Contract ITEA 2 - 07008

MIDAS

Multimodal Interfaces for Disabled and Ageing Society

Project number: ITEA 2 - 07008

ITEA Roadmap application domains:
Major: Home
Minor: Nomadic

ITEA Roadmap technology categories:
Major: Human-Computer Interface
Minor: Data and Content Representation

WP1
Deliverable D1.4:
Technological Survey
and Literature Tool Study

Due date of deliverable: 31/05/2009
Actual submission date: 15/07/2009

Start date of project: 01/10/2008 Duration: 36 Months

Project coordinator name: Laure Chotard

Organisation name of lead contractor for this deliverable: TELEFÓNICA I+D

Editor: Miguel Santos **Revision:** 4.0
Contributors: Fano Ramparany(Orange Labs), Adrien Bruno(Orange Labs), Yazid Benazzouz (Orange Labs), Taner Danisman(CNRS), Jesus Murgoitio(Robotiker), Miguel Santos(TID), Susana Ferreras(TID), David Daroca(TID), Tamer Köşker(Katron), Begoña (ESS), Antonio Maestre (CITIC), Jordi Tenas(Ficosa) , Christophe Leroux (CEA LIST), Patrick Hede (CEA LIST).

TABLE OF CONTENTS

LIST OF FIGURES

0. Executive Summary

1. Introduction

1.1. Purpose of this document

The objective of this document is to identify and define the relevant and potentially useful technologies for MIDAS in the specification, modeling and development of MIDAS metadata framework, service architecture and service platform. The task will take into account the scenarios in order to analyze the potential of the identified technologies and produce a technology survey useful as input and support for the other technical work packages.

This document describes the literature study performed regarding the state of the art on data mining and fusion techniques and methods for modeling of the multi-parametric-recorded data in relation to environmental and other factors. Specific focus concern:

- Gesture modality
- Eye gaze modality
- Speech modality
- Affective modality
- Synchronization and fusion of modalities
- Framework for creation and delivering of multimodal user interfaces
- A framework will allow a model based creation of user interfaces and a simultaneous delivering of multimodal user interfaces to different devices

1.2. Document Overview

This document is split in three parts. The fist one *Theoretical background* that deals with technological background and common concepts; The second one *Technological Framework* is a more detailed explanation about different technologies related whit Midas Project and it is divided in two parts one for each main scenario; and the last One, *Final Conclusions* that provides a conclusion about the technologies used in Midas project and the reason to choose it.

1.3. Editors

Miguel Santos Telefónica I+D

1.4. Change History

Date	Author	Update description	Doc. Version
01-11-2008	TID	Table of Content definition	1.0
02-17-2009	Orange Labs	Context management contribution	1.1
02-17-2009	Orange Labs	SmartCam Localisation contribution	1.2
02-17-2009	Geomobile	SmartCam Localisation Update	1.3
02-17-2009	CNRS	Theoretical Background contribution	1.4
02-18-2009	TID	First Draft compilation	1.5 – 2.0
04-03-2009	CITIC	Sensor System – Glucose Level contribution	2.1
09-03-2009	Robotiker	Graphic interfaces interaction & SOA architecture and Web Services contribution	2.2
03-18-2009	TID	User study: Adaptation to the different profiles contribution	2.3
03-18-2009	TID	Update First Draft	2.4
03-23-2009	Katron	Simulation environment contribution	2.5
03-25-2009	Robotiker	Command recognition technology & Sensor system – ECG contribution	2.6
04-24-2009	ESS	Multimodal interfaces technologies contribution	2.7
06-03-2009	Robotiker	Out car communication contribution	2.8
06-03-2009	TID	System intelligence framework contribution	2.9
06-03-2009	TID	Second Draft compilation	2.9 – 3.0
06-09-2009	Ficotriad	Haptic technology contribution	3.1
06-15-2009	TID	Technological Integration	3.2
06-16-2009	TID	Conclusions included	3.3
06-23-2009	CEA LIST	Image understanding	2.1.6
06-24-2009	Orange Labs	Automated Learning	3.1.2.4

2. Theoretical background

2.1. Affective Computing

Over the last quarter century, there is increased body of research on recognition of emotional expressions on different environments. Emotions are complex psychophysical processes of human behavior that is a part of psychology, neuroscience, cognitive science, and artificial intelligence. On the other hand, emotional understanding is an important issue for intentional behaviors. Since, emotions convey our feeling to others, without emotions we behave like a robot.

Current state-of-art in computer human interaction largely ignores emotion whereas it has a biasing role in human-to-human communication in our everyday life. In the mean time, a successful computer human interaction system should be able to recognize, interpret, and process human emotions. The term “Affective Computing”, first used by Picard (1997) at MIT Media Lab., deals with systems, which can process emotion signals. Affective computing could offer benefits in an almost limitless range of applications such as computer aided tutoring, customer relationship management, automatic product reviews and even card driver safety systems.

2.1.1 Emotional Speech Recognition

Emotional speech recognition is classifying speech segment into a set of predetermined emotional classes. A variety of computer systems can use emotional speech classification including call center applications, psychology and emotion enabled Text to Speech (TTS) engines. Current studies on emotion recognition mainly concentrate on visual modalities, including facial expressions, muscle movements, action units, body movements, etc. However, emotion itself is a multimodal concept and emotion recognition task requires interdisciplinary studies including visual, textual, acoustic, and physiological signal domains.

2.1.2 Voice Activity Detection

Speech vs. non-speech segmentation of audio signals widely used in automatic speech recognition, discrete speech recognition and speaker recognition areas to improve robustness of these systems. Aim of the Voice Activity Detection task is to find the presence or absence of human speech in a given

audio signal. Elimination of non-speech segments within spoken content reduces the computational complexity while improving classification performance. A good speech-vs.-non-speech segmentation method should be successful on unseen data and real world sounds where background noise exists. These methods intended to solve speech classification problem that requires high dimensional feature vectors which is in fact a fusion of a number of different feature sets.

Figure 2-1 shows speech and non-speech segments in an audio signal.

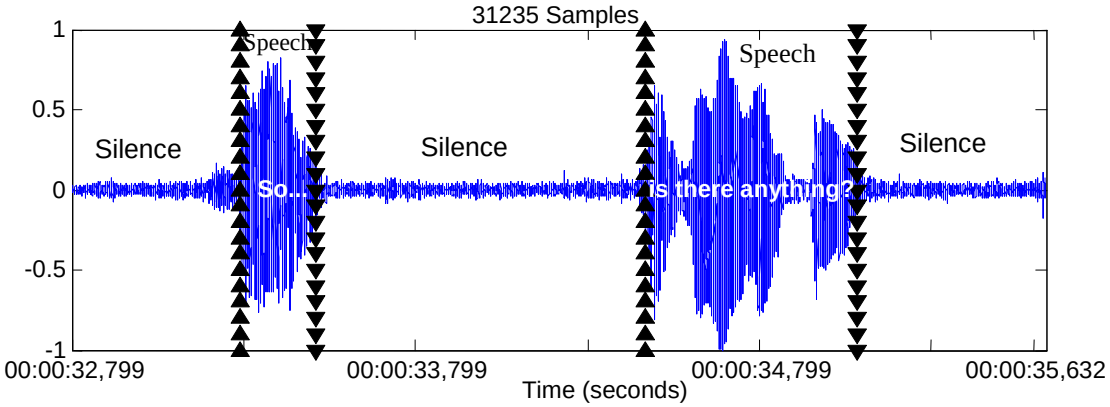

Figure 2-1 Voice activity detection (VAD)

Like other modalities, auditory modality needs the segmentation process. Audio shots or microphone shots are uninterrupted sound recording blocks, which provide boundaries of the speech signal. First audio signal must be cleaned to reduce the noise effect and then it must be segmented into speech, environmental and musical sounds. Continuity on these signals gives more semantic clues about the emotional content. For example, statistical analysis of loudness, brightness, harmonicity, timbre, and rhythm values can give clues about laughter, crowds, water sound, explosions, thunder etc.

According to the research of Murray & Arnott (1993), Table 1 shows the acoustic characteristics of the emotions.

Table 1 Acoustic characteristics of emotions

	Anger	Happiness	Sadness	Fear	Disgust
Speech rate	Slightly faster	Faster or slower	Slightly slower	Much faster	Very much slower
Pitch average	Very much higher	Much higher	Slightly slower	Very much higher	Very much lower
Pitch range	Much wider	Much wider	Slightly narrower	Much wider	Slightly wider
Intensity	Higher	Higher	Lower	Normal	Lower
Voice quality	Breathy, chest tone	Breathy, blaring	Resonant	Irregular voicing	Grumbled chest tone
Pitch changes	Abrupt, on stressed syllables	Smooth, upward inflections	Downward inflections	Normal	Wide downward terminal inflections
Articulation	Tense	Normal	Slurring	Precise	Normal

Previous works on VAD uses Mel Frequency Cepstral Coefficients (MFCC), pitch frequencies as formants, speech rate, and Teager Energy Operator (TEO) for features extraction purposes. Classification techniques used in emotion classification task usually includes Multi-Class Support Vector Machines (MC-SVM), Artificial Neural Networks (ANN), Hidden Markov Models (HMM), Linear Discriminant Analysis (LDA) and K-Nearest Neighbor (K-NN) classifiers.

Vandecatseye & Martens (2003), used GMM and HMM on Hub4 News dataset and their speech detection accuracy is 99.5% while non-speech detection is 76.44%. Shafran & Rose (2003), used Bagging MLP method on SPINE corpus and get 95,8% accuracy. Casagrande, Eck, & Kigl (2005) used AdaBoost method with Haar-like features and smoothing technique on Scheirer-Slaney dataset and get 93% accuracy. Meinedo & Neto (2005) used ANN-MLP on Cost278-BN (Vandecatseye, Martens, & Neto, 2004) dataset and get 97.5% accuracy for speech, 70.6% for non-speech classification. An interesting point in VAD studies is that, classification methods requires large-scale datasets for training purposes (Byrne, Beyerlein, Huerta, Khudanpur, & Marthi et al., 2000), (Foo & Yap, 1997). Difficulties in finding large-scale datasets caused researchers to use limited datasets.

2.1.3 Emotional Speech Studies

Previous works on this area use Mel Frequency Cepstral Coefficients (MFCC) (Shami &Verhelst, 2007), (Altun & Polat, 2007), (Le, Quenot, & Castelli, 2004), pitch frequencies as formants (Zervas, Mporas, Fakotakis, & Kokkinakis, 2006), (Ververidis, Kotropoulos, & Pitas, 2004), (Hammal, Bozkurt, Couvreur, Unay, Caplier, 2005), (Datcu & Rothkrantz, 2005), (Shami & Verhelst, 2007), (Teodorescu & Feraru, 2007), (Lugger & Yang, 2006,2007), (Sedaaghi, Kotropoulos, & Ververidis,

2007), (Altun and Polat, 2007), (Zhongzhe, Dellandrea, Dou, & Chen, 2006), (Sedaaghi et al., 2007), (Pasechke & Sendlmeier, 2000) speech rate (Hammal et al., 2005), zero crossing rate (Lugger & Yang, 2007), Fujisaki parameters (Fujisaki & Hirose, 1984), (Zervas et al., 2006), energy (Zhongzhe, Dellandrea, Dou, & Chen,, 2006), (Ververidis, Kotropoulos, & Pitas, 2004), (Hammal et al., 2005), (Altun & Polat, 2007), (Sedaaghi et al., 2007), (Lugger & Yang, 2007), linear predictive coding (LPC) (Altun & Polat, 2007), (Le et al., 2004) for feature extraction purposes.

(Zhongzhe, Dellandrea, Dou, & Chen,, 2006), (Ververidis, Kotropoulos, & Pitas, 2004), (Sedaaghi et al., 2007) and (Lugger & Yang, 2007) used sequential floating forward selection (SFFS) method to discover the best feature set for the classification.

Classification techniques used in emotion classification task includes Support Vector Machines (SVM) (Hammal et al., 2005), (Shami &Verhelst, 2007), (Altun & Polat, 2007), Neural Networks (NN) (Zhongzhe et al., 2006), Hidden Markov Models (HMM) (Le et al., 2004), Linear Discriminant Analysis (LDA) (Hammal et al., 2005), (Lugger & Yang, 2006), Instance Based Learning (Zervas et al., 2006), Vector Quantification (VQ) (Le et al., 2004), C4.5 (Zervas et al., 2006), (Shami & Verhelst, 2007), GentleBoost (Datcu & Rothkrantz, 2005), Bayes Classifiers (Ververidis, Kotropoulos, & Pitas, 2004), (Hammal et al., 2005), (Lugger & Yang, 2007) and K-Nearest Neighbor (K-NN) (Ververidis, Kotropoulos, & Pitas, 2004), (Hammal et al., 2005), (Shami & Verhelst, 2007) classifiers.

To date, many of studies on this subject employed on DES dataset, and Table 2 provides a quick snapshot of them. Zervas et al. (2006) and Datcu & Rothkrantz (2005) achieved better accuracy than human based evaluation (Engberg & Hansen, 1996) using Instance Based Learning (IBL) and GentleBoost algorithms respectively. Baseline accuracy is computed by classifying all the utterances as the major emotional class in test set. According to Engberg & Hansen (1996), 67% of the emotions are correctly identified by humans on average on DES dataset. Sedaaghi et al. (2007) used sequential floating feature selection (SFFS) for optimizing correct classification rate of Bayes Classifier on DES dataset and get 48.91% accuracy, in average. Le et al. (2004) achieved 55% accuracy for speaker independent study. Their speaker dependent result is between 70% and 80%.

Table 2 Performance of past studies on DES dataset in terms of accuracy.

Study	Classifier	# of Classes	Accuracy %
Baseline	-	5	20.0
Datcu & Rothkrantz (2005)	GentleBoost	5	72.0
Hammal et al. (2005)	Bayes Classifier	5	53.8

Human Eval. (Engberg & Hansen (1996)	-	5	67
Le et al. (2004)	Vector Quantification	5	55.0
Sedaaghi et al. (2007)	Bayes + SFFS + Genetic Alg.	5	48.9
Shami & Verhelst (2007)	ADA-C4.5+ AIBO approach	5	64.1
Shami & Verhelst (2007)	ADA-C4.5+ SBA approach	5	59.7
Ververidis, & Kotropoulos (2004)	Bayes+SFS	5	51.6
Zervas et al. (2006)	C4.5	5	66.0
Zervas et al. (2006)	Instance Based Learning	5	72.9

Table 3 presents squeezed comparison of studies held on EmoDB dataset in terms of classifier type, number of classes and accuracy. As in studies on DES, (Datcu & Rothkrantz, 2005) again used GentleBoost algorithm on EmoDB dataset for six emotion classes out of seven, and achieved 86.3% accuracy. (Altun & Polat, 2007) used SVM for four class emotion classification, (Lugger & Yang, 2007) used linear discriminant analyses for anger, happiness, sadness, and neutral emotions and they reported 81.8% accuracy. Additionally, they have tested Bayes classifier, and achieved 74.4% accuracy for six classes using leave-one-speaker-out method on short utterances. Gender dependent study from (Zhongzhe et al., 2006) achieved 77.3% accuracy for female subjects considering seven classes.

Table 3 Previous studies on EmoDB dataset in terms of accuracy %.

Study	Classifier	# of Classes	Accuracy %
Altun & Polat (2007)	SVM	4	85.5
Baseline		7	23.7
Datcu & Rothkrantz (2005)	GentleBoost	6	86.3
Human Eval. Burkhardt et al. (2005)		7	86.0
Lugger & Yang (2007)	Bayes Classifier	6	74.4
Lugger & Yang (2007)	Linear Discriminant Analyses	4	81.8
Shami & Verhelst (2007)	SVM+ AIBO approach	7	75.5
Shami & Verhelst (2007)	SVM+ SBA Approach	7	65.5
Zhongzhe et al. (2006)	Two-Stage NN	7	77.3

2.1.4 Facial Expression Recognition

Duchenne du Boulogne first expresses facial expressions in 1862. He was a pioneering neurophysiologist and photographer. Most researchers acknowledge their debt to Duchenne and his book "The Mechanisms of Human Facial Expression".

Ekman & Friesen (1978) presented the most important comprehensive study in the content of facial expression recognition, called Facial Action Coding System (FACS). They have defined a method for

describing and measuring facial behaviors and facial movements based on anatomical analysis of facial action. Many of the later studies are based on the initial experiments of the Ekmans' studies.

Measurement unit of the FACS system is Action Units (AUs). They have defined a set of 44 Action Units (AUs) in original work that having a unique numeric code, which represents all possible distinguishable facial movements because of change in muscular actions. 30 of them are related to a specific contraction of muscles and 14 of them are unspecified.

Most of the researchers use six basic "universal facial expressions" corresponding to happiness, surprise, sadness, fear, anger, and last disgust. Figure 2-2 shows sample set from the (Cohen, 2000, pp. 8-30).

Figure 2-2 Sample six basic facial expression data set from (Cohen, 2000)

Ekman studied on video tapes in order to find changes in human face when there is an emotion exists. According to the work, a smile exists if the corners of the mouth lift up through movement of a muscle called zygomaticus major, and the eyes crinkle, causing "crow's feet," through contraction of the orbicularis oculi muscle.

Changes in location and shape of the facial features are observed. Score of a facial expression consists of a set of Action Units. Duration and intensity of the facial expression are also used. Observed raw FACS scores should be analyzed in order to produce behavior that is more meaningful. FACS has four main steps;

Observe movements and then match the AUs with the observed movements.

An intensity score is given for each one of the actions

Determine the action unit's type as asymmetry or joint

Determining the face and facial feature positions during the movement of the face in the sequence.

Interpreting AU is a difficult task. For example, there are six main emotional states exists but each of them has many variations. Figure 2-3 shows two different type of smile of the same person. Therefore, usually each emotional state is represented by a set of action units. Thus most of the action units are additive.

AU12 AU6+12+25

Figure 2-3 Two different types of smile. (Lien, 1998)

One of the limitations of the FACS system is nonexistence of a time element for the action units. Electro-Myo-Graphy (EMG) studies, which are based on the measurement of electrical activity of muscles, showed that facial expressions occur in a time-aligned sequence beginning with application, continuing with release and finally relaxation.

Face tracking is needed to compute the movements of each facial feature. Earlier studies uses facial feature make up and different face tracking algorithms including 3D-based models Cohen, (2000).

Terzopoulos & Water developed a model that tracked facial features in order to observe required parameters for a three-dimensional wire-frame face model (Terzopoulos, & Water, 1993, pp. 569-579). However, their work has a limitation in which the humans facial features should be marked up to robustly track these facial features.

The most difficult task in facial expression recognition is tracking and extracting facial features from a set of image sequence. A huge number of parameters and features should be considered (Cohen, 2000, pp. 8-30). Therefore, it is necessary to decrease the number of points that are required to track facial features thus decreasing computational time. Principal Feature Analysis (PFA) makes

this task and finds the most important feature points that need tracking. (Cohen, 2000, pp. 8-30) used PFA method to find the best facial feature points. Cohen initially marked up the face to be tracked to get robust results and then tracked a video of 60 seconds at 30fps. Figure 2-4 shows example images from the video sequences.

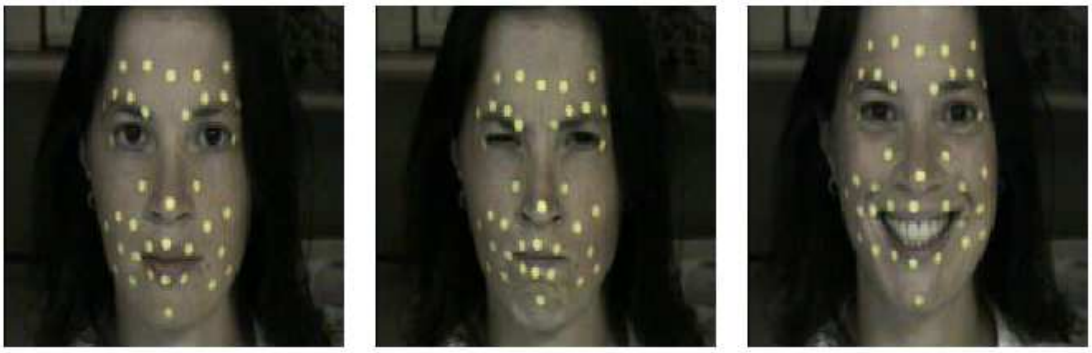

Figure 2-4 Example images from the video sequences. (Cohen, 2000, pp. 8-30)

Cohen has used 40 facial points each having two directions, horizontal and vertical to be tracked. For the PFA, these points divided into two groups, namely upper face (eyes and above) and lower face. Then the correlation matrix computed. After applying the principle feature analysis, the resulting image showed in Figure 2-5 .

In Figure 2-5 , selected feature points are marked by arrows. According to Cohen's work, PFA is able to model complex face motions and reduces the complexity of existing algorithms.

In model based recognition systems, a feature vector should be defined for each expression and a similarity metric should be used to compute the difference between these expressions.

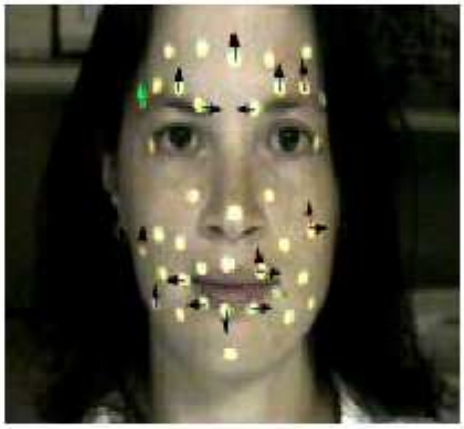

Figure 2-5 Result of PFA method. Arrows shows the principal features chosen

Pantic, & Rothkrantz (2000) developed an Integrated System for Facial Expression Recognition (ISFER) which is an expert system for emotional classification of human facial expressions from still full-face images. The system has two main parts. The first part is ISFER Workbench, used for feature detection and the latter is an inference engine called HERCULES.

First part of the system, ISFER Workbench presents a system for hybrid facial feature detection. In this part, multiple feature detection techniques are applied in parallel. Therefore, it gives a chance to use redundant parts with eliminating uncertain or missing data. It has several modules, each doing different types of pre-processing, detection, and extraction. They have used both frontal view and side view of human faces. Figure 2-6 shows the frontal-view template from their work. **Figure 2-7** shows algorithmic representation of ISFER Workbench. ISFER is complete automated system that is able to extract facial features from digitized still images. It does not deal with image sequences. Automatic encoding of facial Action Units (Ekman & Friesen, 1978) and automatically classifies six basic universal emotional expressions, happiness, anger, surprise, fear, sadness, and disgust.

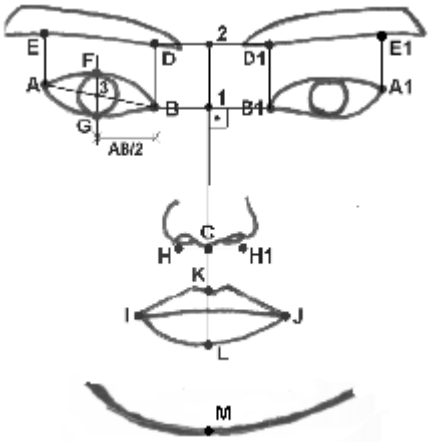

Figure 2-6 Facial points of the frontal-view (Pantic, & Rothkrantz, 2000, pp.881-905)

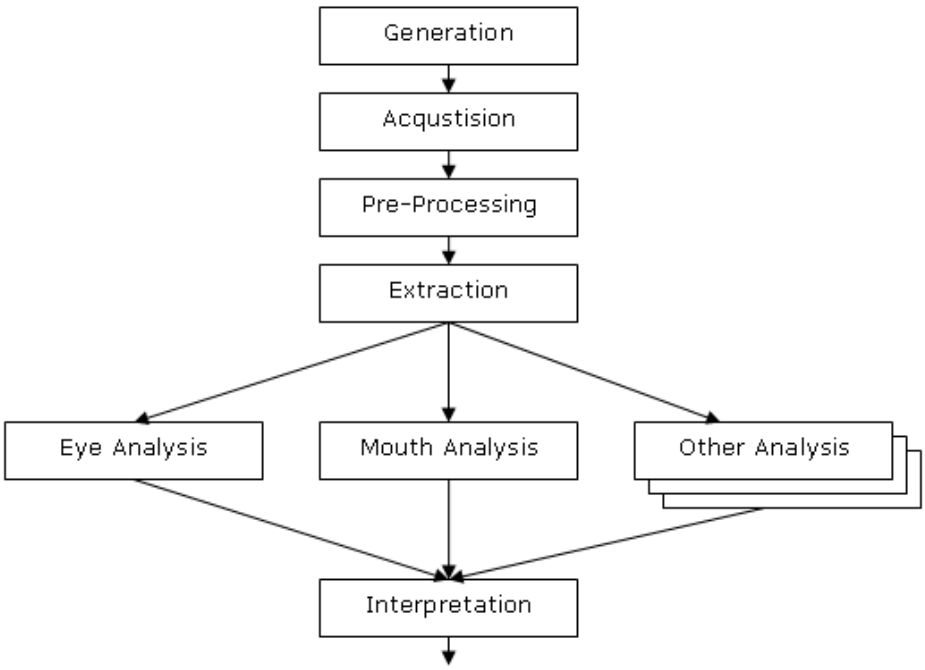

Figure 2-7 Algorithmic representation of ISFER Workbench

The second part of the system, HERCULES, converts low-level face geometry in high-level facial actions. Details of these points are described in Table 4.

