

Classification of Alzheimer’s Disease Patients using Metrics of Oculo-Motors

Wioletta Nowak^{*}, Minoru Nakayama[†], Elzbieta Trypka[‡], Anna Zarowska[§]

^{*} Institute of Biomedical Engineering and Instrumentation
Wrocław University of Science and Technology, Wrocław, Poland 50–370
Email: wioletta.nowak@pwr.edu.pl

[†] Department of Information and Communications Engineering
Tokyo Institute of Technology, Tokyo, Japan 152-8552
Email: nakayama@ict.e.titech.ac.jp

[‡] Wrocław Medical University Rektorat, Wybrzeze Ludwika Pasteura 1, 50-367 Wrocław, Poland
Email: elzbieta.trypka@umed.wroc.pl

[§] Institute of Biomedical Engineering and Instrumentation
Wrocław University of Science and Technology, Wrocław, Poland 50–370
Email: anna.zarowska@pwr.edu.pl

Abstract—Ocular information was observed during a set of dementia tests involving participants with Alzheimer’s Disease (AD), with a mild level of cognitive impairment (MCI), or in a control group. The number of participants was 26. Features of changes in pupil size and in the central position of both eyes of participants of all three types were compared. There are significant differences in some of the metrics between the types, in earlier test sessions. The possibility of classification was confirmed using the extracted features, and the contributions of some features were examined.

I. INTRODUCTION

HUMAN cognitive ability is often affected by disease, such as by Alzheimer’s disease (AD) for example. The decline in cognitive ability of elderly people can be monitored carefully during daily exercise session [1]. In general, behavioural actions requiring some level of workload affect visual information processing and cognitive perception. The symptoms of diseases can be detected using biometrics, such as indices of oculo-motors. Changes in the level of cognitive activity and mental workload handling ability are often assessed using pupillary changes [2], [3], [4]. The symptoms of AD and aged macular disease (AMD) are also detected using pupillary reactions to the pupil light reflex (PLR) [5], [6], [7].

Most diagnostics for AD patients are based on medical consultations using cognitive tests such as the Mini-Mental State Examination (MMSE) [8] or the Montreal Cognitive Assessment (MoCA) [9]. In the test results, participants are classified into AD, mild cognitive impairment (MCI) and no cognitive impairment (NCI), or categories with normal elderly controls (NC) [9]. Since medical consultations are conducted during face-to-face oral questioning, some communication skills such as language ability and the participant’s mental situation may influence the diagnostic result. Therefore, the time

This work was partially supported by the Polish Ministry of Science and Higher Education research grant NN518 405338 and by the International Collaboration Research Promotion Fund, School of Engineering, Tokyo Institute of Technology.

spent by medical practitioners and the workload of participants is not insignificant. Any appropriate procedures which can assist with medical observation using participant’s behavioural metrics to reduce the difficulty of making proper assessments and permit this AD diagnostic procedure to become more widely available.

As mentioned above, ocular responses including pupillary changes and eye movement provide some evidence of symptoms of cognitive activity impairment. Observation of participant’s oculo-motor changes is introduced into medical consultation, and the possibility of detecting cognitive impairment is examined. The following topics are addressed in this paper.

- 1) Pupillary changes and eye movement are measured using oculo-motor indices during a medical consultation to diagnose MMSE, and the metrics of the three diagnostic classes AD, MCI and NC are compared.
- 2) A prediction procedure for diagnostic classes is developed using the metrics measured, and its performance is discussed.

II. EXPERIMENTAL METHOD

During a medical consultation of elderly people for MMSE, ocular metrics were measured.

A. Cognitive test

A cognitive exam, based on MMSE and consisting of the following 6 tests, was administered.

- Test 1: Control condition for measurement without any tasks, or baseline monitoring.
- Test 2: Orientation including simple questions and a conversation session.
- Test 3: Memory and recall tests of 5 words.
- Test 4: Oral calculation such as decrements of 7 starting from 100.
- Test 5: 2nd question and conversation session.

