# A Demo for efficient human Attention Detection based on Semantics and Complex Event Processing

Yongchun Xu[1], Ljiljana Stojanovic[1], Nenad Stojanovic[1], Tobias Schuchert[2]

[1] FZI Research Center for Information Technology
Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, Germany
{name.family name}@fzi.de
[2] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB)
FraunhoferStr. 1, 76131 Karlsruhe, Germany
{name.family name}@iosb.fraunhofer.de

**Abstract.** In this paper we present a demo for efficient detecting of visitor's attention in museum environment based on the application of intelligent complex event processing and semantic technologies. Semantics is used for the correlation of sensors' data via modeling the interesting situation and the background knowledge used for annotation. Intelligent complex event processing enables the efficient real-time processing of sensor data and its logic-based nature supports a declarative definition of attention situations.

**Keywords:** Sensor, Human attention, Complex Event Processing, Ontologies

## 1    Introduction

In this paper we describe a demo, which shows a semantic-based system providing personalized and adaptive experience for the visitor, in which the digital contents react depending on the artwork and the user's engagement/attention state. In the demo we use semantic technologies for the correlation of sensors' data via modeling the so-called interesting situation and use complex-event processing to recognize the attention patterns in the event stream.

## 2    Problem overview

In order to enable an adaptive experience for the visitor to a museum, the demo is constructed around a four-phase OODA (Observe, Orient, Decide, Act) as shown on **Fig. 1.** In the Observe phase, our approach is concerned with the measurement of covert cues that may indicate the level of interest of the user. In order to consider how a user perceives an artwork, different sensors have been considered: The monitoring of visual behavior will allow the system to identify the focus of attention. The acoustic module should provide important information about environmental influences on patterns of visual attention or psychophysiology. Finally, a video-based hand gesture recognition provides an additional input modality for explicit interaction with the system (e.g., for selecting certain visual items, navigating through menus).
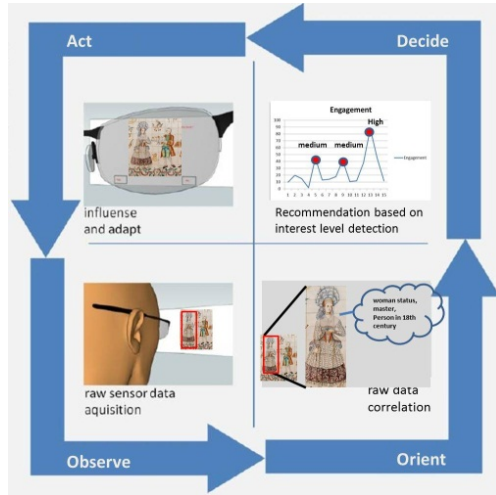
**Fig. 1.** OODA cycle

All data streams are collected and analyzed in real-time in order to yield a dynamic representation of the user attention state (phase Orient). In the Decide phase, covert physiological cues are used to to measure the level of interest or engagement with artwork or with augmented content presented via the AR device. Based on the interpretation of this complex state, the provision of augmented content from a repertoire of available content is made. The presentation of selected content via the AR device (e.g. visual, audio) is subsequently executed during the final Act stage.

## 3 Demo challenge: Semantic-based attention detection

The challenge of this demo is how to detect the attention of the visitor in the museum. In most situations the attention of the visitors can be determined according to the gaze behavior of the visitors. In some cases the observed object is the attention object of the visitor, while in other cases the visitors pay attention to the information behind the observed objects. Thus, we distinguish between visual attention and content-related attention. **Fig. 2** summaries categories of attentions that is relevant in the museum context, including visual attention and content-based attention.
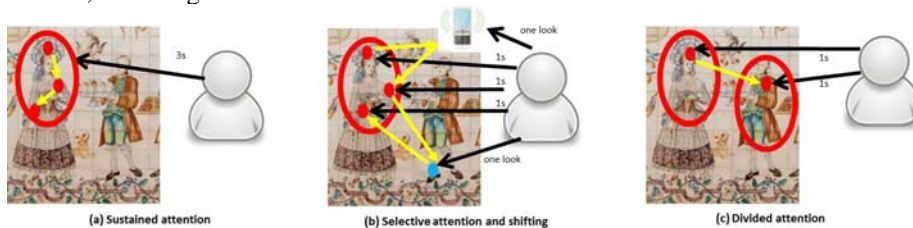


