Project No. 507424
ALLADIN
Natural Language Based Decision Support in Neuro-rehabilitation

SPECIFIC TARGETED RESEARCH PROJECT
PRIORITY 2.3.1.11

DELIVERABLE 4.1:
DATA DEFINITION FOR PATIENTS, TOOLS FOR ELIMINATION OF NOISY DATA AND SOFTWARE FOR DATA PRE-PROCESSING AND DESCRIPTION OF DATA MINING ALGORITHMS

Due date of deliverable: 31/10/2004
Actual submission date: 06/01/2006

Start date of the project: 1/1/2004 Duration: 36 month

Scuola Superiore Sant'Anna

Revision 1

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	PU
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

List of Partners:	Arteveldehogeschool (B) Language and Computing NV (B) Budapest University of Technology and Economics (HU) Univerza v Ljubljani, Fakulteta za Elektrotehniko (SI) Zenon SA, Robotics and Informatics (EL) (University of Wales Cardiff (UK)) Multitel ASBL (B) The Provost Fellows and Scholars of the College of the Holy and Undivided Trinity of Queen Elizabeth near Dublin (IRL) Országos Orvosi Rehabilitációs Intézet (HU) Scuola Superiore di studi universitari e di perfezionamento Sant'Anna Universita' Campus Bio-Medico

Document identifier:	D4.1_rev3.doc
Version:	1.0
Status :	Final
Date:	06/01/2006
Organisation:	SSSA, UCBM, MULTITEL, KUL, BUTE,ULFE, AHS
Workpackage:	4
Task:	4.1
Dissemination:	Public
Authors:	Stefano Mazzoleni, Paolo Dario, Eugenio Guglielmelli, Domenico Campolo, Giuseppe Cavallo, Jo De Lafonteyne, Gert Van Dijck, Marc Van Hulle, Jo Van Vaerenbergh
Implementation Software:	Stefano Mazzoleni, Justin Cinkelj, Jurak Mihaly, Gert van Dijck, Giuseppe Cavallo, Domenico Campolo
Approved by:	Eugenio Guglielmelli, Task Leader – Jo Van Vaerenbergh, Coordinator
Distribution List:	WP4 partners, Project Coordinator, Project manager

RESUME

The aim of this document is twofold:

- *to provide a synthetic presentation of the rationale that guided the design and development of the ALLADIN Pre-Processing Tool (APT) that represents Deliverable D4.1 of the ALLADIN Project;*
- *to describe the functional architecture of the APT and the basic guidelines for operation of the APT Software Package, that is available in the CD ROM enclosed with this document.*

Deliverable D4.1 is the main outcome of Task T4.1 'Data definition for patients, tools for elimination of noisy data and software for data pre-processing', *led by Scuola Superiore Sant'Anna (SSSA).*

The APT has been developed in MATLAB environment. It allows to the user to download from the ALLADIN Global Database the data measured with the ALLADIN Diagnostic Device (ADD) by the three clinical groups during the ongoing ALLADIN clinical trials, and it performs three basic operations:

- *Visualization. All the Force/Torque (F/T) measurements can be easily plotted for visual inspection by the user. This operation is of paramount importance for physiological interpretation of the isometric measurements, and it has been already widely used all along the APT development to guide the selection of the appropriate filtering techniques and of the meaningful features to be extracted;*
- *Filtering. Two-channel parallel low-pass filtering at 40Hz and 2Hz, has been implemented. The 2Hz-channel is mainly devoted to visualization purposes and to the estimation of the time of activation (onset time) of the different sensors, i.e. the time when the sensor starts recording a signal which can be considered related to human voluntary isometric contraction. Measurements data are recorded by the ADD over time windows of a few seconds, so that accurate estimation of the onset time is essential to reduce the amount of data to be pre-processed by using a smaller time window than the whole recording window. Different techniques have been implemented to produce best-estimate of the onset time, and a specific technique has been selected after comparative analysis with sample data manually elaborated by the clinical experts.*
- *Feature extraction. A set of statistical parameters on the resultant force and torque vectors over time, and time of activation of the different sensors within the same task execution are computed for all pre-processed recordings, with the exception of the 2nd Attempt of each task (Imagination), which has been considered of no clinical value. Signal analysis is limited to a time window of 300 ÷500 msec after sensor activation.*

The APT is a research tool which is not specifically meant, at this stage of the project, for on-field clinical use by therapists and other clinical operators not directly involved in the experimental analysis of the ALLADIN clinical data. Nevertheless, the APT functional architecture has been already conceived in view of its possible upgrade to a more user-friendly version, featuring a high-level interface providing access to current APT functions and some additional functions for management of clinical data.

This deliverable has been developed during the second year of the ALLADIN project (Months 13 to 24) by the WP4/T4.1 Team directly joined by five Alladin partners (Multitel, UCBM, AHS, ULFE, BUTE) and also by an external research group at KU Leuven, with specific expertise in data mining and signal processing. According to the ALLADIN multidisciplinary approach, T4.1 has been carried out in tight co-operation with the project co-ordinator but also asking feedback for the validation of the proposed pre-processing techniques by the other ALLADIN clinical partners (AHS, NIMR,TCD). The participation to this effort of the Multitel and KUL research groups was the result of a remediation plan for coping with the termination of the participation of Cardiff University (CU), the former WP4 Leader, to the ALLADIN Consortium. The final approach selected for the design and development for the pre-processing of the clinical data is mainly based on the specific inputs received by Multitel and KUL in the last quarter of the second year of the project. This should guarantee a smooth subsequent application of the selected data mining techniques to the pre-processed data for the extraction of markers and milestones of the recovery process.

D4.1 - Table of Contents

1 Introduction

The ALLADIN approach for assessing the recovery state of stroke patients relies on repeated measurements of motor efforts during mentally simulated movements for specific tasks. The measurements consist in time trajectories of isometric forces and torques obtained simultaneously by the ALLADIN Diagnostic Device (ADD) by using dedicated sensors that relate to eight body parts (thumb, index finder, middle finger, lower arm, trunk, posterior, foot and big toe). Each time a stroke patient is measured, the force and torque trajectories are captured in different situations. First, the patient is placed into the ADD and resting. Secondly, the patient is presented with the video of one Activity of Daily Living (ADL) task and is asked to imagine performing it. Finally, the patient is asked three times to perform himself the proposed ADL task. Six tasks are considered, namely drinking a glass, taking a spoon, turning a key, lifting a bag, reaching a bottle, lifting and carrying a bottle. Each ADD measurement session includes the execution of the above procedure for each of the six ADL tasks. In order to assess the time course of the stroke patient recovery, measurement sessions are repeated regularly.

All the measurement data recorded during the ongoing clinical trials at the three ALLADIN hospitals in Gent, Dublin and Budapest are regularly uploaded to the ALLADIN Global Database.

The goal of Task T4.1 was to produce a software tool (D4.1) which could enable appropriate pre-processing of the clinical data before entering the data mining process for extraction of innovative markers and milestones of the patient recovery process so to derive long-term prognostic indexes that are currently missing in the clinical practice.

In the following, first the general functional architecture of the ALLADIN Pre-Processing Tool (APT) is presented.
Then, the rationale for the design and the implementation of the MATLAB internal modules of the APT is briefly illustrated. Particular attention is devoted to the selection of the features to be extracted by the Force/Torque measurements recordings.
Finally, the detailed format of the output data structure generated by the APT module, which will be the direct input to the Alladin Data Mining Module, is described.

Three appendixes to this document provide some examples of feature extraction on real samples of clinical data, a list of the APT software modules available on the enclosed CD ROM, and a basic guide to the installation and use of the APT software.

2 General functional architecture of the ALLADIN Pre-processing Tool (APT)

The Alladin Pre-processing Tool is a software tool that is meant to automatically derive specific features from the ADD recordings available in the Alladin Global Data Base, convert the type of data when necessary, and re-organise the data into a consistent and non-volatile structure in a desirable format for consequent data mining analysis towards extraction of clinical markers and milestones relevant for functional assessment of the Alladin patients.
The APT also includes a pre-Visualization Module which allows visual inspection of data during the pre-processing operations and that will link to the Alladin Integrated Visualization Tool, which will provide integrated access to the whole set of multimodal clinical data available on the Alladin patients.
The overall functional architecture of the Alladin Pre-Processing Tool is illustrated in Figure 1 below.:

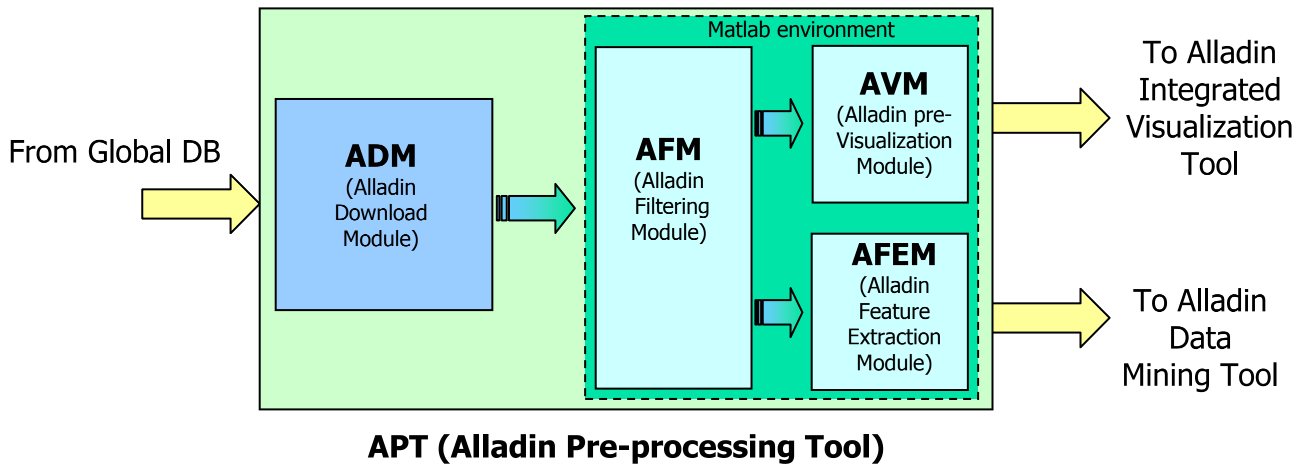


Figure 1. Overall architecture of the APT - Alladin Pre-processing Tool

Responsibilities of WP4 partners for the development of the different modules have been identified as follows:

- AFM – Alladin Filtering Module, Resp. partners: SSSA, UCBM
- AVM – Alladin pre-Visualization Module, Resp. partners: ULFE, SSSA
- AFEM – Alladin Feature Extraction Module, Resp. partners: SSSA, UCBM, KUL, MULTITEL

The implementation of ADM was not included in the development of D4.1 as it is meant for future upgrade of the APT in view of its integration with the other Alladin software tools deriving from other project workpackages.

