Home Monitoring of Mental State With Computer Games

Solution Suggestion to the Mental Modern Pentathlon Scoring Problem

Pál Breuer; Péter Hanák; László Ketskeméty; Béla Pataki
Budapest University of Technology and Economics
Budapest, Hungary
e-mail: {breuer,hanak}@emt.bme.hu, kela@cs.bme.hu, pataki@.mit.bme.hu

Gábor Csukly
Department of Psychiatry and Psychotherapy
Semmelweis University
Budapest, Hungary
csukly.gabor@med.semmelweis-univ.hu

Abstract— **As society is aging, an increasing number of elderly people is affected by cognitive problems. Early detection of mild cognitive impairment (MCI) is crucial for slowing down deterioration at an early stage. Improving detection would allow aging in place and thus more cost effective care. However, detection usually occurs too late. Clinical tests are expensive, not frequent enough, and give only a single snapshot of cognitive performance. Regular home monitoring of the changes in mental state would be important but clinical tests have not been developed for this purpose. In this paper, the use of computer games in measuring and maintaining mental wellness in a regular and voluntary way is proposed. Problems and potential solutions are presented, with special emphasis given on the sensor fusion problem caused by the various games present.**

Keywords-mental wellness; home health monitoring; serious games, Mild Cognitive Impairment (MCI), mixed data fusion.

I. INTRODUCTION

As society is aging (by 2060 more than 28% Europeans will be over 65 [1]), an increasing number of elderly people is affected by cognitive problems. With earlier detection of mild cognitive impairment (MCI) deterioration could be significantly slowed down at an early stage. Slow decay of mental abilities is a normal process, which affects already age group 40 of the population, and which increasingly progress with age. It is not easy to identify the stage at which the process becomes abnormal and the affected person requires serious attention, perhaps medical intervention. Cognitive tests are usually performed only if there are already some concerns in the family about someone's cognitive ability, but due to the natural denying effect (by the elder person, the family and the friends) detection typically comes too late.

Traditional, validated, paper-based tests constitute the gold standard but they have several drawbacks. To begin with, such tests require specialist centers and highly trained professionals. Therefore, there is a growing interest in the development of computerized cognitive assessment batteries [2][3][4]. But clinical tests, using either paper-based or computerized methods, are made quite infrequently, providing too sparse snapshots of the cognitive performance. More frequent and regular population screening would require exceedingly many professionals.

Regular home monitoring of changes in mental state offers a powerful alternative, even if it only allows relatively noisy and less targeted measurements. But, they have the advantage of frequent assessment and so the possibility of evaluating temporal trends. Current clinical tests are not suitable for this purpose. Therefore, new measurement methodology must be developed and validated specifically for this strategy. Given that need, recent years have seen a growing interest in the development of special computer games for cognitive monitoring or training purposes. Few such games have been developed; those aim to monitor and train a specific cognitive domain, e.g., verbal fluency [5], executive functions [6] or perceptual and motor functions [7]. A major challenge in this direction is that entertainment capability and measurement power pose contradictory requirements. There are three approaches to game development for older people:

- well-known, popular games (e.g., chess, tangram or tic-tac-toe [8], memory, freecell [9]),
- slightly modified special clinical tests (e.g., corsicube [9]) transformed into games,
- brand new games specially designed for this purpose [6].

Regular monitoring may be (1) controlled or (2) voluntary. Most elderly persons prefer to lead independent life as long as possible. Moreover, in the early monitoring period (before detecting any problem) they are mentally healthy. Therefore, controlled monitoring seems an undesirable option for them; it is expected to undermine independence and works only for a highly motivated minority. As increasingly more people (even the elderly) use computers, and many of them regularly play computer games, gaming activity could be exploited for measuring their performance in those games. In turn, that performance is related to their cognitive state, according to some experimental studies. The basic idea is the following: with regular use of computer games developed or modified specifically for elderly persons, we may be able to measure their mental changes and tendencies over time in an entertaining way, therefore, regularly and voluntarily.

