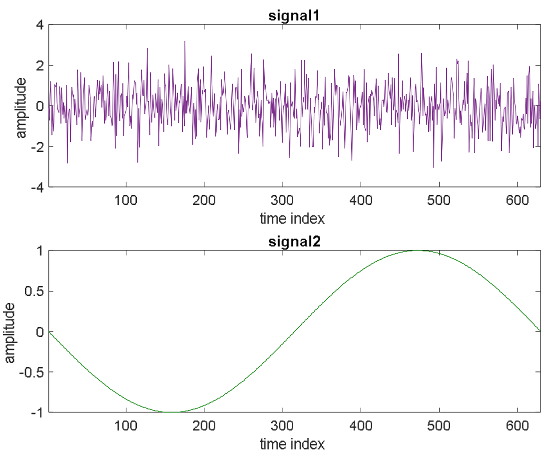


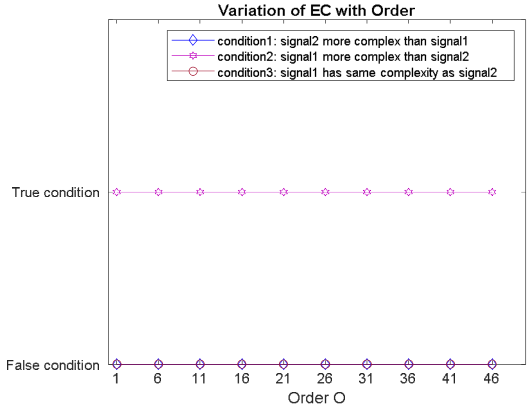
3.2.1. Sine waves





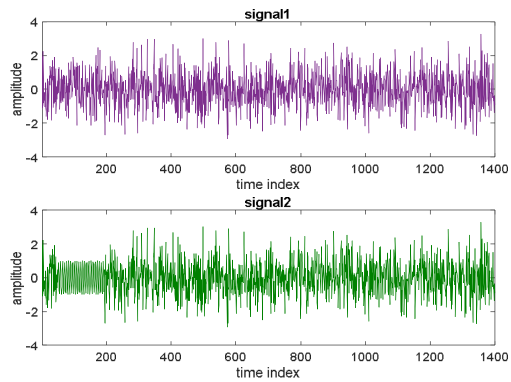
3.3. random signal and sine

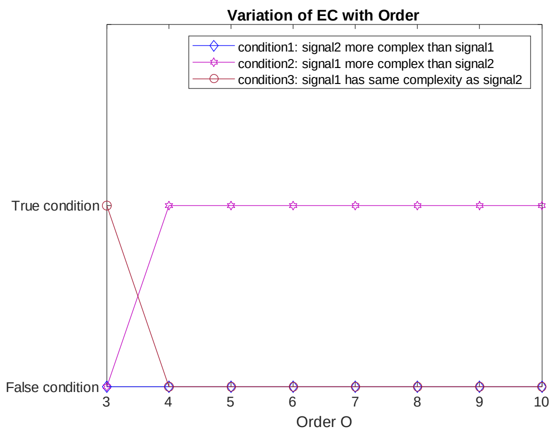




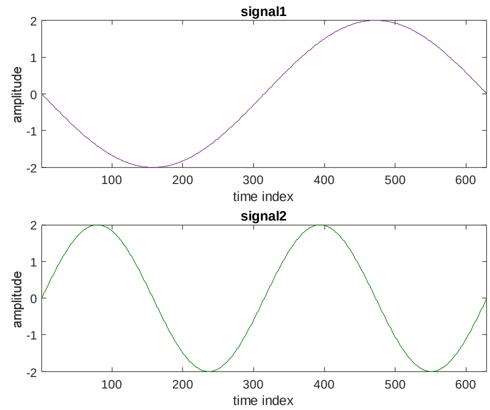
3.4. Tests with random signal

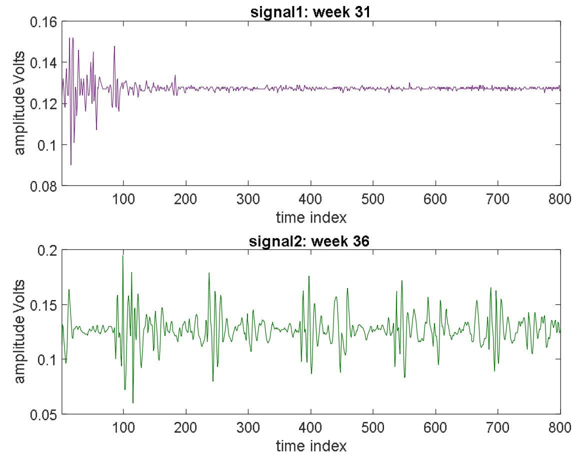
3.4.1. Embedding sine region within random signal

