



**HAL**  
open science

# Pervasive informatics and persistent actimetric information in health smart homes: From Language Model to Location Model

Yannick Fouquet, Jacques Demongeot, Nicolas Vuillerme

► **To cite this version:**

Yannick Fouquet, Jacques Demongeot, Nicolas Vuillerme. Pervasive informatics and persistent actimetric information in health smart homes: From Language Model to Location Model. International Conference on Complex, Intelligent and Software Intensive Systems, Mar 2009, Fukuoka, Japan. pp.935-942, 10.1109/CISIS.2009.81 . hal-00379459

**HAL Id: hal-00379459**

**<https://hal.science/hal-00379459>**

Submitted on 29 Apr 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Pervasive informatics and persistent actimetric information in health smart homes : From Language Model to Location Model

Y. Fouquet, J. Demongeot and N. Vuillerme

**Abstract**—This paper presents an approach of location model deriving from language models existing in speech recognition research. The purpose is to applicate existing model in speech recognition to predict location of an elderly person. Using statistics, the model offers up to 98.03% of good prediction location, considering only the last second of location but distinguishing days of week. Simpler than Polya’s urns derived approach, this approach seems to be a good way of location modelling.

**Index Terms**—smart flats for elderly people, pervasive watching systems, actimetry, localization sensors,  $n$ -grams, Polya’s urns, Markov chain, Mobile and pervasive sensing, Biostatistics and time series of long recording at home, Elderly people watching

## I. Introduction

**E**RRARE humanum est,  
perseverare diabolicum.

In many neuro-degenerative diseases or post-brain stroke disorders, one can meet temporo-spatial disorientation [1], [2], [3] leading to many errors in tasks during the daily activity [4] until a pathologic perseveration [5], which causes a deep handicap in the execution of current vital functions.

The main idea in this paper is to find easy procedures to interpret surveillance at home data in order to get a good persistence (in tasks) parameter that one could use as perseveration index for triggering alarms and counselling diagnosis search for neuro-degenerative diseases or post-brain stroke complications.

This paper presents two approaches for the location modelling. The first one, described more precisely in [6] use Polya’s urns. The second one is a statistical approach based on  $n$ -grams including two back-offs levels.

The section 2 of this paper explains the two modelling approaches. The section 3 details the corpus used for our experiment. The section 4 describes the model proposed. The section 5 presents the experiment while the section 6 shows the results for the whole corpus and for each day of week. The section 7 is a discussion of these results.

Manuscript received October 15, 2008. Pervasive informatics and persistent actimetric information in health smart homes : From Language Model to Location Model. Yannick Fouquet, Jacques Demongeot, and Nicolas Vuillerme.

Y. Fouquet is with the TIMC-IMAG Laboratory, UMR CNRS 5525, Faculty of Medicine of Grenoble, University J. Fourier, 38 700 La Tronche - France, (phone: +33 4 76 63 74 08; fax: +33 4 76 76 88 44; e-mail: Yannick.Fouquet@imag.fr).

J. Demongeot is with the TIMC-IMAG Laboratory, UMR CNRS 5525, Faculty of Medicine of Grenoble, University J. Fourier, 38 700 La Tronche - France, (e-mail: Jacques.Demongeot@imag.fr).

N. Vuillerme is with the TIMC-IMAG Laboratory, UMR CNRS 5525, Faculty of Medicine of Grenoble, University J. Fourier, 38 700 La Tronche - France, (e-mail: Nicolas.Vuillerme@imag.fr).

## II. Two modelling approaches

Among the possible approaches for modelling the actimetric data, two methods have been selected. The first one focus on the Polya’s urns [6] in which the observed activity at time  $t$  is depending on the whole past (since a reset supposed to be made at the beginning of each day). The second one concerns a first order Markov chain approach in which the dependency of the future of  $t$  lies only through the present time  $t$ . In both models, a persistence parameter is defined. For deciding between these two methods, we propose to use the statistics equal to the empirical mean  $E$  of a task remaining (at time  $t$ ) duration, by identifying a task with the location at which it is performed.

### A. Polya’s urns

In the Polya’s urns approach, the location is seen as a colored ball. Each second, a ball is taken from an urn. The balls contained in the urn represent the distribution of probabilities of each location. In order to take into account the persistence in tasks, some balls - from the same color as the one taken - are added in the urn.

The main idea is to considerably simplify the information by giving a color coding number to the different locations (pertinent for the watching), and to follow up the succession of these numbers, *e.g.* by interpreting them as the succession of colors of balls drawn from a Polya’s urn: in this kind of urn, the persistence (or *a contrario* the instability) of an action in a location is represented by adding  $k_i(t)$  balls of color  $i$ , when a ball of color  $i$  has been obtained at time  $t$ .