The methods, problems, and possible solutions to those, that are presented in this paper, are based on a recent research project (M3W, Maintaining and Measuring Mental Wellness [9][10][11]). The final goal of the project is to

 ISBN: 978-1-61208-382-7

develop a method for home monitoring mental state of elderly people, which is a very complex task. Therefore, only some subproblems and suggestions are presented in this paper. The considerations leading to the proposed architecture, the basic conceptual architecture of the system and some of the challenges are discussed. Among the numerous problems, this paper focuses on the special sensor fusion problem, which arises in the voluntary home monitoring scenario when several games (sensors) are present.

Due to the complexity of the project, some important problems, such as motivation of the players, are not addressed here in detail. However, some decisions were indeed influenced by motivational considerations, especially the use of several games, which gave rise to the sensor fusion problem, as noted earlier. Another hard problem just mentioned is game selection. The right balance must be found between entertainment capability and measurement power. Based on our pilot experiments, the game set is still evolving.

In that one year pilot study, more than 50 volunteers registered to take part and help evaluate the framework and the games developed at that time. Due to the voluntary nature of the project only about 20 of them played regularly for nearly one year. Of course, the parallel development of the program package was a drawback for the players. The average age of these regular players was 70.3 years, the standard deviation was 10.9 years. People played at home or in an elderly home. Eleven games had been developed and tested (two card games: FreeCell, Solitaire, one psychology-cal test: Corsi Test, two logical games: Graphs and Rabbits, three attention games: Fowler, Odd One Out, Pick One Out, one retention game: Memory Game, two language-skill games: Word Finder, Word Puzzle). The one year timespan has limited relevance on the timescale of mental aging, but some findings have already surfaced, which are clearly important for the long run as well. Parallel to the home monitoring pilot study, a clinical experiment on patients with mental problems (MCI, Alzheimer's disease, etc.) has also been running; those results are not discussed in this paper.

Section II describes the basic model of the suggested mental state evaluation system and describes the important problems. Section III gives the suggested basic detection method using a single computer game. Section IV addresses the special sensor fusion problem caused by the very different nature of games, and suggests a possible solution. Section V summarizes the findings and gives the directions for further work.

II. BASIC MODEL, PROBLEMS

The basic conceptual architecture of the proposed system is shown in Figure 1. The final goal is to provide appropriate long-term feedback to the user (or to the caregiver, family member, medical expert, etc.). Short-term feedback is for motivation to continue participating in the monitoring ("Well done!", "Play some more games!"). Long-term feedback is the result of the change detection estimation: whether a significant change of mental state has occurred or not.

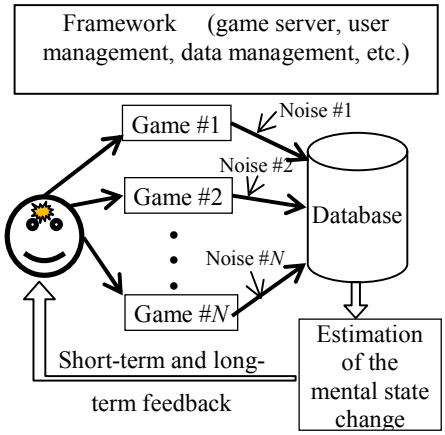

Figure 1. Basic conceptual model of the cognitive state estimation system

Beyond the general problems of such systems (e.g., data privacy concerns), this approach has its special challenges, some of these are given:

1) How to measure the ***cognitive performance*** using computer games?
2) How to cope with the sometimes heavy ***noise*** of the uncontrolled (home) measurement environment?
3) How to ***motivate*** people to take part in the long run?
4) How to compare performance shown in different games, which is basically a special ***sensor-fusion*** problem?

To ***measure the cognitive performance*** three principles are followed:

- To ensure the opportunity of measurements, proper serious games are selected, special ones are developed or clinical tests are modified taking into account the special requirements. Usually, games are modified to improve measurement capability; and tests are modified to be more entertaining. Most of them are logical puzzles, or they need the intensive use of the short-term memory (which is one of the best indicators of MCI), but other important parameters (attention, execution, language skills, etc.) are targeted as well. Two basic parameters are measured: the solving time of the puzzle and the good/bad steps taken during the solution. Currently only successful solutions are measured, for future work there are possibilities in the evaluation of the failed ones as well.
- Because the measurement of the mental state on an absolute scale is very hard, only the change in the person's performance is to be detected. For measuring a change, a reference is needed. There are two possibilities: the performance could be compared to a reference group; or it could be compared to a previously measured reference of the same person. Because the inter-personal comparison is affected by several parameters unknown in this voluntary, uncontrolled method (education, physical abilities, family conditions, profession, environment, etc.) the comparison in time to his/her own previous

 ISBN: 978-1-61208-382-7

performance was chosen. (However, since many people like to compare their own abilities to others' and to compete with others, such functionalities will be offered as well.)