**Fig. 2.** Different attention categories relevant for museums: (a) Sustained attention; (b) Selective attention and shifting; and (c) Divided attention

Sustainable attention (**Fig.2** (a)) means that the attention is focused over extended periods of times. Similarly, if during at least 3 seconds the acoustic noise level is large (e.g. a mobile phone of a visitor is ringing), then selective attention and shifting is detected (see **Fig.2** (b)). Finally, divided attention (see **Fig.2** (c)) means sharing of attention by focusing on more than one relevant object at one time. One possible way of "calculating" similarity is to consider semantics of the topics behind the artworks.

The presented visitor attention model puts some requirements on what has to be modeled: different types of sensors data and their fusion in order to detect visual attention and semantic about artwork in order to profit from content-based attention.

## 4    The role of semantic processing

The demo is based on knowledge-rich, context-aware, real-time artwork interpretation aimed at providing visitors with a more engaging and more personalized experience. Indeed, we propose to combine annotation of artworks with the time-related aspects as key features to be taken into account when dealing with interpretation of artworks. Thus, the aspects of the museums modeled by ontologies are classified into:

- Static aspects which are related to the structuring of the domain of interest, i.e. describing organization of an artwork and assigning the metadata to it;
- Dynamic aspects which are related to how a visitor's interpretation the elements of the domain of interest (i.e. artworks) evolve over time.

## 5    Demo setting

The demo will be performed using following hardware equipment (**Fig. 3**):

- A Poster of Valencia Kitchen in MNAD (Museo Nacional de Artes Decorativas, Madrid Spain) as artwork
- Vuzix Star 1200 AR glasses with camera
- M-Audio Fast Track Pro audio card and BEYERDYNAMIC MCE 60.18 mic
- Bio sensors



**Fig. 3.** The demo setting and equipment

## 6    Demo workflow

**Fig. 4** shows the concrete workflow of the demo. The whole workflow is based on the OODA model.
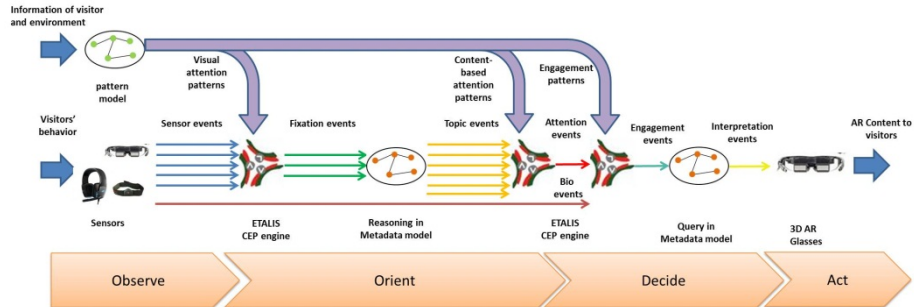
**Fig. 4.** Demo workflow

**Observe phase**: After the visitor got the equipment (e.g. sensors and AR glasses), the patterns have been deployed according to the visitor's information and actual museum environment. When the visitor stands before Valencia Kitchen and starts looking at the kitchen wall, the visual sensor detects the gaze of the visitor. Meanwhile the acoustic sensor monitors the environment sound for possible disturbance and the bio sensors monitors the physiological signals of the visitor. Assuming that the visitor has interest on the "person in 18th century", in the situation of divided visual attention the visitor looks at the head of the Lady for one second firstly, after that the visitor looks somewhere else and then the visitor looks at the servant for another one second. A sequence of gaze events and some acoustic events and bio events is published.

