Available online at www.sciencedirect.com

ScienceDirect

Fusion Engineering and Design 83 (2008) 467–470

Fusion Engineering and Design

www.elsevier.com/locate/fusengdes

First applications of structural pattern recognition methods to the investigation of specific physical phenomena at JET

G.A. Rattá[a,*], J. Vega[a], A. Pereira[a], A. Portas[a], E. de la Luna[a], S. Dormido-Canto[b], G. Farias[b], R. Dormido[b], J. Sánchez[b], N. Duro[b], H. Vargas[b], M. Santos[c], G. Pajares[c], A. Murari[d],

JET-EFDA Contributors

[a] *Asociación EURATOM/CIEMAT para Fusión, Spain*
[b] *Dpto. Informática y Automática—UNED, 28040 Madrid, Spain*
[c] *Dpto. Arquitectura de Computadores y Automática—UCM, 28040 Madrid, Spain*
[d] *Consorzio RFX—Associazione EURATOM ENEA per la Fusione, Padua, Italy*

Abstract

Structural pattern recognition techniques allow the identification of plasma behaviours. Physical properties are encoded in the morphological structure of signals. Intelligent access methods have been applied to JET databases to retrieve data according to physical criteria. On the one hand, the structural form of signals has been used to develop general purpose data retrieval systems to search for both similar entire waveforms and similar structural shapes inside waveforms. On the other hand, domain dependent knowledge was added to the structural information of signals to create particular data retrieval methods for specific physical phenomena. The inclusion of explicit knowledge assists in data analysis. The latter has been applied in JET to look for first, cut-offs in ECE heterodyne radiometer signals and, second, L-H transitions.

Keywords: Structural pattern recognition methods; Specific physical phenomena identification; JET

1. Introduction

At present, thousands of signals are acquired in a JET discharge with a data storage that can be over 10GB. Up to now, data has been exclusively indexed according to shot number. However, a new data retrieval model has been proposed to search for data [1]. It is based on the fact that similar signals represent similar plasma behaviour. Therefore, structural pattern recognition techniques can be used to query databases. This avoids the manual data inspection to look for particular structural shapes and allows high-level queries based on physical criteria instead of shot number.

General purpose methods to search for similar structural shapes have been applied to JET databases. They are founded on structural pattern recognition techniques. Feature extraction is accomplished taking into account the morphological structure of waveforms. Two different techniques allow data retrieval in an intelligent way. First, a search method to retrieve entire waveforms (Section 2). Second, a search for structural forms within signals (Section 3). These techniques have also been applied to study particular physical phenomena as described in Section 4. Specific system knowledge is included in the feature extraction process to optimize searching criteria for the physical properties of the phenomenon. The application of the techniques to ECE signals and L-H transitions must be understood as a proof of principle of the method and not as detailed analyses on both processes.

2. Entire waveform search

This approach allows the search of entire waveforms similar to a given one. The application to JET databases has

---

[*] Corresponding author.
E-mail address: giuseppe.ratta@ciemat.es (G.A. Rattá).

0920-3796/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.fusengdes.2007.09.008

been carried out on temporal evolution signals covering the whole discharge, *i.e.* from plasma start to plasma extinction.

Essential elements to develop the searching system are [1]: feature vectors, a classification system and a similarity measure. Feature extraction is performed by means of the "Haar" wavelet transform. A supervised clustering system based on shot length groups the data [2]. This classification avoids traversing the whole database looking for similar signals. Instead, the similarity computation is only carried out within the waveforms of one cluster, the more likely one to contain similar signals. The similarity measure is computed according to the normalized inner product of the feature vectors [2].

A software application implementing this data retrieval technique is available in a concurrent way on the JET analysis cluster (JAC) computer environment.

3. Pattern search within signals

This approach allows the search of structural shapes (patterns) inside time-series data. Patterns are composed of simpler sub-patterns. The most elementary ones are known as primitives. Primitives are represented by characters, converting the pattern recognition problem into a string matching problem.