In this approach, the persistence in task  $i$  is a parameter denoted by  $\pi_i$  and estimated by the ratio between the number  $k_i(t)$  of balls we add after the  $t^{\text{th}}$  drawing of a ball of color  $i$  in the urn, and the total number of initial balls  $b_0$  :

$$\pi_i(t) = \frac{k_i(t)}{b_0}$$

Let us denote by  $x_i(t)$  the number of times where the ball of color  $i$  has been drawn from the urn at time  $t$ , and by  $p_i(t)$  the probability to get a ball of color  $i$  at the  $(t+1)^{\text{th}}$  drawing. Then we have:

$$p_i(t) = \frac{p_i(0)b_0 + x_i(t)k_i(t)}{b_0 + tk_i(t)}$$

If we suppose the  $k_i(t)$ ’s constant, then we can estimate  $\pi_i$  from the empirical frequencies  $f_i(t)$ ’s to get a ball of color  $i$  at the  $(t+1)^{\text{th}}$  drawing (estimated in a series of days supposed to be independent), whose expectation is  $p_i(t)$ :

$$\overline{\pi_i} = \frac{f_i(0) - f_i(M)}{Mf_i(M) - x_i(M)}$$

where  $M$  is the total number of drawings done during a day. We can then calculate the estimator  $\overline{E}_{i,1}$  of the  $i^{\text{th}}$  task remaining duration by considering the empirical mean (on observed days) of the remaining duration in a day which is defined by:

$$\frac{\sum_{t=0}^M z_i(t)}{M+1},$$

where :

- $y_i(t) = x_i(t) - x_i(t-1)$  is the number (1 or 0) of balls of color  $i$  drawn at time  $t$ ,
- $z_i(t) = \max_{0 \leq m \leq (M-t)} \{m; \prod_{j=0}^m y_i(t+j) = 1\}$  is the length of the sequence of "drawing a ball of color  $i$ " (possibly 0) since a drawing at time  $t$  of a ball of color  $i$

Another way to estimate  $E_i$  consists in calculating the probability  $c_{i,m}(t)$  to have  $m$  consecutive drawing of a ball  $i$  from the drawing  $t$ :

$$\forall m \in \mathbb{N} : 0 \leq m \leq (M-t),$$

$$c_{i,m}(t) = \left(1 - p_i(t+m+1)\right) \prod_{j=0}^m p_i(t+j)$$

with :

$$p_i(M+1) = 0$$

The estimation of  $E_i$  could then be done by replacing the probabilities by the corresponding empirical frequencies:

$$E_i = \frac{\sum_{t=0}^M \sum_{m=0}^M m c_{i,m}(t)}{M+1}$$

Thus,

$$\overline{E}_{i,2} \approx \frac{\sum_{t=0}^M \sum_{m=0}^M m \left[ \left(1 - f_i(t+m+1)\right) \prod_{j=0}^m f_i(t+j) \right]}{M+1}$$

The 95%-confidence interval of  $\overline{E}_{i,2}$  could then be calculated by estimating the 95%-confidence interval of the  $f_i$ 's which is :

$$\left[ f_i \pm \sqrt{\frac{f_i(1-f_i)}{M}} \right]$$

We will reject the null-hypothesis  $H_0$ : "the persistence model is a Polya's urn model", if  $\overline{E}_{i,1}$  does not belong to this interval. Otherwise, this model could be used to represent the persistence in task.

### B. Markov chain

In the Markov chain approach, each location is a node with probabilities of transitions from one location to another one. The succession of locations is seen as a route in a Markov chain. A first order Markov chain takes into account the last location in order to predict the present one. The generalization of such a model offers to represent the probability of a location depending on the historic of locations.

In this approach, let us denote by  $p_{ij}$  the probability (supposed to be constant) to draw a ball of color  $j$  after a ball of color  $i$ . Then  $p_{ii}$  could be the persistence in task  $i$  parameter. If we denote by  $p_j$  the probability (supposed to be constant) to draw a ball of color  $j$ , we have:

$$p_j = \sum_{i=1}^k p_{ij},$$

where  $k$  is the number of colors (*i.e.* of types of task).