Table 4 Details of facial points in Figure 2-6

Poin	Description	Point	Description
B	Left eye inner corner, stable point	F	Top of the left eye, non-stable

B1	Right eye inner corner, stable point	F1	Top of the right eye, non-stable
A	Left eye outer corner ,stable point	G	Bottom of the left eye, non-stable
A1	Right eye outer corner, stable point	G1	Bottom of the right eye, non-
H	Left nostril centre, non-stable	K	Top of the upper lip, non-stable
H1	Right nostril centre, non-stable	L	Bottom of the lower lip, non-
D	Left eyebrow inner corner, non-	I	Left corner of the mouth, non-
D1	Right eyebrow inner corner, non-	J	Right corner of the mouth, non-
E	Left eyebrow outer corner, non-	M	Tip of the chin, non-stable
E1	Right eyebrow outer corner, non-		

Dailey, Cottrell, Padgett, & Adolphs (2002) showed that a simple biologically neural network model, trained to classify facial expressions matches a variety of psychological data into six universal basic emotions. They have considered categorization, similarity, reaction times, discrimination, and recognition difficulty, in both qualitatively and quantitatively. Figure 2-8 shows morphing from happiness to disgust. They have used Morphs software version 2.5.

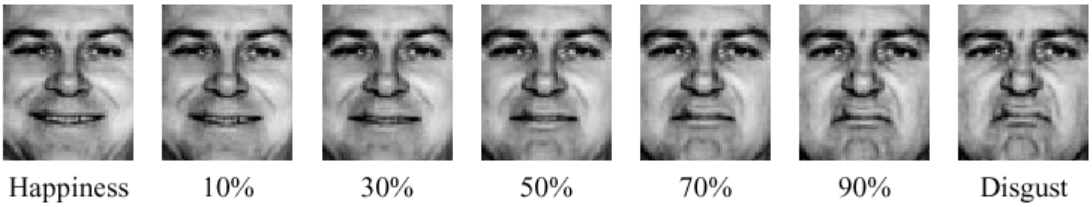

Figure 2-8 Morphs from happiness to disgust (Dailey et al., 2002)

Franco & Treves (1997) inserted a local unsupervised processing stage within a neural network to recognize facial expressions (Franco, & Treves, 1997). They worked with Yale Faces database and their neural net architecture has four layers of neurons. They have success at rate of 84.5% on unseen faces and 83.2% when principal component analysis processing applied at the initial stage.

Another method is use of Hidden Markov Models that solves classification problems especially for speech recognition systems because of its ability to model or classify non-static events. However, compared with other models, time required to solve the problem is significantly higher. Figure 2-9 shows a maximum likelihood classifier for emotion specific HMM.

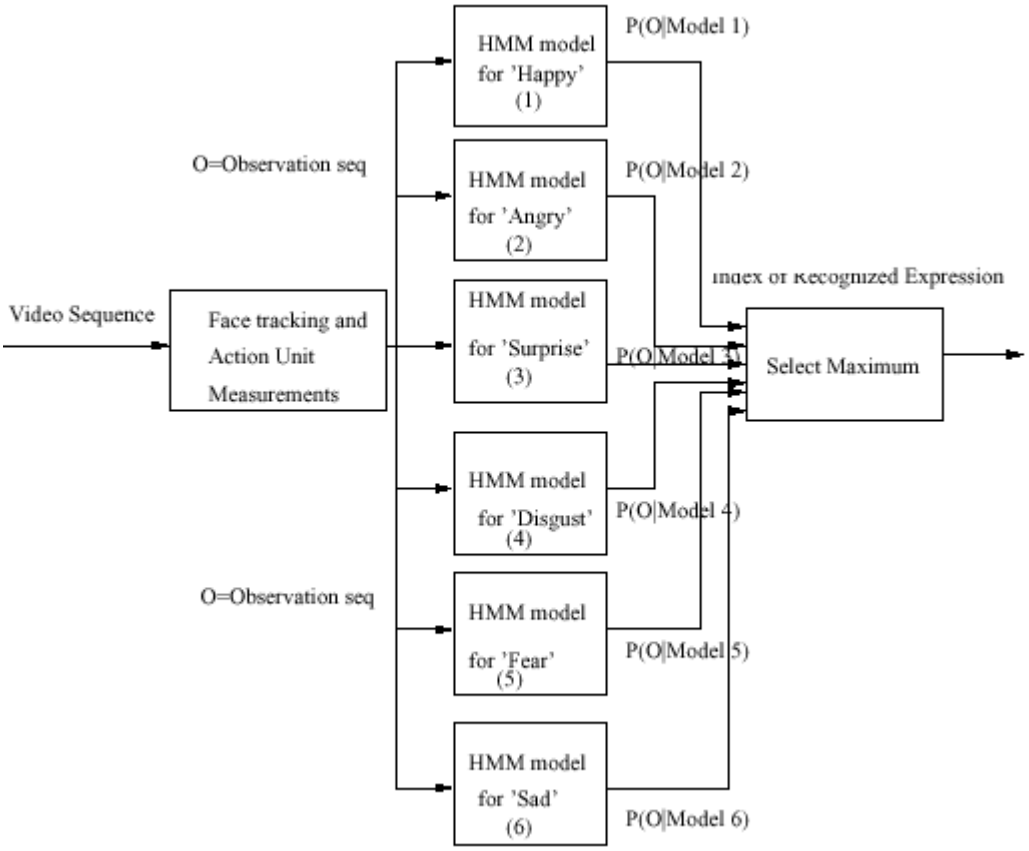

Figure 2-9 Maximum likelihood classifier for emotion specific HMM case (Cohen, 2000)

According to Figure 2-9 , after making face tracking and Action Unit measurements, each one of the six Hidden Markov Model representing six universal emotions produces a result showing the probability of belonging to a specific type of emotion. At the end of the system, the emotion having the maximum probability is chose as observed emotion.

Ioannou et al. (2005) used a robust facial analysis system and a neurofuzzy system which adapts the specific users' facial expression characteristics. They defined facial action paremeters (FAPs) according to the MPEG-4 standard. An overview of their system is shown in Figure 2-10 .

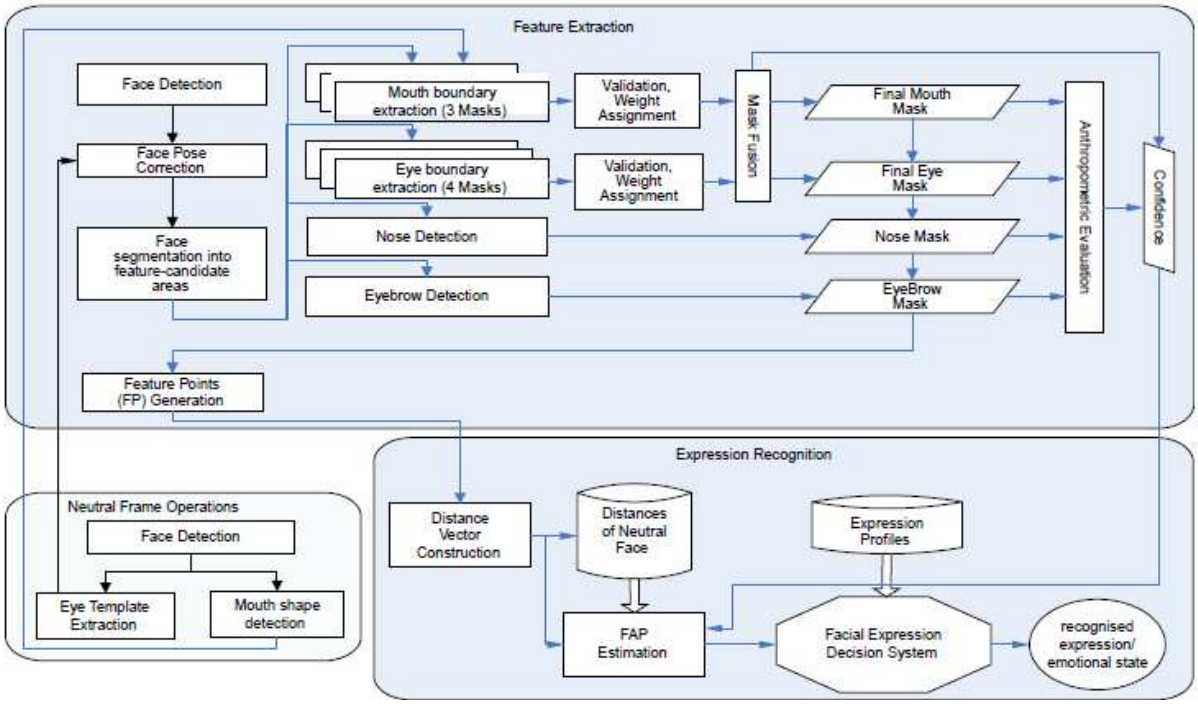

Figure 2-10 Facial Feature extraction and facial expression analysis system Ioannou et al. (2005)

All these studies use different feature selection, feature processing and machine vision algorithms. Popular classification techniques for facial expression recognition are;

- kNN
- Support Vector Machines (linear & RBF kernels)
- Rule-based systems
- Model-based, shape-based, motion template-based systems
- Neural networks, MLP
- Hidden Markov Models
- Bayesian networks

In case of feature processing, optical flow gives the best results for facial expression recognition than Fourier transforms, parametric and appearance models. Yacoob & Davis (1996) and Essa & Pentland (1997) used optical flow algorithm for feature processing and achieved 95% and 98% performance for six and five emotion categories respectively.

2.1.5 Multimodal Emotion Recognition

Multimodal Emotion Recognition (MER) is a need for better classification on real world data. Current studies use primitive dataset for unimodal training and testing. For the audio modality, there exists limited number of speakers where usually one speaker speaks at a time and there is no background voice. Similarly, in visual modality, datasets with static images having single frontal upright faces recorded under studio environments are dominant. In addition, face pictures are taken when they are in silence so that mouth movements do not affect the classification for the video datasets, faces are in frontal upright position, and they do not speak during the emotional state changes. However, in case of video, it is not possible to find faces in frontal upright position or closed mouths all the time.

In spite of all these limitations, because of the semantic complexity of the emotion recognition task, reported state of the art performances for EER are very low. Therefore, researchers moved to study on bimodal and multimodal EER studies. Sebe, Bakker, Cohen, Gevers & Huang (2005) used visual and speech modalities, and Gunes & Piccardi (2007) used face and gesture properties. They reported that bimodal studies perform better than result of single modality. In case of fusing different types of information, models that can handle incomplete and missing values must be used. Bayesian networks, Hidden Markov Models and TBM can be used for this purpose. Current integration methods use probabilistic classifiers such as Hidden Markov Models (HMM) and Bayesian Classifiers Alatan, Akansu, & Wolf (2001), Huang, Liu, Wang, Chen, & Wong (1999), Snoek, & Worring (2005).

De Silva & Ng (2000), studied the six basic emotions on a dataset having 144 image sequences and audio files from 2 subjects using Hidden Markov Models (HMM) and they used a rule based fusion scheme on video by considering both the facial expressions and emotional speech. On visual part they have used optical flow algorithm to find the displacements of facial feature points and on audio part they used pitch values with a HMM classifier. They have used rule based bimodal fusion where the same result is expected from visual and audio modalities, otherwise output of the dominant modality is selected as a final emotional class. Overall, bimodal recognition rate is 72%, which is better than unimodal video and audio results.

Sebe, Bakker, Cohen, Gevers, & Huang (2005), used Bayesian networks for bimodal fusion of audio-visual information containing a set of 38 subjects with 11 affective states. On visual part, they used a face tracker and 3D wireframe model to fit facial features to control points. On audio part, they used logarithm of energy, syllable rate and two pitch values. Their proposed model considers the

speaking state of the speaker. If the subject speaks then recognition process is also affected by speech information. In addition, According to their results, average recognition accuracy is 56% for face-only classifier, 45% for the prosody-only classifier, and 90% for bimodal classifier.

Gunes & Piccardi (2007) automatically extracted the face and upper-body gesture features from video data of 4 subjects with 6 emotions. They selected to use BayesNet classifier for unimodal body and face features with the classification rate of 89.9% and 76.4% respectively. They employed both the feature level and decision level fusion schemes. According to their results, BayesNet classifier with feature level fusion scheme provides 94% accuracy, which is better than single modality results. They used posterior probabilities of unimodal results with sum, product, and weight based criteria for the decision level fusion scheme and obtained 91.1%, 87.3%, 79.7% accuracy for each respectively. Their results showed that early fusion scheme gives better performance than both the late fusion and the unimodal schemes.

Wu, Oviatt, & Cohen (1999) studied on approximating the conditional density functions for speech and gesture modalities. Their approach can be further generalized for other modalities. They also proposed a method to predict the theoretical lower and upper performance bounds of the fused system.

Go, Kwak, Lee, & Chun (2003) studied on a dataset containing facial images and speech signals of 20 people (10 per gender) showing 6 emotions. They have used multi-resolution analysis of wavelets for speech signal and Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for facial feature extraction. Their fusion mechanism considers the maximum of membership value, which is computed by comparing the input features with codebook features. Their bimodal recognition rate is 95% for male and 98% for female subjects.

Hammal, Couvreur, Caplier, & Rombaut (2007) has proposed fusion architecture based on Transferable Belief Model, TBM, where information from different sources can be combined to provide decisions that are more powerful. They showed that, humans are able to detect facial expressions by viewing only the contours of the facial features. In their study, emotions are represented by skeleton of the facial features generated from contour pixels of static and frontal viewed face images. TBM well suits for the fusion problem because TMB can handle both incomplete data and imprecise values.

2.1.6 Image understanding and knowledge representation overview

Introduction

Human and animals have impressive abilities to interact with their environment, navigate themselves in desired path without bumping obstacles, extract the parameters of objects (texture, color and so on) using vision. Simulation of each vision system and understanding objects parameters have been one of the most challenging research branches [SWAIN91]. Image understanding refers to extract 2D-3D information from one or more images. Manifestation of edges, lines, curves, depth, motion vectors and object description are of interest. Images can be obtained from black and white camera, infra-red, laser, synthetic aperture radar (SAR), millimeter wave radar and son on. Integration of two or more sensors with their specific measurements arise sensor fusion problem. [IBLOC95]

One of the goals of computers vision is to automatically interpret general digital images of arbitrary scenes. This goal has produced a lot of research over the last 40 years, yet a good solution to this general problem remains out of reach. A reason for this is that the problem is not complex but very complex as well as some people say that is impossible to solve it by a machine. The problem of visual perception is typically under-constrained. Information like absolute scale and depth is lost when the scene is projected onto an image plane. In fact, there are an infinite number of scenes that can produce the exact same image, which makes direct computation of scene geometry from a single image extremely difficult to recover.

The detection and recognition of objects in images is a key research topic in the computer vision community. Within this area, face recognition and interpretation has attracted increasing attention owing to the possibility of unveiling human perception mechanisms, and for the development of practical biometric systems. [BRUNEL07]

Image understanding

"Image understanding (IU) is the research area concerned with the design and experimentation of computer systems that integrate explicit models of a visual problem domain with one or more methods for extracting features from images and one or more methods for matching features with models using a control structure. Given a goal, or a reason for looking at a particular scene, these systems produce descriptions of both the images and the world scenes that the images represent."[SURV87]

Different methods exist to try to solve the hard problem of segmentation:

Bottom-up processing, Top-down processing, Active vision (snakes), Gradient-decent minimization.

In fact, all this global approach needs each time external information to correctly localize object. In other hand, local approaches working on points or regions avoid segmentation and push away the final decision, so the distinction between background and object became no more a problem.

Object Recognition Survey

Global image descriptors

Color histograms

The histogram provides a compact summarization of the distribution of data in an image. The color histogram of an image is relatively invariant with translation and rotation about the viewing axis, and varies only slowly with the angle of view *[SWAIN91]*.By comparing histograms signatures of two images and matching the color content of one image with the other, the color histogram is particularly well suited for the problem of recognizing an object of unknown position and rotation within a scene. Importantly, translation of an RGB image into the illumination invariant rgb chromacity space allows the histogram to operate well in varying light levels.

The main drawback of histograms for classification is that the representation is dependent of the color of the object being studied, ignoring its shape and texture. Color histograms can potentially be identical for two images with different object content, which happens to share color information. Conversely, without spatial or shape information, similar objects of different color may be indistinguishable based solely on color histogram comparisons. There is no way to distinguish a red and white cup from a red and white plate. Put another way, histogram-based algorithms have no concept of a generic 'cup', and a model of a red and white cup is no use when given an otherwise identical blue and white cup. Another problem is that color histograms have high sensitivity to noisy interference such as lighting intensity changes and quantization errors. High dimensionality(bins) of color histograms are also another issue. Some color histogram feature spaces often occupy more than one hundred dimensions [IMAP02].

An other interesting histogram refinement consist of a simple image analysis algorithm that classifies image pixels as either border or interior (space add on), with a compact logarithmic distance (dLog) for comparing histograms, and a compact representation as a short signature of features extracted from images. [BIC02]

Some of the proposed solutions have been color histogram intersection, color constant indexing, cumulative color histogram, quadratic distance, and last but not least color correlograms and auto-correlograms [CORRELO]

Although there are drawbacks of using histograms for indexing/classifications, using color in a real-time system has several relative advantages. One is that color information is faster to compute, compared to other "invariants." It has been shown in some cases that color can a be an efficient method for identifying objects of known location and appearances (refer to external link for findings in study)[OBDR02].

Further research into the relationship between color histograms data to the physical properties of the objects in an image has shown they can represent not only object color and illumination but relate to surface roughness and image geometry and provide improved estimate of illumination and object color [INFO02].

Usually Euclidean distance, histogram intersection, or cosine or quadratic distances are used for the calculation of the images' similarity rating. Any of these values does not reflect the similarity rate of two images in itself. It is useful only with comparison to other similar values. This is the reason that all the practical implementations of content-based image retrieval must complete computation of all images from the database. It is the main disadvantage of these implementations.

Other approach to representative color image content is 2D-color histogram. 2D-color histogram considers the relation between the pixel pair colors (not only the lighting component) [BASH06]. 2D-color histogram is a two-dimensional array, Cmax*Cmax, where Cmax is the number of colors that was used in the phase of color quantization. These arrays are treated as matrices, each element of which stores a normalized count of pixel pairs, with each color corresponding to the index of an element in each pixel neighborhood. For comparison of 2D-color histograms it is suggested calculating their correlation, because a 2D-color histogram, constructed as described above, is a random vector (in other words, a multidimensional random value). While creating a set of final images, the images should be arranged in decreasing order of the correlation coefficient. Correlation coefficient may be used also for color histograms comparison. Retrieval results with correlation coefficient are better than with other metrics.[COLHIS] An on line Multimedia Search Engine demo that combine many global descriptors describe here with different metrics is available at [PIRIAD].

Local detectors/descriptors:

The Moravec operator was developed by Hans Moravec in 1977 for his research involving the navigation of the Stanford Cart through a clustered environment. Moravec defined the concept of "points of interest" as being distinct regions in images and concluded these interest points could be used to find matching regions in consecutive image frames. This was a vital low-level processing step that allowed him to determine the existence and location of objects in the vehicle's environment. The Moravec operator is considered a corner detector since it defines interest points as points where there is a large intensity variation in every direction. This is the case at corners. However, Moravec was not specifically interested in finding corners, just distinct regions in an image that could be used to register consecutive image frames.

The Plessey operator differs from the Moravec operator in how the measurement of local autocorrelation is estimated. This measurement allows the variation of the autocorrelation (i.e. intensity variation) over all different orientations to be obtained. The rational for the Plessey operator follows from addressing the limitation of the Moravec operator.

HARRIS: This operator was developed by Chris Harris and Mike Stephens in 1988 [HARRIS88] as a low-level processing step to aid researchers trying to build interpretations of a robot's environment based on image sequences. Specifically, Harris and Stephens were interested in using motion analysis techniques to interpret the environment based on images from a single mobile camera. Like Moravec, they needed a method to match corresponding points in consecutive image frames, but were interested in tracking both corners and edges between frames.

Harris and Stephens developed this combined corner and edge detector by addressing the limitations of the Moravec operator. The result is a far more desirable detector in terms of detection and repeatability rate at the cost of requiring significantly more computation time. Despite the high computational demand, this algorithm is widely used in practice.

The literature refers to this detector as both the Harris corner detector and the Plessey corner detector.

Improvement of Harris method is based on scale space: Interest points can be adapted to scale and give repeatable results (geometrically stable) [MIKO01]. Local extrema over scale of normalized derivatives indicate the presence of characteristic local structures. The method first computes a multi-scale representation for the Harris interest point detector. Selected points at which a local measure (the

Laplacian) is maximal over scales. This allows a selection of distinctive points for which the characteristic scale is known. These points are invariant to scale, rotation and translation as well as robust to illumination changes and limited changes of viewpoint. For indexing, the image is characterized by a set of scale invariant points; the scale associated with each point allows the computation of a scale invariant descriptor. Our descriptors are, in addition, invariant to image rotation, of affine illumination changes and robust to small perspective deformations. Experimental results for indexing show an excellent performance up to a scale factor of 4 for a database with more than 5000 images.

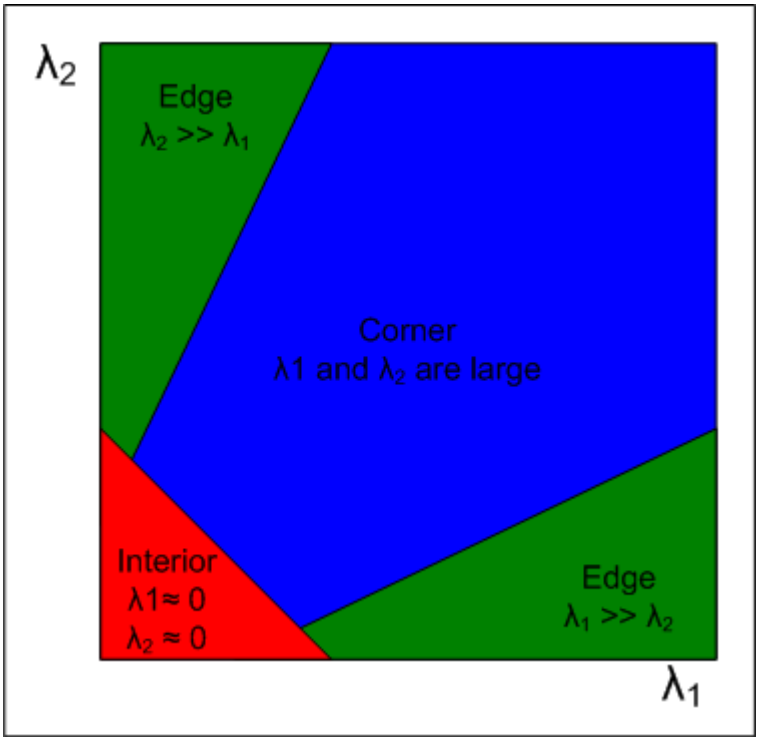

Figure 1: Division of eigenvalue space into distinct feature regions

$$C(x,y) = \det(M) - k(\mathrm{trace}(M))^2$$
$$\det(M) = \lambda_1\lambda_2 = AB - C^2$$
$$\mathrm{trace}(M) = \lambda_1 + \lambda_2 = A + B$$
$$k = \mathrm{constant} \qquad k \text{ around } 0.05$$

Eigenvalues equations for key points selection

SIFT: In 1999 a paper presents a general method for extracting distinctive invariant features from images, which can be used to perform reliable matching between different images of an object or scene. The features are invariant to image scale and rotation, and are shown to provide robust

matching across a a substantial range of affine distortion, addition of noise, change in 3D viewpoint, and change in illumination. The features are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images. Then the method also describes an approach to using these features for object recognition. The recognition proceeds by matching individual features to a database of features from known objects using a fast nearest-neighbor algorithm, followed by a Hough transform to identify clusters belonging to a single object, and finally performing verification through least squares solution for consistent pose parameters. This approach to recognition can robustly identify objects among clutter and occlusion while achieving near real-time performance. [DLOWE03] Evolution Robotics David Lowe start up sale ESPR ViPR software based on SIFT that allowed real time matching [VIPRS]. Many authors have tried to modified and improve SIFT descriptor, an interesting solution called SURF using PCA (principal component analysis) is described in [SURFD].