According to our experiments, the ***noise*** can be modeled using two terms: (1) zero mean low level noise caused by the random differences between the consecutive puzzles and by minor environmental disturbances (2) major disturbances. This first term is eliminated by the averaging effect of the evaluation method (see section III): several game results are evaluated together. The second term is an impulse like noise caused by the physiological, environmental and social disturbances resulting in outliers (for example, the telephone is ringing; the person has to use the bathroom, a storm is arriving, neighbor is coming, etc.). This second problem is solved by a filtering step, the outliers are simply rejected; they are not used in further evaluation steps.

Early detection is the purpose; but the main problem is that nobody knows when the abnormal change will happen; maybe in some persons' life never. Therefore, the ***motivation*** must be managed probably for many years. It is a very complex problem itself; only some aspects are discussed here. Among several other aspects, one basic assumption is that although there is an extrinsic motivation that everybody wants to sustain mental abilities and an independent life of good quality, but generally it is not enough in the long-run. There must be intrinsic motivations too, e.g., entertaining ways of measurement, and short-term feedback (Figure 1) given to the user to encourage further playing (e.g., scoring or encouraging messages such as "Well done!" could generate motivation). Unfortunately, most people do not enjoy the same game for years. Therefore, in different time periods different games will be played by the same person. Not to destroy the level of motivation several games are offered (Figure 1); and the performance measured using different games should be somehow compared to each other (Figure 2). This implies a ***sensor fusion and estimation problem***, where the games are the sensors. It is similar to the modern pentathlon scoring problem, where performances in very different sports (fencing, show-jumping, running, swimming, shooting) have to be measured in one unified scoring scheme. In our case, the problem is even more complex because the same game could be played using different settings (e.g., different number of cards in the well-known memory game, see Figure 2); therefore, each setting creates a new game from the measurement point of view. All these games should be compared to each other. In Figure 2 only the results of a given player in the 3 most frequently played games are shown. (In the figure different games are marked by different colors; different settings of the same game are marked by different symbols.) The proposed solution for solving this problem is detailed in Section IV.

III. DETECTION OF THE MENTAL STATE CHANGE

For detection of the mental state change, the comparison in time to the player's own previous performance was chosen. First the evaluation method is considered when only one game (always with the same settings) is played.

Because the two-term noise is present, the effect of the impulse noise, the outliers should be eliminated first. For that purpose, the time between two consecutive elementary events during the solution (e.g., mouse clicks) is analyzed. Because the impulse noise is usually caused by an extreme interrupt, if the longest time between two such actions is too high in comparison to the average action time, then this game was probably seriously disturbed: it is taken as outlier and is rejected. In Figures 2-5, only the results in outlier-free and successful games are shown.

The other noise term, the small natural fluctuation must be coped with as well. For that reason, the change detection cannot be based on the performance measured in a single game; some sets of parameters should be compared. The goal is to detect the decline of performance, but in some periods improvements can occur as well. The assumption is that the decline is preceded by a period where no improvement is present; the situation is stable or deteriorating very slowly. Therefore, a reference set is selected, which is the group of consecutive games in which the person had stable performance (Figure 3).

It is reasonably assumed that the short-term fluctuations due to tiredness, puzzle-hardness, etc., are zero-mean, stable independent random variables. The puzzle hardness is a zero-mean, stable random variable, because the same game is used with the same parameters, and the current puzzle is selected randomly. The short-term change of cognitive power is again a zero mean random variable, because it models the effects of the random changes of the environment, tiredness and health. The very slow long-term change of the cognitive state is modeled differently. Therefore, if a change is detected in one of the integral characteristics (mean, median, standard deviation) or generally in the distribution of the composite random variable (mental-state plus game-noise), it is caused by the slowly changing component modeling the mental state.