Feature extraction is carried out by dividing the initial waveform into segments and each segment is fitted with a straight line through a least squares minimization process [3]. The segments are encoded according to the slopes of the straight lines. Only a reduced set of five slopes (primitives) is enough to encode the waveforms. The search of patterns is accomplished by means of a relational database which is an optimal system to handle strings of characters.

Two different techniques have been developed to define segments [6]. The first one assumes segments with the same number of samples: equal length segments (ELS). The second technique uses segments with a variable number of samples: variable length segments (VLS).

This searching method has been put into operation on the JET JAC cluster and can be executed in a concurrent way by simultaneous users.

4. Searching for specific physical phenomena

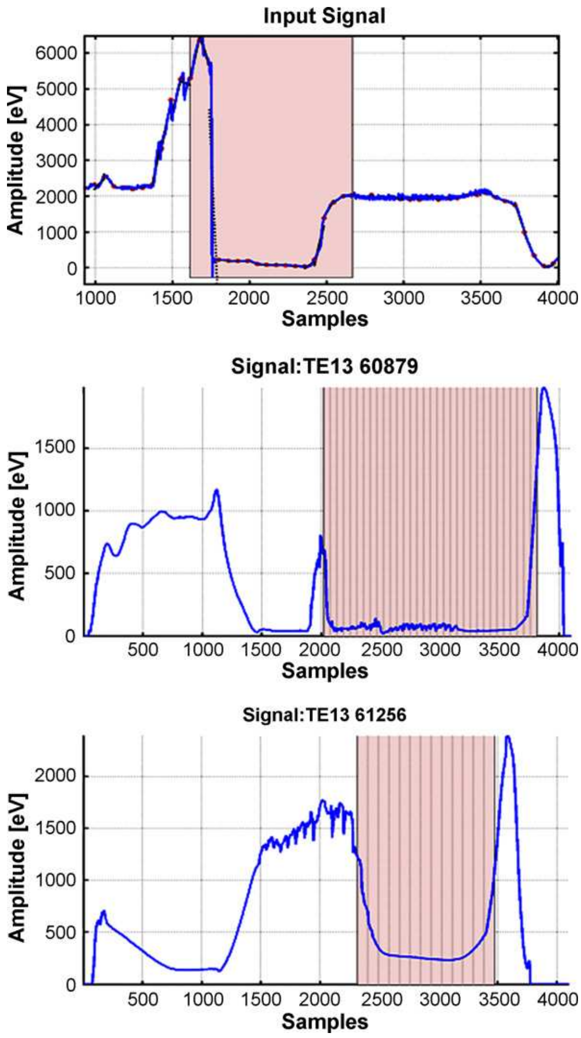
The recognition of structural shapes plays a central role in distinguishing particular physical properties. Sometimes just one structural form (a bump, an abrupt peak or a sinusoidal component), is enough to identify a specific phenomenon. In other occasions it is necessary to check the coincidence of multiple events in more than one waveform, to detect plasma behaviours.

There is not a general rule to describe the structure – or structure combinations – of various phenomena, so specific knowledge about their characteristics has to be taken into account. In other words, signal structural shape may be not enough for a complete description of physical properties. Therefore, domain knowledge has to be added to the structural information provided by ELS or VLS techniques.

In this first approach, we have applied the mentioned techniques to identify ECE cut-offs in temperature signals. It is as an example of physical phenomena recognition by just a single pattern in temperature waveforms. As an application of multiple pattern recognition we applied the searching method to identify regime transitions, combining the analysis of density and Dα signals.

4.1. *ECE cut-offs*

A temperature diagnostic in JET is the ECE heterodyne radiometer [4]. One of the limitations of any ECE system in high-density plasmas is the appearance of cut-off (internal reflection of the ECE radiation). When the density between the ECE antenna (located in the low field side of the plasma) and the emitting region rises above the cut-off density, the radiation is unable to propagate out to the detecting system. When that happens, a sharp reduction of the signal in the central channels is observed, keeping its temporal response low and steady. At a later time in the discharge, when the density diminishes below the cut-off value, the signal reaches the proper level with a sudden upward


Input Signal

Signal:TE13 60879

Signal:TE13 61256

Fig. 1. Input waveform (above), match (middle) and mismatch (below).