**Orient phase**: The CEP engine ETALIS receives the sensor events and use the visual attention pattern. Two short fixations (the first one on the lady and the second one on the servant) are detected and published as events.

The knowledge base receives the two short fixation events and uses SPARQL to find all the related topics of the observed object in the metadata According to the annotation of the Valencia Kitchen (**Fig. 5**) for the fixation on lady three direct topics: "woman status", "master", "person in 18th century" and two indirect topics : "social stratification", "human status" are found. For the short fixation on servant we got direct topics: "man status", "party servant", "person in 18th century" and indirect topics: "human status", "servant", "social stratification". All these topics are published as topic events. ETALIS detects the attention event according to these topic events.
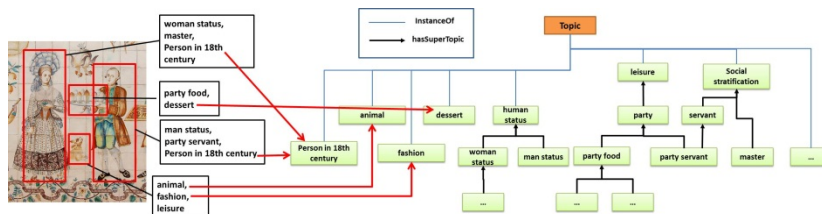


**Fig. 5.** Annotation of the Valencia Kitchen

**Decide phase**: The CEP engine detects the interest and engagement of visitors based on the attention events and the bio signal events. If the bio signal shows the visitor the interesting level is high and meanwhile an attention is detected, we can conclude that the visitor has interest on such topic. In our example the visitor has interest on the topics: "person in 18th century", "human status" and "social stratification". This discovery of engagement is sent out as event by CEP engine.

The knowledge base receives this engagement event and finds the related metadata (guide content) about the topic through reasoning and publishes the metadata as Interpretation event.

**Act phase**: Lastly in the act phase the AR glasses get the interpretation event and show the metadata (guide content) as Augmented Reality content on the glasses to the visitor.

## 7      Demo Implementation

**Fig. 6**. shows the architecture of our system. The following sensors are used: see-through glasses with integrated camera that can track the gaze of visitors and display the augmented reality (AR) content to visitors; acoustic sensor senses the acoustic information surrounding visitors such as environment noise or the content that visitors are listening to, and bio sensor observes the biological signals of visitors like heart rate.
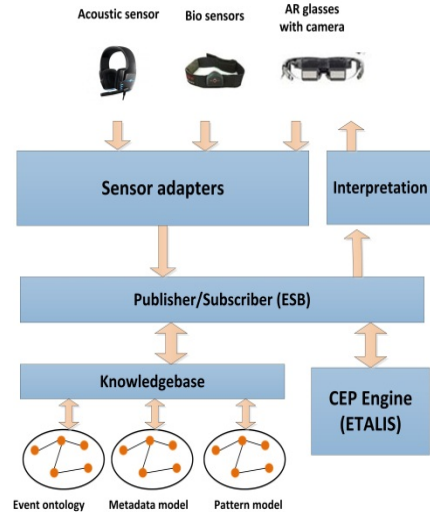


**Fig. 6.** System architecture

All components communicate through ActiveMQ ESB by publishing and/or subscribing to events. The sensor adapters connect to the sensor hardware, collect the physical signals of visitors such as gaze, sound, heart rate and other bio signals from sensors and translate them into meaningful sensor events to be processed by the CEP engine. The complex event processing part detects the situations of interests based on predefined patterns and real-time sensor data. Semantic technologies are used to store the annotation of artworks, semantically-enriched sensor data and patterns. The knowledge base manages the background knowledge and provides the query function to other parts. The interpretation part recommends AR content to visitors based on their engagement and query results