3 The Alladin Download Module (ADM)

The Alladin Download Module, to be developed in Visual Basic, provides user-friendly access to the functions implemented by the other APT modules. It also performs the following additional operations on the input data, retrieved by an operation of synchronization with the Alladin Global Database:

- Data selection, based on patients IDs and required interval of time
- Adverse events handling

Data selection allows to define the patients IDs and/or a time interval of interest (starting and final date), and then to extract the selected data from the ALLADIN database. The data will be the force and torques measurements recorded with the ADDs in the three clinical centres. They are downloaded from the ALLADIN Global Database through an execution of a synchronization, using the Cover Application: the ADM retrieves the data available in the ALLADIN local database and, by default, operates on the full set of patients.
The user can also easily set a flag to choose to operate on the entire set of data for the selected patient.

Adverse Event Handling automatically generates null output values for ADD measurements where the Adverse Event flag is marked. The ADM also offers the possibility of including some of these measurements by manually selecting them from a closed list.

The current implementation of D4.1 does not include the ADM which is meant for future upgrade of the APT to be integrated with the whole Alladin Sofwtare Environment, including the Cover Application Package, the Integrated Visualization Tool, the Data Mining Tool and other software modules to be developed in the remaining lifetime of the project, in view of the Alladin commercial exploitation. Retrieval of data from the Alladin Global Database and suppression of measurements flagged as adverse event are supposed to be performed manually by the user before the APT is launched.

4 The Alladin Filtering Module (AFM)

The Alladin Filtering Module (AFM), developed in Matlab, implements two basic operations:

- *Inconsistent Data Filtering*. Measurements data are checked to verify if the standard procedure has been followed during the clinical trial. If abnormalities are detected in the data, e.g. sensors have not been calibrated before the patient entered the ADD, such data will be not considered for feature extraction and switched to a dedicated area of the local database without further pre-processing;
- *Noise Filtering*. AFM will implement a two-channel parallel low-pass filtering, one featuring cut-off frequency at 40Hz and another with cut off frequency at 2Hz which will provide two separate data sets for subsequent processing. The two cut-off frequencies have been selected taking into account that, on one hand, human muscles can generate mechanical signals up to a maximum frequency of 40Hz (muscle sound), while, on the other hand,. human voluntary movement typically generates signals within the frequency range 0-2 Hz. The 40 Hz-channel is the main channel used for feature extraction, while the 2 Hz-channel is used for visualization and onset time estimation operations.

5 The Alladin pre-Visualization Module (AVM)

The ALLADIN pre-Visualization Module has been developed in order to visualize the ALLADIN measurements, with the aim of supporting the clinical partners in the visual inspection. As a preliminary step, the user has to set the path with the folder containing the force/torque measurements. A file named AUTOLOAD_Base_Directory.m has to be edited. A line containing a string, such as:

```
base_dir = 'C:\MATLAB6p5\work\FTdata';
```

points to the directory where force/torque files are saved on user's pc. In case this is not done, an error message box appears. The user can enter the base directory via menu interface too. Part of the setup procedure is related to the information about side of measurement available to the visualization tool. This information is available in the ALLADIN database and can be performed by launching the program readDB.exe. The output data are saved to a file, named AUTOLOAD_measurementSideFromDB.m. Figure 2 shows the main window of the visualization tool.

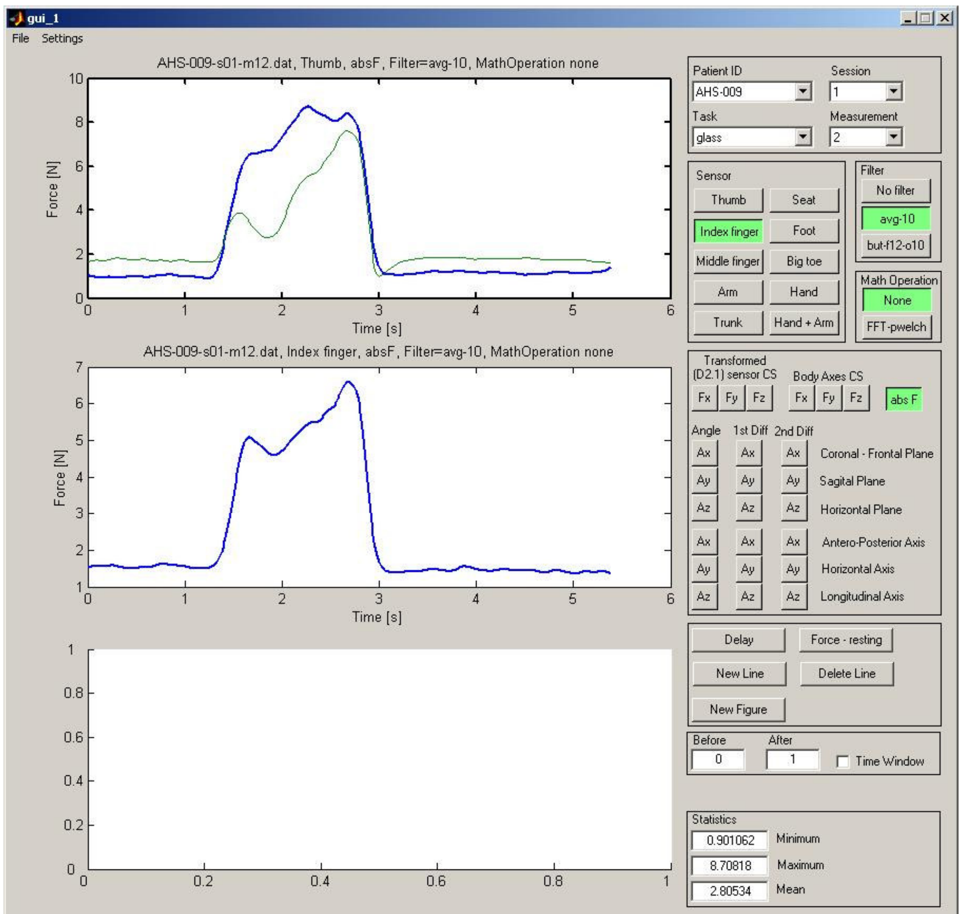


Figure 2. Main AVT window

Through the controls positioned in the main window, the patient ID, session, task and measurement number can be chosen. Also data filtering, some 'mathematical operation' (minimum, maximum, mean) and coordinate transformations can be applied to the measurements, and plotted for inspection. Additional information on the AVT has been already provided in D3.2.

6 The Alladin Feature Extraction Module (AFEM)

The Alladin Feature Extraction Module, developed in Matlab, receives the filtered data from the 40Hz-filtered channel of the AFM and generates the output data containing statistical and temporal features calculated for all the ADD measurements of the input data set.

6.1 Time window of interest

As illustrated in Table 1, the measurement recording time during different ADL tasks goes from a minimum of 2.4 s to a maximum of 5.4 s, depending on the specific ADL task considered.

TASK	Baseline (0)	Video (1)	1st rep. (2)	2nd rep. (3)	3rd rep. (4)
Glass	3	**5.4**	**5.4**	**5.4**	**5.4**
Key	3	**3.7**	**3.7**	**3.7**	**3.7**
Spoon	3	**3.4**	**3.4**	**3.4**	**3.4**
Bag	3	**2.4**	**2.4**	**2.4**	**2.4**
Reaching	3	**4**	**4**	**4**	**4**
Moving	3	**6**	**6**	**6**	**6**

Table1. Duration times ([s]) of the different recordings during a typical ADD session

From a clinical point of view, the data of interest to be extracted by the ADD measurements are conveyed by the very initial part of each recording before the patient adapts to the isometric constraint and the behaviour starts to become not purely physiological.
For instance, at the upper limb level after the initial, physiological onset of the motion pattern, typically the patient starts to exhert an increasing force against the constraint and/or iterates the attempt of moving the limb within the same recording.

Based on these considerations, for extracting features the complete force and torques signals at a given sensor will be considered only within a finite-length analysis frame. It starts from the estimate of the onset time and lasts a few hundreds of milliseconds, corresponding to a finite number N of samples.
The analysis duration frame is a parameter of the AFEM in order to allow testing various values.
Based on a preliminary analysis of normal controls, the duration of the time window of interest can be identified from a minimum of 300 ms to a maximum of 500 ms. Initially, an analysis frame of length 300 ms, 400 ms or 500 ms will be tested, *i.e.* N=30, N=40 and N=50, respectively.
However, the duration of this time window could be easily adapted to cope with possible uncertainties in the estimation of the onset time, as illustrated in the following sub-section.

6.2 Time of Activation of the sensors (Onset time)

In order to properly identify the onset of the motion pattern which shall be selected as the starting point of the time window of interest for each measurement recording a dedicated approach has been implemented.