Let the performance observation based on the game played in time t_k be $\pi(t_k)$, k=1,2,…,K (this could be the score, the number of steps, etc.). Decrease in the values indicates decreasing performance. Significant change in the time series cannot be stated while this seems to be a realization of an independent and identically distributed (i.i.d.) sample. Several statistical tests can be applied for testing the null hypothesis that the data is i.i.d. Such tests are the difference sign test, the turning point test and the rank test [12][13]. If the null hypothesis cannot be rejected, no significant change in the player's performance could be stated.

A less rigorous requirement is that we cannot justify a change, if the time series is weakly stationary; i.e., uncorrelated with constant expected value and variance. This null hypothesis can be tested with the Dickey-Fuller test [14]. If the time series seems to be non-stationary the change of the player's performance is detected.

 ISBN: 978-1-61208-382-7

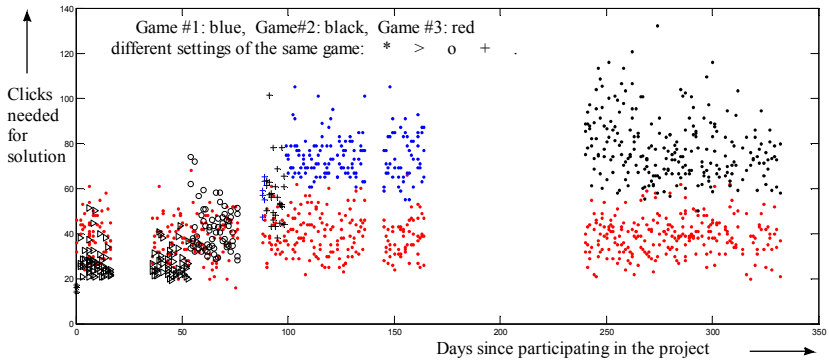

Figure 2. Typical example of a player's performance vs time. The time gaps are caused by travelling to realatives or by other reasons.

Using the Mann-Whitney U or the Kolmogorov-Smirnov two-sample tests, the comparison of the distribution of the reference subset with the distribution of the currently examined subset of the time series could be performed. If we detect a difference between the distributions of the two sub-samples; and the current part of the series has smaller average (of ranks, of scores, etc.), then the player shows performance degradation.

These statistical hypothesis tests were used to check the distribution of the composite random variables. The tests were implemented in Matlab and SPSS. The following findings were obtained:

- The resulting performance parameter is not normally distributed according to the Lilliefors test.
- The time gaps (several users produced 7…60 day gaps) did not change significantly the distribution of the random variable examined (see Table I).
- Several statistical tests were applied to compare the distribution of the reference period data to the current period data of users, who played some hundreds of games in the nearly one year period. (Two-sample Kolmogorov-Smirnov test, Mann-Whitney U test, Wilcoxon signed-rank test). The results confirm that both the stability and the change in the parameters are reliably estimated by the statistical tests. All these tests gave coherent results; later the performance of the different tests should be examined, and the best one should be selected.
- As an alternative to the two-sample statistical tests, a runs test on the sequence of observations was performed to prove the null hypothesis that the values came in random order, against the alternative that they did not. The runs test gave the same result: if there was no significant difference between the distributions of the reference and the current subsets the runs test did not rejected the randomness hypothesis, if there was difference between them, the runs test rejected the hypothesis.
- In some cases, when starting a new game a learning phase occurs, in which the results are improving. The reference is meaningful only when the performance has stabilized The stability could be defined the same way as the stationarity of the current performance. Evaluating these time series has proved that hypothesis testing detected the change of the cognitive performance as well.

In Figure 4, a time series measured during the learning phase is shown. The hypothesis tests accepted the same distribution null hypothesis (the first 30 games' data compared to the current set) for all the sets up to the 187th game; and rejected the null hypothesis for all the sets from the 260th game.