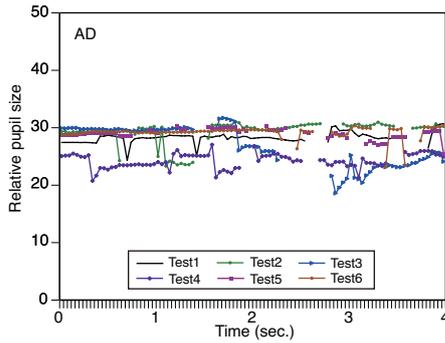


Fig. 1. Pupillary changes of an AD patient

TABLE I
MEAN OBSERVATION DURATION (SEC.) FOR PARTICIPANT GROUPS

Sub. Group	Test					
	1	2	3	4	5	6
AD (N=4)	44.6	58.4	65.7	61.1	57.1	76.4
(SD)	14.0	31.3	10.8	7.3	9.8	32.5
MCI (N=15)	33.4	62.7	53.1	39.6	36.0	42.8
(SD)	3.2	27.4	10.4	15.1	10.5	27.3
Control (N=7)	32.1	61.6	58.5	32.3	32.7	23.1
(SD)	1.1	17.5	24.4	9.5	6.3	9.1

- Test 6: Recall test for words memorised in Test 3.

The tests were conducted at a hospital in Poland and each participant was examined for AD and MCI by an M.D. who was one of the authors. Therefore, test tasks and duration varied by participant according to their medical condition.

B. Participants

In the results of the diagnoses, 26 participants (mean age: 73.7yo, STD: 9.9yo) were classified into three groups. The control group (NC) consisted of people of comparable age.

- AD: 4 (F:4, mean age: 77.3yo, STD: 6.7yo)
- MCI: 15 (F:11, M:4, mean age: 73.4yo, STD 13.1yo)
- NC: 7 (F:5, M:2, mean age: 72.4yo, STD:5.3yo)

Informed consent was obtained from all participants prior to the experiment. One MCI (F,70yo) had data partially missing, and this data was omitted from later analysis.

C. Procedure

Participants wore image tracker eye goggles (VisualEyes 505), and were asked to view the face of an M.D. during the test. Before the test, calibration of viewing points was conducted. Images of both eyes were recorded as 640×240 pixel images at 25Hz using a camera with a lens in each side of the goggles. The images were recorded using the MP4 format. After the experiment, the centre of the pupil and the size of the pupil were extracted. Fig. 1 shows the pupillary changes of an AD patient during Tests 1 through 6. The data measured during blinks was omitted in the analysis which followed. Durations where pupillary changes were dropped indicate data omitted due to blinks.

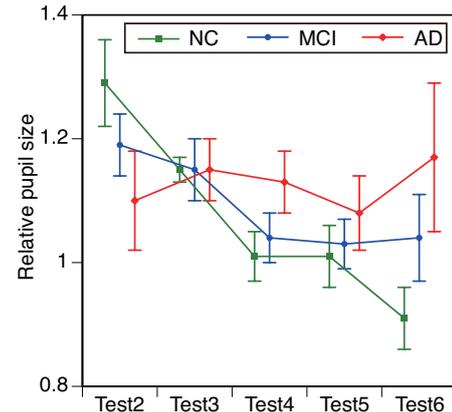


Fig. 2. Mean relative pupil sizes between groups (Error bar: STD Error)

D. Post processing

1) *Pupil response*: Pupil size was calculated as the sum of the pixels of each frame of the black and white video using an image processing technique. Mean pupil sizes were calculated for each duration of observation. Pupillary changes were evaluated as frequency powers of less than 4Hz, as pupils behave like low-pass filters. The frequency powers were calculated as mean power spectrum densities for 0.5~1.6Hz and 1.6~4.0Hz, as they were in the previous study [10]. The lowest frequency power ($f < 0.5\text{Hz}$) was omitted due to signal noise [11].

2) *Eye movement*: During image processing, the centre of the pupil was measured. The measurement specification of the goggles suggests that the shift in the centre of the pupil corresponds with a visual angle of 0.25 deg for both the horizontal and vertical axes. Eye movement data was generated using this method of conversion. The eye movement was classified into fixations and saccades using a threshold of 40deg/sec [12]. For each test observation, saccade frequency, mean saccade lengths, fixation frequency, and mean fixation duration were summarised.

Since the ocular metrics of both eyes were measured, they were evaluated using a procedure which examined repeated individual measurements.

III. RESULTS

A. Observation time

As the duration of test observation depended on the condition of the individual participant, mean durations are summarised in Table I, which consists of test number and participant group. The duration of most tests except for Test 2 was around 30 seconds. While the duration varied between tests, it increased gradually as group participants progressed from NC to MCI to AD. There may have been some difficulty for participants with AD to perform some of the test tasks.

B. Pupillary changes

1) *Mean pupil sizes*: Pupil sizes were standardised in Test 1 using mean pupil sizes. Mean relative pupil sizes for the

TABLE II
TWO-WAY ANOVA: PARTICIPANT GROUPS × TESTS

Source	df	SS	MS	F	Sig.
Sub. Group	2	0.059	0.029	0.43	n.s.
Test	5	1.078	0.216	3.15	$p < 0.01$
Sub-G × Test	10	0.620	0.062	0.91	n.s.