The proposed approach is based on a comparative analysis of the performance of different candidate techniques for onset time estimation with respect to the manual performance of the Alladin clinical experts operating by direct visual inspection. To this aim the experts have been provided with a dedicated onset time estimation tool which has been derived by the AVM, and that is enclosed in the D4.1 software package. The same set of real clinical data (force and torque signals deriving from 100 sample measurements from all Alladin clinical centres) has been used for onset time estimation by the experts and by the candidate automatic techniques.

This algorithm will be applied to each of the recordings of the six force/torque components of each sensor. In such a way, six onset time values will be calculated for each component, but only the minimum of these values will be retained and identified as the Time of Activation of the sensor.

Starting from in-depth review of state-of-the-art techniques and after internal debate between engineers and clinical experts, a shortlist of candidate methodologies for automatic onset time estimation have been identified, namely:

1. To detect the time when the force/torque signal reaches 2% of its peak value;
2. To identify the onset time by using the spectroflatness measure (SFM) of the force/torque signal;
3. To identify the onset time by using PDF/Ks-density measure of the force/torque signal;
4. To identify the onset time by using the 2-nd order derivitative of the force/torque signal (previously low-pass filtered at 3Hz or at 5Hz);

Table 2 presents the results of the comparative analysis among the performances of the different candidate techniques with respect to the reference performance of three clinical experts.
First, the Mean Reference Vector (MRV) has been derived by the experts inputs by computing the mean value of the three onset times estimated by the three experts for each of the measurements.
Then, Mean value, Standard Deviation, Variance and Median of the error vector related to each of the candidate techniques have been calculated (columns 2-5).
Finally, also a non-parametric statistical feature, defined as the *Probability Of Correctness* (POC), has been computed. POC is calculated as the ratio Nc/N, where N is the total number of samples and Nc is the number of samples which fall between the 5th-percentile and the 95th-percentile of the MRV.

Onset technique	Mean value (s)	Standard deviation (s)	Variance (s)	Median (s)	POC
2% rule	**0.5080**	**0.8524**	**0.7266**	**0.3887**	**0,57**
Spectrum flatness	**0.0968**	**0.7870**	**0.6194**	**0.0990**	**0,69**
Ks-density	**-0.0252**	**0.7959**	**0.6335**	**-0.1766**	**0,71**
2nd derivative (filtered 3Hz)	**0.2136**	**0.6223**	**0.3873**	**-0.0188**	**0,89**
2nd derivative (filtered 5Hz)	**0.2044**	**0.6241**	**0.3894**	**-0.0276**	**0,89**

Table 2. Comparative analysis of the candidate techniques for onset time estimation

From the data presented in Table 2 the following conclusions can be derived:

- Techniques based on application of thresholds to the 2nd-derivative of the force/torque signals demonstrate the best performance among the selected candidate techniques;
- All techniques, however, feature a value of the standard deviation which is bigger than the duration time of the window of interest (up to 500 msec);
- Additional reference data from clinical experts shall be produced in order to verify the results presented in Table 2;
- The duration of the time window of interest shall be extended taking into account the uncertainty deriving from the value of the standard deviation of the automatic technique for onset time estimation which will be eventually selected;
- The current release of the APT shall be supervisioned for manual validation of the automaic onset time estimation before data mining processing techniques could be applied.

.

6.3 Feature Extraction

After the proper time windows of interest have been calculated and applied to the original signals of the input data set, various statistical features are extracted from these data.

The choice of this features is the result of an internal debate between clinical partners and datamining experts. The ultimate aim is to identify parameters of some clinical significance which can be at the same time prone to be processed by the datamining algorithms in order to estimate ‘distance from normality’ of the different patients along time.
Also, a compromise between the richness of information provided by the ADD recordings and the need for keeping the computational cost for datamining to acceptable values has been taken into account.

In detail, the selection of the features to be extracted by the AFEM is based on the following basic assumptions/hpyotheses:

1. Features shall be calculated for both forces and torques measurements of all attempts for all the sensors and all the tasks in each session, with the exception of the second attempt (imagination). In fact, estimated accuracy and resolution of the ADD in normal environmental conditions do not allow to perform meaningful measurements during the imagination phase (poor Signal-to-Noise Ratio);
2. Concerning torques, their real usefulness in terms of added-value for effective datamining will be later evaluated. Since the ADD works in isometric conditions, it is expected that torques are nearly proportional to forces, so that probably torques could be of interest for visual and/or datamining investigation only in very limited cases and for some specific sensors, e.g. the seat sensor.
3. Force/Torque resultant vectors spatial series over time could convey most of the useful information about the measurements.
4. Since stroke patients typically demonstrate reduced ability in controlling generated force/torques, both in intensity and in spatial direction. It is expected that this impairment should translate in some kind of “abnormalities” in the force/torque vector direction that could be visible by comparing deviation angles between current force/torque signals and mean force or previous force.
5. In order to reduce the amount of information to be processed for datamining, relevant standard statistical parameters will be extracted from the distribution of such deviation angles over time, such as: maximum, mean, standard deviation, skewness, kurtosis and histogram.
6. additional useful information on the distribution of deviation angles over time could be conveyed by the application of autoregressive models to the available time series. The order and the parameters of such autoregressive models will be extracted as features associated to each measurements. This choice is also related to the type of datamining techniques that are envisaged to be applied after pre-processing. For more details, please refer to the description of the datamining algorithms.
7. the sequence of activation of the different sensors and the relative time delays during the execution of the same task could certainly be of clinical interest for estimating ‘distance to normality’. It is expected that stroke patients will demonstrate abnormal time activation patterns due to limitation in forward model generation, motion planning & supervision, sensorimotor control, etc. Since no a priori knowledge is available on this type of measurements before the ADD platform was developed,

further reasoning on the actual and specific usefulness of this data for clinical markers and milestones estimation could be possible only later in the project.

We define a recording as the set of force and torque measurements at a given measurement site, for a given patient, during a given session and for a given task. Hence, every recording is uniquely identified by the site identifier, the patient identifier, the session number and the task and attempt number. All the gathered recordings represent a large amount of raw data that should be processed in order to capture relevant characteristic features with respect to stroke patient recovery.

Every recording contains discrete-time trajectories of forces and torques for eight sensors with respect to three orthogonal directions. Let define $F_{s,x}[k]$, $F_{s,y}[k]$ and $F_{s,z}[k]$ as the discrete-time force signals along the three orthogonal axes *x*, *y* and z for the *s*-th sensor, *s*=1,…,8. Note the directions of these axes are specific to every sensor. Similarly, $T_{s,x}[k]$, $T_{s,y}[k]$ and $T_{s,z}[k]$ as the discrete-time torque signals around axes *x*, *y* and z for the *s*-th sensor. The sampling rate is equal to 100 Hz, that is, six force and torque values are captured every 10 ms. The data acquisition time is controlled in order to observe the movement initiation part within some context. The measurements have been low-pass filtered at 40hz by the AFM in order to reduce noise effect.

The whole list of statistical and temporal features to be extracted is detailed in the following. However, the Matlab environment will be prone to easy integration and adaptation of the definition of such features based on future revisions of the above listed hypotheses and/or on preliminary datamining results.

6.3.1 Mean Effort Direction

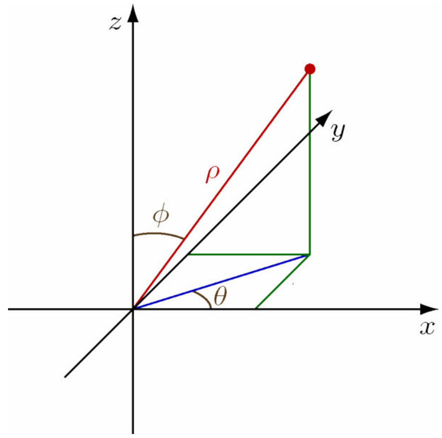
It is assumed that effort direction is more relevant than intensity. This assumption relies merely on that patients are not asked to actually perform the movements but only initiate them according to the ALLADIN protocol design. Given a recording, for the *s*-th sensor, we compute the mean force direction features as the colatitude and azimuth angles of the mean force vector with respect to its referential. The mean force vector is defined by its components $\overline{F}_{s,x}$, $\overline{F}_{s,y}$ and $\overline{F}_{s,z}$ where

$$\overline{F}_{s,x} = \frac{1}{N}\sum_{k=k_0}^{k_0+N-1} F_{s,x}[k]$$

$$\overline{F}_{s,y} = \frac{1}{N}\sum_{k=k_0}^{k_0+N-1} F_{s,y}[k]$$

$$\overline{F}_{s,z} = \frac{1}{N}\sum_{k=k_0}^{k_0+N-1} F_{s,z}[k]$$

with k_0 being the sample index of the estimated onset time. The colatitude $\phi_{F,s}$ is the angle between the *z*-axis of the mean force vector. The azimuth $\theta_{F,s}$ is the angle between the positive *x*-axis and the line from the origin to the end of the mean force vector projected onto the xy-plane. These angles are obtained by converting the cartesian coordinates of the mean force to spherical coordinates, that is,



$$\rho = \sqrt{F_{s,x}^2 + F_{s,z}^2 + F_{s,z}^2}$$

$$\phi_{F,s} = \arccos\left(\frac{F_{s,z}}{\rho}\right)$$

$$\theta_{F,s} = \arctan\left(\frac{F_{s,y}}{F_{s,x}}\right) + \pi\, u_0\left(F_{s,x}\right)\operatorname{sgn}\left(F_{s,y}\right)$$

where $u_0()$ stands for the Heaviside unit step function

$$u_0(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 1 & \text{if } x > 0 \end{cases}$$

and sgn() denotes the signum function

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}.$$

Angle features $\phi_{T,s}$ and $\theta_{T,s}$ can be computed similarly from the mean torque vector to characterize the mean torque “direction”.