MSER: Maximally stable extremal regions (MSER) are used as a method of blob detection in images. This technique was proposed by Matas et al. [MSERM02] to find correspondences between image elements from two images with different viewpoints. This method of extracting a comprehensive number of corresponding image elements contributes to the wide-baseline matching, and it has led to better stereo matching and object recognition algorithms.

Another reconciliation with regions approach is the paper [MSERBI06] that introduces a tracking method for the well known local MSER (Maximally Stable Extremal Region) detector. The component tree is used as an efficient data structure, which allows the calculation of MSERs in quasi-linear time. It is demonstrated that the tree is able to manage the required data for tracking. We show that by means of MSER tracking the computational time for the detection of single MSERs can be improved by a factor of 4 to 10. Using a weighted feature vector for data association improves the tracking stability. Furthermore, the component tree enables backward tracking which further improves the robustness.

3. Technological Framework

3.1. Home Scenario

3.1.1 Multimodal interfaces technologies

Multimodal interfaces are an important research topic nowadays in the field of human-machine interfaces (HMI). A multimodal interface should be able to merge information that can be created by the human actor in many ways. The final objective is to enable the user to communicate with the machine by different actions such as touching a screen, speaking or writing. Each one of these actions would be captured and recognised by a determined interface. The machine would be then informed about the action that the user performed, and it depends on this machine the existence or not of any answer or reaction. In case the machine provides an answer, this can be multimodal as well (it can be shown in a screen, notified by audio, a vibration…)

In general, we can state that multimodal interfaces have a great potential in helping people who is not used to interact with standard interfaces (such as keyboards, mouse…) to interact with a system. As every kind of interaction needs usually a specific interface, in the following we try to identify the most relevant and recent singular technologies that can be combined to reach a multimodal behaviour:

3.1.1.1. Sensorial interfaces

Stereoscopic Vision

- Overview: The stereoscopic vision (also called 3D vision) allows perceiving the deepness levels of some objects in an image. Although there are some ways of achieving the effect, the most used technique is based on the so called binocular disparity, for which each eye perceives a slightly different image due to the slightly different observation point. These techniques use polarized glasses, obstruction glasses, virtual reality helmets, or holograms.
- Problem that solves: It enhances the visual perception by adding depth information.
- Current state and prospects: It is a relatively old technology that is rising again thanks to the appearance of new techniques that substantially enhance the image quality, the user comfort and the application to computer and TV screens.

Augmented reality glasses

- Overview: Augmented reality is an interaction paradigm that allows visualizing an environment that combines elements of virtual reality with real elements. The simplest implementation consists on glasses that include in one of the lens a small text (or graphics) display that the user can see, having the perception that this text or image is floating before him.
- Problem that solves: It can provide the user with information about the environment, or about the user's state.
- Current state and prospects: Emerging.

Binaural sound

- Overview: Binaural sound allows locating the origin of sounds in the space surrounding the user, simply by using headphones. It is based on obtaining the answer of a sound in a determined location, applying to it the so called Head Related Transfer Function (HRTF) that takes into account how the head modifies differently the sound coming from the front side and the back side of the user. This information is used to modify a sound depending on the relative position of the user and the source of the sound.
- Problem that solves: It allows finding the source of a sound in a 3-dimensional space, and this is useful for people with visual deficiency.
- Current state and prospects: Being exploited. Its most extended use is for creation of immersive environments in video games, while its usage as location technology is less common

Tactile feedback

- Overview: This technology implements haptic (active touch-sense) displays, related with the skin perception of temperature, pressure, etc. The devices implementing this technology are pin matrices (tactile pin arrays), but there are also actuators which vibrate, heat, blow some pressurized air, change their shape, or create some small currents through the skin.
- Problem that solves: They complement or stand in for the visual and auditive channels of the user. They can add information without diverting their attention out from the task that he is performing. They also help people with visual deficiency to access to information as graphics or tables.

- Current state and prospects: The idea of using tactile feedback dates from 1960, and it has been always considered to have a great potential as an alternative or complementary interfaces, although it has not been very successful due to its technical complexity. Nowadays the most popular devices implementing tactile feedback are installed in video-game pads, but they are very limited.

Force feedback

- Overview: It allows implementing haptic interfaces related with the strength, weight, outlines, etc. This technology is oriented to limit the possible movements of the user that would be handling a remote controlled device. There are many types of these devices, and they are differenced by the zones of the body in which they are used, as well as by the contact points and degree of liberty for the user's movement.
- Problem that solves: It is the key technology for touching and handling 3-dimensional virtual objects, so it is used for remote controlling of robots.
- Current state and prospects: Being exploited. It is a more mature technology that the previous one, since it has been long used in robotic applications. However, the current investigation trends focus on handling more complex virtual models, with more contact points and/or more freedom of movement for the user.

Positional force feedback

- Overview: The positional force feedback implements a robotic arm with sensors that detect the position and orientation of its peak, and returns forces and torsions to the user. It can be understood as a 3-dimensional mouse that returns these feelings to the user depending on the position and the tactile properties of the virtual objects he is touching.
- Problem that solves: It allows to touch and feel small 3-dimensional objects. It si very useful for people with visual deficiency.
- Current state and prospects: Being exploited. In the latter, the price of the low-range devices has been considerably reduced. An example of it is the Phantom OMNI device, from the manufacturer Sensable.

Smelling interface

- Overview: These systems act as outputs, and they emit a determined smell from a bank of available substances. There are some techniques for storing the smells: in liquids, gels, and in microcapsules. The last ones are broken in the moment of generating the smell, while in the other two the smell passes directly to the air. To eliminate a smell that has been released is a hard-to-solve problem.
- Problem that solves: It can enhance multimodal systems by increasing the sensorial perception and by adding information channels that reduce the cognitive load of the user. It helps to create environments and to provide information about different events.
- Current state and prospects: Emerging technology.

3.1.1.2. Motor interfaces

Tactile screen (touch screen)

- Overview: These screens can recognise the contact with the finger of the user (or a pointer), and specify in which point of the area, the screen has been contacted. This feature makes them a powerful input device, eliminating the need for a mouse of joystick. There are many touch screen technologies, and the most popular are APR(Acoustic Pulse Recognition Touch), SAW(Surface Acoustic Wave), and the capacitive/resistive screens.
- Problem that solves: It simplifies the user interface, making it more intuitive, compact and robust.
- Current state and prospects: Being exploited. They are very popular as the input interface of PDAs and POS terminals. Lately, they are being installed in mobile phones as well.

Multitouch screens

- Overview: They are used in touch screens and touch pads, in order to recognize several contact points simultaneously. It makes easier the direct handling of some systems by means of recognizing intuitive gestures performed with some fingers. The most popular technology for implementing these interfaces is the so called Frustrated Total Internal Reflection.
- Problem that solves: It provides a control interface much more intuitive and natural, reducing the required cognitive load of the user, and enhancing the speed of interaction.
- Current state and prospects: It is an emerging technology, although there already exist some successful commercial solutions, as the Apple IPhone.

3D Movement tracker

- Overview: A 3D tracker is a device that informs about its location and orientation with respect to a reference point. These devices are the basis for the virtual and augmented reality systems, and some video game systems.
- Problem that solves: They are really useful for developing adapted interfaces for people with movement disabilities, and surveillance applications.
- Current state and prospects: Being exploited. The most popular device implementing this technology is the Wiimote from Nintendo's Wii console.

Sight tracker

- Overview: It registers in real time the eye movement. The most used technique nowadays is by using a video camera and image processing, in order to identify where the user is looking to. One of its most known applications is the control of the pointer in a screen by tracking the users' sight. There is being researched also multimodal displays that depend on the users' sight, the so called gaze-continent interfaces.
- Problem that solves: Since it is a natural and immediate way of communicating, it requires very low manual effort and cognitive load. For this, it is very interesting for people with motor disabilities.
- Current state and prospects: Emerging. Currently it is being researched how to extract additional information from the users' sight, and how to filter subconscious movements. There are commercial devices but they are not affordable for the vast majority, and they are not accurate and usable enough for being introduced in the market.

3.1.1.3. Data input interfaces

Voice recognition

- Overview: Voice recognition is the process of converting digitalized voice, captured with a microphone, in a set of phonemes. After it, a grammatical process allows to identify words by taking the flow of received phonemes and identifying them by comparison with a predefined set of words. These words can be used for giving commands to the computer, inserting information in a database, or dictating to a text processor.

- Problem that solves: It is a natural interface that can substitute the usual ones in a computer (keyboard and mouse). It can solve some interaction problems for people who are not used to deal with computers, or who have some motor disabilities

- Current state and prospects: Being exploited. The command recognition uses to require a previous training and to be in a not noisy environment.

Handwriting recognition

- Overview: This technology is used for recognising intelligible manual writing with a special kind of pen or pointer.

- Problem that solves: It is an alternative system for data entry.

- Current state and prospects: Being exploited. Although it does not substitute the keyboard as data input device, it is very useful for small devices that cannot integrate a keyboard, and also in some video game consoles as Nintendo's DS Lite

Voice synthesis

- Overview: Voice synthesis is the process of artificially generating of speech. The employed technique depends on the application requirements. It can be based on recorded samples or in a Text To Speech (TTS) system that generates voice by concatenating some phonetic symbols.

- Problem that solves: It can be used for answering questions, notifying events, and reading documents. It solves some problems of interaction with the machine, for people with visual disabilities. The natural interaction adds an important affective component to this interface.

- Current state and prospects: It is being exploited. As an example of a successful voice synthesis commercial system, we can mention Loquendo.

3.1.2 User study: Adaptation to the different profiles

3.1.2.1. User Profiling

The User Profiling functionality is in charge of the characterization of each user based on a selection of his/her known information. This data consists of the personal parameters and the measures taken from the home devices and sensors which are stored in an updatable database and which have been demonstrated to be significant for the objective of characterization. User Profiling will require an initial training where a division of patients in groups is generated from the most important features. The creation of groups will be done automatically by means of clustering techniques and will be afterwards revised by an expert. Each group will have an associated general profile which characterizes its members. The User Profiling can be considered as a user modeller which determines the profile of a patient assigning him to the created groups of users with similar characteristics. The users of the same group will be treated in a similar way and some of the recommendations will be the same. New patients can be included in the already identified profiles and, in addition, current users with new information and modifications throughout time because of the progress of their disease can automatically change their characterization and, therefore, some of the resulting actions of the AmIE system. The resulting groups and profiles will be later used in the knowledge rule bases included in the reasoning engines of other components, like the Adaptive System Intelligence, in order to optimize their size and performance.

3.1.2.2. Adaptive System Intelligence

The objective of **Adaptive System Intelligence** is to give automatic answers according to the existing context: profile of the patient, specific events, etc. The answers include recommendations, communications, alarms... It interacts with the user whenever it concludes that is needed and it will have the assistance of the **Adaptive User Interaction** in order to communicate the answer in a personalized mode. The Adaptive System Intelligence functionality takes advantage of the dynamic user profiles and complements them with real-time information about environment and historical information. These will make possible the generation of predictions and the application of the rules defined in the reasoning engine. The main part of the Adaptive System Intelligence is the reasoning engine which accesses the knowledge rule base and the working memory. The first one is composed of inference rules specified by means of an ontology language and proposed by experts including significant actions to be taken when necessary. The second one stores the current facts, assertions and events about patients and their environment. The engine determines which rules are relevant and choose which one(s) to apply: the ones that have antecedents satisfied by the past and present users'

and events' information. Adaptive System Intelligence is also in charge of the predictions based in the medical and behavioural history of each user and their present behaviour and health status. It will be capable of detecting trends and thus give information about future problems of the elderly person.

3.1.2.3. Context management

The most simple definition of context is the one of the philosopher Martin Heidegger "Context isn't something that describes a setting; it's something that people do, the horizon within which the user makes sense of the world" (Heidegger, 1927).

The networked home system shall provide innovative application services to end-users. Such services shall in particular account for the user's situation, according to both the technological environment and the user's will. This issue is known as context-awareness, which should be dealt with at the middleware layer, regarding both context management and realization of middleware functions.

The term 'context' is overloaded with a wide variety of meanings, depending on the purpose of the particular application and/or the specific research community standpoint. In this project, we adopt the following general definition of context, extending it to include device-to-device communication:

"Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves"(Dey, Abowd & Salber, 2001:97).

The following categories of context are defined:

- Device context: contextual information related to devices. Examples are available memory, computation power, networks (and their quality), codecs, etc.
- User context: context information that describes a person, further decomposing into:
    - Personal context: health, mood, schedule, activity, etc.,
    - Application context: email received, Web sites visited, preferences, etc.
    - Social contexts: group activity, social relationship, people nearby etc.
- Physical context: contextual information related to the physical environment of an entity (device, room, building, user). Examples are location, time, weather, altitude, light.

The role of the a Context Manager is to acquire information coming from various sources, ranging from physical sensors to Internet applications, combine these pieces of information into "context information", and make this context information available to other services so as to enable these services to become context-aware. Applications may then be context-aware, as they can get contextual information from the context manager and use it for their specific application purposes.

3.1.2.4. Data Mining and Automated learning

Data mining is the process of extraction of information from a certain quantity of data. Information is analyzed from many points of view to get knowledge and relations between data. Data by itself may not contribute a lot, but data-mining techniques helps us to obtain specific information about our interests.

The field of Data mining appeared a lot of years ago and slowly, it has been developed and researched to become one of the most useful technologies nowadays, due to the high number of different applications that it can have: bank analysis, science, genetic, medicine, insurances, trades...

Other important secret for data mining's success is the number of techniques to be applied. Some of them are specially designed for getting a high accuracy, some to obtain the system behavior, some to group or cluster and so on. It is necessary to have an expert to determine the most interesting algorithm to apply in every situation, in the same way that an expert in the concrete field will be needed to work in cooperation.

The exploitation of data by learning methods supposes an analysis of the history of context information, collected along the time by sensors and applications in the environment.

Since context data might be represented in RDF, several techniques from the graph data mining domain can be applied to detect similarities or to find recurrent schemes. These require persistent storage and querying methods. Currently, there exist several projects addressing RDF storage problems and many solutions are already available that can be used or adapted.

Then, it is possible to discover relations between relevant contexts and applications or services using a second stage of learning methods, such as reinforcement algorithms. These will lead to the construction of adaptation models.

There are three types of approaches for learning contextual data according to their representation: numerical, symbolic and structural.

The numerical learning of the context handles only quantitative values to build contextual states. PlaceLab [44] is a home lab fully equipped with sensors. The data of more than 900 sensors (interrupter, indicator, detector, RFID) are annotated and analyzed by two classification algorithms: decision tree and naive Bayes to detect activities (folding linen, brushing teeth, etc.) In [48][49], time series data produced by sensors (accelerometer, microphone, etc) are used for the detection of situations while using a mobile device, for example (walking, running, etc).

Symbolic learning of context seeks to find relationship between contextual states in order to construct higher level context. The main advantage of this method is its semantic contribution. The high level context expresses interpretations of what had happened in the ambient environments. For example, considering the following contextual states (sitting, TV-on, in the living room), one can infer that the person is watching TV. A high level context can also results from temporal series of contextual states for example (TV-off, standing up, walking, opening the entry door) means that the person is leaving. In [47], symbols are associated to contextual states. Obtaining high level context is obtained by two methods: multi-source fusion and temporal fusion. Multi-source fusion consists of generating high level context starting from symbols coming from several sources at a given moment. Temporal fusion is a sequential fusion of the symbols. The algorithm used is called "Symbol Clustering Map (SCM [45]").

Structural learning of context differs from symbolic learning of context from the fact that contextual states are structural information, which is semantically richer. The goal is to construct high level context by using learning methods for structural information.

In the literature, learning from complex data was addressed for two potential problems: ontology learning as in [50][51] and graph learning as in [46]. Graph learning is a very active research domain. It gathers several methods of classification and grouping. Also, other techniques such as grammar learning or ILP (inductive logic programming) can be used.

MIDAS structure has been designed to study the user and help him in daily life. The platform will have a high amount of data, collected every day, about the behavior and actuation of the elderly or disabled people. Patterns and future trends will be extracted to determine the way to act in a dynamic system.

Thus, if user has not gone outside in last four days, system will detect it and will act in consequence, for instance giving the patient an advice or informing to his relatives.

3.1.2.4.1. Data Organization

Internet, computers, video surveillance, globalization… actual situation is favorable to have big amounts of data. Size of database and the number of queries and accesses to be done will establish the technical requirements of storage.

Traditional databases are not so useful in big situations, so data warehouses offer interesting solutions according to integrate all the information in it. A central repository will organize data and will facilitate the access from users to have a global vision of the situation, adding flexibility and scalability. This technology includes On-Line Analytical Processing (OLAP) operations which can describe and summarize but do not obtain knowledge, as we do with data mining techniques.

Apart from that, some other databases can be useful depending on the kind of problem we are working with: spatial databases (with geographic information about cities, images…), temporally databases (for intervals of time), documentary databases (which save descriptions for text documents) and multimedia databases (with audio, video and images, with big file sizes).

3.1.2.4.2. Software

One of the main advantages of data mining is the simplicity and speed to apply some techniques to real problems, thanks to the software existing which have a lot of algorithms and different techniques implemented. Two main groups are created according to the source of the system: commercial or academic, but we distinguish here three kinds of tools according to their use.

- **Libraries:** Libraries in data mining assert methods and functions specifically for data mining, which define the models to apply. *Xelopes* and *MLC++* are the most used libraries by programmers nowadays.
- **Suites:** They collect techniques for data processing, analysis and graphic support so they can be used without programming knowledge (thanks to their interfaces). SPSS Clementine is multi-platform, commercialize by SPSS and especially interesting because of their functionalities. On the other hand, *WEKA* represents the open source option, created by the University of Waikato (New Zealand). Also, *DBMiner*, *YALE* or *ODMS* are other options in data-mining.
- **Specific tools:** These tools are specialized for one model; they are not as generic as suites are. It is no necessary to have programming knowledge. *CART*, *AutoClass*, *NeuroShell* or *See5/C5.0* are good solutions in this field.

3.1.3 System intelligence framework

3.1.3.1. System Based in rules

Humans have the ability to process stimuli and decide on the appropriate response based on the knowledge they have (Figure 1). However, for software systems to simulate that behaviour, knowledge has to be converted into a computer understandable form.

A classic approach to the problem is to implement the rules directly in the code using logical conditions in the control of flow. One problem of hard coded rules is their lack of flexibility and scalability. Most of the time, the rules cannot be reutilized as the logic code which implements the rules is specific for a concrete project. Talking about efficiency, there are some disadvantages of the rules directly coded in the source; first of all, there are bigger costs while making changes and building the project as there are more lines of code. Another disadvantage is the apparition of delays in execution caused by the sequential code and flow control that continually re-evaluates the rules.

Instead of representing knowledge in a relatively declarative, static way, rule-based systems represent knowledge in terms of a set of rules that tell you what you should do or what you could conclude in different situations.

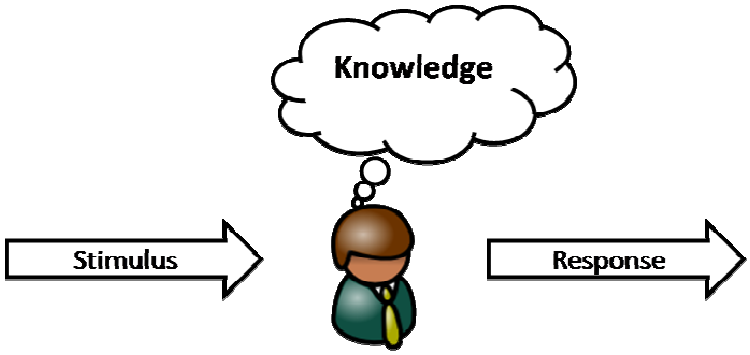

Figure 3-1 The way humans respond to stimuli (Adopted from [34])

For application systems of great scale a rule-based approach to “injecting” and implementing knowledge in a system has many advantages (as opposed to a procedural approach). The "rule" is really a data structure that sometimes it is referred to as having a head and a body, sometimes a predicate and an action or more simply a conditional if… then… structure [34, 33]. Figure 2 demonstrates the main advantages of the rule-based approach:

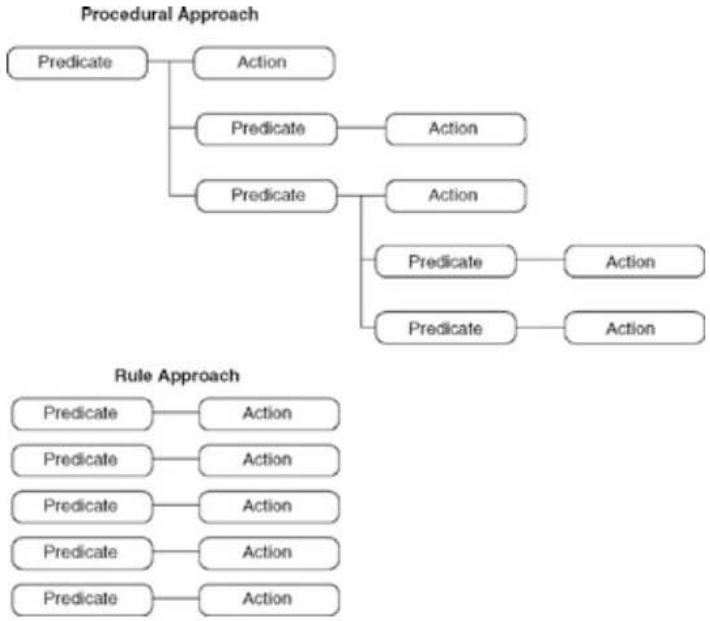

Figure 3-2 Procedural versus Rule-based approach to application creation [34].

Business Rule Systems use rules defined outside the code; this increases the flexibility and reduces the costs of including changes in the application. The language used to define the rules is specific and adapted to this issue, so it is easier to learn and others than programmers can implement new rules. In fact, there are even graphic editors that simplify the edition process. Another enhanced feature of Business Rules Management Systems is the speed of rules execution as they are not considered in a sequential way but like a set of nodes forming a network. This algorithm is called RETE and is the basis of many BRMS implementations like Drools, Jess, BizTalk or CLIPS. The main disadvantage of RETE algorithm is the increase of memory use comparing to hard coded rules systems.

There are two broad kinds of rule system [3]: *forward chaining* systems, and *backward chaining* systems. In a forward chaining system you start with the initial facts, and keep using the rules to draw new conclusions (or take certain actions) given those facts. In a backward chaining system you start with some hypothesis (or goal) you are trying to prove, and keep looking for rules that would allow you to conclude that hypothesis, perhaps setting new sub-goals to prove as you go. Forward chaining systems are primarily data-driven, while backward chaining systems are goal-driven.

Whether to use forward or backwards reasoning to solve a problem depends on the properties of your rule set and initial facts [4]. Sometimes, if you have some particular goal (to test some hypothesis), then backward chaining will be much more efficient, as you avoid drawing conclusions from irrelevant facts. However, sometimes backward chaining can be very wasteful - there may be many possible ways of trying to prove something, and you may have to try almost all of them before

you find one that works. Forward chaining may be better if you have lots of things you want to prove (or if you just want to find out in general what new facts are true); when you have a small set of initial facts; and when there tend to be lots of different rules which allow you to draw the same conclusion. Backward chaining may be better if you are trying to prove a single fact, given a large set of initial facts, and where, if you used forward chaining, many rules would be eligible to fire in any cycle.

To verify the conditions of the rules and manage the execution of the programmed actions, a rule engine is employed.

Rule engines are software that provides the ability to register, define, classify, and manage rules. In addition, this kind of tools helps to verify consistency of rule definitions and link the execution of rules to the software application.

Rule engines can follow the “Recognise-Act” style, which involves a cycle of three stages to process rules:

- **Match rules:** The engine finds the rules which are satisfied/not satisfied depending on the data of interest.
- **Select rules:** At this stage the engine determines which rules are more relevant to the particular context and decide upon their suitability for execution.
- **Execute rules:** The rule causes some reaction which can result in changing data or interfacing with the end user.

