# A Web-of-Things solution to enrich TV viewing experience using Wearable and Ambient sensor data

Prabhakar Srinivasan, Cisco SPVTG
Email: prabhasr@cisco.com

## Introduction

A combination of specialized sensors embedded in the form of 'smart' wearable or ambient devices and advanced pattern-recognition and predictive data-science models come together to form the ingredients of many Web-of-Things (WoT) applications. So too does this solution from Cisco's SPVSS group. Wearable technology is increasingly appearing in convenient form-factor and becoming more pervasive. Companies are packing more sensors into the wearable devices that perform highly specialized functions. These devices communicate with each other using Radio Frequency (RF) protocols like Bluetooth, Wi-Fi or Zigbee. We at Cisco, are evolving a unique WoT proof-of-concept that uses data from these wearable and ambient sensors to predict TV viewing behaviour and to enrich it. Modelling of TV viewing behaviour using biometric and ambient devices, cross-domain recommendations, user-content affinity mapping, portable user profile and comprehension of audience retention as a cognitive process are some interesting use-cases that this solution addresses.

Data from wearable devices that measures biometric attributes like heart-rate, blood pressure, skin galvanic response, electro-encephalograph (EEG), posture etc., has the inherent quality of being very close to defining the truth about the state of mind of a person. This data also represents a near-instantaneous reaction to events in the environment of a person. Use of biometric data to inform on user behaviour has been proven to be successful in the past but using it to inform on TV viewing behaviour is a novel research angle with many potential applications. The ability to define a viewer's reaction to content (Linear, Video-on-Demand content and Advertisements) viewed on TV using an unprecedented vocabulary of biometric metadata enables Cisco to build very interesting applications for video. This concept has the potential to be generic enough to not be limited to TV viewing use-cases and can be extrapolated to scenarios where users interact with a general class of consumer electronic devices. This research gives an insight into what goes on in the mind of a person when they actively seek certain types of content or they actively avoid some types of content. We endeavour to answer the eternal question of what goes on in the mind of a viewer a few seconds prior to them changing the channel or switching off their TVs. Knowing this information empowers us to work towards retaining the audience and enriching their TV viewing experience.

Fig 1.1 Biometric Sensors

Also, as a human-computer interaction story, some of the sensors act as a Brain Computer Interface (BCI) and provide the viewer with the capability to drive the TV-viewing experience just with thought-waves. This is a leap from traditional HCI techniques and this also doubles as an accessibility story. These and more such use-cases are within the scope of this current paper. This paper addresses the communication protocols between the sensors, the system design, Cloud-centric Big Data architecture and most importantly the predictive models that solve these specialized use-cases.

The following use-cases are within scope of this research:

1. Build a database of biometric metadata corresponding to content. E.g., Fear Factor rating of 4/5 for a horror movie based on how elevated viewers' average heart rate is
2. Understand mood-content correlation for each user and across demography segments
3. Acquire biometric frame-level correlation data with content. E.g., If someone feels excited when a celebrity appears on a talk show 10 minutes into the show we can detect that and correlate the mood with the event
4. Establish content similarity and segments based on biometric responses
5. As an affective computing use-case, infer the mood of the viewer and based on the mood recommend content
6. Understand the biometric markers that are precursor to the viewer ending a TV viewing session
7. Classification and tagging of content based on stricter parental control to prevent children from harmful content that evokes negative reactions like Fear, Anger etc.,
8. Facilitate a virtual profile that travels with the viewer wherever they go. E.g., in-flight entertainment recommender uses the virtual profile of user
9. Create a multi-agent system of autonomous collaborative intelligent agents that provide cross-domain recommendations. E.g., On entering a retail store like Walmart, a shopping software agent recommends an online deal for a light- sabre toy when someone's virtual profile knows that they enjoyed watching a Star Wars movie.
10. Power-saving initiative: Detect lack of attention or sleep cycle using posture and head position and heart rate and tend customer devices to idle state including dimming lights
11. As an Accessibility and HCI use-case provide an alternate scheme for navigating content using wearable devices like neuro-headsets. E.g., A paraplegic patient in a wheel-chair can activate the TV menu using just eye blinks or facial muscles or thought-waves. A smile could switch on the TV
12. The correlation between attention-level and advertisement content is evaluated. E.g., high attention-level on an Audio car advertisement would trigger an agent to respond with a recommendation for test ride at a nearby Audi dealership
13. The biometric signals leading up a few seconds prior to a channel change are recorded and the moods like boredom, frustration, lack of attention etc., are attributed to the events. Key influencers for triggering a user to end a session are detected. Onset of similar biometric signals is predicted.