Moreover :

$$P(z_i = 0) = (1 - p_i)$$

and

$$\forall k \in \mathbb{N} : 1 \leq k \leq M, P(z_i = k) = p_i(1 - p_i)(p_{ii})^{k-1},$$

then the expectation of the  $i^{\text{th}}$  task remaining duration  $E_i$  could be calculated as:

$$E_i = \sum_{k=0}^M \frac{k(k+1)}{2k} P(z_i = k) = \sum_{k=0}^M \frac{k+1}{2} P(z_i = k)$$

Thus,  $E_i$  can be estimated by :

$$\overline{E}_{i,3} = \sum_{k=0}^M \frac{k+1}{2} f_i(1 - f_i)(f_{ii})^{k-1}$$

The 95%-confidence interval of  $\overline{E}_{i,3}$  could be calculated by estimating the 95%-confidence interval of the  $f_i$ 's and  $f_{ii}$ 's which are respectively :

$$\left[ f_i \pm \sqrt{\frac{f_i(1-f_i)}{M}} \right]$$

and :

$$\left[ f_{ii} \pm \sqrt{\frac{f_{ii}(1-f_{ii})}{M}} \right]$$

The 95%-confidence interval could also be more accurate by empirically calculus using min and max values of  $\overline{E}_{i,3}$  :

$$\left[ \min_{1 \leq i \leq k} \left( \overline{E}_{i,3} \right) \dots \max_{1 \leq i \leq k} \left( \overline{E}_{i,3} \right) \right]$$

We will reject the null-hypothesis  $H_0$ : "the persistence model is a first order Markov chain model", if  $\overline{E}_{i,1}$  does not belong to this interval. Otherwise, this model could be used to represent the persistence in task.

If both tests above are concluding to the acceptance, one prefers the first order Markov chain due to its simplicity. If both tests above are concluding to the rejection of the null-hypothesis, we retain the model having the closest distance between  $\overline{E}_{i,1}$  and the confidence interval of  $\overline{E}_{i,j}$  ( $j = 2, 3$ ).

The Polya's urn model has been precisely described in [6]. We will focus now on the description of the first order Markov chain model and its generalization by the  $n$ -grams approach proposed.

### III. Statistical Location Prediction

People use to settle in daily routines. Barely awake, they go up, prepare the coffee, wash, take the coffee, go to the toilet and so on. Each person as its own procedure. When people become more aged, their procedure is more and more important. Activity prediction could be helpful in order to detect variations in their comportment which could be abnormal and need further medical assistance.

The problem of describing such a procedure is difficult. People's habits are highly variable. A first dimension of variation is the difference in usages between individuals. Different people do different things at different moments of the journey, week, month, etc. The second variation is over time. Habits change as does the world. The third variation is over people state. Depending on the location or activity history as well as their health, people will act differently. Without the need to modelize all the possible variations, there is, however, the promise of a solution.

In the speech recognition domain, such problem exists for word prediction. Indeed, most of recent speech recognition systems are based on acoustic model associated to language model. For the language model, statistical approaches are often based on  $n$ -grams [7] of words for word prediction [8], [9].

In human-machine dialog as in speech translation [10], such problem exists for speech act prediction. Indeed, some systems are based on speech acts derived from works in language philosophy [11], [12], [13], [14], [15]. It assumes that a dialog can be described by means of a limited but open set of dialog acts. For helping the dialog management, statistical approaches have been proposed [16]. Some of them are based on  $n$ -grams of acts and act prediction as an indicator for both act interpretation and dialog management [17], [18].

The purpose here is to use this kind of modelling and apply  $n$ -grams theory to our location modelling problem.

### IV. Corpus description

Since 12 years, many experiments have been achieved for watching dependent people at home, in particular elderly and handicapped persons [19], [20], [21], [22], [23], [24]. Some of important things to be done are localizing a person. For acquiring data necessary to permit this localization, lots of sensors haven been invented. This sensors networks permit to represent the location of a person in a flat room (Figure 1). Recording timestamped locations permits us to create a corpus for experiments [25].

The corpus describes the location of an elderly person within his/her home environment in time. It has a structure as in the Table I: the columns represent successively the time (with the day, the month, the year, the hour, the minute and the second of the recording) and the activity-station-code corresponding to the location of the watched person at this time.

Thus, the corpus is on the form of a timestamped location. Timestamps are space separated numerals representing day of month, month, year, hour, minutes, seconds of the location captured. The location itself is a code :