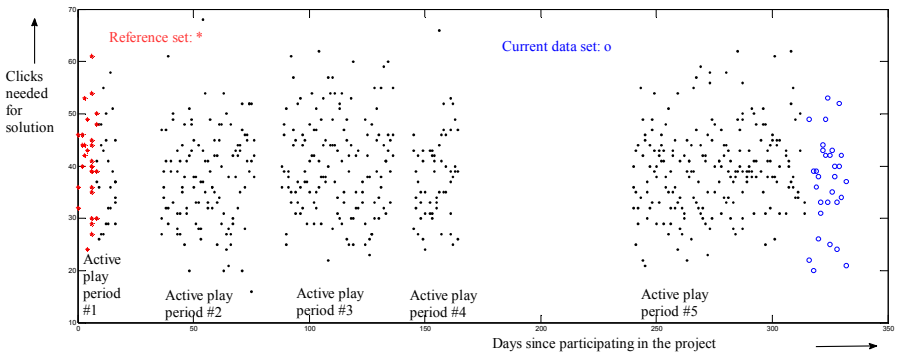

Figure 3. The current performance is always compared to the reference set

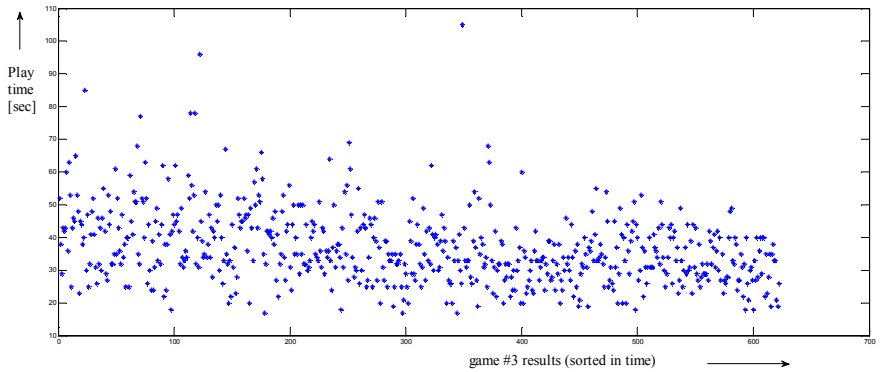

Figure 4. Nonstationary series of play times in the learning phase

IV. SOLVING THE SENSOR FUSION PROBLEM

Computer games are proposed for detecting change in mental state as soon as possible. For motivational purposes several different games should be offered (and different settings of the same game could be used). Because of the voluntary nature there is no guarantee that the same person will play with the same game in the long-run. In our pilot only a few voluntary participants played continuously the same game for this nearly one year period. (Results shown in Figure 4 belong to a participant who played about 2 games per day with the same game for nearly one year!) Most of them changed the game or at least changed the settings of a given game (harder or easier). Therefore, in different – overlapping and non-overlapping – time periods different sensors (games) are available to measure some parameters connected to mental state. Because the more data we have the more reliable the detection of the cognitive state; therefore, every effort is worth to keep all the data.

In this section, a possible solution of that sensor-fusion problem is proposed. The basic idea is that proper linear normalization of the performance measures results in parameters, which are compatible with the normalized parameters of other games. The normalization is based on the reference set of the current game. Let the performance observation using game m in time t_k be $\pi_m(t_k)$, the average of the performance measures of this game's reference set be denoted by $\text{avg}(\pi_{m\text{REF}})$, the standard deviation of this reference set is $\text{std}(\pi_{m\text{REF}})$. The normalization:

$$\pi_{mn}(t_k) = (\pi_m(t_k) - \text{avg}(\pi_{m\text{REF}}))/\text{std}(\pi_{m\text{REF}}),\ m=1,...,N \quad (1)$$

After normalizing all the parameters of the different games the combined time series is constructed by simply sorting the data in time.

$$\{\pi_{COMBn}(t_1),\ldots,\pi_{COMBn}(t_k)\} = \{\pi_{m1n}(t_1),\ldots,\pi_{mkn}(t_k)\}$$

$$t_1 < t_2 < \ldots < t_k \quad (2)$$

The block diagram of the suggested idea is shown in Figure 5. The resulting combined time series derived from the data of Figure 2 is shown in Figure6

Using the time series of combined data gives very similar results as using the data of one game only. In Table I the null hypothesis of having the same distribution of the data subsets compared are shown in two ways. In both evaluations the reference set comes from the first 30 observations of the active play period 1, the comparison is made to the first 30 observations of the 2nd, 3rd, 4th, 5th active play periods, respectively. The difference is that in the first experiment only the Game#3 data are used, and in the second experiment the combined data are used.