## Metadata

Metadata about video content has intrinsic qualities like the video quality, duration, cast etc., and ascribed qualities like genre, ratings etc., Most often the ascribed or inferred metadata attributes like genre have the limitations that it is very subjective and can be noisy and does not have a universal quality that crosses regional tendencies.

Biometric metadata on the contrary have the quality of being universal and easily interpreted across various populations. This data is also more representative of the user's perception and reaction to an experience than an ascribed attribute like Rating. So the premise is that applications like recommendation engines built using biometric metadata would be more accurate in modeling user-content affinity than those built using ascribed metadata.

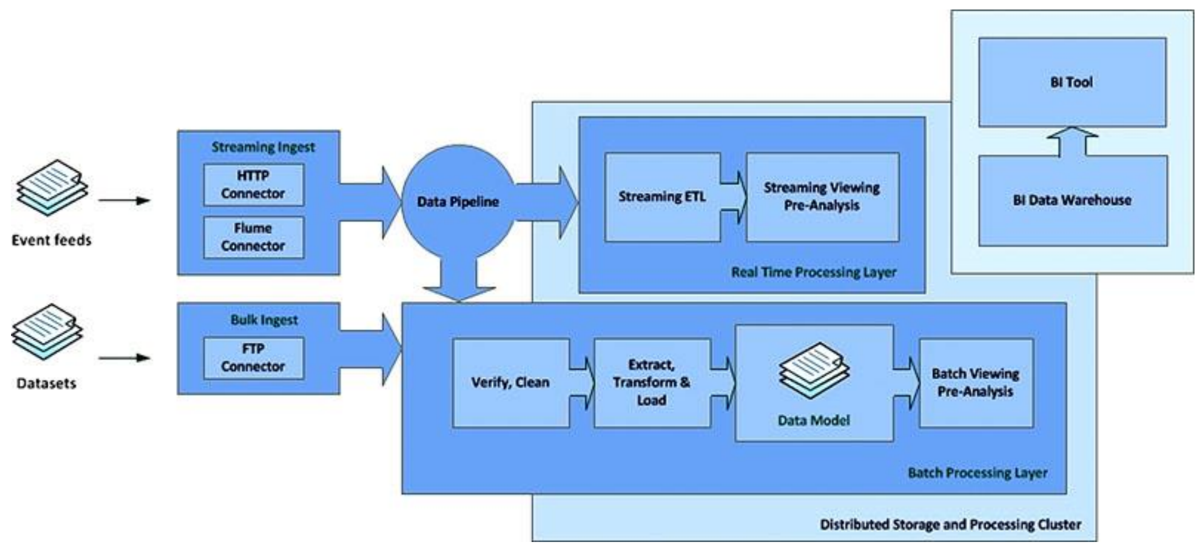
The various commodity biometric sensors embedded in wearable technology could report on many metrics like Head Position, Facial Expressions, Posture, Airflow Sensor, Electrocardiogram, Electromyography, Galvanic Skin Response, Body Temperature, Pulse Oxygen in Blood, Blood Pressure and EEG.

## Experiment

This experiment is a proof-of-concept. We create a TV-viewing room environment where all the commodity hardware with biometric and ambient sensors is fitted to a participating control group viewer. An inventory of content from diverse genres and categories is pre-processed and kept ready for viewing. Each control group viewer's identity is anonymous, except for their gender and age and their biometric responses to the content at a frame-level are recorded.

The test participant is free to choose from the catalog of content and is free to initiate terminate or change the session to view other content at will. The session event data along with the time-stamp and content identifiers are recorded. Intra-content metadata like audio characteristics and video features are recorded. From the participant who is fitted with the array of biometric sensors the biometric metadata is also collected and correlated with the session event identifiers. All the data which comes as streaming ingest of event feeds is provisioned on a big-data infrastructure for later processing

Fig 1.2 Acquiring and provisioning big data

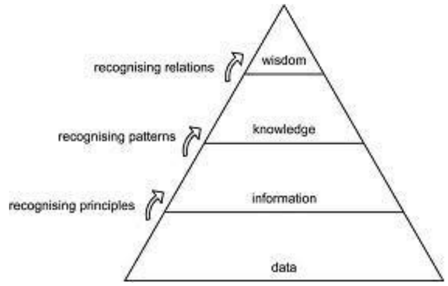



## Knowledge Disovery

A recurring pattern in WoT solutions is the knowledge pyramid. The raw sensor data is aggregated at the bottom of the pyramid. Interesting patterns are detected and relationships between patterns are inferred at the tip of pyramid which constitutes the wisdom. This process is known as Knowledge Discovery.