TABLE III
ANOVA ANALYSIS FOR EACH SUBJECT GROUP

Sub. Group	df	SS	MS	F	Sig.
AD	5	0.154	0.031	0.35	n.s.
MCI	5	0.816	0.163	2.14	$p < 0.10$
NC	5	1.298	0.260	6.33	$p < 0.01$

three groups of participants are summarised in Fig. 2. Error bars indicate standard errors. The mean size of NC decreased gradually from Tests 2 and 6. Both task difficulty and order effect may have affected pupil sizes. However, mean pupil size of the AD group remained relatively large throughout the tests, and the deviations were smaller than the deviations of other groups. The means for MCI also decreased along with the tests, and the deviation was smaller than for the NC group.

Two-way ANOVA was conducted in order to examine the contributions of the factors of participant groups and test results. These results are summarised in Table II. The test factor is significant, but participant groups and the interaction of the two factors are not significant.

In further analysis, an F-test of the participants of each group was conducted. The results are summarised in Table II. Test factors were significant for both NC ($p < 0.01$) and MCI ($p < 0.05$). The results may reflect the pupil responses of participants of the groups.

2) *Pupillary oscillation:* As pupillary changes between tests and groups of participants was confirmed, frequency analysis was also applied. As mentioned in the section above, power spectrum densities (PSD) in the frequency ranges from 0.5 to 1.6Hz and from 1.6 to 4.0Hz were compared. There were no significant changes between tests or groups of participants.

A typical change was observed during Test 1. The PSD of pupillary changes of each group is summarised in Fig. 3. As the figure shows, there are some differences between groups of participants. The results of F-tests of the means of PSDs are summarised in Table IV. There is a significant difference ($p < 0.01$) between groups in the high frequency range (1.6~4.0Hz), and the tendency for a difference to exist ($p < 0.10$) in the low frequency range (0.5~1.6Hz).

This result suggests that differences in PSDs appear during controlled conditions, such as in Test 1.

C. Eye movement

Features of eye movement, which were tracked using the above mentioned procedure, are summarised for mean saccade frequency in Fig. 4. The error bars indicate STD Errors.

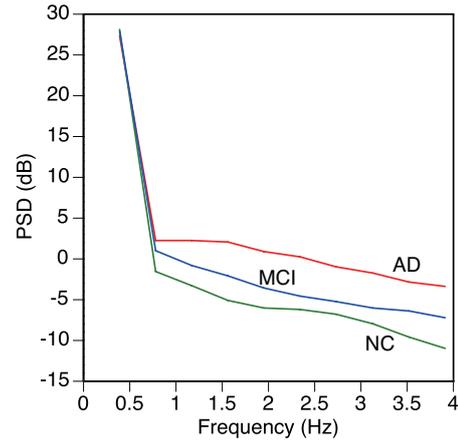


Fig. 3. PSD of pupillary change in Test 1

TABLE IV
ANOVA FOR PARTICIPANT GROUPS IN TEST 1

Frequency	df	SS	MS	F	Sig.
0.5 < f < 1.6Hz	2	6.887	3.444	2.62	$p < 0.10$
1.6 < f < 4Hz	2	2.062	1.031	5.62	$p < 0.01$

In Fig. 5, there are few differences between participant groups, but mean frequencies increase gradually during medical consultation after Test 1. The change in frequencies during complete tests was examined using two-way ANOVA, and the factor for tests was significant ($F(5, 294) = 28.4, p < 0.01$), however the factor for groups of participants was not significant ($F(2, 294) = 1.8, p = 0.16$). The effect of groups of participants on each test was examined, the while factor in Test 1 ($F(2, 49) = 7.7, p < 0.01$) was significant, there were significant differences between groups except for one pair of MCI and NC. Even in Test 2, the factor for groups of participants showed a tendency to be different ($F(2, 49) = 2.48, p < 0.10$), but overall there were no significant differences between groups.

Changes in mean saccade lengths are summarised in Fig.

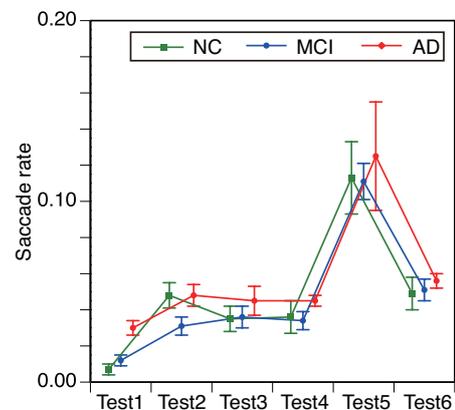


Fig. 4. Saccade frequency (deg./sec.)