- 0 for Entry hall
- 1 for Living Room

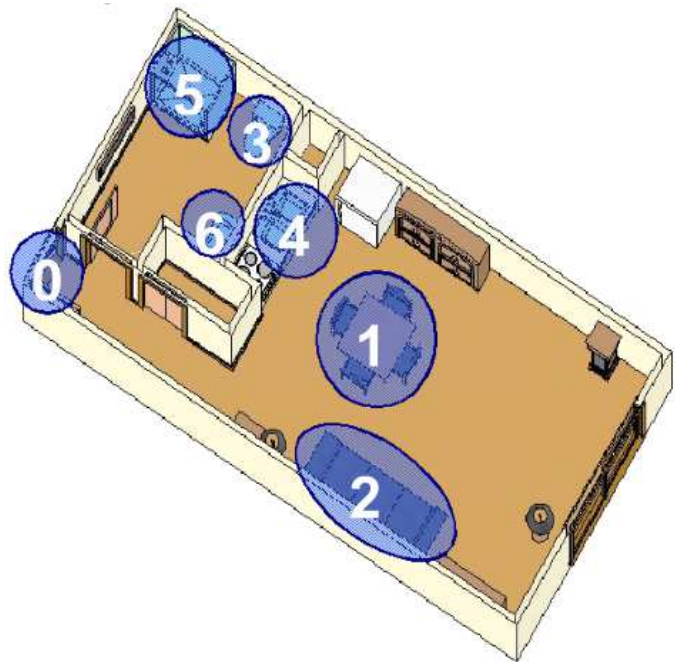


Fig. 1. Architecture of the experimental health smart home. Location sensors are placed at different places in the apartment, allowing the monitoring of individual's successive activity phases within his/her home environment: 0. Entry hall - 1. Living room - 2. Bedroom - 3. WC - 4. Kitchen - 5. Shower - 6. Washbasin.

TABLE I  
The times and locations of the records

Day	Month	Year	Hour	Minute	Second	activity-station-code
24	03	2005	17	37	36	1
24	03	2005	17	37	37	2
24	03	2005	17	37	38	2
24	03	2005	17	37	40	1
24	03	2005	17	37	48	1
24	03	2005	17	37	49	4
24	03	2005	17	37	50	4
24	03	2005	17	37	51	1
24	03	2005	17	37	53	4
24	03	2005	17	37	54	4
24	03	2005	17	37	55	4
24	03	2005	17	37	56	2
24	03	2005	17	37	57	2
24	03	2005	17	37	58	4
24	03	2005	17	37	59	1
24	03	2005	17	38	00	1
24	03	2005	17	38	02	2
24	03	2005	17	38	03	2
24	03	2005	17	38	04	2

- 2 for Bedroom
- 3 for WC
- 4 for Kitchen
- 5 for Shower
- 6 for Washbasin
- 9 for an error

An example of a line of the corpus is 18 07 2007 11 27 48 4, which suits as : on 18<sup>th</sup> of July 2007, at 11:27,48", subject was in the kitchen.

The files treated bring together the data recorded in the flat of the elderly people in a period of 10 months from the 22<sup>th</sup> of March 2005 until the 24<sup>th</sup> of January 2006 and a period of 6 months from the 18<sup>th</sup> of July 2007 to the 15<sup>th</sup> of January 2008.

From these records, a statistical approach has been applied, using  $n$ -grams location probabilities.

## V. Statistical Location Prediction model

For the location prediction a statistical method has been implemented to predict the next location on the basis of the location history [16]. Currently,  $n$ -grams location probabilities are used to compute the most likely follow up location. To predict the  $i^{\text{th}}$  location  $a_i$ , we use the  $n - 1$  previously uttered locations and determine the most probable location by computing:

$$a_i = \operatorname{argmax}_a P(a|a_{i-1}, a_{i-2} \dots, a_{i-n+1})$$

To estimate this probability the standard estimations using relative frequency techniques are used.

Otherwise, our corpus is a real-collection and, like in many real-situations, it was not possible to collect a large amount of data to properly estimate the statistics. This implies that it is not reasonable to use classical smoothing techniques. We need a solution for the two following problems:

- 1) unexpected input: the location model based on  $n$ -grams location sequences can not be used in case unexpected input occurs,
- 2) lack of training data: the  $n$ -grams model predict several locations with the same probability.

The treatment of these cases consists in using the  $(n - 1)$ -grams model, recursively.

## VI. Experimental procedure

For this experiment, the corpus has been reformatted in order to represent the location of the person, each second. A line of this 'new' corpus (example Table II) represents a journey as a series of location, each second. It is on the form of a space separated locations as a code as explained above.

For example, "s 2 2 2 ... 2 2 3 3 3 ... 3 3 4 4 4 ... e" suits as : since  $s$  the start of day, the person was in the bedroom (2), after  $x$  seconds ( $x$  is the number of successive 2), the person passed in the toilet (3), then after  $y$  seconds ( $y$  is the number of successive 3), she passed in the kitchen (4), etc. The end of journey is represented by  $e$ .

It permits us to apply our  $n$ -grams model with  $(n - 1)$  last seconds used to predict the  $n^{\text{th}}$  one. We choose to set  $n$  up to 10 so that we watch for the 9 last seconds in order to predict the 10<sup>th</sup>. The corpus has been cut into 80% for learning model, 20% for testing it. Tests have been done for an history of location set from 1 to  $n$  (10 here).