There are many probabilistic and machine-learning models that helps us in realizing the use-cases. Some of these are supervised learning algorithms like Artificial Neural Networks (ANN) and Support Vector Machines (SVMs) which classify entities like user or content into categories. Some of the models are unsupervised learning algorithms like Clustering and Self-Organizing Maps (SOMs) which segment the viewers and content based on their biometric responses to content.

Fig 1.3 Knowledge Pyramid




Both the machine-learning approaches feed into a larger context of a multi-agent framework and provide the much-needed learning component to make the agents 'intelligent'. These multi-agents are autonomous and collaborate with each other sharing their world-view of knowledge allowing for cross-domain recommendations. The user-content affinity mapping is maintained as a hypergraph ontology structure. This forms a semantic web of context to provide the 'world-view' to the agents. Gaming-inspired algorithms like Behavior Trees are also used within the agent framework. The Agents could also provide a high degree of personalization using Bayesian Belief Networks. The agents also use a spatial and temporal context in their world-view

Global trends are analyzed and used to build models that inform on local events. Collaborative filtering allows for interesting recommendations. Eg., "*Others who watched this Horror movie also watched the following movies which elevated their average heart rate in similar manner…*". Particle swarm optimization algorithms are also considered to inform on this global to local phenomenon.

## Privacy & Security

Security is a fundamental requirement in the WoT ecosystem. Both the raw-data and the patterns recognized from the data need to be secured.

This proof-of-concept uses commodity hardware assembled to form the data acquisition layer. A bespoke security model is proposed as part of this project to ensure that that the sensors are paired with the users. This bespoke security model could be aligned with WoT committee recommendations and standardization efforts.

The raw-data is secured as follows:

- The wearable devices are paired with the user. The user goes through a sequence of steps to accomplish the pairing. The sensors are also registered with a home gateway. The handshake is enabled by security certificates issued by a security server in the Cloud
- The home gateway server which acts as the data acquisition hub for the event feeds from the sensors authenticates the feeds and associates them with identities. This allows for multiple viewers' data correlated to a single content to be concurrently and securely acquired and processed.

There is a need to secure the virtual profile of the viewers that goes where the viewer goes (along with the wearable devices). The viewers can access their virtual profiles by signing in through the authentication framework. This ensures that the person who is using the wearable devices is who they claim they are.

A default strict security profile comes into effect upon signing on. Subsequently, viewers may also configure various levels of security and privacy settings for the data and the patterns in the data.

## Conclusions

This experiment is meant to prove that biometric and ambient sensor data can be correlated with consumer device interactions like TV viewing behavior for an enriched experience. This is a Web-of-Things use-case. With the advances in the Wearable computing domain it is now possible to conceive of pragmatic solutions to acquire and process the sensor data. Cisco could participate in defining this emerging landscape of devices and formulate policies, protocols and standards for these 'smart' devices to securely acquire and share data for processing.

## References


[1] The Foundation for Intelligent Physical Agents: Specifications, Available from http://www.fipa.org, 2006.

[2] C. Conati, "Probabilistic assessment of user's emotions in educational games", Journal of Applied Artificial Intelligence, 16(7-8):555–575, 2002

[3] T. Bylander, "The Computational Complexity of Propositional STRIPS Planning," Artificial Intelligence, pp. 165–204, 1994

[4] M. Dean and G. Schreiber, "OWL Web Ontology Language Reference", Technical report, W3C, February 2004

[5] JADE, Java Agent Development Framework, http://jade.tilab.com/ , Online, Accessed August 31, 2011

[6] GOAL Agent Programming Language, https://mmi.tudelft.nl/trac/goal/ , Online, Accessed August 31, 2011

[7] Wijayasekara, D., Manic, M., "Human machine interaction via brain activity monitoring", The 6th International Conference on Human System Interaction (HSI), 2013

[8] Wright, Franklin Pierce, "Emotional instant messaging with the Epoc headset", M.S thesis., University of Maryland, Baltimore County, 2010

[9] e-Health Sensor Platform:
http://www.cooking-hacks.com/documentation/tutorials/ehealth-biometric-sensor-platform-arduino-raspberry-pi-medical

[10] Brain Computer Interface:
http://emotiv.com/ideas/epoc.php

[11] EEG Headset:
https://www.youtube.com/watch?v=40L3SGmcPDQ
https://www.youtube.com/watch?v=YTvXtaOSfuo

[12] Emotiv Programming Interface:
https://github.com/ozancaglayan/python-emotiv