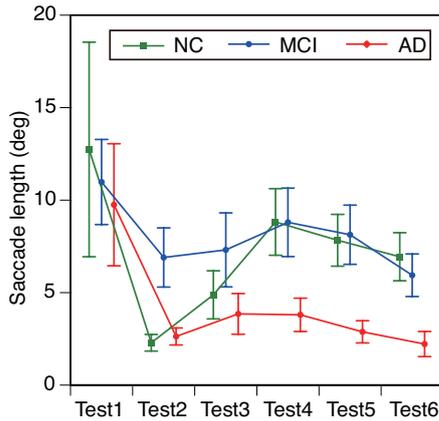


Fig. 5. Saccade length (deg.)

TABLE V
CLASSIFICATION RESULTS USING TESTS 1 AND 2

Sub Grp	(1)	(2)	(3)	Recall	F1	Pred. pairs
(1)AD	6	2	0	0.75	0.80	ad-ad:3/4
(2)MCI	0	25	3	0.89	0.85	mci-*:14/14
(3)NC	1	4	9	0.64	0.69	nc-*:5/7
Precision	0.86	0.81	0.75	P-mean:0.81; R-mean:0.76		
Rate of Lower Triangular Matrix: 0.90						

5. The effect of groups of participants was only examined in Test 2 ($F(2, 48) = 3.0, p < 0.10$), but there were no significant differences between groups. Also, a significant different tendency for groups of participants to be different was observed in the fixation frequency of Tests 1 and 2. Therefore, typical eye movement responses may appear in initial tests, as they did in Tests 1 and 2.

In this study, conversations during medical consultation were not analysed, though some mental workloads were detected in oculo-motor indices during Tests 2 through 6. The results suggest that there are significant differences between the metrics of Test 1 and the metrics of the other tests. However, the deviations in metrics measured during tests may be not large. Therefore, the metrics can be noted as means of duration for the tests.

D. Prediction of groups of participants

Since there were some significant differences in the metrics of the groups of participants extracted, the possibility of predicting the class of the group was examined using the features extracted.

1) *Prediction using the observed metrics:* Three groups of participants were estimated in 6 tests using 6 metrics (mean pupil size, mean PSD of pupil oscillation, mean saccade frequency, mean saccade length, mean fixation frequency, and mean fixation duration). An estimation model was trained using the Random Forest method.

Estimation was conducted using the 6-dimensional data from Tests 1 to 6, and performance was examined using trial and error. As a qualitative evaluation, performance during

TABLE VI
CLASSIFICATION RESULTS USING TESTS 1 TO 6 WITH DATA EXTENSIONS

Sub Grp	(1)	(2)	(3)	Recall	F1	Pred. pairs
(1)AD	8	0	0	1.00	0.64	ad-ad:4/4
(2)MCI	7	20	1	0.71	0.71	mci-*:13/14
(3)NC	2	5	7	0.50	0.64	nc-*:4/7
Precision	0.47	0.71	0.88	P-mean:0.69; R-mean:0.74		
Rate of Lower Triangular Matrix: 0.98						

TABLE VII
CLASSIFICATION RESULTS USING TESTS 1 AND 2 WITH DATA EXTENSIONS

Sub Grp	(1)	(2)	(3)	Recall	F1	Pred. pairs
(1)AD	8	0	0	1.00	0.64	ad-ad:4/4
(2)MCI	6	21	1	0.77	0.82	mci-*:12/14
(3)NC	3	2	9	0.64	0.75	nc-*:6/7
Precision	0.47	0.91	0.90	P-mean:0.76; R-mean:0.80		
Rate of Lower Triangular Matrix: 0.98						

Tests 1 and 2 was better. The results are summarised in the contingency table shown in Table V. The vertical cells represent groups of participants, and the horizontal cells represent predictions. As the predictions were conducted using both eyes, the overall numbers are twice the number of participants. The prediction performance recall rate, precision rate and F1 indices were calculated. Also, correct prediction by either eye is summarised in the rightmost cell of Table V.

The results of the table suggest the possibility of predicting a participant's group using the metrics of eye observations. The purpose is a targeted diagnostic procedure which can predict the most-advanced levels of AD. Therefore, the number of predictions of the lower triangular matrix may be another index of performance. The highest recall rate appears with the MCI group, as the number of participants is the largest. The number of participants may be an influence bias regarding prediction performance.