6.3.2 Angular Deviation to Mean Effort

Beside features characterizing mean direction of efforts, the angular deviation of every effort sample within the analysis frame from the mean effort is computed. It is assumed that the distribution of these angular deviations depicts some specific pattern (sudden variations, lack of smoothness, etc) in the stroke patient movements. Given a recording, for the s-th sensor, the angular deviation $\delta_{F,s}[k]$ between the k-th force sample ($F_{s,x}[k], F_{s,y}[k], F_{s,z}[k]$), within the analysis frame $k = k_0, \cdots, k_0 + N - 1$, and the mean force ($\overline{F}_{s,x}, \overline{F}_{s,y}, \overline{F}_{s,z}$) is computed as the inverse cosine of the normalized scalar product, *i.e.* the dot product of the corresponding unit-norm vectors,

$$\vec{a} = \left(\overline{F}_{s,x}, \overline{F}_{s,y}, \overline{F}_{s,z}\right)$$
$$\vec{b} = \left(F_{s,x}[k], F_{s,y}[k], F_{s,z}[k]\right)$$
$$\delta_{F,s}[k] = \arccos\left(\frac{\vec{a}}{\|\vec{a}\|}\cdot\frac{\vec{b}}{\|\vec{b}\|}\right) = \arccos\left(\frac{\overline{F}_{s,x}F_{s,x}[k]+\overline{F}_{s,y}F_{s,y}[k]+\overline{F}_{s,z}F_{s,z}[k]}{\sqrt{\overline{F}_{s,x}{}^2+\overline{F}_{s,y}{}^2+\overline{F}_{s,z}{}^2}\sqrt{F_{s,x}[k]^2+F_{s,y}[k]^2+F_{s,z}[k]^2}}\right).$$

Several features are computed in order to characterize the distribution of the angular deviations $\delta_{F,s}[k]$, $k = k_0,\cdots,k_0+N-1$. The angular deviations can take values between 0 to π. First, the maximum value $Max(\delta_{F,s})$ is computed in order to characterize the support of the distribution. Next, the mean value $Mean(\delta_{F,s})$ and the standard deviation $Std(\delta_{F,s})$ are estimated in order to characterize the central tendency and the dispersion of the distribution, respectively. Then, the skewness $Skew(\delta_{F,s})$ and the kurtosis $Kurt(\delta_{F,s})$ are estimated in order to characterize the asymmetry and the peakedness of the distribution. Finally, the probability density function of the angular deviations is estimated using kernel-based method $KS(\delta_{F,s})$. The ksdensity is a continuous function and for that reason it has been preferred to the histogram which is a discrete function. All the maximum and minimum values of the ksdensity are extracted as relevant features.

$$Max(\delta_{F,s}) = \underset{k=k_0,\cdots,k_0+N-1}{\arg\max}\left(\delta_{F,s}[k]\right)$$
$$Mean(\delta_{F,s}) = \frac{1}{N}\sum_{k=k_0}^{k_0+N-1}\delta_{F,s}[k]$$
$$Std(\delta_{F,s}) = \left(\frac{1}{N-1}\sum_{k=k_0}^{k_0+N-1}\left(\delta_{F,s}[k]-Mean(\delta_{F,s})\right)^2\right)^{1/2}$$
$$Skew(\delta_{F,s}) = \frac{n}{(n-1)(n-2)}\frac{\sum_{k_0}^{k_0+N-1}\left(\delta_{F,s}[k]-Mean(\delta_{F,s})\right)^3}{\left(\sum_{k=k_0}^{k_0+N-1}\left(\delta_{F,s}[k]-Mean(\delta_{F,s})\right)^2\right)^{3/2}}$$
$$Kurt(\delta_{F,s}) = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\frac{\sum_{k=k_0}^{k_0+N-1}\left(\delta_{F,s}[k]-Mean(\delta_{F,s})\right)^4}{\left(\sum_{k=k_0}^{k_0+N-1}\left(\delta_{F,s}[k]-Mean(\delta_{F,s})\right)^2\right)^2} - 3\frac{(n-1)^2}{(n-2)(n-3)}.$$

Ksdensity function computes a probability density estimate of the input vector. A typical probability density function (PDF) is:

$$PDF(y) = \frac{1}{N\cdot A}\cdot\sum_{i=1}^{N}K\left(\frac{y-X_i}{h}\right)$$

where N is the number of the samples of PDF, A is a normalization factor, the function K(…) is a Gaussian, X_i are 100 samples between y_min and y_max and h is a value equivalent to a covariance calculated according to the number of the samples.

Besides characterizing the statistical distribution of the angular deviations of the sequence of force samples to the mean force within the time region of interest, the feature extraction aims also at modelling the time correlation of the sequence of angular deviations. Such information can be provided in a compact form as the coefficients of an auto-regressive (AR) model fitting to the sequence of angular deviations. This model assumes that every angular deviation can be merely predicted by the past values, that is,

$$\delta_{F,s}[k] = \sum_{p=1}^{P} - a_p \delta_{F,s}[k-p] + \varepsilon[k]$$

where the coefficients $(a_1, \cdots, a_P)$ denote the AR coefficients. These parameters are classically estimated by minimising the mean square error between the observed angular deviations and their predicted values over the entire analysis frame,

$$(\hat{a}_1, \cdots, \hat{a}_P) = \arg\min_{a_1, \cdots, a_P} \sum_{k=k_0+P}^{k_0+N-1} \left(\delta_{F,s}[k] + \sum_{p=1}^{P} a_p \delta_{F,s}[k-p] \right)^2 .$$

The AR parameters can be obtained as the solutions to the set of Yule-Walker linear equations by estimating the correlation coefficients of the angular deviation sequence up to the P-th order and applying the Levinson-Durbin recursive algorithm. Clearly, the goodness-of-fit improves as the number of parameters increases. The AR model order is chosen according to the Akaike information criterion (AIC) in order to find the best tradeoff between goodness-of-fit and model complexity, and to avoid overfitting the model to the data,

$$AIC = \frac{2P}{N} + \log \sum_{k=k_0+P}^{k_0+N-1} \left(\delta_{F,s}[k] + \sum_{p=1}^{P} a_p \delta_{F,s}[k-p] \right)^2 - \log N.$$

Actually, BIC is very similar to AIC and it will not make big difference. Anyway, this criterion is used to have a rough idea about the AR model order to be used. True the exact order estimation is treated somewhat in the same way by the different techniques. They will all underestimate the true order. For the limited sample size one can expect an underestimation in 60% of the cases (under the hypothesis that the time series corresponds with an AR process). Significant differences between different information criteria only occur when enough samples are available (e.g. at least 200). In that case BIC tends to outperform AIC (and other methods).

Same features are extracted for the angular deviations of the sequence of the torque samples to the mean torque vector.

6.3.3 First-order Angular Deviation of Effort Series

Additional information on stroke patient ability in controlling generated forces/torques is expected to be found in the angular deviations between successive effort samples within the analysis frame. Given a recording, for the s-th sensor, the angular deviation $\varphi_{F,s}[k]$ between the k-th force sample $(F_{s,x}[k], F_{s,y}[k], F_{s,z}[k])$ and the (k-1)-th force sample $(F_{s,x}[k-1], F_{s,y}[k-1], F_{s,z}[k-1])$, within the analysis frame $k = k_0 + 1, \cdots, k_0 + N - 1$, is computed as the inverse cosine of the normalized scalar product, *i.e.* the dot product of the corresponding unit-norm vectors,

$$
\begin{aligned}
\vec{a} &= \left(F_{s,x}[k], F_{s,y}[k], F_{s,z}[k]\right) \\
\vec{b} &= \left(F_{s,x}[k-1], F_{s,y}[k-1], F_{s,z}[k-1]\right) \\
\delta_{F,s}[k] &= \arccos\left(\frac{\vec{a}}{\|\vec{a}\|}\cdot\frac{\vec{b}}{\|\vec{b}\|}\right) \\
&= \arccos\left(\frac{F_{s,x}[k]F_{s,x}[k-1] + F_{s,y}[k]F_{s,y}[k-1] + F_{s,z}[k]F_{s,z}[k-1]}{\sqrt{F_{s,x}[k-1]^2 + F_{s,y}[k-1]^2 + F_{s,z}[k-1]^2}\sqrt{F_{s,x}[k]^2 + F_{s,y}[k]^2 + F_{s,z}[k]^2}}\right).
\end{aligned}
$$

Several features are computed in order to characterize the distribution of the angular deviations $\varphi_{F,s}[k]$, $k = k_0 + 1, \cdots, k_0 + N - 1$. The angular deviations can take values between 0 to π. First, the maximum value $Max(\varphi_{F,s})$ is computed in order to characterize the support of the distribution. Next, the mean value $Mean(\varphi_{F,s})$ and the standard deviation $Std(\varphi_{F,s})$ are estimated in order to characterize the central tendency and the dispersion of the distribution, respectively. Then, the skewness $Skew(\varphi_{F,s})$ and the kurtosis $Kurt(\varphi_{F,s})$ are estimated in order to characterize the asymmetry and the peakedness of the distribution. Finally, the probability density function of the angular deviations is estimated using kernel-based method $KS(\varphi_{F,s})$. Again, all the maximum and minimum values of the ksdensity are extracted as relevant features.