The information employed to determine which rule should be executed is stored in the working memory (Figure 3). This information evolves through the time having different values then the rule set is executed the information in the working memory is executed to determine which conditions are fulfilled.

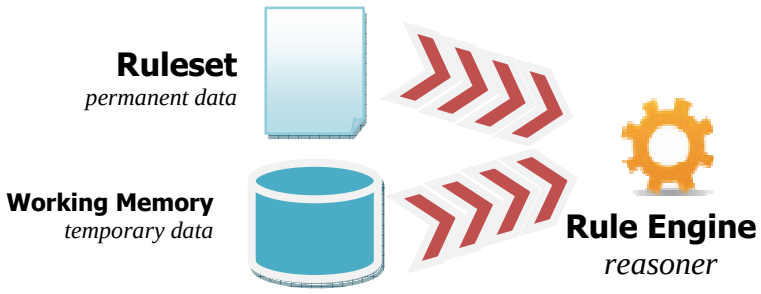

Figure 3-3 Rule Based System main components.

Although many different techniques have emerged for organizing collections of rules, all rule-based systems share certain key properties:

- They incorporate practical human knowledge in conditional if-then rules,
- Their "skill" increases at a rate proportional to the enlargement of their knowledge bases,
- They can solve a wide range of possibly complex problems by selecting relevant rules and then combining the results in appropriate ways,
- They adaptively determine the best sequence of rules to execute, and
- They explain their conclusions by retracing their actual lines of reasoning and translating the logic of each rule employed into natural language. [33]

The Business Rules Group is actively researching the application of rules in business environments. They generically define "business rules" as statements which define or constrain some aspect of the business. Despite the business-driven approach of the BRG it is clear that ultimately the bulk of rules will have to be used as an automatic control mechanism for the underlying information system. Therefore the principles defined by the BRG for the categorisation and definition of rules can hold for any rule-based information system.

According to the BRG [37] business rules fall into one of four categories:

- **Definitions of business terms:** The most basic element of a business rule is the language used to express it. The very definition of a term is itself a business rule that describes how people think and talk about things. Thus, defining a term is establishing a category of business rule. Terms have traditionally been documented in glossaries or as entities in a conceptual model.
- **Facts relating terms to each other:** The nature or operating structure of an organization can be described in terms of the facts that relate terms to each other. To say that a customer can place an order is a business rule. Facts can be documented as natural language sentences or as relationships, attributes, and generalization structures in a graphical model.
- **Constraints:** Every enterprise constrains behaviour in some way, and this is closely related to constraints on what data may or may not be updated. To prevent a record from being made is, in many cases, to prevent an action from taking place.
- **Derivations:** Business rules define how knowledge in one form may be transformed into other knowledge, possibly in a different form.

3.1.3.2. State of the art in rule languages

3.1.3.2.1. Rule Markup Language (RuleML)

Rule Markup Language (RuleML) [5] is a markup language developed to express both forward (bottom-up) and backward (top-down) rules in XML for deduction, rewriting, and further inferential-transformational tasks. It is defined by the Rule Markup Initiative, an open network of individuals and groups from both industry and academia that was formed to develop a canonical Web language for rules using XML markup and transformations from and to other rule standards/systems.

RuleML covers derivation, transformation and reaction rules, so it can specify queries and inferences in Web ontologies, mapping between Web ontologies, and dynamic Web behaviors of workflows, services and agents.

RuleML implementation efforts are focused on the “axioms” which can be used with the taxonomy created to form an ontology structure. Defining a uniform ontology language with description-logic taxonomy and Horn-logic-like rules is required since large-scale taxonomies require a rule system to use certain implicit information that is not defined by the taxonomy alone.

RuleML is integrated within the Semantic Web Rule Language (SWRL), combining sublanguages of the OWL Web Ontology language (OWL DL and Lite) with RuleML Datalog. However, the expressive power provided by SWRL comes with the price of decidability, and hence of practical, sound and complete reasoning, which is a critical requirement for the real-world applications that SWRL and RuleML aims to implement. [6]

3.1.3.2.2. Semantic Web Rule Language (SWRL)

Semantic Web Rule Language (SWRL) [7] is a proposal for a Semantic Web rules-language, combining sublanguages of the OWL Web Ontology Language (OWL DL and Lite) with those of the Rule Markup Language (Unary/Binary Datalog).

Compared with DLP (Description Logic Programs), another relatively recent proposal in the Semantic Web community for integrating rules and OWL, SWRL takes a diametrically opposed integration approach. DLP is the intersection of Horn logic and OWL, whereas SWRL is (roughly) the union of them. In DLP, the resultant language is a very peculiar looking description logic and rather inexpressive language overall. It is hard to see the restrictions are either natural or satisfying.

Contrariwise, SWRL retains the full power of OWL DL, but at the price of decidability and practical implementations.

Rules in SWRL are of the form of an implication between an antecedent (body) and consequent (head). The intended meaning can be read as: whenever the conditions specified in the antecedent hold, then the conditions specified in the consequent must also hold.

3.1.3.2.3. Jena Rules

Jena is a Java framework for building Semantic Web applications. It provides a programmatic environment for RDF, RDFS and OWL, SPARQL and includes a rule-based inference engine [8].

Jena is open source and grown out of work with the HP Labs Semantic Web Programme.

The Jena Framework includes:

- A RDF API
- Reading and writing RDF in RDF/XML, N3 and N-Triples
- An OWL API
- In-memory and persistent storage
- SPARQL query engine

JenaRules is based on RDF(S) and uses triple representation of RDF descriptions (also known as N3 Notation and Turtle syntax). The abstract syntax of JenaRules can be found at [9].However Jena deals with data sources represented in RDF and OWL which is expressed in XML syntax, JenaRules provides a textual syntax for defining rules. When dealing with computerized data represented as XML, this may require translators between different languages.

3.1.3.2.4. Drools

Drools is a business rule management system (BRMS) with a forward chaining inference based rules engine, more correctly known as a production rule system, using an enhanced implementation of the Rete algorithm.

Drools is based on the JSR-94 standard for business rules engine and enterprise framework for the construction, maintenance, and enforcement of business policies in an organization, application, or service.

Drools lets you express your business logic rules in a declarative way. You can write rules using a non-XML native language that is quite easy to learn and understand. Drools also have other advantages. It is:

- Supported by an active community
- Easy to use
- Quick to execute
- Gaining popularity among Java developers
- Compliant with the Java Rule Engine API (JSR 94)
- Free

3.1.3.2.5. Prolog

Prolog is a logic programming language. It is a general purpose language often associated with artificial intelligence and computational linguistics. It has a purely logical subset, called "pure Prolog", as well as a number of extralogical features.

Having its roots in formal logic, and unlike many other programming languages, Prolog is declarative: The program logic is expressed in terms of relations, and execution is triggered by running queries over these relations.

The language was first conceived by a group around Alain Colmerauer in Marseille, France, in the early 1970s, while the first compiler was written by David H. D. Warren in Edinburgh, Scotland. Prolog was one of the first logic programming languages, and remains among the most popular such languages today, with many free and commercial implementations available. While initially aimed at natural language processing, the language has since then stretched far into other areas like theorem proving, expert systems, games, automated answering systems, ontologies and sophisticated control systems, and modern Prolog environments support the creation of graphical user interfaces, as well as administrative and networked applications.

3.1.3.2.6. R2ML – The Reverse I1 Markup Language

R2ML is a XML rule format developed by REWERSE [11] that allows interchanging rules between different systems and tools, enriching ontologies by rules and connecting another rule system with R2ML based tools for visualization, verbalization, verification and validation.

R2ML integrates

- Object Constraint Language (OCL)
- Semantic Web Rule Language (SWRL)
- Rule Markup Language (RuleML)

and includes four rule categories: derivation, production, integrity and ECA/reaction rules. [12] R2ML can be used with Strelka [13], an URML-based visual rule modeling tool, to model the rules visually.

3.1.3.3. Ontology

An ontology is a shared and common understanding of a domain that can be communicated between people and application systems [35]. It is a formal and agreed conceptualisation (a model) of that particular domain represented as objects and the roles played between them, using a predefined vocabulary. Formality arises from the fact those relations can be defined by axioms stated in a formal logic, or a language which can be translated to logic. Agreement implies that the terms in the vocabulary used to represent the concepts and relations have an agreed meaning and structure, at least between the stakeholders in the domain of interest.

The uses of ontologies range from knowledge representation, to enterprise integration, database design, natural language translation and many more [31]. An in-depth investigation on ontologies and their applications is beyond the scope of this document and therefore the user is referred to the seminal work of Ushold and Gruninger [29] and also [31], [35].

Ontologies are the result of an engineering process, involving the elicitation of facts about a domain, their analysis and manipulation to create the domain model and their commitment to applications. Depending on the intended use of ontology, the ontology engineer can choose between various representation methods to model the domain. To facilitate the communication between humans (e.g. through the disambiguation of terms between technical and non-technical people), a more graphical and natural language approach may be necessary. If that domain knowledge also needs to be processed by software agents (e.g. in the case of a semantic search engine or a data entry form), a more formal language must be used, represented in a computer-understandable notation, which will inevitably reduce expressiveness and human readability. At the other extreme, for communication between software agents, the domain knowledge has to be shared in a formal language, with a strict semantics. This format is usually not meant to be processed by human readers.

3.1.4 Medical related technologies

The aim of the project is to allow **real-time and asynchronous monitoring of patients**. Therefore, it is necessary to collect information from patients through sensors attached to their bodies or from the measuring devices used and send this information to medical control center. In addition, it is necessary to build a communication system that will let these transactions of the patient's vital data. The system has to be capable of working both if the patient is at home and out of home, therefore it must support **fixed and on the move technologies**. The key equipments involved in the communication system are the following:

3.1.4.1. Bio-medical sensors and Bluetooth

Today's mobile telecommunications infrastructure makes it possible to take medical monitoring technologies developed for space and military applications to the consumer market. This means that more and more user friendly and cheap technologies are available at present allowing the patient himself to be involved in the healthcare process.

The biomedical device field has changed (and is changing) rapidly, introducing also a telemedicine product line, which means products (of course coming from different manufacturers), able to acquire and transfer different medical data coming from sensors directly applied to the patient. Even if the use of these systems was at the beginning for professional applications (Hospitals), at present technologies become also available for personal and home use. The industrial development of this kind of sensors, of course, has been supported by the same development in terms of standardization at European and worldwide level. The combination of medical sensors with communication capability, found its maximum fulfillment with the introduction of wireless communication and in particular introducing the use of Bluetooth technologies. Hundreds of millions of people worldwide who suffer with chronic illnesses, who are considered as person at risk or elderly, should benefit of Telemedicine applications for patient monitoring at home should benefit from Bluetooth technology for the following reasons:

- Compliance with a global standard in wireless communication.
- Wide selection of chip devices for different applications.
- Secured communication – built-in encryption and frequency hopping ensures data security and integrity.
- Flexibility – a variety of Access Points to choose from PDA, mobile phone, PC, etc.

There are many devices available on the market enabling monitoring (active or passive) of parameters. There is actually quite an extensive diversity of available device from those, which are the

most simple and have only some memory to keep previous results to those enabling wireless transmission of the test results in real-time mode to personal or home-based terminal. Because of its worldwide availability, existing pervasiveness in mobile phones and laptops along with low power, low cost and security features, Bluetooth technology is ideally suited for medical devices that will provide a better quality of life for patients while reducing the cost of healthcare.

Biomedical sensors examples

Oxymeter

A pulse oximeter is a medical device that indirectly measures the oxygen saturation of a patient's blood. This device displays the percentage of arterial haemoglobin in the oxyhemoglobin configuration. Because of their simplicity and speed (they clip onto a finger and display results within a few seconds), pulse oximeters are of critical importance in emergency medicine and are also very useful for patients with respiratory or cardiac problems, or for diagnosis of some sleep disorders such as apnea and hypopnea.

Blood pressure device

This device is used to measure patient’s blood pressure. Blood pressure refers to the force exerted by circulating blood on the walls of blood vessels, and constitutes one of the principal vital signs. The systolic arterial pressure is defined as the peak pressure in the arteries, which occurs near the beginning of the cardiac cycle; the diastolic arterial pressure is the lowest pressure (at the resting phase of the cardiac cycle).

Weight scale

It measures patient’s weight.

Electrocardiogram

An electrocardiogram (ECG or EKG, abbreviated from the German Elektrokardiogramm) is a graphic produced by an electrocardiograph, which records the electrical activity of the heart over time.

Glucose sensor

It gives the glucose rate.

Biomedical sensors manufacturers

RDSM, Corscience; NONIN, IEM…..

Current limitations and main innovation

The raw data captured by the sensors must be processed in order to calculate the medically relevant vital parameters. Therefore, signal processing system with a high performance and Bluetooth communication capabilities is needed. This should be directly integrated into the wearable device to minimize communication overhead by data reduction. Very low power consumption has to be achieved, so that a high degree of mobility can be ensured. Although Bluetooth enabled medical devices currently exist, the method of Bluetooth communications used is proprietary and therefore not usually interoperable with devices from other manufacturers. To become a central component of the way we manage health, personal health and medical devices must be fully interoperable with each other and with other information sources. Because broad interoperability has yet to be achieved, we will work on emerging communication standard such as those developed by the Continua Health Alliance, a powerful initiative of the most important companies in the health and IT sector (http://www.continuaalliance.org/home).

3.1.4.2. Home gateway

A Home gateway has to be a dedicated homecare unit, synchronizing with the biomedical sensors.

When patients are at home, it must ensure data transmission to the **control centre using terrestrial networks**. Videoconferencing **for live interaction** with medical professionals, and educational multimedia health content delivery bring added-value capabilities to the patient, and should be considered during the system definition process.

The Home gateway specifications will address:

- Operating system, Hardware design
- Interface with terrestrial telecommunication, network, indoor environment and complex environment. Given the variety of Access Network technologies, the Home Gateway Architecture has to be flexible in terms of providing the support for different WAN side interfaces types, although there will only be one WAN connection at a time,
- Secured Data transfer (VPN, SSL, X509),

- Identification to the Electronic Patient Record using a patient unique identifier, as defined by IHE initiative recommendations,
- Automatic routing mechanism for live transmission of data coming from the portable device,
- Push mechanism to receive multimedia contents from the Expert Centre,
- Content management mechanism to exploit educative/preventive contents on patient's TV,
- Videoconferencing…

Integration of the Home gateway in the end to end system will require study of crucial performances and networking aspects like:

- Required communication standards.
- Quality of service.
- Security aspects.

3.1.4.3. Telemedicine technologies and standards

Five principal standardization organisations, a promising open source initiative and a major interoperability initiative have been selected for detailed analysis in this survey as they can be expected to play a leading role in further telemedicine standards development:

- **ISO**, the International Organisation for Standardisation, as the largest developer of world-wide standards,
- **CEN**, the European Committee for Standardisation as the principal SDO in Europe,
- **IHTSDO**, the International Health Terminology SDO, as the developer of the fairly widely adopted SNOMED-CT terminology standard,
- **HL7**, Health Level 7, as the developer of the most widely used standards for electronic messages in healthcare,
- **DICOM**, Digital Imaging and Communications in Medicine, as a de facto standard for electronic medical imaging,
- **OpenEHR** as a promising open source activity for electronic health records,
- **IHE**, Integrating the Healthcare Enterprise, as a major e-health systems interoperability initiative.

Organisation name	Acronym	Domain	Principal e-health standards developed
International Standardisation Organisation	ISO	General standards development	ISO/TR 18307

Organisation name	Acronym	Domain	Principal e-health standards developed
European Committee for Standardisation	CEN	General standards development	ENV 13606 (parts 1-5), HISA
International Health Terminology Standards Development Organisation	IHTSDO	Terminology	SNOMED
Health Level 7	HL7	Communication and architecture	HL7 v2.x, HL7 v3.0, CDA, RIM, CCOW
Digital Imaging and Communications in Medicine	DICOM	Imaging	DICOM
openEHR	openEHR	EHR architecture	openEHR
Integrating the Healthcare Enterprise	IHE	Standards frameworks	Integration profiles

HL7,

Health Level Seven (HL7) is a not-for-profit, multi-national standards development organisation with headquarters in Ann Arbor, Michigan, US. It is accredited at the American National Standards Institute. Founded in 1987, HL7 specialises in standards development for clinical and administrative data. The number 7 stands for the highest level in the Open Systems Interconnection reference model for implementing computer protocols, the "application level"[24]. The mission of HL7 is to "create standards for the exchange, management and integration of electronic healthcare information" [25]. This includes the aim to promote "the use of such standards within and among healthcare organisations to increase the effectiveness and efficiency of healthcare delivery" [25]. HL7 has national bureaus in 29 countries across the world and is continuously growing.

DICOM

Digital Imaging and Communications in Medicine (DICOM), or the DICOM Standards Committee, was established in 1993 and has its headquarters in Rosslyn, US. The Diagnostic Imaging and Therapy Systems Division of the US National Electrical Manufacturers Association [26] are responsible for the development, maintenance, and governance of the DICOM standard. Many countries have local DICOM subsidiaries or national representatives. The objective of DICOM is to "ensure the interoperability of systems used to: produce, store, display, process, send, retrieve, query or print medical images and derived structured documents as well as to manage related workflow."[27]

ISO

The International Organisation for Standardisation (ISO) is the world's largest developer of international standards [28]. It was founded in 1947 with a Central Secretariat in Geneva, Switzerland, and had 153 employees at the end of 2007. ISO is a network of the national standards institutes of 157 countries. ISO standards are being developed for a wide range of activities in areas such as manufacturing, trade, legislation, innovation, and consumer protection. Thus telemedecine standards are only part of ISO's work.

CEN

Founded in 1961, the European Committee for Standardisation (Comité Européen de Normalisation, CEN) is a not-for-profit organisation headquartered in Brussels. It contributes to the objectives of the European Union and European Economic Area with voluntary technical standards which promote, among other items, interoperability of networks [36]. The objective of CEN is to "foster the European economy in global trading, the welfare of European citizens and the environment" [37]. ICT standards in the health sector are only one part of CEN's work.

CEN, CENELEC and ETSI together constitute the "European standardisation system". While CENELEC and ETSI were also included in the EC's e-health mandate issued in March 2007,1 they deal with standards for medical devices that are not subject of this report. Thus CENELEC and ETSI are not described in detail here.

IHTSDO

The International Health Terminology Standards Development Organisation (IHTSDO) was established in 2006 with a main office in Copenhagen, Denmark [38]. It is responsible for the ongoing development, maintenance and governance of a standard named "Systematized Nomenclature of Medicine – Clinical Terms" (SNOMED-CT) as well as of other healthcare terminology standards. IHTSDO promotes and enables the uptake and correct use of SNOMED-CT in health systems, services and products around the world. Before 2006, the standard was owned and developed by the College of American Pathologists and the UK's National Health Service (NHS). The decision to form IHTSDO was made to allow "other countries the opportunity to take a leading role in the ownership, development, maintenance, and promotion of the SNOMED-CT clinical terminology" [39].

[1] See section 2.3.2 for a brief description of this mandate.

SNOMED-CT seeks to ensure that clinical staff has consistent and easy to understand information about a patient's medical history, illnesses, treatments, and test results immediately available. It seeks to provide a single and comprehensive system of terms, centrally maintained and updated for use in all National Health Service organisations as well as in research. The standard is meant to improve the communication consistency of patients' clinical records.

OpenEHR

OpenEHR is a not-for-profit foundation, which was formed in 2002 by the University College London (UCL) and the Ocean Informatics company after their collaborative work on the Good European Health Record (GEHR) project. OpenEHR is not a formal SDO as it does not have balloting and consultation processes implemented like the other organisations described in the preceding sections. The aim of openEHR is to make EHRs “adaptable and future-proof”[40] through the use of a technology independent architecture. openEHR seeks “to improve the clinical care process by fostering the development and implementation of open source, interoperable EHR components. These components should be based on internationally agreed requirements and address the need for privacy and security, while supporting the development of interoperable and evolving clinical applications”.[41] The open source aspect of openEHR methodology refers to the development characteristic of “distributed peer review and transparency of process”.[42]

IHE

Integrating the Healthcare Enterprise (IHE) is “an initiative by healthcare professionals and industry to improve the way computer systems in healthcare share information”.[43] IHE’s objective is to facilitate interoperability of healthcare ICT. It promotes “the coordinated use of established standards such as DICOM and HL7 to address specific clinical needs in support of optimal patient care”.[43] IHE thus does not develop standards itself but it provides a framework for the adoption of certain standards.

IHE was created in 1998 through the leadership of the Healthcare Information and Management Systems Society (HIMSS) and the Radiology Society of North America (RSNA). HIMSS and RSNA together with the American College of Cardiology (ACC) are the principal sponsors of IHE. The organisation thus has its origin and main pillar in the US but there is also strong support from Europe and Japan.

3.1.5 Video Analytics Technology

Video Analytics (Video Analysis) is a new and growing market. The first objective of VideoAnalytics for video surveillance is to make easy the work of video surveillant. A large amount of information is provided by video. It is an advantage but all these data are very difficult to treat by a human being. Consequently, all can be done to automatise the video analyse. The Midas objective is to recognise and to track dependent person at home to prevent health analyses or to propose them more adapted services. Located in a camera or in a box (developed by Orange Labs), the system analyses the video to transcript it in semantic data (XML technology).

3.1.5.1. Video Analysis:

Analytics Video company are usually issued from research like INRIA or CNRS in France like Keeneo or FoxStream or from Defense Research like Object Video the leading company financed by the Defense Advanced Research Projects Agency (DARPA) or like Mate from Israel. Through advanced computer vision science, these company bring an unmatched set of capabilities to operational challenges, including those in critical infrastructure, retail, banking, transportation and gaming. These softwares enable automatic detection of a wide range of events and activities, such as perimeter breaches, loitering, unauthorized entry/exit, and theft of items. Our main objective for Midas Project is to detect and define how this technology could be used for elderly at home.

What *different types* of video analytic technologies are available?

Video analytics can be real-time: using technology configured to track and provide alerts to specific incidents as they happen; or post event: retrospectively searching for incidents that have already occurred.

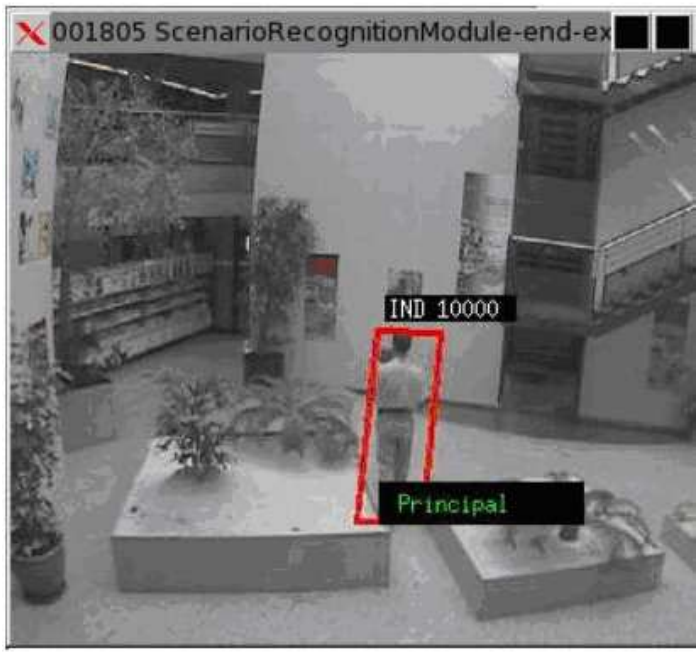

Figure 3-4 Video analysis example to track people

Face recognition: This technology first detects the face and can recognise afterwards the person using characteristics from a database. It is not our main objective. It will be a mean to recognize persons to be monitored. This technology should be implemented at the end of MIDAS project.

Behaviour recognition: This technology can recognize a person from the frame background. The main company works in 2 dimensions losing the field depth considering people far-away on the same plan. Few technologies are proposed in which persons are located in 3 dimensions relatively to the room configuration. We think that a 3 dimensions calculation are necessary to make the difference to detect a person and to measure his activity or non activity for preventing or alerting.

Other technologies: In the domain of behaviour recognition, a zone can be forbidden near the windows or the doors related to a vocal alert. Furthermore, important information can be detected related to the movement of a person. For instance, the information of how a person is standing from a seated position on a chair is very important. Elderly have to learn how to get up from a chair safely to prevent fall injuries. With the aging process, we lose strength in our hip and knee extensors (the muscles that help straighten our legs). We rely on these muscles to get up from a chair. For this reason, the elderly have difficulty standing from the seated position. Therefore, this information is very important to know the health evolution of a person. INRIA is working on this objective. This technology will be provided in the near future.