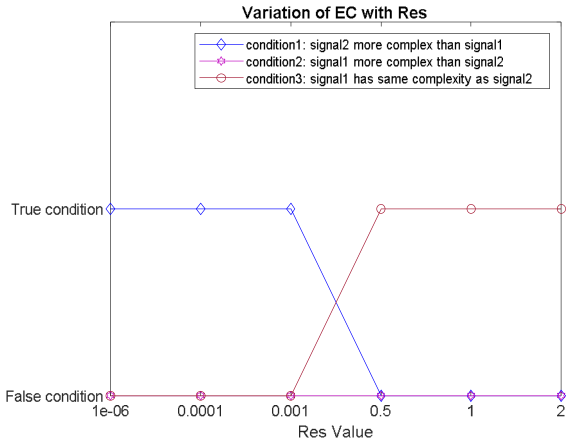


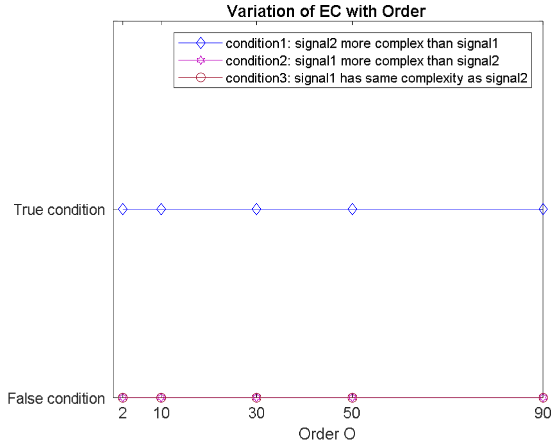


3.5. Effect of Res (Fitting Precision)







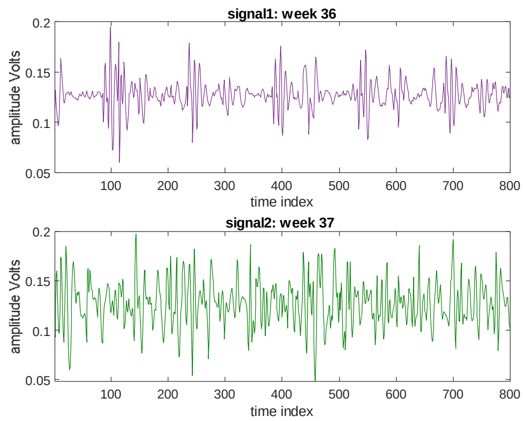


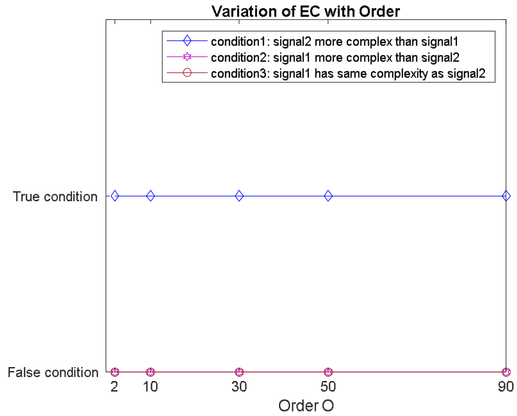
3.6. Real World Signals

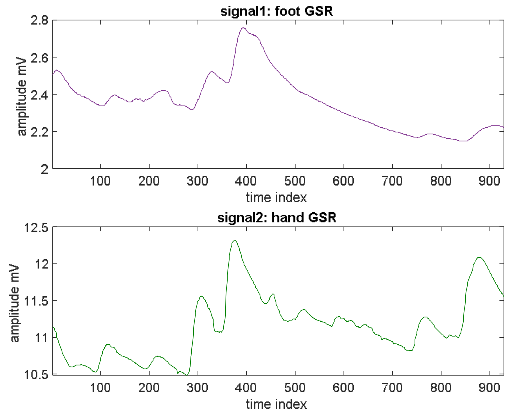
3.6.1. Phonocardiographic (PCG) signals over gestational weeks

The comparison of PCG signal complexity variation over 3 gestational weeks is performed [6][7][8].

With a Res (acceptable tolerance) set to 10^-3, we get the following results. Discussion on medical inference is beyond the scope and is not the objective of the current work.