$$
\begin{aligned}
Max(\varphi_{F,s}) &= \underset{k=k_0+1,\cdots,k_0+N-1}{\arg\max}\left(\varphi_{F,s}[k]\right) \\
Mean(\varphi_{F,s}) &= \frac{1}{N}\sum_{k=k_0+1}^{k_0+N-1}\varphi_{F,s}[k] \\
Std(\varphi_{F,s}) &= \left(\frac{1}{N-1}\sum_{k=k_0+1}^{k_0+N-1}\left(\varphi_{F,s}[k] - Mean(\varphi_{F,s})\right)^2\right)^{1/2} \\
Skew(\varphi_{F,s}) &= \frac{n}{(n-1)(n-2)}\frac{\sum_{k=k_0+1}^{k_0+N-1}\left(\varphi_{F,s}[k] - Mean(\varphi_{F,s})\right)^3}{\left(\sum_{k=k_0+1}^{k_0+N-1}\left(\varphi_{F,s}[k] - Mean(\varphi_{F,s})\right)^2\right)^{3/2}} \\
Kurt(\varphi_{F,s}) &= \frac{n(n+1)}{(n-1)(n-2)(n-3)}\frac{\sum_{k=k_0+1}^{k_0+N-1}\left(\varphi_{F,s}[k] - Mean(\varphi_{F,s})\right)^4}{\left(\sum_{k=k_0+1}^{k_0+N-1}\left(\varphi_{F,s}[k] - Mean(\varphi_{F,s})\right)^2\right)^2} - 3\frac{(n-1)^2}{(n-2)(n-3)}.
\end{aligned}
$$

Besides characterizing the statistical distribution of the first-order angular deviations of the sequence of force samples within the time region of interest, the feature extraction aims also at modelling time correlation. Such information can be provided in a compact form as the order and the coefficients of an auto-regressive (AR) model fitting to the sequence of angular deviations, as described previously.

Same features are extracted for the angular deviations between successive torque samples.

6.3.4 Cumulative Sum of Effort Series

The integrals of the effort signals are expected to convey some information on the velocity of the imaginary movements, thereof on the stroke patient ability to perform some movement velocity patterns. More especially, the norm of the integral of the force/torque sample sequence is used. Given a recording, for the *s*-th sensor, the norm $\left\|\vec{\gamma}_{F,s}[k]\right\|$ of the integral vector $\vec{\gamma}_{F,s}[k]$ of the force sample vector sequence at the *k*-th time instant, within the analysis frame $k = k_0, \cdots, k_0 + N - 1$, is computed as the norm of the cumulative sum of the force sample vector from the k_0-th time instant up to the *k*-th time instant,

$$\gamma_{F,s,x}[k] = \sum_{l=k_0}^{k} F_{s,x}[l]$$

$$\gamma_{F,s,y}[k] = \sum_{l=k_0}^{k} F_{s,y}[l]$$

$$\gamma_{F,s,z}[k] = \sum_{l=k_0}^{k} F_{s,z}[l]$$

$$\vec{\gamma}_{F,s}[k] = \left(\gamma_{F,s,x}[k], \gamma_{F,s,y}[k], \gamma_{F,s,z}[k]\right)$$

$$\left\|\vec{\gamma}_{F,s}[k]\right\| = \sqrt{\gamma_{F,s,x}[k]^2 + \gamma_{F,s,y}[k]^2 + \gamma_{F,s,z}[k]^2}.$$

Several features are computed in order to characterize the distribution of the norms of the integral force vectors $\left\|\vec{\gamma}_{F,s}[k]\right\|$, $k = k_0, \cdots, k_0 + N - 1$. First, the mean value $Mean(\left\|\vec{\gamma}_{F,s}\right\|)$ and the standard deviation $Std(\left\|\vec{\gamma}_{F,s}\right\|)$ are estimated in order to characterize the central tendency and the dispersion of the distribution, respectively. Next, the skewness $Skew(\left\|\vec{\gamma}_{F,s}\right\|)$ and the kurtosis $Kurt(\left\|\vec{\gamma}_{F,s}\right\|)$ are estimated in order to characterize the asymmetry and the peakedness of the distribution. Finally, the probability density function of the angular deviations (note that an angle is not defined in this paragraph) is estimated using kernel-based method $KS(\left\|\vec{\gamma}_{F,s}\right\|)$.

$$Mean\left(\left\|\vec{\gamma}_{F,s}\right\|\right)=\frac{1}{N}\sum_{k=k_0}^{k_0+N-1}\left\|\vec{\gamma}_{F,s}[k]\right\|$$

$$Std\left(\left\|\vec{\gamma}_{F,s}\right\|\right)=\left(\frac{1}{N-1}\sum_{k=k_0}^{k_0+N-1}\left(\left\|\vec{\gamma}_{F,s}[k]\right\|-Mean\left(\left\|\vec{\gamma}_{F,s}\right\|\right)\right)^2\right)^{1/2}$$

$$Skew\left(\left\|\vec{\gamma}_{F,s}\right\|\right)=\frac{n}{(n-1)(n-2)}\frac{\sum_{k=k_0}^{k_0+N-1}\left(\left\|\vec{\gamma}_{F,s}[k]\right\|-Mean\left(\left\|\vec{\gamma}_{F,s}\right\|\right)\right)^3}{\left(\sum_{k=k_0}^{k_0+N-1}\left(\left\|\vec{\gamma}_{F,s}[k]\right\|-Mean\left(\left\|\vec{\gamma}_{F,s}\right\|\right)\right)^2\right)^{3/2}}$$

$$Kurt\left(\left\|\vec{\gamma}_{F,s}\right\|\right)=\frac{n(n+1)}{(n-1)(n-2)(n-3)}\frac{\sum_{k=k_0}^{k_0+N-1}\left(\left\|\vec{\gamma}_{F,s}[k]\right\|-Mean\left(\left\|\vec{\gamma}_{F,s}\right\|\right)\right)^4}{\left(\sum_{k=k_0}^{k_0+N-1}\left(\left\|\vec{\gamma}_{F,s}[k]\right\|-Mean\left(\left\|\vec{\gamma}_{F,s}\right\|\right)\right)^2\right)^{2}}-3\frac{(n-1)^2}{(n-2)(n-3)}.$$

Besides characterizing the statistical distribution of the norms of the integral force vectors within the time region of interest, the feature extraction aims also at modelling time correlation. Such information can be provided in a compact form as the order and the coefficients of an auto-regressive (AR) model fitting to the sequence of angular deviations, as described previously.

Same features are extracted for the norm of the integral torque vector. Note that we cannot give the real interpretation of movement while the objects are fixed. A constant force implies a linear increase in speed under the imagined situation of free moving objects, hence imaginary movement. This parameter would probably very interesting with free moving objects. The usefulness of this parameter can be doubtful in the situation of fixed objects, but we can give it a try anyway. It will serve as a kind of lowpass filtering on the data and should probably be motivated as such.

6.3.5 Cross-Sensor Time Delay Estimation

Different approaches can be taken here. Either measure the difference in time between the initiation points from the 2 different sensors. This is very simple to calculate, while the movement initiation points are already estimated from a particular method. Another more statistically founded method is to calculate the delay between different sensors under which the mutual information between different sensors is maximized. Inspiration should be found here from image registration. In order not to calculate the mutual information between every possible component (X, Y or Z) with every other component from another sensor, one could e.g. correlate the energy or magnitude between different sensors. With means of the mutual information the optimal delay $a_{optimal}$ can be found as:

$$a_{optimal}=\max_{a} I(\left\|\vec{F}_{s1}(k)\right\|,\left\|\vec{F}_{s2}(k-a)\right\|)$$

The mutual information under this optimal delay $a_{optimal}$ could be a useful feature:

$$I_{a\_optimal}=I(\left\|\vec{F}_{s1}(k)\right\|,\left\|\vec{F}_{s2}(k-a_{optimal})\right\|)$$

This mutual information can be a measure of how good synchronization is between different movements from different sensors.
Mutual information can be computed from historgrams or kernel density estimation. The mutual information can be computed as:

$$I(\|\vec{F}_{s1}(k)\|, \|\vec{F}_{s2}(k-a)\|)$$

$$= \sum_{\|\vec{F}_{s1}\|} \sum_{\|\vec{F}_{s2}\|} p(\|\vec{F}_{s1}\|, \|\vec{F}_{s2}\|) \ln\left(\frac{p(\|\vec{F}_{s1}\|, \|\vec{F}_{s2}\|}{p(\|\vec{F}_{s1}\|) p(\|\vec{F}_{s2}\|)} \right)$$

$$= \sum_{k=k\min}^{k\max} p(\|\vec{F}_{s1}(k)\|, \|\vec{F}_{s2}(k-a)\|) \ln\left(\frac{p(\|\vec{F}_{s1}(k)\|, \|\vec{F}_{s2}(k-a)\|}{p(\|\vec{F}_{s1}(k)\|) p(\|\vec{F}_{s2}(k-a)\|)} \right)$$

It should be noted that the time series are windowed between kmin and kmax. The joint probability distribution based on kernels can be computed as:

$$p(\|\vec{F}_{s1}(j)\|, \|\vec{F}_{s2}(j-a)\|) = \frac{1}{k\max - k\min + 1}$$

$$\sum_{i=k\min}^{k\max} \frac{1}{h_1 h_2} \mathrm{K}\left(\frac{\|\vec{F}_{s1}(j)\| - \|\vec{F}_{s1}(i)\|}{h_1}, \frac{\|\vec{F}_{s2}(j-a)\| - \|\vec{F}_{s2}(i-a)\|}{h_2}\right)$$

K is the kernel, h1 and h2 are the kernel bandwidths for the respective sensors.
Note that the data in the mutual information estimation is used twice: once in building the density estimate and secondly in evaluating the density at these data points.

7 Output data

The APT generates, through the AFEM module, the following output data structure:

F.mat

with the following format:

F =

Identification:	[1x1 struct]
PreProcessing:	[1x1 struct]
Features:	[1x1 struct]

It is a data structure containing the all the features of interest for the recording, given as input to the AFEM module. The AFEM module is implemented by a Matlab file, named afem.m, which has the following syntax:

[F]=afem(filename,n)

The filename contains the full path of the measurement (e.g. C:\MATLAB6p5\work \FTdata\AHS-011\AHS-011-s13-m34.dat') on which the features will be extracted, n is the index of the time window: 1 is for the 300ms time window, 2 is for the 400ms and 3 is for the 500ms.