3.1.5.2. SmartCam:

We named smartcam, cameras which embed the video analytics software in order to fulfil the frame analyse and to send only the main information (position, alert…). The goal is twice. First, we want to reduce wires and informatics clutter: only the smartcam will appear at home. Second, the objective is to make easy the apparatus installation.

Video surveillance systems have existed for some 25 years, starting out as 100% analog systems and gradually becoming digitized. Today's systems have come a long way from the early analog tube cameras connected to a VCR. They now use network cameras and PC servers for video recording in a fully digitized system. In the past year, the number of camera manufacturers and products has increased dramatically, with new camera designs addressing the security, military, semiconductor, and machine-vision markets. Broadcast-standard analog cameras still dominate the low-cost security markets, but FireWire, USB 2.0, Camera Link, and Gigabit Ethernet sensors have come to dominate machine-vision applications. In the design of such cameras, vendors have incorporated the latest CCD or CMOS devices, analog front ends, and off-the-shelf interfaces to produce low-cost products.

As soon as cameras have become digital, they acquired the potential to treat informations. A network camera combines a camera and computer in one unit, which includes the digitization and compression of the video, as well as a network connector. The video is transported over an IP-based network, via network switches, and recorded to a standard PC with video management software. This represents a true network video system, and is also a fully digital system, where no analog components are used.

Resolution is increasing in camera becoming megapixel. These ones are also Intelligent in order to compress the megapixel frames: intellio, mobotix, axis…various examples exist about the importance to work with high resolution cameras. Low resolution does not provide enough information to recognise the person face. Moreover, the cameras are never oriented in the good direction. Recent camera, proposed on the market, have megapixel sensors and hemispherical all-round view for complete room coverage. The high resolution provides digital, continuous zooming and panning (virtual ptz). The cost remains high (500 to 700€) but one megapixel camera can replace up to four fixed cameras. The device is equivalent to one ptz camera without mechanically moving parts. Moreover, this camera works without loosing part of the scene because all is saved. Furthermore, these cameras solutions and their non-intrusive aspect (difficult to recognise the camera) seem to be very adapted for elderly assistance context.

Figure 3-5 PC Intelligent Camera

3.1.5.3. Orange Labs proposition:

Our objective is to provide hardware and software in order to detect people behaviour at home.

About the software, we propose a commercial one embedded by Orange Labs in a smartcam. The frames analyse will detect the exact position of a person in a room. The technology works in 3 dimensions. The furniture and all main object positions will be indicated in a 3 dimensions view. Theoretically, the system is able to measure how long the person walk in a day at home. And, it can detect if a person remain too much time without moving. An interface will be developed to collect all data needed. So, the information will be sent to a centralised system to be semantically analysed with others data provided from other sensors.

About the hardware, we propose different type of cameras from classical IP cameras to the intelligent one. The intelligent cameras are able to embed all the treatment needed to detect people behaviour. If a simple camera is used, frames analyse will be performed in a near box working like a pc. From the box or the intelligent camera, the data quantity is reduced because only semantic information is sent to the central system. No video are transmited. If video verification is needed to confirm the sent alert, a high bit rate connexion is needed to transfer a set of frames. However, the

video verification can be done through others technologies than the cameras one, using webcam on the robot or on the TV set.

3.1.6 Indoor Localisation:

Indoor Localization is the determination of the position of a person in an indoor environment. While the localization in outdoor environments can in most cases efficiently and accurately be performed by GPS. Geomobile is providing a commercial location based service. This solution is based on GSM and GPS localization. Both techniques allow localizing in indoor environment. But GSM is not accurate as localization result is a circle with a radius starting from 80 meter to 6kms. GPS is not indoor localization as satellite signal is not sufficient for a receiver to be detected. Even if GPS receiver can find a position, it is not accurate as satellite signals are reflected by surrounding buildings may be detected through windows.

Therefore, the investigation of position detection technologies suitable for indoor environments is a current research of Geomobile through MIDAS project.

Indoor localisation for elderly people allows to detect strange behaviour at home and to report relevant alarms. Goal is to help elderly people in their daily life through a service which can save their life. Goal is to answer to following questions: why does he/she stay to much time in …? why does he/she have a strange moving at home ?

Continuous localisation between indoor and outdoor environment is an interesting technology to assist medical employees in charge of elderly people who suffer of cognitive troubles. Goal is to answer to the following question: I'm a nurse and I would like to be informed when my patient is going outside after 10:00pm from home or from hospital and to know where he/she is on a map in real time to find her/him immediately.

3.1.6.1. Theoretical Background:

Knowing where a device or a person is in the environment becomes of major importance for applications providing services based on the position. The evolution of the technology and the advent of the GPS, lead to have a good estimate of the position of a mobile outside buildings. Even if GPS is dominating outdoor positioning, no technology flooded the indoor positioning market yet. Many WiFi networks are deployed inside buildings. Signal strength (RSS) information are present for the roaming inside the network. The RSS fingerprinting technique can determine the position of a mobile as long as a signal strength database (RSS coverage) is available.

Time based techniques are less constraining than fingerprinting.

A thesis written by Frédéric Evennou “Techniques et technologies de localisation avancées pour terminaux mobiles dans les environnements in-door” describes the major localisation techniques as (ultra sound, video, wifi, Bluetooth, RFID, UWB, infra red).

The analysis of 2 techniques was interesting. Both are based on the measure of received signal strength estimating the distance between the receiver and the transmitter.

The first solution is based on a model of signal propagation and direct measures.

The second named “fingerprinting” is based on a database, which is making a correspondence between the received signal strength and the position of a person.

Fingerprinting technique has a major drawback as database needs to be fed by plenty of measurements to deliver an accurate position. Setup of this solution becomes more complex.

These techniques are mainly performed by radio based technology. But these solutions remains long and difficult to install and to setup. Therefore, some technologies propose a zero pre-installed infrastructure tracking. These inertial systems based technologies do not require any previously installed wiring or equipment in the target building, such as it is required by wifi or RFID-type systems. In the following, we will talk about three type of technologies, the one that will be developed in the Midas project.

3.1.6.2. Wifi-based technology:

The WiFi technology is widely spread in indoor environments. Networks deployed are mostly used for communication purposes. With the increasing needs of having positioning information to develop new services, different systems tried to use these networks to fulfill their location needs. Contrary to the GPS, no timing information can be valuable indoor. Most positioning systems based on this technology use the Received Signal Strength Information (RSSI) to estimate the position of WiFi mobile clients present in the network.

As the indoor propagation of RF waves is difficult to model, making the use of multi-lateration unreliable, the most common method is to create a database of Received Signal Strength (RSS) tuples, for a sampling grid of positions in the building. The principle of the fingerprinting technique is then to find a best match between the RSS tuple measured on the mobile device and those stored in the database, and estimate a position from this match. An example of hybrid solution is the one provided

by Cisco which uses a combined approach of RSS fingerprinting and signal propagation modeling. This system uses a detailed map of the environment to estimate the expected RSS indications and locate the RSS tuple.

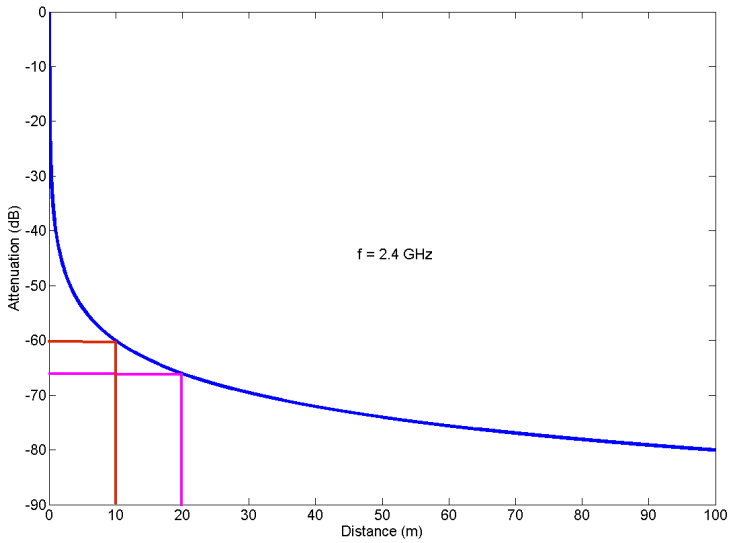

Figure 3-6 Signal's attenuation vs. the distance
The principle of WiFi use for IPS purposes is explained by the curve on Figure 3-6 above.

This figure presents the attenuation of the signal along the distance. It is noticeable that the highest variations are present in the left hand side of the curve. It means that the RSSI significantly fluctuates in that zone (less than 30m). Combined with the impact of the presence of walls, localization based on the RSSI information is possible in indoor environments. Such technique is not applicable in outdoor environments. Indeed, for long distances (higher than 50 m), the attenuation does not fluctuate enough to allow discriminating the positions thanks to the measurements (which are noisy). This is the main reason why some vendors propose to combine the RSSI positioning system with a T(D)OA or AOA based system for outdoors.

Different vendors propose a WPS (WiFi Positioning Systems) either based on a terminal centric approach (measurements done by the terminal) or a network centric approach (measurements done by the network). Both have some advantages and drawbacks according to the test case they have to fulfill.

In the solution proposed by Aeroscout, some Aeroscout Exciters are available. These are equipments working in the 125 kHz frequency band. They are working like RFID readers, broadcasting all the time their Identity (ID) data. If a tag is in the vicinity of the exciter, it catches the ID of the neighbouring exciter and sends this information to the positioning engine which corrects the position of the tag by forcing it to be at the position occupied by the exciter. To work properly, the

positioning engine must be entered the position of the different exciters being in the environment. These special equipments are suitable when a special event or the presence of a tag shall absolutely be detected.

We can find some actors, which based their business model on a positioning system.

Typical independent vendors proposing a WiFi positioning system are listed in Table below:

PRODUCT	PLAYER
Ekahau Site Survey + Ekahau Positioning Engine + Vision (http://www.ekahau.com/)	Ekahau
Formerly Bluesoft, Aeroscout System Manager + Aeroscout Analyzer + Aeroscout Engine + Mobile View (http://www.aeroscout.com/)	Aeroscout
Formerly Pango, (http://www.innerwireless.com/)	Innerwireless

– Typical independent vendors of WiFi Positioning Systems

3.1.6.3. Zigbee-based technology:

Another active tag-based solution was developed by Chipcon [5], in which the location of a ZigBee transmitter is located based on RSS measures. The solution was analyzed in the CAP-M2M project in 2007.

☺ **Advantages from ZigBee-based solution:** Low consumption, positioning precision (from 3 to 5 meters depending on the number of readers and using RSSI; less than 2 meters depending on the number of readers and using TDOA), secured protocol.

☹ **Drawbacks from ZigBee-based solution:** Tags size, need specific network setup, low indoor range (from 20 to 100 meters) involving many readers to be set-up, still high cost.

Development kit based TI CC2431DK allows to evaluate this technology based on TI propogation model. Technique based a propagation model offers a simple solution and can be easily setup.

This component is using an algorithm based on the measurement of received signal strength (RSSI: Received Signal Strength indicator). This value decreases with the distance. The position is estimated from referenced nodes.

Main objective is to find the location of zigbee module called “blind node” in a grid of known nodes called “reference nodes”. Each reference node sends to blind node his position and his corresponding received signal strength. From these data, the “blind” nodes evaluate his position.

The grid can be composed up to 16 referenced nodes. Each reference node is spaced around 10 meters. The accuracy of the position is about 3-5 meters in a real environment. Low leakage ensures to be a low power device around 1-3 years. (GPS device needs to be charged every day).

3.1.6.4. Inertial-based technology: (inertial navigation systems: INS):

When satellite signals are available, the localisation of a pedestrian is fairly straightforward. However, in cities or indoors, dead reckoning systems are necessary. By coupling a magnetic compass with a low-cost gyroscope in a decentralized Kalman filter configuration, the advantage of one device can compensate the drawback of the other. If we compare the rate of change of both signals while measuring the strength of the magnetic field, it is possible to detect and compensate magnetic disturbances. In the absence of such disturbances, the continuous measurement of the azimuth allows to estimate and compensate the bias and the scale factor of the gyroscope. The reliability of indoor and outdoor navigation improves significantly thanks to the redundancy in the information. Numerous tests conducted with different subjects and in various environments validate this approach. Experience shows that the main source of error in position comes from the errors in the determination of the azimuth of walk.

A nice aspect of human walking is the freedom of motion. The worst hypothesis in modelling human trajectories is precisely this liberty of movements. Such aspect will play a major role in the filtering of the azimuth. Sudden rotations measured by a magnetic compass can be caused either by the movement itself or by a magnetic disturbance. Intuitively, if there is a disturbance, the total value of the earth magnetic field changes too. Some examples show that this condition might be sufficient but not necessary to reliably determine a disturbance. In order to improve the reliability of the azimuth determination, a gyroscope will be used.

On several occasions, a swiss society, Vectronix, has demonstrated that a “wearable personal navigation system” is feasible and able to deliver reliable results. Together with system integrators and/or end-users, Vectronix forsees that customized solutions will be developed. The Vectronix DRC Dead Reckoning Compass weighs less than 35 grams and can complement in principle any GPS receiver. Although, this technology success have given more satisfaction to help blind walking in a town, the project was abandoned due to a weak market and to the device cost (5000€).

Figure 3-7 The Vectronix device can track a person in a building.

In fact the market depends on one main compound, the MEMS. Microelectromechanical system (MEMS) is the technology of the very small. MEMS are made up of components between 1 to 100 micrometers in size and MEMS devices generally range in size from 20 micrometers to a millimeter. They usually consist of a central unit that processes data, the microprocessor and several components that interact with the outside such as microsensors.

The main use of this technology is in image stabilising systems in camera and mobile phone.

All the leading smartphone makers, including Nokia, Samsung Electronics, Motorola, LG Electronics, Sony Ericsson, Apple and many more, are looking forward to introducing MEMS gyroscope technology in their smartphones that are scheduled for launch in 2009. MEMS gyroscope is also an innovative technology used by the Nintendo Wii to sense rotational motion and enhance the gaming experience. MEMS sensors breathe life into Wii controller. The Wii-mote includes STMicroelectronics' three-axis acceleration sensors. The sensors, which are based on the company's Micro Electro-Mechanical Systems (MEMS) technology, measure tilt to allow users to move characters, while three-axis sensing transforms the controller into a virtual sword, gearshift, or musical instrument. Now we can track our arm when playing Wii, detect its position, and orientation. Called IDG-600 multi rate gyroscopes, according to InvenSense, the company behind this gizmo, the gizmo delivers high performance silicon based MEMS rate gyroscopes.

Apple has announced that it will introduce this technology in its 2009 iPhones. However, it's not very clear to what extent the MEMS gyroscopes will be implemented and for what purpose. The Apple iPhone already features an accelerometer to detect the motion, but the MEMS gyroscope is expected to enhance this functionality and help in stabilising the digital images captured via the iPhone's camera. According to the top smartphone makers, this new technology will make their mobile devices appeal to more consumers.

Furthermore, MEMS gyroscopes will not increase the production cost too much and will account for only 3 percent of the total production cost. The vendors who use MEMS G-sensors in their smartphones are also looking forward to shift to the new technology.

Handset vendors, which have already incorporated MEMS G-sensors into smartphones, will be more willing to add MEMS gyroscopes into their mobile devices once prices drop to more affordable levels, said the sources.

If the prices of MEMS goes to more affordable levels, we will observe that this technology can be used for others applications. For instance, Xsens Technologies (www.xsens.com) B.V. offers small and highly accurate 3D motion tracking products based upon miniature MEMS inertial sensor technology. Applications include control, stabilization and navigation of unmanned vehicles, robots and other objects, as well as human motion capturing for character animation, training and simulation, and movement science.

Figure 3-8 Motion tracking

3.1.6.5. Geomobile and Orange Labs proposition:

Orange Labs have developped solutions on wifi-based localisation. Based on this knowledge, we would like to propose in Midas project this mean of localisation and so to evolve to others solutions like zigbee technology or inertial navigation systems. We want to test MEMS solutions even if the technology is not ready. To avoid the localisation drift, we will test the coupling with RFID, wifi or zigbee localisation. The first tests will be performed using a technology like the xsens one and we expect to use MEMS integrated in smart phone for the end of this project.

Geomobile Study mainly focus on low power solution as zigbee solution which is a low power solution and is interesting to work on inertial based technology which is more simple to be setup but needs to be evaluated.

The main constraint to provide a seamless solution for indoor and outdoor localization based on the same device is a low power consumption.

3.1.7 Context management related technologies

There exist various approaches for designing context aware systems. Here, we will focus on approaches where a context management service or infrastructure makes it possible for context information to be acquired, modelled and put at the disposal of context aware applications and systems. In particular, we don't consider approaches where the context is identified and exploited by the system itself. A thermostat is a prototypical example of such a system where the switch is put under the direct control of the current temperature. For a one shot, non-reusable implementation of such a system you could embed the context acquisition and modelling functions directly into the application and don't really need a context management infrastructure.

A growing number of context management infrastructures are available. Some properties of these infrastructures are largely the same. For example, most infrastructures support both synchronous (query), as well as asynchronous (publish/subscribe) types of interaction between a context source and a context consumer. Differences exist in e.g. the distribution used; some infrastructures use a strongly centralised approach, such as the CUMULAR Context Server[CUMULAR] that stores the context data in a relational database, while others chose a more distributed approach, such as JCAF [JCAF], the Context Toolkit [CTK] or Amigo CMS [CMS]. The data format used to describe context information can be roughly divided into the categories: RDF, PIDF, or XML. Some infrastructures use either a complete proprietary format (binary of not), or simply use (attributes of) Java objects for that purpose.

A more elaborate analysis of (properties of) different context management infrastructures can be found in [Poortinga].

Compared to existing context management systems, the Amigo CMS design is unique in adopting a more Service Oriented approach by using Web Services for communication between the different components and not prescribing the platform or language to use for implementing its components.

Compared to existing work, this Service Oriented Approach to designing a context management system fits the dynamic nature of ambient intelligence environments or more generally pervasive computing environments, where objects and humans appears and disappears and interact in a loosely and ad-hoc fashion.

This also means that no specific run-time platform is necessary; the only requirement is support for Web Services. Within the Amigo project two deployment platforms have been used; .NET and an OSGi implementation [OSGi]. Libraries are provided for both .NET and OSGi (the latter in the form of OSGi bundles), that implement common functionalities for context sources and consumers, such as discovery of a context broker, registration with the context broker (for context sources), etc. The libraries also provide an abstraction of the context sources, such that clients can treat context sources as if they are local objects.

But the use of .NET or Java/OSGi is not mandatory. For example: in collaboration with the ITEA Trust4All project ([T4A-1],[T4A-2]), an Amigo CMS (Bluetooth) context source was used as input to the Trust4All trust management infrastructure. The client of the CMS in this case was implemented in C using the gsoap and libxml2 libraries.

So what is the best Context Management technology? Clearly there isn't any universal ideal solution. All depends on your needs and the resource you would allocate to build your context management system.

For a reusable solution, we recommend a decentralized solution. First of all, the context management system and the applications will generally be deployed on different runtime platform. Secondly, the context management system will be more flexible if the entities that acquire context information and provide it to the applications could run themselves on different runtime platform. This guarantees execution platform independence (OS and programming language agnostics). Furthermore this improve the dependability (tolerance to sensors or client faults), as a generic advantage of decentralization.

We recommend the use of the Amigo CMS solution for the following pragmatic reason:

- It is available
- It support most of the functional requirements identified for MIDAS

- It provides a sound basis for extending the solution to implement the functions not supported today (context learning, context reasoning,…) that will be necessary for some MIDAS scenarios such as the "daily activities monitoring" one.

Starting with the Amigo CMS solution provides also the following advantages (non exhaustive list):

- it is made available with a development infrastructure (tutorial, examples)
- it is open source and we expect the community of users to grow substantially in the near future
- additional SW modules could be easily integrated as OSGi bundles.

3.2. Driving Scenario

3.2.1 Simulation environment

Desktop Simulators are entry level trainers. These systems can be used as stand-alone units or networked with other simulators. The hardware consists of custom or commercially available driving controls (steering wheel, gearshift, foot pedals for gas and brake), and powered speakers for sound. Both single and 3-channel display models can be networked with others. Desktop simulators are not designed for skills training but are geared toward behaviour and decision making abilities. The same software infrastructure is used by desktop, part-task and full-mission systems. This allows for a low cost upgrade path to full-mission simulators.

Part-Task Simulators feature either an open air, simplified driving station or enclosed (generic or actual) driving stations. The part-task trainer is equipped with a single forward display ranging from a 20-inch computer monitor to a full-size (60-inch diagonal) display. The main difference between a part-task and a full-mission simulator is the number of displays; part-task simulators typically have one forward display and full-mission simulators typically have three displays (left, forward and right). Part-task trainers are not simply a shifting simulator; they use the same software infrastructure as full-mission simulators allowing for training which is limited only by the single forward display. Part-task trainers can be used as stand-alone units or can be networked with other simulators.

Full-Mission Simulators feature either a fabricated or an actual cabin and a panoramic field-of-view (FOV) from the driver's seat. The full-mission simulator has, at least, three channels (displays) providing 180-225 degrees FOV and can be used stand alone or networked with other systems.

3.2.1.1. General Design Guidelines

KaTron's driving simulators simulate the dynamics of the vehicle. The driver operates and feels the response of the actual steering mechanism, the accelerator pedal, the brake pedal, and other operator interfaces. Sensors transmit driver actions to a high-speed computer to create predictions of the resulting motion of the vehicle on the roadway. These predictions are transmitted to the graphics computer and audio equipment to generate and display high-resolution imagery, the feeling of motion and sound. This approach allows for realistic training by duplicating the feel of operating the actual vehicle under the conditions that are either selected by the instructor or programmed into the scenario.

- **Scope of Training** KaTron's driving simulators support training in the following areas:
  - Basic driving skills

- Practice in driving curves
- Practice in braking
- Practice in accelerating
- Unexpected obstacles and hazards
- Visual detection (fog)
- Pre-crash avoidance maneuvering
- Change in road adhesion (rain, snow, ice)
- Slow-moving vehicles
- Component failures (blow-out, power steering failure)
- Driving in hostile environments
- Driving in different conditions (low, middle, high traffic)
- Time of the day (darkness, dusk, day)

The systems also include comprehensive database management software that allows the user to establish and maintain all the trainee records. In addition, this software allows a tracking system which provides a practical means to assess and evaluate performance of the trainee, along with monitoring of the progress. The software is designed with easy to use menu-screen displays and requires no special computer or software training. The printer provides a hard-copy record of the exercise. A wide variety of custom reports can be generated and printed on the laser printer supplied with the trainer.

3.2.1.2. Configuration

KaTron's driving simulators also have, as an option, an Instructor Station located outside the simulator cabin. The simulators consist of the following major components:

- Driver Compartment
- Computer System and I/O
- Control Loading System
- Sound System
- Instructor Station
- Visual System
- Communication System

- **Driver Compartment.** The Driver Compartment consists of a simulated environment that closely replicates the vehicle to be simulated, with a seat for the driver and the controls, indicators, and other equipment contained in the front seat of the vehicle. The simulator interface package provides the interface between the front panel, instruments, steering wheel, accelerator, brakes, clutch, gear shift, and other equipment located in the compartment, and the computer system.