TABLE II  
The succession of locations

$t_0$	$t_1$	$t_2$	...	$t_k$	$t_{k+1}$	...	$t_{n-1}$	$t_n$
s	2	2	...	3	3	...	2	e
s	2	2	...	4	5	...	2	e
s	2	2	...	2	3	...	2	e
s	2	2	...	6	6	...	2	e
s	2	2	...	4	4	...	2	e
s	2	2	...	4	3	...	2	e
s	2	2	...	3	3	...	2	e
s	2	2	...	6	5	...	2	e
s	2	2	...	5	5	...	2	e
s	2	2	...	5	5	...	2	e

TABLE III  
Prediction result for the whole corpus

n	Number of correct predictions	Number to predict	Good prediction rate
1	983209	2011554	48.88 %
2	<b>1958535</b>	2011554	<b>97.36 %</b>
3	1958535	2011554	97.36 %
4	1958524	2011554	97.36 %
5	1958481	2011554	97.36 %
6	1958380	2011554	97.36 %
7	1958297	2011554	97.35 %
8	1958221	2011554	97.35 %
9	1958062	2011554	97.34 %
10	1957773	2011554	97.33 %

## VII. Results

### A. Global results

A first test was made with the whole corpus without data distinction (day of week, day of month, month, hour of journey, etc.).

Table III shows a best prediction with  $n = 2$ . Indeed, approximatively the same performance is obtained with  $n > 2$  (with 2 digits after the decimal point). But  $n$  does not need to be bigger than 2. The last second location is sufficient to predict the next one. Raw performance seems to decrease while  $n$  increase. This result seems to indicate that accuracy by watching too far in the past is not a good way to predict the future location of a person.

### B. Results with day of week distinction

Other tests were made considering that day of week influences activity of an elderly person. For instance, on Sunday, she wakes later and prepares herself going to church while Wednesday is dedicated to shopping.

TABLE IV  
Prediction for the Sunday

n	Number of correct predictions	Number to predict	Good prediction rate
1	154846	300105	51.60 %
2	291248	300105	97.05 %
3	291250	300105	97.05 %
4	291244	300105	97.05 %
5	<b>291258</b>	300105	<b>97.05 %</b>
6	291190	300105	97.03 %
7	291192	300105	97.03 %
8	291143	300105	97.01 %
9	291121	300105	97.01 %
10	291043	300105	96.98 %

TABLE V  
Prediction for the Monday

n	Number of correct predictions	Number to predict	Good prediction rate
1	126858	267983	47.34 %
2	261253	267983	97.49 %
3	<b>261254</b>	267983	<b>97.49 %</b>
4	261252	267983	97.49 %
5	261222	267983	97.48 %
6	261212	267983	97.47 %
7	261181	267983	97.46 %
8	261148	267983	97.45 %
9	261113	267983	97.44 %
10	261057	267983	97.42 %

For Sunday, table IV shows a best prediction with  $n = 5$  (using the 4 last locations to predict the next one). Best approximated rate (with 2 digits after the decimal point) appears since  $n = 2$ .

For Monday, table V shows a best prediction with  $n = 3$  (using the 2 last locations to predict the next one). Best approximated rate (with 2 or 3 digits after the decimal point) appears since  $n = 2$ .

For Tuesday, table VI shows a best prediction with  $n = 2$  (using the last location to predict the next one).

For Wednesday, table VII shows a best prediction with  $n = 3$  (using the 2 last locations to predict the next one). Best approximated rate (with 2 or 3 digits after the decimal point) appears since  $n = 2$ .

For Thursday, table VIII shows a best prediction with  $n = 2$  (using the last location to predict the next one).

TABLE VI  
Prediction for the Tuesday

n	Number of correct predictions	Number to predict	Good prediction rate
1	123530	289876	42.61 %
2	<b>284163</b>	289876	<b>98.03 %</b>
3	284162	289876	98.03 %
4	284156	289876	98.03 %
5	284129	289876	98.02 %
6	284105	289876	98.01 %
7	284089	289876	98.00 %
8	284057	289876	97.99 %
9	284002	289876	97.97 %
10	283950	289876	97.96 %

TABLE VII  
Prediction for the Wednesday

n	Number of correct predictions	Number to predict	Good prediction rate
1	197499	427006	46.25 %
2	416983	427006	97.65 %
3	<b>416984</b>	427006	<b>97.65 %</b>
4	416980	427006	97.65 %
5	416953	427006	97.65 %
6	416909	427006	97.64 %
7	416889	427006	97.63 %
8	416841	427006	97.62 %
9	416759	427006	97.60 %
10	416687	427006	97.58 %

TABLE VIII  
Prediction for the Thursday

n	Number of correct predictions	Number to predict	Good prediction rate
1	136005	232471	58.50 %
2	<b>226140</b>	232471	<b>97.28 %</b>
3	226129	232471	97.27 %
4	226106	232471	97.26 %
5	226099	232471	97.26 %
6	226085	232471	97.25 %
7	226070	232471	97.25 %
8	226047	232471	97.24 %
9	226024	232471	97.23 %
10	225961	232471	97.20 %