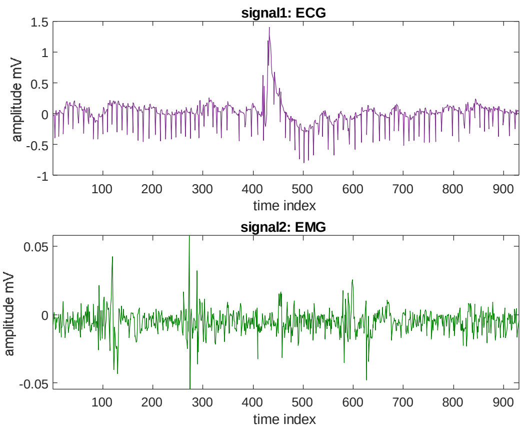


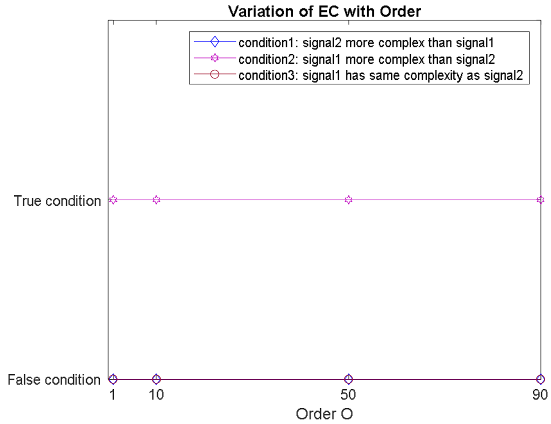


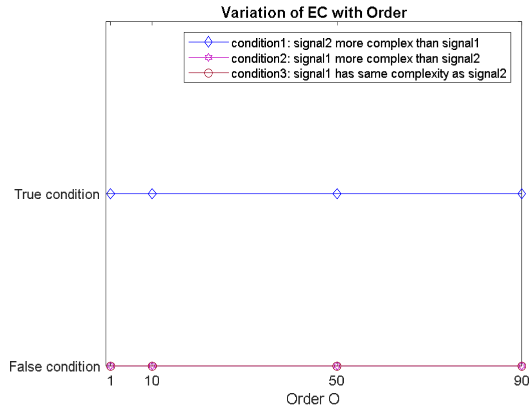


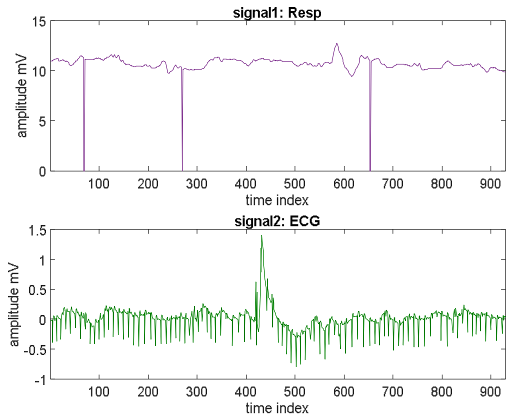
3.6.1. Driver's Data

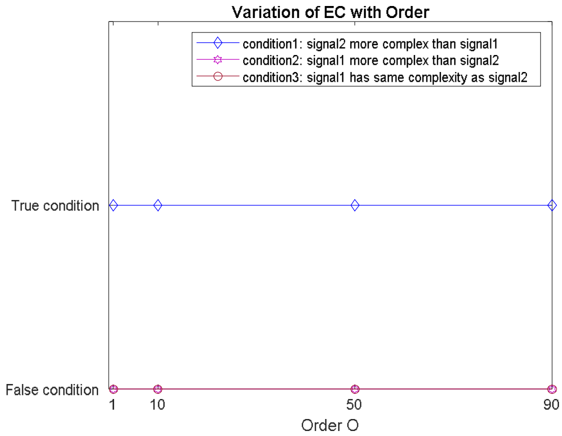
A set of signals fetched from a driver's real time recordings [6][9] are compared in three pairs.











IV. DISCUSSION

This measure is practical enough, as shown from the results obtained, to assist in identification of more complex signal, among any two given signals represented as time series. Relevance of this method in comparing binary sequences is yet to be studied, so are the modifications required in incorporating comparison of multiple signals with different lengths.

REFERENCES

[1] N. Nagaraj, K. Balasubramanian and S. Dey, "A new complexity measure for time series analysis and classification," The European Physical Journal Special Topics, vol. 222, no. 3–4, p. 847–860, 2013.

[2] B. Fadlallah, B. Chen, A. Keil and J. Prıncipe, "Weighted-permutation entropy: A complexity measure for time series," PHYSICAL REVIEW, vol. 87, no. 2, February 2013.

[3] C. Bandt and B. Pompe, "Permutation Entropy: A Natural Complexity Measure for Time Series," Physical Review Letters, vol. 88, no. 17, April 2002.

[4] X. Liu, A. Jiang, N. Xu and J. Xue, "Increment Entropy as a Measure of Complexity for Time Series," Entropy, vol. 18, no. 1, 2016.

[5] R. Zhou, C. Yang, J. Wan, W. Zhang, B. Guan and N. Xiong, "Measuring Complexity and Predictability of Time Series with Flexible Multiscale Entropy for Sensor Networks," Sensors, 2017 April.

[6] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23): e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13).

[7] Cesarelli M, Ruffo M, Romano M, Bifulco P. Simulation of foetal phonocardiographic recordings for testing of FHR extraction algorithms. Comput Methods Programs Biomed 2012 Sep;107(3):513-23.

[8] Ruffo M, Cesarelli M, Romano M, Bifulco P, Fratini A. An algorithm for FHR estimation from foetal phonocardiographic signals. Biomedical Signal Processing and Control 2010 Jan; 5:131-141.

[9] Healey JA, Picard RW. Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions in Intelligent Transportation Systems 6(2):156-166 (June 2005).