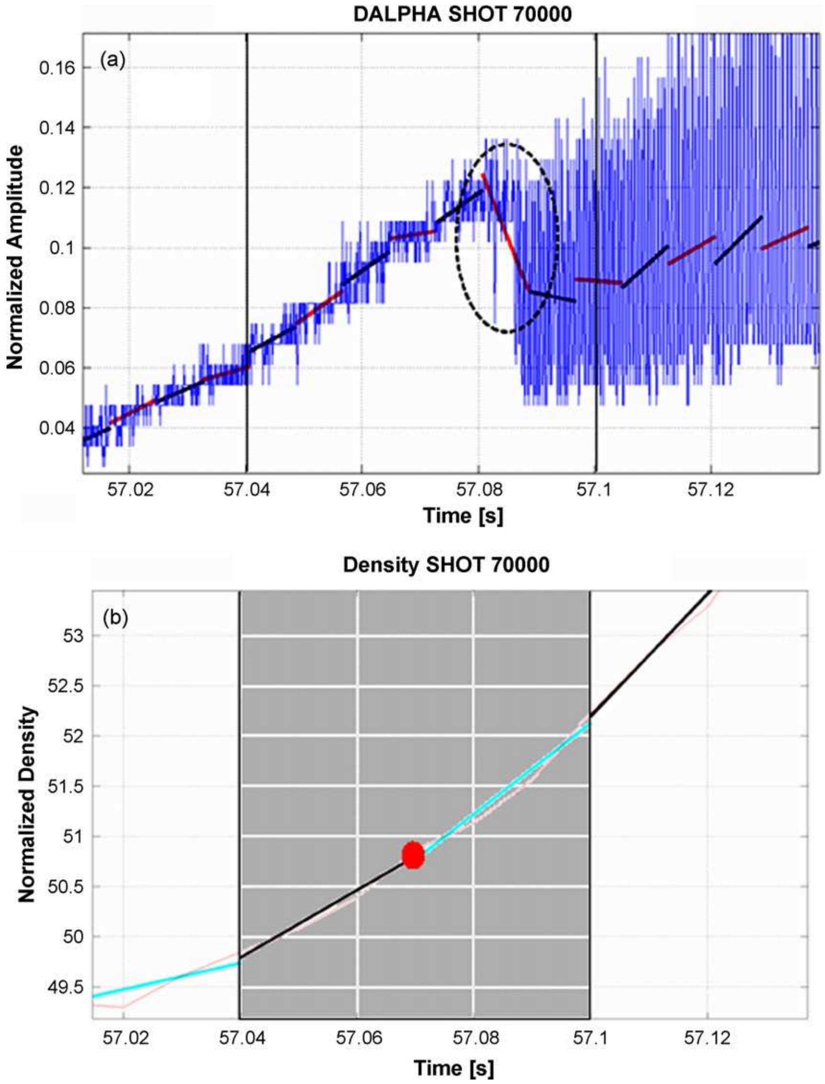


Fig. 2. L-H transition recognition. (a) In a Dα signal and (b) density signal.

transition. The time interval of the cut-offs varies differently in each shot, so the pattern to search must be “flexible” in this respect.

A large set of JET signals (from shot 60,628 to 68,749) was used to create the test database. After the ELS computation of the primitives, the searching procedure was slightly changed: the database must return all the strings that show a quick drop, subsequently a “flexible” time of flat slope (channel in cut-off) and finally a high positive slope. As outputs the starting and ending times of the cut-off and their corresponding densities are estimated for each shot.

About 92% of the signals presenting this behaviour were detected. However such pattern does not correspond univocally to ECE cut-offs because it could be produced by abrupt changes in the plasma temperature. So, some of those similar shapes were also recognized as due to the onset of cut-offs and that is the reason of the mismatch (bottom image of Fig. 1). Also, it is possible to observe in Fig. 1 that the cut-off time length in the input signal is different to the retrieved ones.