TABLE IX  
Prediction for the Friday

n	Number of correct predictions	Number to predict	Good prediction rate
1	137179	260158	52.73 %
2	<b>252605</b>	260158	<b>97.10 %</b>
3	252597	260158	97.09 %
4	252596	260158	97.09 %
5	252594	260158	97.09 %
6	252550	260158	97.08 %
7	252519	260158	97.06 %
8	252445	260158	97.04 %
9	252386	260158	97.01 %
10	252325	260158	96.99 %

TABLE X  
Prediction for the Saturday

n	Number of correct predictions	Number to predict	Good prediction rate
1	91473	199767	45.79 %
2	194207	199767	97.22 %
3	194207	199767	97.22 %
4	194200	199767	97.21 %
5	194194	199767	97.21 %
6	194192	199767	97.21 %
7	194174	199767	97.20 %
8	194126	199767	97.18 %
9	194121	199767	97.17 %
10	194085	199767	97.16 %

For Friday, table IX shows a best prediction with  $n = 2$  (using the last location to predict the next one).

For Saturday, table X shows a best prediction with  $n = 2$  (using the last location to predict the next one).

### C. Summary of results with day of week distinction

Table XI shows a best prediction approximated rate with  $n = 2$  (using the last location to predict the next one). Performance seems to decrease with  $n$  increasing. The real best performance, in bold, shows that results differ according to day of week but a good approximation could be made with  $n = 2$ . Moreover, with  $n = 2$ , results of good prediction differ according to the day of week from 97.05% on Sunday to 98.03% on Tuesday. It seems to show that day of week is an important factor of variation.

TABLE XI  
Good prediction rate (%) depending on day of week.

n	sun	mon	tue	wed	thu	fri	sat	total
1	51.60	47.34	42.61	46.25	58.50	52.73	45.79	48.88
2	97.05	97.49	<b>98.03</b>	97.65	<b>97.28</b>	<b>97.10</b>	<b>97.22</b>	<b>97.36</b>
3	97.05	<b>97.49</b>	98.03	<b>97.65</b>	97.27	97.09	97.22	97.36
4	97.05	97.49	98.03	97.65	97.26	97.09	97.21	97.36
5	<b>97.05</b>	97.48	98.02	97.65	97.26	97.09	97.21	97.36
6	97.03	97.47	98.01	97.64	97.25	97.08	97.21	97.36
7	97.03	97.46	98.00	97.63	97.25	97.06	97.20	97.35
8	97.01	97.45	97.99	97.62	97.24	97.04	97.18	97.35
9	97.01	97.44	97.97	97.60	97.23	97.01	97.17	97.34
10	96.98	97.42	97.96	97.58	97.20	96.99	97.16	97.33

TABLE XII  
Empirical frequencies  $f_i$  (%)

0	1	2	3	4	5	6	9	e
10.27	21.2	<b>53.18</b>	1.29	6.13	1.67	5.13	1.12	0.01

## VIII. Discussion

First results tend to show best performances occurring in the first order Markov case with  $n = 2$ , and a degradation of performances with  $n$  increasing up to 10. This seems to indicate that watching more far in time is more accurate but a bad way to predict the future location of the person. For the confirmation of this trend, experiment should be applied for  $n = 60$ , watching for the whole last minute in order to predict the 60<sup>th</sup> second.

Moreover, performance seems to differ for each day of week. This factor of variability should be taken into account when designing a system using a location model. New experiments should be made for other comparisons. The distinction of each day of the month could show that some days, as 1<sup>st</sup> of the month for example, are particular. The comparison between each month should show that activity differs from summer to winter, and so on.

It could then be interesting to develop a new model with a continuum approach considering estimations (interpolation) between data observed.

The model of the location of a person seems to be well approximated by a Markovian process. A first order Markov chain is sufficient in order to represent the probabilities of transitions from locations to other locations. The empirical means  $E_{i,j}$  of tasks remaining duration should now be calculated.

The Table XII shows the frequencies  $f_i$  empirically calculated from the 20% learning part of the corpus.

The Table XIII shows the frequencies  $f_{i,j}$  empirically calculated from the 20% learning part of the corpus.