The extracted features for every recording is provided in a MATLAB file and the data are stored in a hierarchical structure of strings, arrays and cell arrays containing identification information as well. The format is as much self-explanatory as possible, that is, every stored feature presents a description and a value. For example, considering the raw data file AHS-010-s01-m1.dat, the corresponding feature file AHS-010-s01-m1.fea contains the following data structure:

F.Identification.RawDataFile.Description

Raw data filename used to compute the features

F.Identification.RawDataFile.Value

string vector

F.Identification.SiteID.Description

Identifier of the measurement site (possible values are [0] AHS, [1] NIMR, [2] TCD)

F.Identification.SiteID.Value

integer scalar

F.Identificiation.PatientID.Description

Identifier of the stroke patient

F.Identificiation.PatientID.Value

integer scalar

F.Identificiation.SessionID.Description

Identifier of the measurement session

F.Identificiation.SessionID.Value

integer scalar

F.Identificiation.TaskID.Description

Identifier of the task (possible values are [0] drinking a glass, [1] taking a spoon, [2] turning a key, [3] lifting a bag, [4] reaching a bottle, [5] lifting and carrying a bottle)

F.Identificiation.TaskID.Value

integer scalar

F.Identificiation.AttemptID.Description

Identifier of the attempt (possible values are [0] resting, [1] video, [2,3,4] repetitions)

F.Identificiation.AttemptID.Value

integer scalar

F.PreProcessing.SampleRate.Description

Raw data sampling rate in Hz

F.PreProcessing.SampleRate.Value

double scalar

F.Preprocessing.LowPass.Description

Low-pass filtering for noise reduction (value corresponds to low-pass frequency, infinite value means not low-pass filtering)

F.Preprocessing.LowPass.Description

double scalar

F.Features.Force.Description

Cell array of features computed on the force measurements (s=1,...,6)

F.Features. Force{s}.MeanDirection.Phi.Description

Colatitude angle of the mean force vector within the time region of interest

F.Features. Force{s}.MeanDirection.Phi.value

double scalar

F.Features. Force{s}.MeanDirection.Theta.Description

Azimuth angle of the mean force vector within the time region of interest

F.Features. Force{s}.MeanDirection.Theta.value

double scalar

F.Features. Force{s}.AngleToMeanDirection.Max.Description

Maximum value of the angular deviation between the force sample vector and the mean force vector within the time region of interest

F.Features. Force{s}.AngleToMeanDirection.Max.Value

double scalar

F.Features. Force{s}.AngleToMeanDirection.Mean.Description

Mean value of the angular deviation between the force sample vector and the mean force vector within the time region of interest

F.Features. Force{s}.AngleToMeanDirection.MeanValue

double scalar

F.Features. Force{s}.AngleToMeanDirection.Std.Description

Standard deviation value of the angular deviation between the force sample vector and the mean force vector within the time region of interest

F.Features. Force{s}.AngleToMeanDirection.Std.Value

double scalar

F.Features. Force{s}.AngleToMeanDirection.Skew.Description

Skewness value of the angular deviation between the force sample vector and the mean force vector within the time region of interest

F.Features. Force{s}.AngleToMeanDirection.Skew.Value

double scalar

F.Features. Force{s}.AngleToMeanDirection.Kurt.Description

Kurtosis value of the angular deviation between the force sample vector and the mean force vector within the time region of interest

F.Features. Force{s}.AngleToMeanDirection.Kurt.Value

double scalar

F.Features. Force{s}.AngleToMeanDirection.KS.Description

Maximum/minimum values of the PDF.
In the first and the second columns there are respectively
the maximum/minimum point (index) and its relative value.
The last column contains a 1 for a maximum, a 0 for a minimum
and a -1 for an unidentified value.

F.Features. Force{s}.AngleToMeanDirection.KS.Value

double vector

F.Features. Force{s}.AngleToMeanDirection.AROrder.Description

Order of AR model of the angular deviation between
the force sample vector and the mean force vector
within the time region of interest

F.Features. Force{s}.AngleToMeanDirection.AROrder.Value

integer scalar

F.Features. Force{s}.AngleToMeanDirection.ARCoefficients.Description

Coefficients of AR model of the angular deviation between
the force sample vector and the mean force vector
within the time region of interest

F.Features. Force{s}.AngleToMeanDirection.ARCoefficients.Value

double vector

F.Features. Force{s}.AngleToPreviousDirection.Max.Description

Maximum value of the angular deviation between
a force sample vector and the previous force vector
within the time region of interest

F.Features. Force{s}.AngleToMeanDirection.Max.Value

double scalar

F.Features. Force{s}.AngleToPreviousDirection.Mean.Description

Mean value of the angular deviation between
a force sample vector and the previous force vector
within the time region of interest

F.Features. Force{s}.AngleToPreviousDirection.MeanValue

double scalar

F.Features. Force{s}.AngleToPreviousDirection.Std.Description

Standard deviation value of the angular deviation between
a force sample vector and the previous force vector
within the time region of interest

F.Features. Force{s}.AngleToPreviousDirection.Std.Value

double scalar

F.Features. Force{s}.AngleToPreviousDirection.Skew.Description

Skewness value of the angular deviation between a force sample vector and the previous force vector within the time region of interest

F.Features. Force{s}.AngleToPreviousDirection.Skew.Value

double scalar

F.Features. Force{s}.AngleToPreviousDirection.Kurt.Description

Kurtosis value of the angular deviation between a force sample vector and the previous force vector within the time region of interest

F.Features. Force{s}.AngleToMeanDirection.Kurt.Value

double scalar

F.Features. Force{s}.AngleToPreviousDirection.KS.Description

Maximum/minimum values of the PDF.
In the first and the second columns there are respectively the maximum/minimum point (index) and its relative value.
The last column contains a 1 for a maximum, a 0 for a minimum and a -1 for an unidentified value.

F.Features. Force{s}.AngleToPreviousDirection.KS.Value

double vector

F.Features. Force{s}.AngleToPreviousDirection.AROrder.Description

Order of AR model of the angular deviation between a force sample vector and the previous force vector within the time region of interest

F.Features. Force{s}.AngleToPreviousDirection.AROrder.Value

interger scalar

F.Features. Force{s}.AngleToPreviousDirection.ARCoefficients.Description

Coefficients of AR model of the angular deviation between a force sample vector and the previous force vector within the time region of interest

F.Features. Force{s}.AngleToPreviousDirection.ARCoefficients.Value

double vector

F.Features. Force{s}.NormOfIntegral.Mean.Description

Mean value of the norm of the integral of the force sample vectors within the time region of interest

F.Features. Force{s}.NormOfIntegral.MeanValue

double scalar

F.Features. Force{s}.NormOfIntegral.Std.Description

Standard deviation value of the norm of the integral of the force sample vectors within the time region of interest

F.Features. Force{s}.NormOfIntegral.Std.Value

double scalar

F.Features. Force{s}.NormOfIntegral.Skew.Description

Skewness value of the norm of the integral of the force sample vectors within the time region of interest

F.Features. Force{s}.NormOfIntegral.Skew.Value

double scalar

F.Features. Force{s}.NormOfIntegral.Kurt.Description

Kurtosis value of the norm of the integral of the force sample vectors within the time region of interest

F.Features. Force{s}.NormOfIntegral.Kurt.Value

double scalar

F.Features. Force{s}.NormOfIntegral.KS.Description

Maximum/minimum values of the PDF.
In the first and the second columns there are respectively the maximum/minimum point (index) and its relative value.
The last column contains a 1 for a maximum, a 0 for a minimum and a -1 for an unidentified value.

F.Features. Force{s}.NormOfIntegral.KS.Value

double vector

F.Features. Force{s}.NormOfIntegral.AROrder.Description

Order of AR model of the norm of the integral of the force sample vectors within the time region of interest

F.Features. Force{s}.NormOfIntegral.AROrder.Value

integer scalar

F.Features. Force{s}.NormOfIntegral.ARCoefficients.Description

Coefficients of AR model of the norm of the integral of the force sample vectors within the time region of interest

F.Features. Force{s}.NormOfIntegral.ARCoefficients.Value

double vector

F.Features.CrossSensorDelay.Force.Description

Mutual information (MI) for the norm of the force vector: Value1 is the matrix with values of MI, Values2 contains the matrix with the values of optimal delay

F.Features.CrossSensorDelay.Force.Value1

matrix

F.Features.CrossSensorDelay.Force.Value2

matrix

F.Features.Torque.Description

Cell array of features computed on the torque measurements (s=1,…,6)

F.Features.Torque{s}.MeanDirection.Phi.Description

Colatitude angle of the mean torque vector within the time region of interest

F.Features.Torque{s}.MeanDirection.Phi.value

double scalar

F.Features.Torque{s}.MeanDirection.Theta.Description

Azimuth angle of the mean torque vector within the time region of interest

F.Features.Torque{s}.MeanDirection.Theta.value

double scalar

F.Features.Torque{s}.AngleToMeanDirection.Max.Description

Maximum value of the angular deviation between the torque sample vector and the mean torque vector within the time region of interest

F.Features.Torque{s}.AngleToMeanDirection.Max.Value

double scalar

F.Features.Torque{s}.AngleToMeanDirection.Mean.Description

Mean value of the angular deviation between
the torque sample vector and the mean torque vector
within the time region of interest

F.Features.Torque{s}.AngleToMeanDirection.MeanValue

double scalar

F.Features.Torque{s}.AngleToMeanDirection.Std.Description

Standard deviation value of the angular deviation between
the torque sample vector and the mean torque vector
within the time region of interest

F.Features.Torque{s}.AngleToMeanDirection.Std.Value

double scalar

F.Features.Torque{s}.AngleToMeanDirection.Skew.Description

Skewness value of the angular deviation between
the torque sample vector and the mean torque vector
within the time region of interest

F.Features.Torque{s}.AngleToMeanDirection.Skew.Value

double scalar

F.Features.Torque{s}.AngleToMeanDirection.Kurt.Description

Kurtosis value of the angular deviation between
the torque sample vector and the mean torque vector
within the time region of interest

F.Features.Torque{s}.AngleToMeanDirection.Kurt.Value

double scalar

F.Features.Torque{s}.AngleToMeanDirection.KS.Description

Maximum/minimum values of the PDF of the angular deviation between
the torque sample vector and the mean torque vector
within the time region of interest.
In the first and the second columns there are respectively
the maximum/minimum point (index) and its relative value.
The last column contains a 1 for a maximum, a 0 for a minimum
and a -1 for an unidentified value.