- **Computer System and Input/Output Controller.** The computer system is the brain of the trainer. The computer system configuration includes a PC, hard disk, memory, and the software needed to control the driving simulator and perform simulation tasks required for training purposes. The functions of the input/output controller are to monitor the steering, accelerator, gear position, controls, and to drive the vehicle indicators and instruments.
- **Control Loading System.** The purpose of the control loading system is to simulate the feel of the steering wheel in both the power steering and non-power steering modes (failures). An electric motor driven, force feedback control loading system is provided. This digital computer control system will enhance the realism to provide a higher level of transfer of training.
- **Instructor Station.** Each simulator contains one Instructor Seat/Station as an option, located outside the cabin, which provides the interface between the instructor and the trainee. The instructor station provides the instructor with total overriding control of the simulation. It consists of a PC, visual system repeater display, a graphic CRT, a keyboard input device, a mouse, and a primary console. In addition, the instructor station contains an intercom system, providing communication with the driver. The instructor station provides visibility control, failure control, environmental control, driving scenario control, and positioning of the vehicle. It also has the capability to freeze the vehicle position, and can quickly change the position of the vehicle to a new position. The instructor can record data to be replayed during debriefing. A high quality printer provides the hard copy records and reports.
- **Visual System.** The training requirements for basic visual driving skills, as well as more advanced techniques such as handling changing road conditions and maximum performance turning, require that the driver be given good visual cues. KaTron’s driving simulators meet this requirement by providing the customer with a high-resolution, computer-generated image display system. This visual system is computer-generated, not video-based, therefore, it is flexible and allows real terrain display representing the actual environment to be simulated. The flexibility to build exercises representing real life driving hazards and conditions gives the customer a training platform rather than a limited trainer. Simulation of weather conditions, full day, dusk, and night, along with lighting models, provides the user with the ability to establish customized training scenarios. Moving objects and signs can be positioned as desired by the instructor.
- **Sound System.** KaTron's simulated sound system has proven in the past to be an extremely valuable adjunct to guaranteeing training effectiveness, due to its realistic simulation of the sounds heard by the driver. These sounds include engine, drive train, tires, wind, different pavements, and surrounding vehicles.

- **Communication System** The communication system handles communication between the student and the instructor. The selected frequency will be displayed at the instructor station. This feature allows for instructor/trainee communications, as well as simulation of control station messages and any special directions that may be pertinent to emergency vehicles enroute to an incident site. The capability to accomplish this task in multiple languages provides an added feature to the simulators.

3.2.1.3. Performance

Accurate and fast vehicle dynamic simulation is the key to customer satisfaction. All factors affecting the performance of the vehicle are modeled. Software modules are iterated and sequenced for optimum performance and minimum cue delays.

- **Tire Model.** Because motion of the vehicle is dependent on forces and moments applied to the vehicle from the road, a realistic and accurate tire model is required. The tire model uses kinematic parameters of the tire to calculate the force and moment, including tire deflection, camber angle, slip angle, and longitudinal slip. These parameters are calculated from the position of the wheel assembly and the velocity of the wheel center. Different road profiles (smooth road, uneven road) are user-selected.
- **Suspension, Spring-Damper-Force Model.** Suspension links and joints and key suspension forces influence vehicle dynamics and are also modelled. Elements consisting of a suspension spring, a shock absorber, or a force actuator are required by the equation of motion. Saturation limits, suspension bump stops, and other non-linearities are also provided.
- **Brake Model.** The purpose of the brake model is to calculate the braking torque that is generated by the brake subsystem. The brake configuration is dependent on the vehicle chosen to be simulated. Provision is made to incorporate differential and algebraic state equations that define the brake subsystem to account for power assisted devices. Time derivatives of brake subsystem variables are transmitted to the integration subsystem for a solution with the equations of motion of the vehicle system.
- **Power Train Model.** Engine, transmission, and drive train models are incorporated. Output torque to the drive wheels is calculated as a function of accelerator input, transmission setting, and wheel angular velocity.
- **Steering Model.** The model for steering dynamics includes the transformation of steering wheel motion induced by the driver, from kinematic inputs to steering linkages that control road wheel steer angle. Due to the importance of torque feedback through the steering column to the driver,

models of steering compliance and active feedback mechanisms are included. The torque commands to be applied to the steering column in the simulated vehicle are transmitted to the optional electric control loading system.

- **Aerodynamic Load Model.** An aerodynamic load model is provided to compute the forces and moments that act on the vehicle due to air flow relative to the chassis. Aerodynamic coefficients are supported, including front lift, rear lift, roll moment, side force, yaw moment, and drag. These forces and velocities are transmitted to the equation of motion.

3.2.1.4. Configuration Models and Types

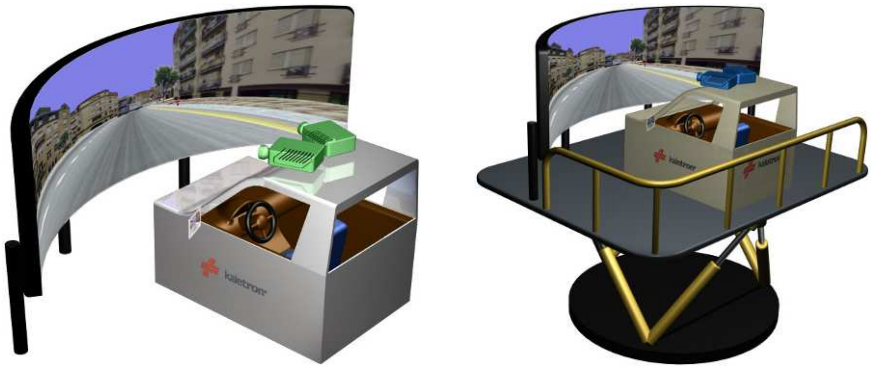

Figure 3-9 Simulators platform examples

3.2.2 Graphic interfaces interaction

Now there are a wide range of graphical technologies but not all of them are applicable in the automotive industry, either for price, for size, for weight or for basic technical characteristics. Among the technologies that nowadays have current representativeness in vehicles in production, or could have it in a near future are:

- LCD-TFT
- HUD
- VFD
- SMELL
- Tactile screen
- OTHERS

LCD - TFT

TFT-LCD (Thin Film Transistor-Liquid Crystal Display) it is a variant of liquid glass panel (LCD) that uses technology (TFT) of thin movie transistor to improve their image quality. The panels TFT-LCD is a type of LCD of active matrix. They are used in televisions, displays of plane screen and projectors. TFT monitor is displacing the technology of CRT (cathodic rays), and they are commonly available in multiple sizes from 7" to more than 42".

The quality of the TFT screens is determined depending features such as the number of dead pixels, the homogeneity of the image, time of answer (maximum deviation in the same production line and in the same day: + / - 2ms) and the general quality of the product. They are classified in the following way:

- V. Screens of worst quality.

- S. Intermediate quality.
- P. Best quality (professional applications).

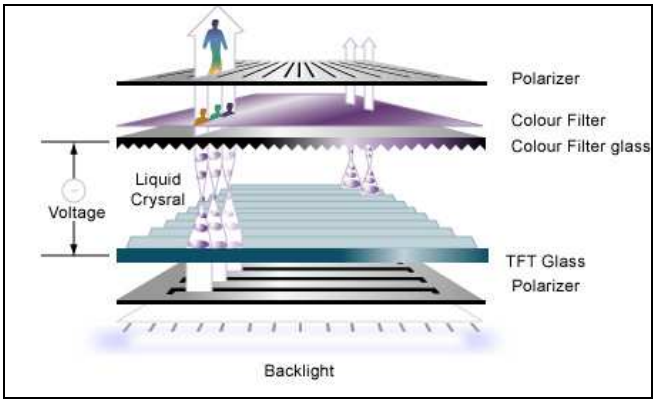

Figure 3-10 Basic scheme of crystal liquid display (TFT/LCD)

Types of colour devices:

- *Twisted Nematic + Film (TN+Film).* It is cheap to produce and has a low pixel answer time. One of the disadvantages of the screens based on TN is their scarce angle of vision, and in general, the colour reproduction and angle of vision of the type panels TN is poor.
- *In Plane Switching (IPS).* It was developed by Hitachi to solve two of the problems of TN panels, the low angles of vision and the bad reproduction of the colour. The technology IPS initial has been overcome by successive improved variants (Super-IPS (S-IPS),…).
- *Multi Domain Vertical Alignment (MVA).* In MVA (Multi Vertical Domain Alignment) the molecules of liquid glass are usually aligned in right angle regarding the substratum, rotating 90º to be parallel in the presence of an electromagnetic field.
- *Patterned Vertical Alignment (PVA).* It was developed by Samsung like an alternative to the MVA. The PVA panel offers lightly improved angles of vision, contrast ratios, improves, and repeatability of much bigger quality than the rest of panels. In infotainment applications with images that change quickly, can present noises and strange effects. This is the weak point of PVA panels.
- *Dual View Display*. Developed by Sharp Corporation in 2005. It is based on a LCD that can show simultaneously different images depending on the angle of vision. In a car, the monitor could be showing a map of highways to the driver while the passengers can be watching a movie. The technology of the monitor is denominated "parallax barrier". The bottom light is divided in two directions that makes possible to see two different things depending on the angle of vision. In the current Dual View Display, half of the pixels are seen by the right

observer and the other half by the left observer, so the resolution of the image is reduced in a half.

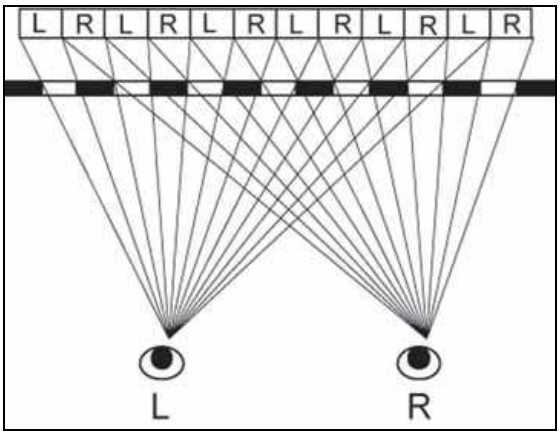

Figure 3-11 Pixels layout in Dual View Display

HUD

The Head-Up Display (HUD) is an electronic system that projects information in the windshield allowing the driver to maintain the attention in the highway while he receives information that would usually require looking to the dashboard or the instrument panel. It can show a wide range of information regarding to the conduction in front of the driver's eyes, so it seems that it floats about 3 or 4 meters for before him.

It consists on a transmitter located on the instrument panel and directed toward the windshield. This light is reflected by a layer of the windshield (of polymeric material, PVB), usually with some specific dimensions so that it is done with clarity. The emission source in the traditional HUD systems has been the tube of cathodic rays (CRT), but in automotive industry it is used most the LCD technology although is possible to find others like LED, LCoS, DMD, SMELL, DLP and laser.

In cars the HUD systems have passed from showing simple data as the speed, to show all type of information with several colours (speed, trajectories to continue, radio information, etc.). Most of the systems allow the user to personalize the information shown, or to disable the system totally. The advanced HUD systems show sailing information via satellite to the driver, and with the help of an anti-collision system supplemented with cameras (infrared for example) can be obtained an improved vision of the roadway in cases of bad time or precarious vision.

Figure 3-12 Citroën HUD

The use of a HUD can reduce up to 0.5 seconds the time the driver consults the instrumentation. On the other hand, their use requires a period of adaptation and although many drivers adapt quickly, others finish discarding it because it distracts or confuses them. HUD technology is a great step in the task of reducing the perceptive load and therefore to increase the security. As HUD system evolve and settle down as a standard, they will substitute many of the indicators and instrumentation that at the moment are in the vehicles

HUD types:

- *HUD refractive*. The rays have parallel form thanks to collimation lenses. The parallel rays are projected toward the support (usually the windshield), and they are reflected to the driver's eyes. An advantage of the refractive HUD is the possibility for the driver to move the head while continues to see the image shown.
- *HUD reflective*. In this design the collimation lenses are integrated in a combinatory transparent lens. The technique is based on the reflection of the rays in this lens. The curved lens prepares the field of vision in such way that they offer a field of vision of 30 grades. The lens reflects the rays in a different orientation from the one received. The rays reflected are only visible in a certain angle, what requires maintaining the head in a quite static position. The production of the lens is complex because they should have the appropriate form so that the rays of the exterior are not distorted. The great disadvantage of the reflective HUD is therefore the difficulty of production in terms of materials and necessary engineering. The advantages are a high capacity of shine, an attenuation of the light of the minimized external scene, and the possibility of saving the space of the collimation lenses since they are unnecessary. In automotive industry they are not used because they are difficult to apply.

The disadvantages of HUD devices are its reading difficulty when the external brightness is very high, or if dirt exists in the windshield and the technical difficulties related with the optic aberrations produced due to the nature of the windshield

VFD

VFD (Vacuum Fluorescent Display) devices have been used in many electronic machines from the 80s, as much in electronic devices as in applications for the automobile. On the contrary than LCD, VFD emits a very brilliant light with a great contrast and it can easily have elements of several colours. VFD types:

- *Tube or module*. In this VFD you can simply select the fluorescent tube, or the tube with the whole necessary electronics to illuminate it forming a module.
- *CIG (Chip in Glass)*. This technology introduces electronics chips inside the hole of the tube. Reduces the dimensions, and also the number of connections what facilitates the assembling.
- *Active matrix*. This technology allows some bigger resolutions and shine comparing to the more traditional construction. The technology of active matrix also allows the representation of graphics with flowing movements and it can reproduce any image inside its range of colours.
- *Rib Grid*. The Rib Grid technology has allowed a bigger flexibility to show colours and to represent forms in VFD. This technology avoids the limitations that VFD causes on metallic meshes.

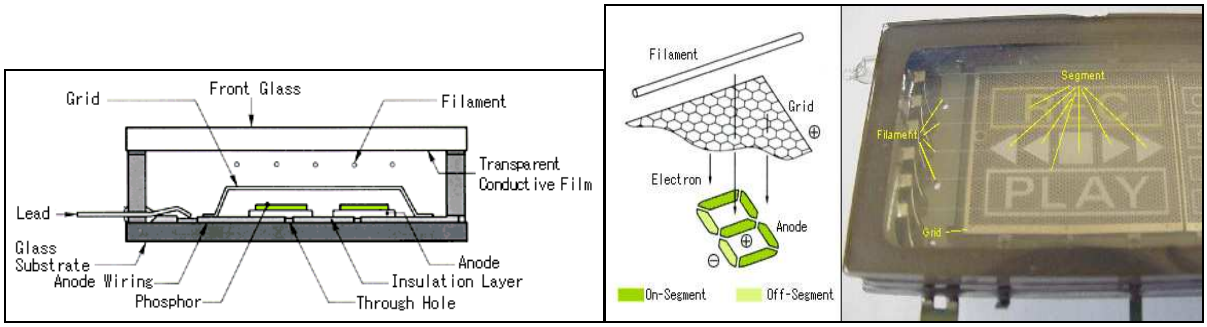

Figure 3-13 VFD working scheme and y components

The main advantages are:

- High shine, up to 35,000cd / cm2.
- Very wide angle of vision.
- Very good visual answer.
- High reliability. Minimum of 30.000 hours to 50% of shine (for panels of 700cd/cm2 of shine)
- Wide margin of temperatures of operation -40°C to +85°C.

- They don't need back illumination.

The main disadvantage of this technology is that is focused for messages and colours. There exist VFD that can show information that is more dynamic but the quality of the image cannot be comparable to a LCD-TFT for example. The electric consumption can also be high compared with other technologies.

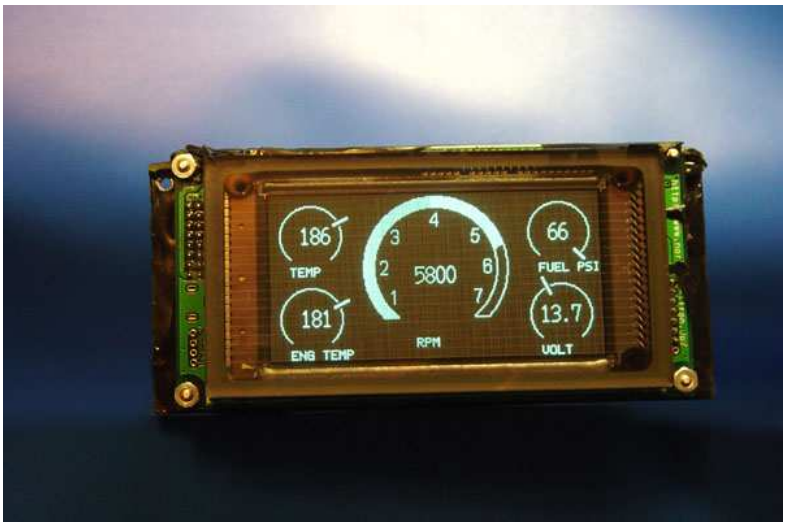

Figure 3-14 Active matrix VFD with digital instrument panel

OLED

OLED (Organic Light-Emitting Diode) is a diode that is based on a luminescent electro layer formed by a movie of organic components that react, to a certain electric stimulation, generating and emitting light. There are as many OLED technologies as many diverse structures (and materials) that have been able to devise, contain and maintain the luminescent electro layer, as well as according to the type of organic components used. The main advantages of OLEDs are: smaller cost, bigger scalability, bigger range of colours, more contrasts and shines, bigger angle of vision, smaller consumption (they don't need back illumination) and, in some technologies, flexibility. But the degradation of OLEDs materials has limited their use at the moment.

Types:

- SM-OLED (Small-molecule OLED)
- PLED (Polymer Light-Emitting Diodes)
- TOLED (Transparent OLED)
- SOLED (Stacked OLED)
- IOLED Inverted OLED (IOLED)

- POLED Patternable organic light-emitting device (POLED)
- PMOLED (Passive-matrix OLED)
- AMOLED (Active-matrix OLED)
- FOLED (Flexible OLED)

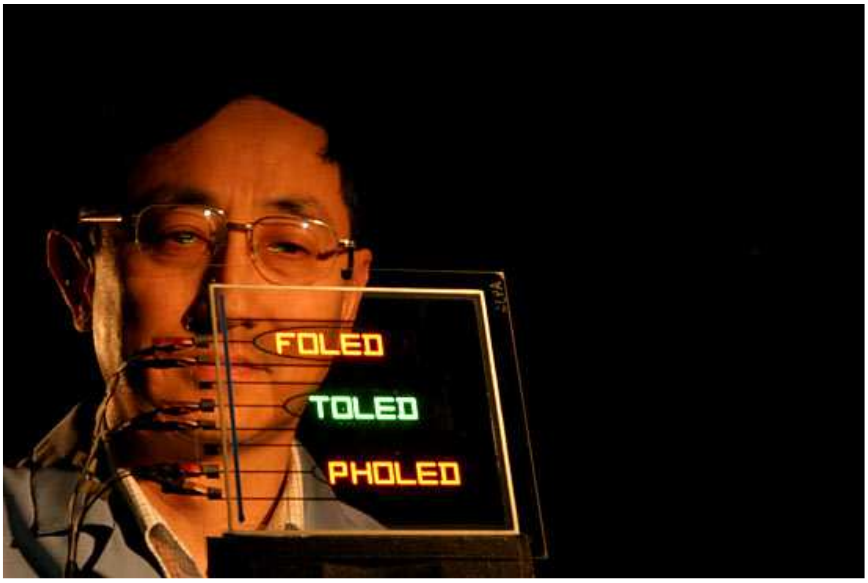

Figure 3-15 Prototype built with OLED technology

The OLEDs offers many advantages in comparison with the LCDs, LEDs and screens of plasm.

- Thinner and more flexible.
- More economic, in the future.
- More shine and contrasts. The OLED pixels emit light directly. For this reason, concerning LCDs, they give a bigger range of colours, more shine and contrasts, and more angle of vision.
- Less energy consumption. The OLEDs doesn't need the backlight technology.

The main inconveniences are:

- Short time of life. The green and red OLED layers have longer time of life (10.000 at 40.000 hours) than the blue one (only 1.000 hours).
- Expensive process of production. At the moment most of OLED technologies are in research process, and the processes of production (mainly initially) are economically high.
- The water easily can ruin the OLEDs permanently.
- Environmental impact. The organic components (molecules and polymers) are difficult to recycle (high cost, complex technical).

Touch Screen

A tactile screen is a screen that through tactile contact on its surface allows the entrance of data and orders to the computer. This contact can also be carried out with pencil or other similar tools.

The tactile screens have been of common use since the invention of the tactile electronic interface in 1971 for the Dr. Samuel C. Hurst. They have ended up being common in TPV, automatic cashiers and PDAs. The popularity of these devices is increasing the demand and acceptance. There are different technologies of touch-screens:

- Resistive.
- Superficial Acoustic Wave.
- Capacitive
- Infrareds.
- Extensiometric gauge
- Optic image
- Dispersive signal
- Recognition of acoustic pulse

Figure 3-16 VW touch-screen

Conclusions

The OLED technologies presented in this report have a promising future (with their different options), since it allows low consumptions and costs of production, with great versatility (transparent, flexible screens, etc). The technology of flexible panels FOLED could have many uses inside the vehicle, as it gives the chance to adapt the presentation of independent graphic information from the surface to use.

The HUD systems are used at the moment by several car brands but its generalization in the future is quite doubtful. At the moment the visual representation obtained with HUD is of low quality, although it is expected new developments in the near future.

3.2.3 Haptic technology

What does haptic mean?

Haptic denomination comes from greek háptō, meaning to touch. Haptic are all those things referred to touch sense, mostly when it is used actively. Some authors have extended the meaning of this word to name all group of not visual and not acoustic feelings experienced by an individual.

Haptic technology is a denomination in which we find technological interfaces, which somehow interact with human through the touch of sense. Origin of this specific denomination comes from 90's, where emerged strongly in some scientific applications. Along last two decades, this technology has suffered a spectacular evolution, being currently one important technology under study in different sectors like automotive, medicine and multimedia.

Which haptic devices are living with us today?

Haptic technologies current development status, as many experts conclude, is just in a starting point. In fact, we have seen how they have been inserted into marked during last years through mobile technology, in particular by means of vibrations in mobile devices. However, this technology has been spread rapidly not only through mobile technology even through joysticks in videogames like Sony Play Station or Nintendo Wii. Next, some related current market technologies are described:

Videogames

Figure 3-17 Force Feedback steering wheel by Thrustmaster

Figure 3-18 Force Feedback Joystick by Thrustmaster

Figure 3-19 Wii videogame console by Nintendo

Figure 3-20 PlayStation3 videogame console by Sony

Figure 3-21 FALCON haptic device by NOVINT

Medicine

Figure 3-22 Medical's Endoscopy AccuTouch System IMMERSION

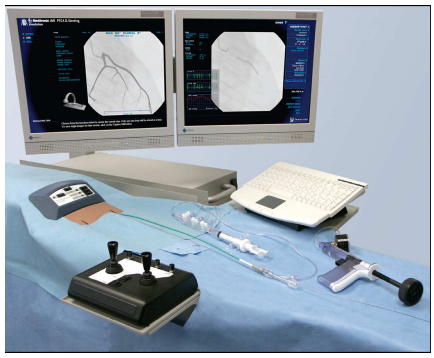

Figure 3-23 The AccuTouch system *graphics and forcefeedback action IMMERSION

Mobile devices

Figure 3-24 Samsung P3 Media Player haptic feedback on the touchscreen by IMMERSION

Figure 3-25 Samsung SCH-n330Tactile Feedback Give Games More Appeal by IMMERSION

Haptic automobile devices

Figure 3-26 Lexus RX450h *Force Feedback Enhances Driver Experience and Safety IMMERSION*

Figure 3-27 iDrive BMW de IMMERSION

Figure 3-28 Mercedes S-Class Rotary COMAND Controller de IMMERSION

Figure 3-29 Advanced Control Panels from Visteon IMMERSION

Others

Figure 3-30 Haptic lever for visual handicapped people (autor not desccribed)

Figure 3-31 Cue Acoustics Radio r1 Radio by IMMERSION

Figure 3-32 3M Touch Systems MicroTouch Capacitive TouchSense® Systemby IMMERSION

Which benefits give haptic devices in automobile?

Most current haptic devices developed in this sector provide some vibrations to users through visual and tactile display interfaces or even through static buttons placed over car central consoles. This in-car integration has been partially a relevant point of interest to users, which have seen how mobile technologies have been inserted in a not common of use environment. In this case, haptic technology finally provide a new channel of information to users, making possible to use this devices smartly reducing user's inattention over full visual display interfaces and consequently to drive in a safe manner.

Figure 3-33 Haptic Device

Currently and progressively, automobile command devices called “By Wire” are replacing mechanical ones. Best improvements of By Wire technology are related to reduction of in-car integrated device package and to the fast response provided by electronics compared with common mechanical devices. Actually, the trend in automobile to use a full car electronic architecture is causing this haptic technology to spread over elements which where usually mechanical.

Not only tactile feedback is acquiring an important paper in this sector. Kinestesic feedback is evolving and providing new findings. That is to say, some studies approaching kinestesic command devices to car’s environment demonstrate and prove advantages over common used mechanical ones. For instance, guiding a car with a haptic steering wheel could maintain car inside a line if somnolence in driver is detected. In addition, if using a pedestrian detector, if system detects one pedestrian the haptic accelerator pedal could advice driver using some vibrations and applying force limitations by itself as to not give the opportunity to driver to increase car speed.