TABLE XIII  
Empirical frequencies  $f_{ij}$  (%)

	0	1	2	3	4	5	6	9	e
0	<b>99.64</b>	0.12	0.04	0.01	0.03	0.10	0.06	0.00	0.00
1	0.07	<b>97.48</b>	0.19	0.01	2.20	0.03	0.02	0.00	0.00
2	0.01	0.08	<b>99.89</b>	0.00	0.00	0.01	0.00	0.00	0.00
3	0.17	0.14	0.08	<b>81.29</b>	0.06	3.00	15.26	0.00	0.00
4	0.05	7.62	0.03	0.01	<b>92.24</b>	0.03	0.01	0.01	0.00
5	0.56	0.54	0.54	1.84	0.13	<b>86.50</b>	9.10	0.79	0.00
6	0.07	0.02	0.02	4.02	0.01	2.81	<b>90.63</b>	2.42	0.00
9	0.06	0.04	0.01	0.00	0.07	1.51	10.74	<b>87.57</b>	0.00
s	0.19	0.58	9.24	0.19	0.13	0.06	0.45	0.26	<b>88.89</b>

TABLE XIV  
Expectation  $\overline{E_{i,3}}$  estimation and  $f_i$  frequencies

$i$	0	1	2	3	4	5	6	9	e
$\overline{E_{i,3}}$	3596.89	133.15	102829.88	0.20	4.97	0.48	2.91	0.38	0.01
$f_i$	0.10	0.21	0.53	0.01	0.06	0.02	0.05	0.01	0.00

As said before,  $M$  is the number of locations recorded during a journey. The sampling frequency is 1 second. Thus,  $M = 60 \times 60 \times 24 = 86400$ .  $E_i$  can be estimated by :

$$\overline{E_{i,3}} = \sum_{k=0}^{86400} \frac{k+1}{2} f_i (1-f_i) f_{ii}^{k-1}$$

The Table XIV shows the  $\overline{E_{i,3}}$  estimation calculated for each  $i$  and the  $f_i$  frequencies.

The Table XV shows the confidence interval for  $\overline{E_{i,1}}$ .

TABLE XV  
Confidence interval for  $\overline{E_{i,1}}$

$i$	confidence interval
0	[0.101 ... 0.103]
1	[0.210 ... 0.213]
2	[0.530 ... 0.534]
3	[0.012 ... 0.013]
4	[0.060 ... 0.062]
5	[0.016 ... 0.017]
6	[0.050 ... 0.052]
9	[0.010 ... 0.012]
e	[ $0.8e^{-4}$ ... $1.6e^{-4}$ ]

The mean of remaining time in task  $i$ ,  $\overline{E_{i,1}}$ , consists in calculating, for each observing time  $t$ , time remaining in task  $i$ , divided by the number of times observed (which is equal to  $M + 1$  if the observation start from 0 to  $M$ ). It express persistence in task  $i$ , but is not equal to the mean of past time in  $i$  (it should be half the preceding one).

One can now distinguish two particular cases. If  $i$  was never observed :

$$\overline{E_{i,1}} = 0$$

If  $i$  was always observed :

$$\overline{E_{i,1}} = \frac{(M+1)(M+2)}{M+1} = \frac{M+1}{2} = \frac{86401}{2} = 43200.50$$

For the other cases, some works have to be done now in order to calculate  $\overline{E_{i,1}}$ . It should be calculated for the Polya's urns approach and for the Markov chain approach. Then, it could verify each hypothesis.

If  $\overline{E_{i,1}}$  is in the confidence interval of  $\overline{E_{i,3}}$ , then we should use this Markovian model due to its simplicity (despite Polya's urns approach is available [6]). If it is not the case, the same work has to be done for Polya's urns approach.

## IX. Conclusion

This paper presents two mathematical approaches to location modelling : Polya's urns and a statistical model derived from Markov chain models. This last approach is already used in language modeling as in parts of natural language processing.

The first results are convincing but must be refined in order to be more precise. 98.03% of good prediction could be obtained using only the last second of location to predict the next one. This result indicate that taking day of week into account offer better performance (97.36% without considering day of week). Further works should be made to investigate in this way.

For the validation of each model, a persistence parameter has been defined which is the mean of a task remaining duration. This parameter differs from the mean of the time passed in task (It should be by a factor of 1/2). For deciding between the two methods proposed, further works should be made using the statistics equal to the empirical mean  $E$  of a task remaining duration and their confidence interval.

## References

- [1] A. Schnider, A.-C. von Däniken, and K. Gutbrod, "Disorientation in amnesia. a confusion of memory traces," *Brain*, vol. 119, pp. 1627–1632, 1996.
- [2] A. Schnider, "Spontaneous confabulation and the adaptation of thought to ongoing reality," *Nature Rev. Neurosciences*, vol. 4, pp. 662–671, 2003.
- [3] S. Joray, F. Herrmann, R. Mulligan, and A. Schnider, "Mechanism of disorientation in alzheimer's disease," *Eur. Neurol.*, vol. 52, pp. 193–197, 2004.
- [4] M.-V. Sebastian, J. Menor, and R. Elosua, "Patterns of errors in short-term forgetting in AD and ageing," *Memory*, vol. 9, pp. 223–231, 2001.
- [5] M.-V. Sebastian, J. Menor, and M.-R. Elosua, "Attentional dysfunction of the central executive in AD: Evidence from dual task and perseveration errors," *Cortex*, vol. 42, pp. 1015–1020, 2006.
- [6] J. Demongeot, N. Noury, and N. Vuillermé, "Data fusion for analysis of persistence in pervasive actimetry of elderly people at home," in *Proc. IEEE CISIS' 08 & APPLIMS' 08*, Piscataway, 2008, pp. 589–594.