F.Features.Torque{s}.AngleToMeanDirection.KS.Value

double vector

F.Features.Torque{s}.AngleToMeanDirection.AROrder.Description

Order of AR model of the angular deviation between the torque sample vector and the mean torque vector within the time region of interest

F.Features.Torque{s}.AngleToMeanDirection.AROrder.Value

integer scalar

F.Features.Torque{s}.AngleToMeanDirection.ARCoefficients.Description

Coefficients of AR model of the angular deviation between the torque sample vector and the mean torque vector within the time region of interest

F.Features.Torque{s}.AngleToMeanDirection.ARCoefficients.Value

double vector

F.Features.Torque{s}.AngleToPreviousDirection.Max.Description

Maximum value of the angular deviation between a torque sample vector and the previous torque vector within the time region of interest

F.Features.Torque{s}.AngleToMeanDirection.Max.Value

double scalar

F.Features.Torque{s}.AngleToPreviousDirection.Mean.Description

Mean value of the angular deviation between a torque sample vector and the previous torque vector within the time region of interest

F.Features.Torque{s}.AngleToPreviousDirection.MeanValue

double scalar

F.Features.Torque{s}.AngleToPreviousDirection.Std.Description

Standard deviation value of the angular deviation between a torque sample vector and the previous torque vector within the time region of interest

F.Features.Torque{s}.AngleToPreviousDirection.Std.Value

double scalar

F.Features.Torque{s}.AngleToPreviousDirection.Skew.Description

Skewness value of the angular deviation between a torque sample vector and the previous torque vector within the time region of interest

F.Features.Torque{s}.AngleToPreviousDirection.Skew.Value

double scalar

F.Features.Torque{s}.AngleToPreviousDirection.Kurt.Description

Kurtosis value of the angular deviation between a torque sample vector and the previous torque vector within the time region of interest

F.Features.Torque{s}.AngleToMeanDirection.Kurt.Value
double scalar
F.Features.Torque{s}.AngleToPreviousDirection.KS.Description

Maximum/minimum values of the PDF of the angular deviation between a torque sample vector and the previous torque vector within the time region of interest. In the first and the second columns there are respectively the maximum/minimum point (index) and its relative value. The last column contains a 1 for a maximum, a 0 for a minimum and a -1 for an unidentified value.

F.Features.Torque{s}.AngleToPreviousDirection.KS.Value

double vector

F.Features.Torque{s}.AngleToPreviousDirection.AROrder.Description

Order of AR model of the angular deviation between a torque sample vector and the previous torque vector within the time region of interest

F.Features.Torque{s}.AngleToPreviousDirection.AROrder.Value

interger scalar

F.Features.Torque{s}.AngleToPreviousDirection.ARCoefficients.Description

Coefficients of AR model of the angular deviation between a torque sample vector and the previous torque vector within the time region of interest

F.Features.Torque{s}.AngleToPreviousDirection.ARCoefficients.Value

double vector

F.Features.Torque{s}.NormOfIntegral.Mean.Description

Mean value of the norm of the integral of the torque sample vectors within the time region of interest

F.Features.Torque{s}.NormOfIntegral.MeanValue

double scalar

F.Features.Torque{s}.NormOfIntegral.Std.Description

Standard deviation value of the norm of the integral of the torque sample vectors within the time region of interest

F.Features.Torque{s}.NormOfIntegral.Std.Value

double scalar

F.Features.Torque{s}.NormOfIntegral.Skew.Description

Skewness value of the norm of the integral of the torque sample vectors within the time region of interest

F.Features.Torque{s}.NormOfIntegral.Skew.Value

double scalar

F.Features.Torque{s}.NormOfIntegral.Kurt.Description

Kurtosis value of the norm of the integral of the torque sample vectors within the time region of interest

F.Features.Torque{s}.NormOfIntegral.Kurt.Value

double scalar

F.Features.Torque{s}.NormOfIntegral.KS.Description

Maximum/minimum values of the PDF of the norm of the integral of the torque sample vectors within the time region of interest.
In the first and the second columns there are respectively the maximum/minimum point (index) and its relative value.
The last column contains a 1 for a maximum, a 0 for a minimum and a -1 for an unidentified value.

F.Features.Torque{s}.NormOfIntegral.KS.Value

double vector

F.Features.Torque{s}.NormOfIntegral.AROrder.Description

Order of AR model of the norm of the integral of the torque sample vectors within the time region of interest

F.Features.Torque{s}.NormOfIntegral.AROrder.Value

integer scalar

F.Features.Torque{s}.NormOfIntegral.ARCoefficients.Description

Coefficients of AR model of the norm of the integral of the torque sample vectors within the time region of interest

F.Features.Torque{s}.NormOfIntegral.ARCoefficients.Value

double vector

F.Features.CrossSensorDelay.Torque.Description

Mutual information (MI) for the norm of the torque vector: Value1 is the matrix with values of MI, Values2 contains the matrix with the values of optimal delay

F.Features.CrossSensorDelay.Torque.Value1

matrix

F.Features.CrossSensorDelay.Torque.Value2

matrix

8 Appendix A - Examples of feature extraction on real sample clinical data

The following features have been computed on a sample measurement from a patient (TCD-014-s01-m23.dat)

The features described in par 7.3.1 (Mean effort direction) and par. 7.3.2 (Angular deviation to mean effort) are computed using the angdevmean.m Matlab script.
It presents the following syntax:

```
function [paramF, paramT, arF, arT, MinMaxPDF_F, MinMaxPDF_T]=angdevmean(ton,n,p)
```

Input parameters

- ton is the onset time to be passed as input from the onset detection algorithm
- n is the index of the time window: 1 is for the 300ms time window, 2 is for the 400ms and 3 is for the 500ms
- p is the order of the AR model

Output parameters

paramF is the vector containing the features calculated on the angular deviations force vector

paramT is the vector containing the features calculated on the angular deviations torque vector

arF, arT contain the estimates of a p-th AR model coefficients using the Yule-Walker method

MinMaxPDF_F, MinMaxPDF_T contain the maximum and minimum values of the PDF, computed on the vectors of angular deviations.

The paramF and paramT vectors have the following format:

```
paramF=[MaxdF MeandF StddF SkewdF KurtdF colatF azimF];
paramT=[MaxdT MeandT StddT SkewdT KurtdT colatT azimT];
```

where MaxdF and MaxdT are the maximum values, MeandF and MeandT are the mean values, StddF and StddT are the standard deviations, SkewdF and SkewdT are the values for the skewness, KurtdF and KurtdT are the values for the kurtosis, all computed on the vector of angular deviations from the mean, for the forces and for the torques signals, respectively.

colatF and colatT corresponds to the colatitude angles, as defined in paragraph 7.3.1 of this document.

azimF and azimT corresponds to the azimuth angles, as defined in paragraph 7.3.1 of this document.

Example:
[paramF, paramT, arF, arT, MinMaxPDF_F, MinMaxPDF_T]=angdevmean(80,2,3)

paramF = 0.0770 0.0125 0.0211 1.5783 4.2876 2.7993 0.1398
paramT = 0.1319 0.0185 0.0308 1.5487 4.4981 1.4096 0.5769
arF = 1.0000 -0.9529 0.0291 0.0066
arT = 1.0000 -0.9286 0.0073 0.0001
MinMaxPDF_F = 0.0068 1.0000
MinMaxPDF_T = 0.0117 1.0000

The features described in par 7.3.4 (First order angular deviations) are computed using the angdevmean.m Matlab script. It presents the following syntax:

function [paramF, paramT, arF, arT, MinMaxPDF_F, MinMaxPDF_T]=angdev (ton,n,p)

Input parameters

- ton is the onset time to be passed as input from the onset detection algorithm
- n is the index of the time window: 1 is for the 300ms time window, 2 is for the 400ms and 3 is for the 500ms
- p is the order of the AR model

Output parameters

paramF is the vector containing the features calculated on the angular deviations between successive effort samples for force vector

paramT is the vector containing the features calculated on the angular deviations between successive effort samples for the torque vector

arF, arT contain the estimates of a p-th AR model coefficients using the Yule-Walker method

MinMaxPDF_F, MinMaxPDF_T contain the maximum and minimum values of the PDF, computed on the vectors of angular deviations between successive effort samples.

The paramF and paramT vectors have the following format:

paramF=[MeandF StddF SkewdF KurtdF];

paramT=[MeandT StddT SkewdT KurtdT];

where MeandF and MeandT are the mean values, StddF and StddT are the standard deviations, SkewdF and SkewdT are the values for the skewness, KurtdF and KurtdT are the values for the kurtosis, all computed on the vector of angular deviations between successive effort samples for the forces and for the torques signals, respectively.