4.2. Regime transitions

Time instants for L-H transitions can be identified in an automated way [5]. However, structural pattern recognition techniques can also be used (as a first approach) easily.

The L-H transition is determined by a rapid drop in the Dα signal. However, this behaviour is difficult to identify because its amplitude changes considerably signal by signal. Also, the waveform is corrupted with noise and it presents sudden amplitude variations; so, if it is independently analyzed, there remains the risk of missing the real transition time. Consequently, the analysis of these signals needs extra accuracy, achieved using the VLS method. An example of the codification is depicted in Fig. 2a. Notice that when the transition takes place approximately at 57.08 s, there is a high negative slope and the length of the primitive is bigger than the average. The searching criterion, defined with this additional knowledge, was the following: the first of the primitives with a high negative slope and with a length at least two times bigger to the average would mark the estimated transition instant.

As it was mentioned before, it is necessary to look simultaneously for other patterns in order to minimise the errors.

A sudden increase in the plasma density could evidence a confinement improvement, so we studied this waveform to confirm the regime change. However, a density augment can be also a consequence of other factors (neutral beam injections, gas injection or particle recycling). Thus, the precise selection of the encoding angles was essential to identify only the required phenomena. Once again, to improve the accuracy and "tune" the search, the VLS computation of primitives was selected. The transition time was determined by the first primitive with a high positive gradient, and the error bars were set to the temporal length of its adjacent primitives (Fig. 2b).

In summary, the recognition of the L-H transition is accomplished by the combination of two patterns: a drop in the Dα and a simultaneous increment of the density.

The technique was applied to 50 shots picked from JET database. When the identified instant in the Dα signal (57.08 s in the example) was inside the error bars (grey highlighted in Fig. 2b) an L-H transition was recognized, setting this Dα instant as the estimated transition time. That happens in the 76% of the cases.

Comparing them to the real L-H mode times, the average error was 34 ms, with a maximum error of 89 ms and a standard deviation of 19.2 ms.

5. Discussion

In the present article the results of structural pattern recognition methods applied to JET signals are shown. General techniques (entire waveforms identification and structural recognition within signals) are already available on the JAC Linux Clusters of JET.

To recognize particular physical phenomena, the specific knowledge about the behaviour of the waveforms involved in each case was exploited, guiding the searching methodologies.

These *ad hoc* modified methods were applied, as a proof of principle, to analyze two specific phenomena in JET plasmas. Firstly, we tuned the technique to identify ECE cut-offs. Secondly, we determined the L-H transition time by means of multiple patterns searching, combining the structural shapes of density and Dα waveforms.

References

[1] J. Vega and JET EFDA Contributors, Intelligent methods for data retrieval in fusion databases, these proceedings, in press.

[2] J. Vega, A. Pereira, A. Portas, S. Dormido, G. Farias, R. Dormido, et al., Data mining technique for fast retrieval of similar waveforms in Fusion massive databases, Fusion Eng. Des, submitted for publication.

[3] S. Dormido-Canto, G. Farias, J. Vega, R. Dormido, J. Sánchez, M. Santos, et al., Search and retrieval of plasma waveforms: structural pattern recognition approach, Rev. Sci. Instrum. 77 (2006) 10F514.

[4] E. de la Luna, G. Conway, J. Fessey, R. Prentice, D.V. Bartlett, J.M. Chareau, et al., Electron cyclotron emission radiometer upgrade on the JET tokamak, Rev. Sci. Instrum. 75 (10) (2004) 3831–3833.

[5] A. Murari, G. Vagliasindi, M.K. Zedda, R. Felton, C. Sammon, L. Fortuna, et al., Fuzzy logic and SVM approaches to regime identification in JET, IEEE Trans. Plasma Sci. 34 (3) (2006).

[6] S. Dormido-Canto, G. Farias, R. Dormido, J. Sánchez, N. Duro, H. Vargas, et al., Comparison of structural pattern recognition methods for fusion databases, these proceedings, submitted for publication.