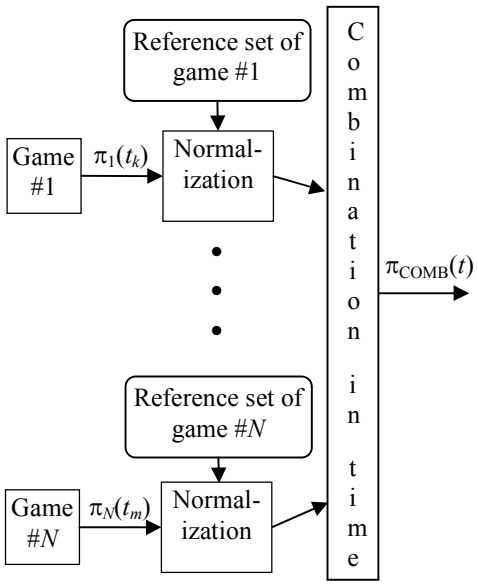

Figure 5. The normalized performance parameters of the different games are combined to form one composite time series

 ISBN: 978-1-61208-382-7

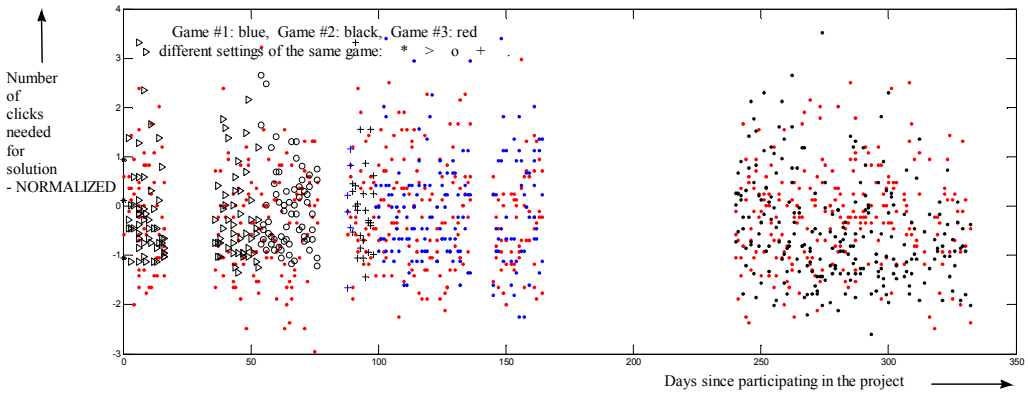

Figure 6. Normalized and combined data

In Table I, the acceptance or rejection (on the p=0.05 level) of the null hypothesis are shown.

TABLE I. RESULTS OF TWO-SAMPLE KOLMOGOROV-SMIRNOV TESTS: REFERENCE SET SHOWN IN FIGURE 2 (FIRST 30 OBSERVATIONS OF ACTIVE PLAY PERIOD 1) COMPARED TO THE FIRST 30 DATA OF EACH ACTIVE PLAY PERIOD

	Game #3 data only		***Combined data***	
Reference: active play period 1 compared with	***Null hypothesis accepted:0, rejected: 1***	***Probability value: p***	***Null hypothesis accepted:0, rejected: 1***	***Probability value: p***
Active play period #2	0	0.43	0	0.11
Active play period #3	0	0.76	0	0.20
Active play period #4	1	0.03	1	0.01
Active play period #5	0	0.54	0	0.06

Although in the first experiment only Game#3 data were used and in the second one combined data were used, they resulted in the same acceptance/rejection scheme although the pure one-game only data gave higher probability values.

V. CONCLUSION AND FUTURE WORK

Home monitoring of changes in mental state using computer games was proposed in a regular, voluntary scenario; some of the problems were analyzed and solutions were proposed. The system assumes voluntary participation; therefore, several different games were developed to sustain motivation in the long run. In the game battery there are both well-known, popular games and modified clinical tests, it is continuously evolving.