- [7] C.-E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 & 623–656, July & October 1948.
- [8] R. Rosenfeld, "Adaptative statistical language modeling : A maximum entropy approach." Thesis for the degree of Doctor of Philosophy, Carnegie Mellon University, 1994.
- [9] —, "A maximum entropy approach to adaptive statistical language modeling," *Computer, Speech and Language*, 1996.
- [10] L. Besacier, H. Blanchon, Y. Fouquet, J.-P. Guilbaud, S. Helme, S. Mazenot, M. D., and D. Vaufreydaz, "Speech translation for french in the nespole! european project," in *Proc. IEEE Eurospeech 2001*, Scandinavia, 2001.
- [11] J.-L. Austin, *How To Do Things With Words*. Oxford U.P., 1962.
- [12] J.-R. Searle, *Les Actes De Langage : Essai De Philosophie Du Langage*. Paris: Hermann, 1972, (French Trad. by H. Pauchard).
- [13] D. Vanderveken, *La logique illocutoire*. Bruxelles: Mandarga éd, 1990.
- [14] J.-F. Allen and C.-R. Perrault, "Analyzing intention in utterances," *Artificial Intelligence*, vol. 15, 1980.
- [15] A.-J. Greimas, *Sémantique structurale*. Paris: Seuil, 1966.
- [16] N. Reithinger and E. Maier, "Utilizing statistical dialogue act processing in verbmobil," *Verbmobil*, VM-Report 80, 1995.
- [17] Y. Fouquet, "Un modèle de dialogue par les attentes du locuteur," in *Proc TALN 2002*, Nancy, France, 2002.
- [18] —, "Modélisation des attentes en dialogue oral." Thèse de doctorat informatique, Université J. Fourier (Grenoble I), 2004.
- [19] P. Couturier, A. Franco, J.-F. Piquart, J. Mansotte, C. Montani, C. Suarez, A. Mollier, C. Gucher, M. Frossard, L. Nicolas, G. Jasso Mosqueda, M.-C. Mouchet, A.-M. Argentier, J.-L. Bosson, C. P., and J. Demongeot, "Telegerontology: from teleassistance to teleconsultation of elderly people. mythe or reality?" *Rev. gériatrie*, vol. 21, pp. 23–31, 1996.
- [20] J. Demongeot, G. Virone, F. Duchêne, G. Benchetrit, T. Hervé, N. Noury, and V. Rialle, "Multi-sensors acquisition, data fusion, knowledge minig and alarm triggering in health smart homes for elderly people," *Comptes Rendus Biologies*, vol. 325, pp. 673–682, 2002.
- [21] G. Virone, N. Noury, and J. Demongeot, "The health integrated integrated smart home information system (HIS2) : monitoring the nycthemeral circadian rhythms divergences," in *Proc. IEEE Healthcom 2002*, J.-P. Thomesse et al. eds, Ed., LORIA, Nancy, 2002, pp. 57–62.
- [22] G. Virone, D. Istrate, M. Vacher, N. Noury, J.-F. Sérignat, and J. Demongeot, "First steps in data fusion between a multichannel audio acquisition and an information system for home healthcare," in *Proc. IEEE EMBC 2003*, Piscataway, 2003, pp. 1364–1367.
- [23] G. Virone, B. Lefebvre, N. Noury, and J. Demongeot, "Modeling and computer simulation of physiological rhythms and behaviors at home for data fusion programs in a telecare system," in *Proc. IEEE Healthcom 2003*, Piscataway, 2003, pp. 111–117.
- [24] G. Virone, N. Noury, J.-P. Thomesse, V. Rialle, and J. Demongeot, "A home health information system based on the can fieldbus," in *Proc. IFAC FeT'2003 Aveiro, 5th IFAC Int. Conf. On Fieldbus Systems and their applications*. Amsterdam, The Netherlands: Elsevier Science, 2003, pp. 270–281.
- [25] G. Le Bellego, N. Noury, G. Virone, M. Mousseau, and J. Demongeot, "A model for the measurement of patient activity in a hospital suite," *IEEE Transactions ITB*, vol. 10, pp. 92–99, 2006.
- [26] N. Yankelovich, G.-A. Levow, and M. Marx, "Designing speech acts issues in speech user interfaces," in *Proc. ACM CHI Conference on Human Factors in Computing Systems*, 1995, pp. 369–376.