Concerning to handicapped people, some applications have to be defined. For instance, having a haptic device to control the car makes people with disabilities in lower extremities possible to use car normally, offering the possibility to set up turning angles, forces to apply when turning, braking or accelerating. Even if driver has problems of mobility in upper extremities, a haptic steering device could easily permit to drive with a command configuration adapted to this disability.

3.2.4 Command recognition technology

Voice interface is the less distractive interface for driver interaction with de HMI, nevertheless it could already be distractive if dialog does not fit the actual needs. A voice subsystem can be divided in two main categories: **resources** and **dialog**.

Resources will cover the well now technologies of Text to Speech (TTS) and Automatic Speech Recognition (ASR). These technologies are however in this way of evolution with several challenges that must be considered:

- **Text to Speech** is the base technology in the "Information to speech" module. It has in a car environment the challenges of understandability on a noisy environment and limited hardware working. For a limited hardware working, voice unit selection strategy will be based on corpus technology that could be vocabulary stretched for a limited hardware operation. For understandability signal process will base their synthesis models in sinusoidal models with phase and synthesis analysis (SINUF_AS). Using these strategies and synthesis models, it will be available synthetic voices adapted to limited hardware with optimal function that will allow output control of reading speed, tone and amplitude with minimal audio distortion.
- **Automatic Speech Recognition** (ASR) is the base technology in the "Speech recognition" module. It has, in a car environment, also the same challenges of TTS: understandability on a noisy environment and limited hardware working. However command recognizing understandability is a key feature for dialog and voice interaction services. Afterward an ASR basic architecture is shown.

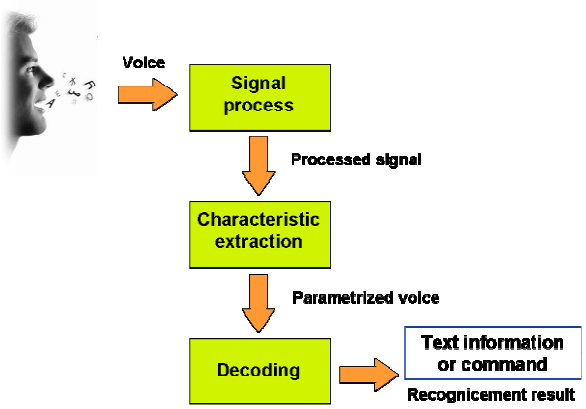

Figure 3-34 ASR architecture

First bloc "Signal process" is used to cancel noise effect over a digital audio base. Noise could be classified as:

- Environmental noise (direct noise): Noise produced for other passengers, motor, wheels, conditioned air, traffic, etc. Most of this noise could be reduced using directive microphones and close-talk microphones.
- Reverberation noise (indirect noise): Noise produced for voice and another sounds reflection on the vehicle.

Noise could be fight with:

- Microphone Arrays: Using several microphones "delay-and-sum beamformer" could be used where voice signal is properly joined considering different delays produced by different distance of each of them to user mouth, obtaining an improved voice signal in comparison to noise produced from other sources. However this technique is too specific and related to development environment.
- Spectral subtraction: Noise is estimated during speech inactive periods and subtracted from the power spectrum of the current frame resulting in the power spectrum of the speech. Generally Spectral subtraction is suitable for stationary or very slow varying noises (so that the statistics of noise could be updated during speech inactive periods).
- Spectral normalization: Acoustical qualities of an input speech waveform are mapped onto that of a desired neutral voice. Such a method can be effective in reducing the impact of speaker variability such as accent, stress, and emotion for speech recognition.

In any case, main developments are applied to "Characteristic extraction" and "Decoding" modules.

Dialog has as main function interact with user as a copilot will do helping driver to access information or services. Therefore, dialog needs access to all information sources: originated by the car, by pluggable gadgets or by external information systems. Dialog would be able also to set or control any of the resources like ordering the reading of an SMS or initiate a phone call.

All this functionalities will be available through an OSGI Framework and any HMI Manager to help service to be improved from vocal services to multimodal services.

Main functionalities of dialog system are:

- Activate speech recognition when "push to talk" button is pressed.
- Chose appropriate vocabulary to be recognized considering all available information that could be categorized with semantic ontologies in different driving states.
- Determine when is appropriate for a text to speech message considering driving states. Messages to confirm commands or generated by alarms inside or outside
- Request information to the driver required by services for no ambiguity command.
- Cooperate with other subsystems like graphical and tactile interfaces for multimodal services.
- Voice dialog configuration through external xml service configuration file.
- Control of service and interaction priority

There are several technologies available or on definition for:

- Vocabulary definition (like JSGF)
- Voice services and multimodal services definition (VoiceXML, SALT, X+V)
- Semantic information, context control and knowledge processing (OWL)

Selection of technology to be used will be done in the future. In any case only solution based on "Commands recognition" would be considered in order to obtain a multimodal HMI easy to use by elderly people.

3.2.5 SOA architecture and Web Services

3.2.5.1. SOA architecture

SOA is an architectural style whose goal is to achieve loose coupling among interacting software agents. A service is a unit of work done by a service provider to achieve desired results for a service consumer. Both provider and consumer are roles played by software agents on behalf of their owners. A SOA actually achieve loose coupling among interacting software by employing two architectural constraints:

- A small set of simple and ubiquitous interfaces to all participating software agents. Only generic semantics are encoded at the interfaces. The interfaces should be universally available for all providers and consumers.
- Descriptive messages constrained by an extensible schema delivered through the interfaces. No, or only minimal, system behavior is prescribed by messages. A schema limits the vocabulary and structure of messages. An extensible schema allows new versions of services to be introduced without breaking existing services.

The following figure illustrates a basic service-oriented architecture. It shows a service consumer at the right sending a service request message to a service provider at the left. The service provider returns a response message to the service consumer. The request and subsequent response connections are defined in some way that is understandable to both the service consumer and service provider. A service provider can also be a service consumer.

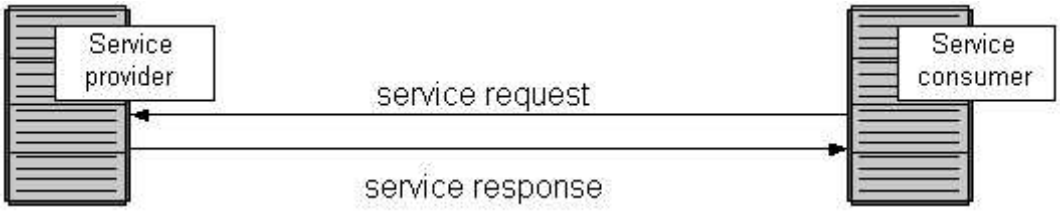

Figure 3-35 Basic-service oriented architecture

Since only a few generic interfaces are available, it is necessary to express application-specific semantics in messages. Any kind of message can be sent over the interfaces, but there are a few rules to follow before it can be said that the architecture is service oriented:

- Messages must be descriptive, rather than instructive, because the service provider is responsible for solving the problem. This is like going to a restaurant: you tell your waiter what you would like to order and your preferences but you do not tell their cook how to cook your dish systematically.
- Service providers will be unable to understand a request if messages are not written in a format, structure, and vocabulary that is understood by all parties. Limiting the vocabulary and structure of messages is a necessity for any efficient communication. The more restricted a message is, the easier it is to understand the message, although it comes at the expense of reduced extensibility.
- Extensibility is vitally important. It is not difficult to understand why. The world is an ever-changing place and so is any environment in which a software system lives. Those changes demand corresponding changes in the software system, service consumers, providers, and the messages they exchange. If messages are not extensible, consumers and providers will be locked into one particular version of a service. Despite the importance of extensibility, it has been traditionally overlooked. At best, it was regarded simply as a good practice rather than something fundamental. Restriction and extensibility are deeply entwined. Both are needed, and increasing one comes at the expense of reducing the other. The trick is to have a right balance.
- A SOA must have a mechanism that enables a consumer to discover a service provider under the context of a service sought by the consumer. The mechanism can be flexible, and it does not have to be a centralized registry.

To implement a SOA, organizations need a service architecture, an example of which is shown in the next figure:

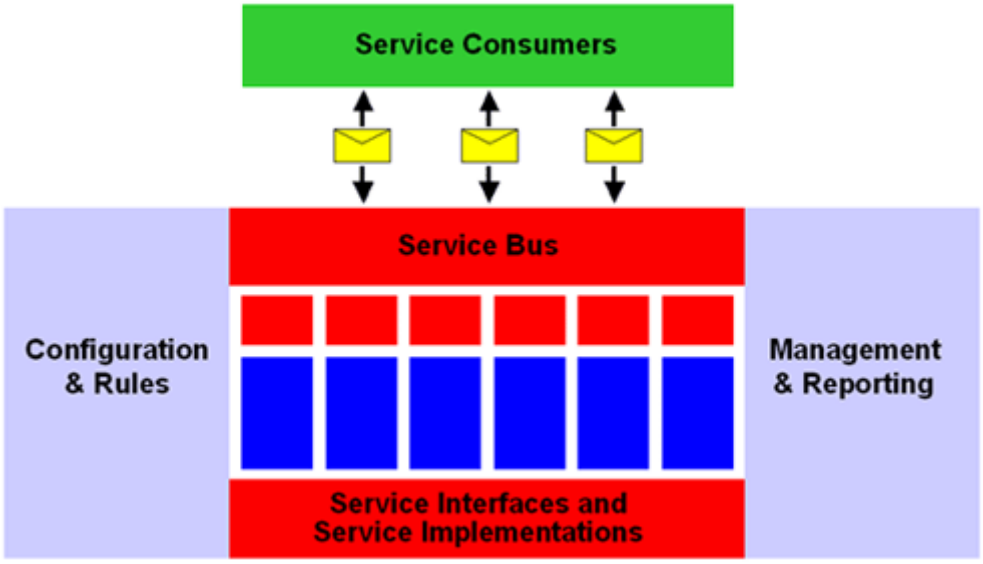

Figure 3-36 Sample service architecture

In the schema above, several service consumers can invoke services by sending messages. These messages are typically transformed and routed by a service bus to an appropriate service implementation. This service architecture can provide a business rules engine that allows business rules to be incorporated in a service or across services. The service architecture also provides a service management infrastructure that manages services and activities like auditing, billing, and logging. In addition, the architecture offers organizations the flexibility of having agile business processes, better address the regulatory and change individual services without affecting other services.

Service-oriented architectures have the following key characteristics:

- SOA services have self-describing interfaces in platform-independent XML documents. Suitable standard languages and notations need to be used to describe the services.
- SOA services communicate with formally defined messages. Communication among consumers and providers or services typically happens in heterogeneous environments, with little or no knowledge about the provider. Messages between services can be viewed as key business documents processed in an enterprise.
- SOA services are maintained in the enterprise by a registry that acts as a directory listing. Applications can look up the services in the registry and invoke the service. Also in this case a suitable standard language is needed for describing service registries.
- Each SOA service has a quality of service (QoS) associated with it. Some of the key QoS elements are security requirements, such as authentication and authorization, reliable messaging, and policies regarding who can invoke services.

Furthermore, a typical Service Oriented Architecture consists of three main entities:

- *Service Providers*, which create services and publish them to the outside world by registering the services with service brokers.
- *Service Brokers* that maintain a registry of published services
- *Service Consumers*, which find required services by searching the service broker's registry.

Service Consumers then bind their applications to the Service Provider to use particular services. The next figure shows the interaction between service providers, service brokers, and service consumers in the publication, discovery, and consumption of services.

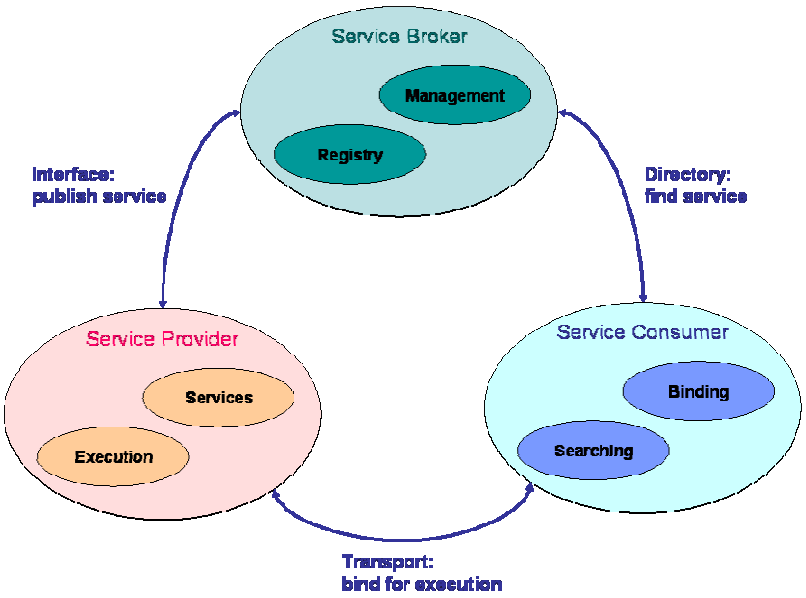

Figure 3-37 Actors and roles in a Service Oriented Architecture

While the service-oriented architecture concept is fundamentally not new, SOA differs from existing distributed technologies in that most vendors accept it and have an application or platform suite that enables SOA:

- SOA, with a ubiquitous set of standards, brings better reusability of existing assets or investments in the enterprise and lets you create applications that can be built on top of new and existing applications.
- SOA enables changes to applications while keeping clients or service consumers isolated from evolutionary changes that happen in the service implementation.
- SOA enables upgrading individual services or services consumers; it is not necessary to completely rewrite an application or keep an existing system that no longer addresses the new business requirements.
- SOA provides organizations better flexibility in building applications and business processes in an easy manner by leveraging existing application infrastructure to compose new services.

Service Oriented Architectures are seen by many organizations as the next evolution in providing agile technology solutions. It promises a world where software is constructed as services and used by more roles than in traditional software development. In principle, applications can be assembled in a safe, quick and easy manner by non-technical people. It is essentially the next step beyond mass customization. Organizations still need to switch paradigms into the world of software-as-services. Understanding the benefits of services and assembly will not be easy initially, but as technology becomes more pervasive and standards proliferate more among platform vendors, return on investment and new business opportunities will clear up the under-standing.

3.2.5.2. Web Services

A Web Service is a self-contained, modular application that can be described, published, located, and invoked over a network (World Wide Web, LAN, etc.). It represents an interface placed between the application and the user of that application and it acts as an abstraction layer, separating the platform and programming-language-specific details of how the application code is actually invoked. This standardized layer means that any programming language that supports the web service can access the application's functionality.

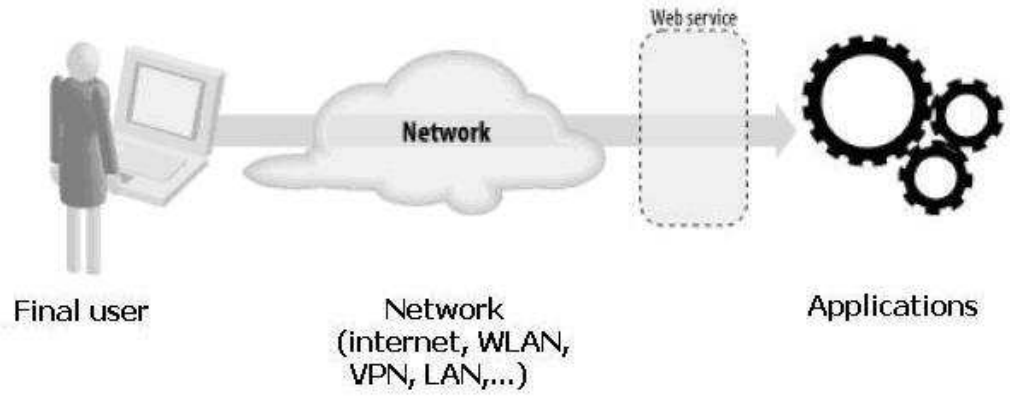

Figure 3-38 A web service allows access to applications using standard Internet technologies.

The Web Services architecture describes three roles: service provider, service requester and service broker; and three basic operations: publish, find and bind. A network component can play any or all of theses roles. The main reasons to choose this technology are:

- *Open Technology*. It facilitates the integration of applications independently of the platform used to develop them and it is based on standards (XML…) and not on proprietary solutions.
- *Small investment in development*. It makes use of the existing infrastructure and any developed component (EJB, Servlets, COM objects, etc.) can be converted easily into a Web Service.

- *Multiplatform*. Web Services are multiplatform as they are based on protocols that are extensively accepted.
- *XML format*. The use of XML for the data format allows any device having a XML parser (PDAs, mobile phones, GPRS / UMTS, Thin Clients, etc.) to use the Web Services.

Web Services are small programs built with various components that are invoked for other programs via http, generating an XML response. Web Service Technology is a compilation of the following technologies:

- *XML (eXtensible Mark-Up Language)*. XML is a W3C purpose mark-up language that supports a wide variety of applications.
- *SOAP (Simple Object Access Protocol)*. SOAP is a lightweight protocol for exchanging structured information in a decentralized, distributed environment and independent from the operative system. It is an XML based protocol that consists of three parts: an envelope that defines a framework for describing what is in a message and how to process it, a set of encoding rules for expressing instances of application-defined data types, and a convention for representing remote procedure calls and responses.
- *WSDL (Web Services Description Language)*. It is an independent platform and language used to describe the Web Services (interface, functions, location…) using XML. WSDL describes four types of data:
    - Information of the interface that describes of the public functions.
    - Information of the types of data for all the request and answer messages.
    - Information about the transport protocol to be used.
    - Information to locate a specific service.
- *UDDI (Universal Discovery, Description, and Inventory)*. UDDI register constitutes a kind of "yellow pages" that allow developers to publish their Web Services so they are accessible to anyone.

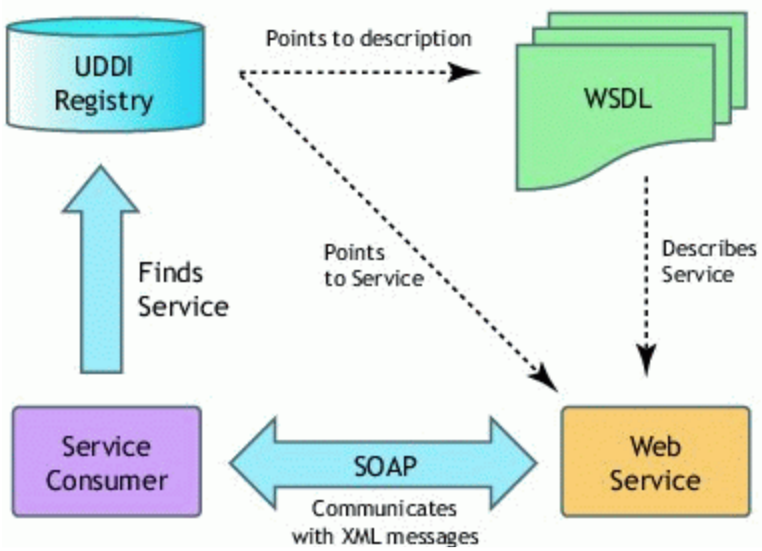

Figure 3-39. Web Service data flow.

All these components facilitate the creation of distributed applications which use different software components, developed in different platforms, with different languages and located in ubiquitous machines.

3.2.6 Sensor system - ECG

Background

In order to fill in the background related to using the ECG signal for car driver monitoring, some basic concepts will be previously explained.

Two anatomical different structures are used as physiological indicators of workload measures: Central Nervous System (CNS, it includes the brain, brain stem, and spinal cord cells), and Peripheral Nervous System (PNS) measures. The PNS can be divided into the Somatic Nervous System (SoNS, it's concerned with the activation of voluntary muscles) and Autonomic Nervous System (AuNS, it controls internal organs and is autonomous because AuNS innervated muscles are not under voluntary control). The AuNS is further subdivided into the Parasympathetic Nervous System (PaNS, to maintain bodily functions) and the Sympathetic Nervous System (SyNS, for emergency reactions):

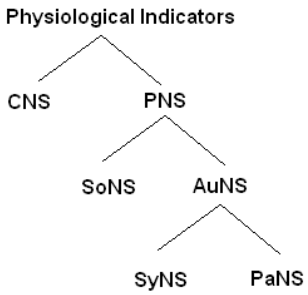

Figure 3-40 Anatomical structures

Most organs are dually innervated both by SyNS and PaNS, and both can be coactive, reciprocally active, or independently active. Heart rate is an example of AuNS measures.

Heart Activity

As mentioned before, the heart is innervated both by the PaNS and SyNS. Each heart contraction is produced by electrical impulses that can be measured in the form of the ECG (Electrocardiogram). The following figure shows typical register of heart electrical activity:

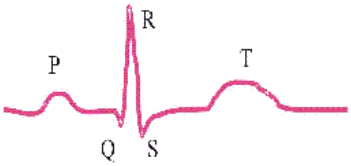

Figure 3-41 Heart Activity

Time domain, frequency and amplitude measures can be derived from the ECG signal:

- **Time domain**.

In the time domain the R-Waves of the ECG are detected, and the time between these peaks (IBI: Inter Beat Interval) is calculated. IBI is directly related to Heart Rate (HR), however, this relation is non linear and IBI is more normally distributed in samples compared with HR. Then, IBI scores should be used for detection and testing of differences between mean HR. IBI scales is less influenced by trends than the HR scale.

According to some scientific works, average heart rate during task performance compared to rest-baseline measurement is a fairly accurate measure of metabolic activity, and not only physical effort affects heart rate level, emotional factors, such high responsibility or the fear of failing for a test, also influence mean heart rate. Other factors affecting cardiac activity are speech and high G-forces. The effect of sedative drugs and time-on-task resulting in fatigue is a decrease in average HR, while low amounts of alcohol are reported to increase HR.

In the time domain, HRV is also used as measure of mental load. HRV provides additional information to average HR about the feedback between the cardiovascular systems and CNS structures. In general HRV decrease is more sensitive to increases in workload than HR increase, although there have been several reports of both HR and HRV unaffected. One of the causes for finding no effect of mental load on HRV lies in the globalness of the measure and its sensitivity to physical load. Some works showed that an increase in physical load decreased HRV and increased HR, while an increase in mental load was accompanied by a reduced HRV and no effect on HR. Fatigue is reported to increase HRV while low amounts of alcohol decrease HRV.

- **Frequency domain**.

In frequency domain, HRV is decomposed into components that are associated with biological control mechanisms. Three frequency bands have been identified: a low frequency band (0.02-0.06 Hz) believed to be related to the regulation of the body temperature, a mid frequency band (0.07-0.14 Hz) related to the short term blood-pressure regulation and a high frequency band (0.15-0.50 Hz) believed to be influenced by respiratory-related fluctuations:

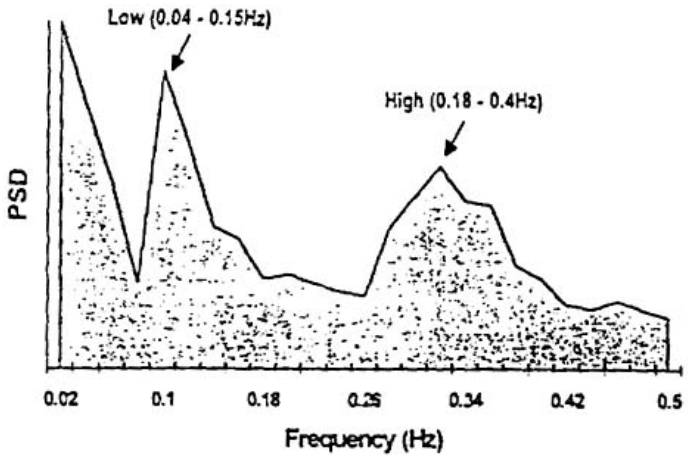

Figure 3-42 HRV: frequency analysis (PSD=Power Spectral Density)
(The John D. & Catherine T.MacArthur Fundation)

A decrease in power in the mid frequency band ("0,10 Hz" component) and in the high frequency band have been shown to be related to mental effort and task demands (Jorna, 1992; Backs & Seljos, 1994; Paas et al.-1994).

- **Amplitude domain**

Finally, amplitude information from the ECG signal can be used to obtain information about workload. The amplitude of the T-wave (TWA) is said to mainly reflect SyNS (Furedy, 1996) and decreases with increases in effort.

Driver mental workload

The reason to consider interesting to integrate driver ECG signal within the car system is closed related to the "driver mental workload" measurement and the relation with some physiological indices. One of them is the heart activity and the changes produced on some parameters calculated from the ECG signal.