For cognitive performance change detection, the within-subject comparison is proposed. The reference set of performance results are to be compared to the current set of results using statistical hypothesis tests. The null hypothesis is that the two sets came from the same distribution. Until the null-hypothesis cannot be rejected, the stability of mental state could be assumed.

Because of the large number of diverse games, the problem of unifying the different data should be solved as well. This problem is basically a sensor fusion one. Proper linear normalization using the reference set of each game is proposed for producing a mixed time series that could be used for our detection purposes as well.

In the future

- a pilot has to be launched to validate the method using further clinical tests,
- the most appropriate games for both entertainment and measurement should be investigated,
- the feasibility of multiplayer games is to be analyzed,
- the potential in the evaluation of the failed games should be investigated,
- the best statistical hypothesis test is to be identified.

ACKNOWLEDGMENT

This research was performed in the Maintaining and Measuring Mental Wellness (M3W) project, supported by the AAL Joint Programme (ref. no. AAL-2009-2-109). The authors also gratefully acknowledge the contributions of their project partners in Greece, Luxembourg, Switzerland and Hungary.

REFERENCES

[1] Population structure and ageing. Eurostat, May 2014, http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Population_structure_and_ageing [retrieved: January, 2015]

[2] Cantab test battery, Cambridge Cognition, http://www.cambridgecognition.com/ [retrieved: January, 2015]

[3] MindStreams, http://www.mind-streams.com/ [retrieved: January, 2015]

[4] T. Dwolatzky, "The Mindstreams computerized assessment battery for cognitive impairment and dementia", PETRA'11 May 2011, pp. 501-504, ISBN: 978-1-4503-0772-7.

[5] H. Jimison, M. Pavel and T. Le, "Home-Based Cognitive Monitoring Using Embedded Measures of Verbal Fluency in a Computer Word Game" 30th Annual International IEEE EMBS Conference, Aug. 2008, pp. 3312-3315. doi: 10.1109/IEMBS.2008.4649913

[6] A. López-Martínez et al, "Game of gifts purchase: Computer-based training of executive functions for the elderly", IEEE 1st International Conf. on Serious Games and Applications

 ISBN: 978-1-61208-382-7

for Health (SeGAH), Nov. 2011, pp. 1-8, Print ISBN:978-1-4673-0433-7 DOI: 10.1109/SeGAH.2011.6165448

[7] K. Ogomori, M. Nagamachi, K. Ishihara, S. Ishihara and M. Kohchi, "Requirements for a Cognitive Training Game for Elderly or Disabled People", Int. Conf. on Biometrics and Kansai Engineering, Sept. 2011, pp 150–154, E-ISBN 978-0-7695-4512-7, Print-ISBN 978-1-4577-1356-9, DOI: 10.1109/ICBAKE.2011.30

[8] V. Menza-Kubo and A. L. Morán, "UCSA: a design framework for usable cognitive sstem for worried-well", Personal Ubiquitous Comput. Vol. 17, Issue 6, Aug. 2013, pp. 1135-1145. ISSN:1617-4909. DOI: 10.1007/s00779-012-0554-x

[9] Maintaining and Measuring Mental Wellness (M3W) AAL Joint Programme project (ref. no. AAL-2009-2-109) https://m3w-project.eu/ [retrieved: January, 2015]

[10] E. Sirály et al., "Differentiation between mild cognitive impairment and healthy elderly population using neuropsychological tests", Neuropsychopharmacol Hung. (3), Sept. 2013, pp. 139-46.

[11] P. Hanák et al., "Maintaining and Measuring Mental Wellness", Proc. of the XXVI. Neumann Kollokvium, Nov. 2013, pp. 107-110 (in Hungarian).

[12] P. Brockwell and R. Davis, Introduction to Time Series and Forecasting, Springer, New York, 1996.

[13] M. G. Kendall, and A. Stuart, "The Advanced Theory of Statistics", Vol. 3, Griffin, London, 1976.

[14] D. A. Dickey, W. A. Fuller, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root". J. of the American Statistical Association 74 (366), Jun. 1979, pp. 427–431.

 ISBN: 978-1-61208-382-7