Example:
[paramF, paramT, arF, arT, MinMaxPDF_F, MinMaxPDF_T]=angdev(80,2,3)

paramF = 0.0048 0.0018 -0.0366 2.0280
paramT = 0.0078 0.0054 0.6539 2.2084
arF = 1.0000 -0.7664 -0.0499 -0.1207
arT = 1.0000 -1.0820 0.1783 -0.0283
MinMaxPDF_F = 0.0044 1.0000
MinMaxPDF_T = 0.0043 1.0000

The features described in paragraph 7.3.5 (Cumulative sum of effort series) are computed using the csum.m Matlab script. It presents the following syntax:

function [paramF, paramT, arF, arT]=csum (ton,n,p)

Input parameters

- ton is the onset time to be passed as input from the onset detection algorithm
- n is the index of the time window: 1 is for the 300ms time window, 2 is for the 400ms and 3 is for the 500ms
- p is the order of the AR model

Output parameters

paramF is the vector containing the features calculated on the angular deviations between successive effort samples for force vector

paramT is the vector containing the features calculated on the angular deviations between successive effort samples for the torque vector

arF, arT contain the estimates of a p-th AR model coefficients using the Yule-Walker method

The paramF and paramT vectors have the following format:

paramF=[MeandF StddF SkewdF KurtdF];

paramT=[MeandT StddT SkewdT KurtdT];

where MeandF and MeandT are the mean values, StddF and StddT are the standard deviations, SkewdF and SkewdT are the values for the skewness, KurtdF and KurtdT are the values for the kurtosis, all computed on the vector of angular deviations between successive effort samples for the forces and for the torques signals, respectively.

Example:
[paramF, paramT, arF, arT]=csum(80,2,3)

paramF = 519.4633 291.8396 -0.0042 1.8233
paramT = 36.5332 20.3733 -0.0635 1.7917
arF = 1.0000 -0.9794 0.0006 0.0171
arT = 1.0000 -0.9829 0.0006 0.0192

Matlab scripts for the calculation of the cross sensor time delay (paragraph 7.3.6) and for the identification of the best order for the AR model are in progress and have to be refined in some parts, in order to ensure the maximum reliability.

In a first implementation of the method for the calculation of the cross sensor time delay, the computation of the estimate of the probability density function for each sensor and the estimate of the joint probability density function is computed. The function which computes the estimate of the mutual information has the following format:

function [MI_est_f, a_opt_f, MI_est_t,a_opt_t]=mutualinfo(in1,in2,ton_f,ton_t,n)

where in1 and in2 are the data from two different sensors, ton_f and ton_t are the onset times, calculated using one of the four proposed onset detection methods (for the preliminary results presented in the present document, the 2% rule has been used) and n is corresponding to the time window of interest.

MI_est_f, MI_est_t , a_opt_f and a_opt_t are a 8by8 symmetric matrices.

MI_est_f and MI_est_t contain the values of the mutual information, according to the following meaning:
MI_est_f (i,j) and MI_est_t (i,j) are the value of the mutual information between the sensor i and the sensor j for the force vector and the torque vector, respectively, according to the definition given in the Section 8 of this document.

a_opt_f and a_opt_t contain the values of the time delay corresponding to the related value of the mutual information, according to the following meaning:
a_opt_f (i,j) and a_opt_t (i,j) are the value of the time delay corresponding to the related value of the mutual information between the sensor i and the sensor j for the force vector and the torque vector, respectively

The calculation of the best order to be given to the AR models can be performed according the following method: for different orders, from 1 to N, the variance estimate is computed using the aryule.m function which uses the Yule-Walker method, also called the autocorrelation method. It fits a pth order autoregressive (AR) model to the windowed input signal by minimizing the forward prediction error in the least-squares sense. This formulation leads to the Yule-Walker equations, which are solved by the Levinson-Durbin recursion.

The variance estimate is given as input to the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC) to be computed. The order which corresponds to the minimum value for AIC and BIC criteria is the best order.

The function findbestorder.m has the following format:

[AIC,BIC,P]=findbestorder(in, N)

where in is the windowed input signal, N is the maximum order on which the analysis is to be performed. AIC and BIC are the vectors of values for the AIC and BIC criteria, computed for the different orders and P is the best order.

9 Appendix B – List of APT software modules

The CD ROM enclosed with this accompanying document contains all the software modules which have been developed for the implementation of the APT organized in the following folders:

/Feature extraction
/Visualization tool
/Onset detection tool

The folder /Feature extraction/ contains the following files:

/FTData	folder containing a reference dataset which has been used to generate examples and for comparative analysis for the identification of the best technique for onset time estimation
afem.m	Main Matlab module which implements the AFEM module for feature extraction. It calls the following Matlab sub-modules:
angdev.m	Calculation of features on the angular deviations between successive effort samples
angdevmean.m	Calculation of features on the angular deviations from the mean effort
csum.m	Calculation of features on the cumulative sum of effort series
mutualinfo.m	Computation of the mutual information based on a density estimate
calc_PDF.m	Calculation of the onset time according to the ks-density
onsetmitrule.m	Calculation of the onset time according to the 2% rule
entropy.m	Sub-module used by mutualinfo.m
marginal.m	Sub-module used by mutualinfo.m
getCachedFTfile.m	Sub-module used for reading F/T data
transformCS.m	Sub-module used for the coordinate transformation
F.mat	Example of output data structure containing all the features, extracted on a sample measurement

The folder /Visualization tool/ contains the following files:

matlab visualization tool-v06.rar	Matlab visualization tool (Matlab version)
matlab visualization tool-v06 compiled.rar	Matlab visualization tool (Compiled version)

The folder /Onset detection tool/ contains the file:

onsetDetection.rar	Matlab onset detection tool (Matlab version) which has been made available to the ALLADIN clinical experts

10 Appendix C – Quick reference guide to installation and use of the APT

The Matlab script which computes and stores the features on a given measurement data set has the following format:

F = afem(filename,n)

It has two input arguments: “filename” and “n”: “filename” contains the full path of the measurement data set extracted by the Global Alladin Database (e.g. 'C:\MATLAB6p5\work\onsetDetection\FTdata\AHS-011\AHS-011-s13-m34.dat') on which the features will be computed; “n” is the index of the time window: 1 is for the 300ms window, 2 is for the 400ms and 3 is for the 500ms.

F is the output data structure containing the features, in the same format as previously described in Section 8 of this document.

By default, the onset time is calculated using the 2%-threshold rule method: once the final results from the comparative analysis with the data prepared by the clinicians will be completed, it will be replaced by the best onset detection method.

By default, the AR model order is kept as constant (p=3). Best order estimation can be performed off-line, according to the AIC and BIC criteria.

The following features are proposed and are computed by the Matlab script for the features extraction:

F.Features.CrossSensorDelay.Force.Description = 'Mutual information (MI) for the norm of the force vector: the field with Value1 is the matrix with values of MI, the field with Values2 contains the matrix with the values of optimal delay';

F.Features.CrossSensorDelay.Force.Value1 = MI_est_FORCE;
F.Features.CrossSensorDelay.Force.Value2 = A_OPT_FORCE;

F.Features.CrossSensorDelay.Torque.Description = 'Mutual information (MI) for the norm of the torque vector: the field with Value1 is the matrix with values of MI, the field with Values2 contains the matrix with the values of optimal delay';

F.Features.CrossSensorDelay.Torque.Value1 = MI_est_TORQUE;
F.Features.CrossSensorDelay.Torque.Value2 = A_OPT_TORQUE;

where MI_est_FORCE, MI_est_TORQUE , A_OPT_FORCE and A_OPT_TORQUE are a 8by8 symmetric matrices.

MI_est_FORCE and MI_est_TORQUE contain the values of the mutual information, according to the following meaning:
MI_est_FORCE(i,j) and MI_est_TORQUE(i,j) are the value of the mutual information between the sensor i and the sensor j for the force vector and the torque vector, respectively,

according to the definition given in the ALLADIN WP4 report "Feature Extraction for Force-Torque Measurement Based Therapy Assessment".

A_OPT_FORCE and A_OPT_TORQUE contain the values of the time delay corresponding to the related value of the mutual information, according to the following meaning:
A_OPT_FORCE(i,j) and A_OPT_TORQUE(i,j) are the value of the time delay corresponding to the related value of the mutual information between the sensor i and the sensor j for the force vector and the torque vector, respectively, according to the definition given in the ALLADIN WP4 report "Feature Extraction for Force-Torque Measurement Based Therapy Assessment".

WARNING: for each measurement, extracted features are stored in a structure named F. After each computation, rename the F structure and save in a separate folder. If you forget to do it, the successive computation will overwrite the F structure: features extracted on the previous measurement would be lost!

The features are stored in a structure, called F, as *F.mat*, with the following format:

F =

```
Identification:    [1x1 struct]
PreProcessing:     [1x1 struct]
Features:          [1x1 struct]
```

In order to read the features, the structure must be loaded into the Matlab environment, typing the following command in the command window:

>>load F

After this command, the structure is loaded into the workspace and can be read.

In order to read a the value or the description of a feature, the corresponding command, according the syntax given in the “WP4 ALLADIN Feature extraction” document, must be typed in the command window:

For instance, launching the *featurextract.m* with the filename 'C:\MATLAB6p5\work\FTdata\AHS-042\AHS-042-s01-m24.dat' as first argument and one index, chosen among 1,2 or 3, as second argument (time window), the command

>> **F.Identification.AttemptID.Description**

will cause the following output:

ans =
Identifier of the attempt (possible values are [0] resting, [1] video, [2,3,4] repetitions)

The command

>> **F.Identification.AttemptID.Value**

will allow to evaluate the identifier of the attempt for the measurement given as input. It will give the following output:

ans =

4

The same approach must be used to evaluate the other features, for the given measurement.

GUI_1