Driving is a very dynamic task in a changing environment. Moreover, the driving task is large influenced by drivers themselves. Nowadays, there are factors that may even lead to increased human failure in traffic:

- The number of vehicles on the road is increasing, so increased road intensity leads to higher demands on the human information processing system and an increased likelihood of vehicles colliding.
- People continue to drive well into old age. Elderly people suffer from specific problems in terms of divided attention performance, a task that is more and more required in traffic. One of the causes of these increased demands is the introduction of new technology into the vehicle.
- Drivers in a diminished state endanger safety on the road (longer journeys, night time driving, and so on). Driver fatigue is currently an important factor in accident causation.

The above mentioned factors and situations have in common that in all cases driver workload is affected. Although there are several definitions and models to explain it, “mental workload” could be defined as a relative concept; it would be the ratio of demand to allocated resources. From this point of view, several scientific works have demonstrated that some parameters obtained from physiological measures (pupil diameter, heart rate and respiratory, electro dermal activity, EEG, electro-oculography, ...) could help to know the driver mental workload and one of them is the ECG. Due to his low level invasive characteristic, ECG information seems very interesting information to increase safety in driving tasks. Main idea is to use laboratory methods considered in traffic research and based on ECG signal, and integrate in a car to improve safety and/or wellness.

Sensor alternatives

Several alternatives and related transducers will be evaluated in order to obtain typical ECG signal or at least the corresponding R wave mentioned previously.

The first one will consider directly the electrical signal and impulses produced by each heart contraction measured using typical 3 leads pre-amplifiers. The second one will be based on the micro-vibrations induced by the same electrical signal produced by heart activity.

Direct electrical measurements

Measurement of heart rate is not very complex, the ECG signal needs little amplifying (about 10 to 20 times less as EEG) and electrode placement is not very critical if measurement is limited to R-wave detection and registration.

So no continuous ECG signal could be acquired from the steering wheel because hands position is changing, nevertheless it would be possible to sample this information and find out how some particular parameters are changing during the driving activity. According to this schema, heart rate (HR) may provide an index of overall workload, and spectral analysis of heart rate variability (HRV) would be more useful as index of cognitive, mental workload

In any case, in order to improve the ratio “Good signal time/Total acquisition time” a “two electrodes” pre-amplifier could be evaluated, or the mentioned “three electrodes” but with only two of them integrated in the steering wheel and the third one integrated in the driver seat (this last option will be more driver clothes dependent).

How it works

The pre-amplifier of EKG sensor detects and amplifies the small electrical voltage that is generated by the heart muscle when it contracts. The EKG signal is measured in microvolts (uV).

The standard EKG sensor placement requires that the negative sensor (electrode) be placed on the right side (hand), the positive sensor (electrode) be placed on the lower centre or left side (hand or leg) and the ground sensor (electrode), on the left side (hand).

EKG measures muscle activity by detecting and amplifying the tiny electrical impulses that are generated by muscle fibres when they contract. Since all the muscle fibres within the recording area of the surface sensor contract at different rates, the signal detected by the sensor is a constantly varying difference of potential between its positive and negative sensors (electrodes). The number of muscle fibres that are recruited during any given contraction depends on the force required to perform the movement. Because of this, the intensity (amplitude) of the resulting electrical signal is proportional to the strength of contraction.

3.2.7 Out car communication

Out car communications, possible technologies for MIDAS project will be based on 3G solutions. As this communication solution (out car) must be focused on communicate home and driving scenarios, UMTS (Universal Mobile Telecommunications System) 3G technology will be used.

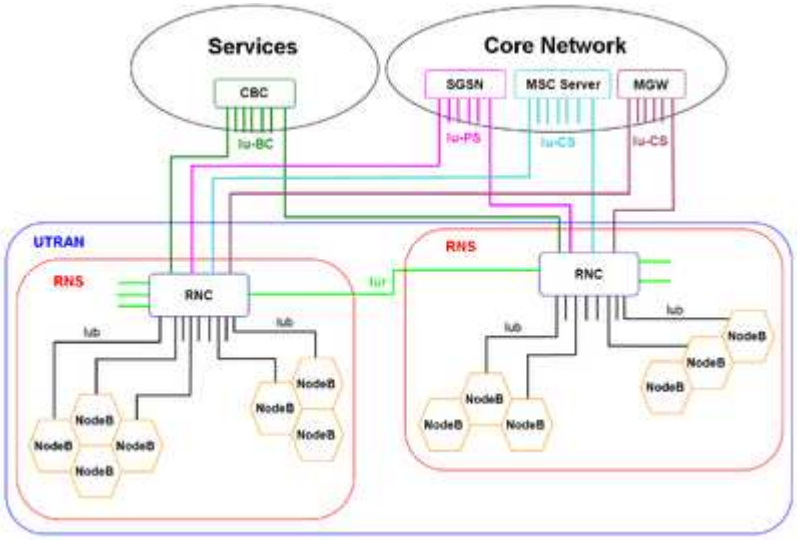

Figure 3-43 UMTS Network Architecture

Universal Mobile Telecommunications System (UMTS) is one of the third-generation (3G) mobile telecommunications technologies, which is also being developed into a 4G technology. Currently, the most common form of UMTS uses W-CDMA as the underlying air interface. UMTS and its use of W-CDMA is standardized by the 3GPP, and is the European answer to the ITU IMT-2000 requirements for 3G cellular radio systems. To differentiate UMTS from competing network technologies, UMTS is sometimes marketed as 3GSM, emphasizing the combination of the 3G nature of the technology and the GSM standard which it was designed to succeed.

UMTS, using W-CDMA, supports up to 21 Mbit/s data transfer rates in theory (with HSDPA, High Speed Downlink Packet Access), although at the moment users in deployed networks can expect a transfer rate of up to 384 kbit/s for R99 handsets, and 7.2 Mbit/s for HSDPA handsets in the downlink connection. This is still much greater than the 9.6 kbit/s of a single GSM error-corrected circuit switched data channel or multiple 9.6 kbit/s channels in HSCSD (14.4 kbit/s for CDMAOne), and—in competition to other network technologies such as CDMA2000, PHS or WLAN—offers access to the World Wide Web and other data services on mobile devices.

Since 2006, UMTS networks in many countries have been or are in the process of being upgraded with HSDPA, sometimes known as 3.5G. Currently, HSDPA enables downlink transfer speeds of up to 21 Mbit/s. Work is also progressing on improving the uplink transfer speed with the High-Speed Uplink Packet Access (HSUPA). Longer term, the 3GPP Long Term Evolution project plans to move UMTS to 4G speeds of 100 Mbit/s down and 50 Mbit/s up, using a next generation air interface technology based upon Orthogonal frequency-division multiplexing.

The first consumer UMTS networks launched in 2002 with a heavy emphasis on telco-provided mobile applications such as mobile TV and video calling. The high data speeds of UMTS are now most often utilised for Internet access: experience in Japan and elsewhere has shown that user demand for video calls is not high, and telco-provided audio/video content has declined in popularity in favour of high-speed access to the World Wide Web - either directly on a handset or connected to a computer via Wi-Fi, Bluetooth, Infrared or USB.

All of the major 2G phone manufacturers (that are still in business) are now manufacturers of 3G phones. The early 3G handsets and modems were specific to the frequencies required in their country, which meant they could only roam to other countries on the same 3G frequency (though they can fall back to the older GSM standard). There are almost no 3G phones or modems available supporting all 3G frequencies (UMTS850/900/1700/1900/2100MHz). However, many phones are offering more than one band which still enables extensive roaming. For example, a tri-band chipset operating on 850/1900/2100MHz, such as that found in Apple's iPhone, allows usage in the majority of countries where UMTS is deployed.

Using a cellular router, PCMCIA or USB card, customers are able to access 3G broadband services, regardless of their choice of computer (such as a tablet PC or a PDA). Some software installs itself from the modem, so that in some cases absolutely no knowledge of technology is required to get online in moments. Using a phone that supports 3G and Bluetooth 2.0, multiple Bluetooth-capable laptops can be connected to the Internet. The phone acts as gateway and router, but via Bluetooth rather than wireless networking (802.11) or a USB connection.

UMTS phones (and data cards) are highly portable—they have been designed to roam easily onto other UMTS networks (assuming your provider has a roaming agreement). In addition, almost all UMTS phones are UMTS/GSM dual-mode devices, so if a UMTS phone travels outside of UMTS coverage during a call the call may be transparently handed off to available GSM coverage.

Compared to GSM, UMTS networks initially required a higher base station density. For fully-fledged UMTS incorporating video on demand features, one base station needed to be set up every 1–1.5 km (0.62–0.93 mi). This was the case when only the 2100 MHz band was being used, however with the growing use of lower-frequency bands (such as 850 and 900 MHz) this is no longer so. This has led to increasing rollout of the lower-band networks by operators since 2006. Even with current technologies and low-band UMTS, telephony and data over UMTS is still more power intensive than on comparable GSM networks.

So, any solution oriented to connect home and driving scenarios within the MIDAS project will take UMTS into account, mainly to implement other different technologies on the application layer as web-services, SOAP, and so on, over the basic communication layers provided by this 3G alternative.

3.3. Technological integration

3.3.1 Technological integration between the different technologies at home

MIDAS system involves many different devices, services and resources. The compilation of all these elements forms a global and heterogeneous system that solves the problems and situations identified in the Use Cases and Requirements document generated in the task 1.1 of MIDAS project plan [1]. One of the main technological challenges of MIDAS development is to integrate all the different modules and make them work together as a single, coordinated system. One example of the collaboration between elements is represented in Figure 3-44; a Service gets information from a medical device and processes the data. Then it recovers some knowledge, from the User Modelling ontology, about user preferences; and sends the information to a multitouch table which represents data in a simple, user-friendly way according to user profile. This simple workflow should be managed with some communication technology in order to integrate the different technologies.

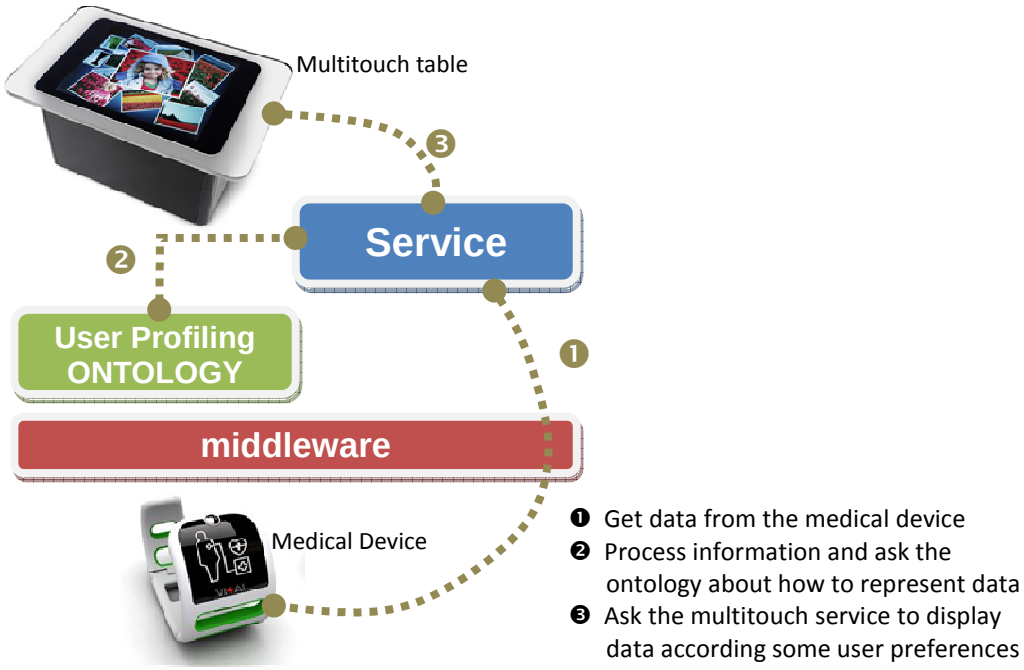

Figure 3-44.- Example of coordination of services within MIDAS system

- **RMI and CORBA,** RMI (Remote Method Invocation) [2] and CORBA (Common Object Request Broker Architecture) [14] are traditional reference technologies to implement distributed systems. The basis of these technologies is to use methods available in a remote service just by having a local representation of the remote interface.

CORBA services are described by an interface, written in the Interface Definition Language (IDL). IDL mappings to most popular languages are available, and mappings can be written for languages written in the future that require CORBA support. CORBA allows objects to make requests of remote objects (invoking methods), and allows data to be passed between two remote systems. Remote method invocation, on the other hand, allows Java objects to be passed and returned as parameters. This allows new classes to be passed across virtual machines for execution (mobile code). CORBA only allows primitive data types, and structures to be passed - not actual code.

Remote Method Invocation allows Java developers to invoke object methods, and have them execute on remote Java Virtual Machines (JVMs). Under RMI, entire objects can be passed and returned as parameters, unlike many remote procedure calls based mechanisms which require parameters to be either primitive data types, or structures composed of primitive data types. That means that any Java object can be passed as a parameter. In order to adapt RMI to CORBA, RMI-IIOP [15] was released to provide connection capabilities between services developed in other technologies different than Java.

The weakest point of these technologies is the lack of flexibility and code reusability. Many of the solutions proposed using these connection technologies are *ad hoc* and don't allow dynamic evolution of the system.

- **Jini** [16] technology is the natural evolution of RMI and RMI-IIOP giving the next step towards service oriented architecture. Jini defines a programming model which both exploits and extends Java technology to enable the construction of secure, distributed systems consisting of federations of well-behaved network services and clients. Jini technology can be used to build adaptive network systems that are scalable, evolvable and flexible as typically required in dynamic computing environments.

  Jini provides service discovery capabilities implemented in a lookup service that manages the resources deployed in the system. Jini also provides facilities for dealing with some problems of system evolution, security and the dynamic assembly of service components.

- **JEE** stands for Java Enterprise Edition and is the reference platform for implementing server-centric applications proposed by SUN [17]. It is centred in the idea of sharing different software components within the same container. But JEE provides a wide set of coordinated technologies that reduce the cost and complexity of developing, deploying, and managing applications. Many of the technologies used are related to communication to other subsystems

that could be servers or clients. JEE allows the use of previous technologies such as RMI or Jini and other new features such as Web Services or JMS (Java Message Service).

- **DOT NET,** The .NET Framework [18] is Microsoft's platform for building applications. It includes a large library of coded solutions to common programming problems including user interface, data and data access, database connectivity, cryptography, web application development, numeric algorithms, and network communications. .NET allows compatibility with C/C++ and Visual Basic programming languages as code is precompiled to an intermediate language (CIL – Common Intermediate Language) that runs on platform specific runtime CLR (Common Language Runtime). As .NET relies over the Windows operating system, it provides some features such as memory management very interesting for developers.

- **OSGi** [19] is a proposed standard to interconnect several devices and services. At the beginning, OSGi was conceived to integrate devices in an intelligent house environment. Thus, it has features such as service discovery or universal plug and play to allow interactive detection of devices and service discovery.

  The OSGi Service Platform provides the functions to change the composition dynamically on the device of a variety of networks, without requiring restarts. To minimize the coupling, as well as make these couplings managed, the OSGi technology provides a service-oriented architecture that enables these components to dynamically discover each other for collaboration. The OSGi Alliance has developed many standard component interfaces for common functions like HTTP servers, configuration, logging, security, user administration, XML and many more. Plug-compatible implementations of these components can be obtained from different vendors with different optimizations and costs. However, service interfaces can also be developed on a proprietary basis.

  The strongest point of OSGi is the development of small software pieces that coordinated together, forms a complete application. All the pieces are independent and can be replaced by new updated versions, giving flexibility and dynamicity to the global system.

3.3.2 Technological integration whit Car communications

Car industry has evolved fast in the last decades. Some years ago, everything was mechanical and electricity was used in few components mainly aimed at elimination. Nowadays electronic devices controls car behaviour in many ways: engine ECU, sensors, airbag, ESP… the next step is include ICTs to improve the quality experience of the user in the car. Many new devices can be embedded in

the car to provide entertainment, social network capabilities, health assistance… OSGi platform is a good reference framework to coordinate these new services; in fact some car manufacturers have begun studies to include this architecture in their next generation of automobiles (BMW is a good example of this [20], [21]).

3.3.3 Technological integration between home and driving scenarios

As OSGi framework seems to be the most suitable reference architecture for MIDAS project, the connection problem between home and car scenario is reduced to the connection of two different OSGi frameworks.

OSGi allows the implementation of heterogeneous services, some of them can act as extremes in the communication between different OSGi frameworks. Many technologies can be used to communicate two modules, but the most suitable is the Web Service technology [22]. It provides a simple way to use remote services just using HTTP as transport protocol. Web Services is a mature technology and easy to implement in the two extremes of communication.

Other solution is the use of a distributed implementation of OSGi [23]. At first, OSGi standard was conceived for embedded applications but now it is being used in enterprise applications. Distributed OSGi is a new component of the upcoming OSGi 4.2 specification and was developed in the OSGi Enterprise Expert Group as RFC 119. The concept of distributed software extends the use of the framework to another JVM or address space.

4. Conclusions

With this technological survey we identified the relevant technologies potential to be utilized in the specification, modelling and development of MIDAS. And some interesting technologies have been identified for both scenarios.

4.1.1 Home Scenario

In this scenario many different interfaces human-computer has been tested, including image and voice, and touch interfaces. Some examples are like tactile screen, voice synthesis or Sight tracker. Some technologies related to systems intelligence are also studied and these studies provide a good input to the work packages two and three.

Video Analytics technologies have been studied to make easy the work of video surveyor, in this section 'Video Analysis' and 'Smartcams' technologies are tested and some propositions have obtained.

Finally the localization in home environment has been studied. Wi-Fi, Zigbee and other technologies have been considered in this section.

4.1.2 Car Scenario

In this scenario some technologies have been studied, many related to interfaces in vehicles (LCD, HUD, OLED…) or simulation. Medical technologies also have been studied in this scenario but more related to sensors installed in a car.

4.1.3 Integration between scenarios

Many technologies have been studied to integrate both scenarios, RMI, Corba , OSGi an others similar technologies are include in this study.

5. References

[All Partners must include references]

[1] MIDAS Deliverable 1.1

[2] RMI White paper http://java.sun.com/javase/technologies/core/basic/rmi/whitepaper/index.jsp

[3] Ajith Abraham: Rule-based Expert Systems. Oklahoma State University, Stillwater, OK, USA: http://www.softcomputing.net/fuzzy_chapter.pdf

[4] Rule Based System description and comparison of forward chaining versus backward chaining systems: http://www.macs.hw.ac.uk/~alison/ai3notes/section2_4_4.html

[5] RuleML, 2009, p. http://www.ruleml.org/

[6] H. Bonley, S. Tabet and G. Wagner, "Design Rationale of RuleML: A Markup Language for Semantic Web Rules," Proc. Semantic Web Working Symposium (SWWS'01), 2001.

[7] B. Parsia, E. Sirin, B. C. Grau, E. Ruckhaus, D. Hewlett, "Cautiously Approaching SWRL," ESWC 2006, 2006

[8] Jena 2 Inference Support, 2009, p. http://jena.sourceforge.net/inference/

[9] Jena 2 Inference Support, 2009,p. http://jena.sourceforge.net/inference/#RULEsyntax

[10] RIF W3C Working Group, 2009, p. http://www.w3.org/2005/rules/wiki/RIF_Working_Group

[11] REWERSE, 2009, p. http://www.rewerse.net/

[12] R2ML – The REWERSE I1 Rule Markup Language, 2009, p. http://oxygen.informatik.tu-cottbus.de/rewerse-i1/?q=node/6

[13] Strelka – An URML-Based Visual Role Modeling Tool, 2009, p. http://oxygen.informatik.tu-cottbus.de/rewerse-i1/?q=Strelka

[14] CORBA http://www.omg.org/corba-e/index.htm

[15] RMI-IIOP http://java.sun.com/products/rmi-iiop/ http://java.sun.com/j2se/1.4.2/docs/guide/rmi-iiop/

[16] Jini web page: http://www.jini.org/wiki/Main_Page

[17] JEE overview http://java.sun.com/javaee/technologies/javaee5.jsp

[18] .NET overview: http://msdn.microsoft.com/es-es/library/zw4w595w.aspx

[19] OSGi Technology: http://www.osgi.org/About/Technology

[20] OSGi Vehicle workshop in Detroit: http://www.osgi.org/blog/2007/01/osgi-vehicle-workshop-in-detroit.html

[21] BMW http://www.osgi.org/wiki/uploads/News/040929_OSGiSinFronteras.pdf + Infortainment: http://www.bmw.es/es/es/owners/accessories/seasonal_campaigns/spring_2009/infotainment.html

[22] Web Services: http://www.w3.org/2002/ws/

[23] Distributed OSGi http://www.osgi.org/download/osgi-4.2-early-draft.pdf including D OSGi (RFC 119)

[24] See http://webopedia.internet.com/quick_ref/OSI_Layers.asp.

[25] See http://www.hl7.org/.

[26] See http://medical.nema.org/dicom/geninfo/Strategy.pdf.

[27] See http://medical.nema.org/dicom/geninfo/Brochure.pdf.

[28] See http://www.iso.org/

[29] Uschold, M. and M. Gruninger (1996). Ontologies: principles, methods and applications, University of Edinburgh, Artificial Intelligence Applications Institute.
[30] Gu, T., et al., An Ontology-based context model in intelligent environments, in Communication Networks and Distributed Systems Modeling and Simulation Conf. 2004, Soc. for Modeling and Simulation Int'l
[31] Guarino, N. (1998). Formal Ontologies and Information Systems. Formal Ontology in Information Systems, Amsterdam, IOS Press.
[33] Heyes-Roth, F., (1985) Rule-based Systems, Communications of the ACM, 28(9): p. 12
[34] McComb, D., (2004) Semantics in Business Systems: The Savvy Manager's Guide, Morgan Kaufmann Publishers, ISBN 1558609172.
[35] Hepp, M., et al.,(Eds.) (2007) Ontology Management: Semantic Web, Semantic Web Services, and Business Applications, Springer, ISBN 978-0-387-69899-1
[36] See http://www.cen.eu/.
[37] See http://www.cen.eu/cenorm/aboutus/generalities/strategy/censtrategy2011.pdf.
[38] See http://www.ihtsdo.org/.
[39] See http://http://www.ihtsdo.org/accessed October 2007).
[40] See http://www.e-health-insider.com/news/item.cfm?ID=1726.
[41] See http://www.cancerinformatics.org.uk/Documents/OpenEHR.pdf).
[42] See http://www.opensource.org/.
[43] See http://www.ihe.net/About/ihe_faq.cfm.
[44] http://www.placelab.org/, PlaceLab", consulted on Nov 2007
[45] J. A. Flanagan, "Unsupervised Clustering of Symbol Strings", Journal Article, Nokia Research Center, 2003
[46] D. J. Cook and L. B. Holder "Mining graph data", WILEY, Book, 2007
[47] J. Himberg and al., "Towards context awareness using symbol clustering map", July 22, 2003
[48] J. Himberg, K. Korpiaho, H. Mannila, J. Tikanmaki, "Time series segmentation for context recognition in mobile devices", Nokia Research Center, 2003
[49] J. Mantyjarvi, J. Himberg, P. Kangas, U. Tuomela, P. Huuskonen, "Sensor signal data set for exploring context recognition of mobile devices", Nokia Research Center, 2003
[50] J. Zhang and al., "Learning Ontology-Aware Classifiers", Artificial Intelligence Research Laboratory, Iowa State University, USA, 2005
[51] J. Zhang, "Learning ontology aware classifiers", Iowa State University, thesis, 2005

Consortium

ORANGE LABS
France
www.orange.com

Project Coordinator :
Laure Chotard, Orange Labs
laure.chotard@orange-ftgroup.com

CEA LIST
France
www.cea.fr

CITIC
Spain
www.citic.es

CNRS
France
www.lifl.fr

ENERGY SISTEM SOYNTEC
Spain
www.energysistem.com

FICOTRIAD
Spain
www.ficosa.com

GEOMOBILE
France
www.whereru.eu

I&IMS
Spain
www.ims.es

INTUILAB
France
www.intuilab.com

KALETRON
Turkey
www.kaletron.com

KIT
South-Korea
http://www.kitvalley.com/

LI2G
France
www.chu-grenoble.fr

MORGAN CONSEIL
France
www.morganconseil.com

MOVIQUITY
Spain
www.moviquity.com

PHILIPS
The Netherlands
www.philips.com

ROBOSOFT
France
www.robosoft.fr

ROBOTIKER-TECNALIA
Spain
www.robotiker.es

SIEL BLEU
France
www.sielbleu.org

TELEFONICA I+D
Spain
www.tid.es

THALES ALENIA SPACE
France
www.thalesaleniaspace.com