2) *Prediction with data extensions:* The medical consultation data set seems insufficient because it is not easy to gather a large and balanced number of participants for each group. A data extension technique was introduced in order to obtain a set of prepared data. Here, all metrics were hypothesised to deviate using Gaussian distribution $N(\mu, \sigma)$, and the range of re-sampled data was controlled between $\mu \pm k\sigma$ ($k = 1, 1.5, 2$). As a result, 300 sets of data were prepared in total, with 100 sets of data being generated for each participant group.

The prediction models were trained with the data generated using the Random Forest technique once again, and actual measured data was tested. Recall performance between k parameters was compared, and $k = 1.5$ produced the highest level of performance.

The test results are presented in Table VI using the same format as the previous table, where all generated data for Tests 1 to 6 is employed as the set of trained data. The results show that for all participants in the AD group, AD was able to be predicted using both eyes. The data extensions may have compensated for the unbalanced data samples. However, incorrect predictions for MCI and NC groups increased, and

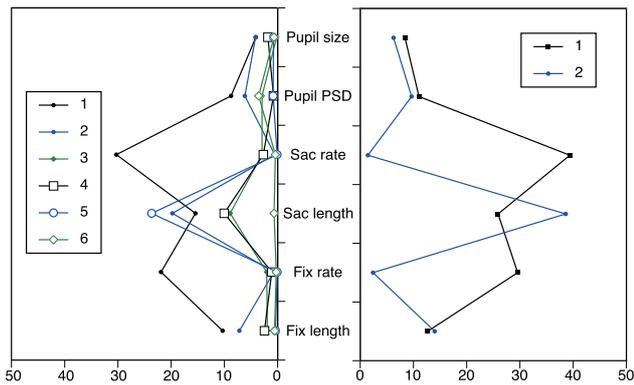


Fig. 6. Comparison of Gini coefficients of two conditions. Left side: Tests 1~6, Right side: Tests 1~2

some indices of performance decreased slightly. The number of predictions of the lower triangular matrix was sufficiently high, and thus the possibility of using prediction as a diagnostic procedure was confirmed.

In the previous section, features of metrics of patients in the early stages of consultation were suitable for prediction. A second set of results is summarised in Table VII, where features of Tests 1 and 2 are limited. In these results, NC group prediction performance improved, and some of the performance metrics showed the highest rates of correct prediction.

For a detailed analysis of performance, the contributions of features of observed metrics between the two models are compared in Tables VI and VII. The degree of contribution of prediction as an index of importance (the Gini coefficient) is summarised in Fig. 6. The left side represents coefficients in Tests 1 to 6, and the right side represents coefficients in Tests 1 to 2. The figure suggests that the coefficients for Test 1, in particular saccade frequency, saccade length and fixation frequency, present a larger contribution than do the features of pupil responses. The coefficients on the left side suggest that most coefficient in Tests 3 to 6 were relatively small. These results confirm the importance of using metrics in the early stages of the consultation, as the changes between the initial cognitive load and the control session may represent the

The results show that there were differences in measured metrics between groups of participants, and that significant differences were observed during the early stages of testing, including during the control session which was the first stage of the consultation. The prediction procedure was conducted using the Random Forest and data extension techniques. The accuracy of the prediction procedure was confirmed.

situation of individual participants.

IV. SUMMARY

In this paper, changes in pupil responses and eye movement between AD, MCI and NC participants during cognitive tests were compared. The tests consisted of using MMSE as a form of medical consultation. Also, the possibility of estimating participant groups was examined.

Validation of the procedure and improvement of prediction performance will be subjects of our further study.

REFERENCES

- [1] K. Aoki, T. T. Ngo, I. Mitsugami, F. Okura, M. Niwa, Y. Makihara, Y. Yagi, and H. Kazui, "Early detection of lower MMSE scores in elderly based on dual-task gait," *IEEE Access*, vol. 7, pp. 40 085–40 094, 2019.
- [2] J. Beatty, "Task-evoked pupillary response, processing load, and the structure of processing resources," *Psychological Bulletin*, vol. 91, no. 2, pp. 276–292, 1982.
- [3] J. Kuhlmann and M. Böttcher, Eds., *Pupillography: Principles, Methods and Applications*. Munchen, Germany: W. Zuckschwerdt Verlag, 1999.
- [4] M. Nakayama and M. Katsukura, "Development of a system usability assessment procedure using oculo-motors for input operation," *Universal Access in Information Society*, vol. 10, no. 1, pp. 51–68, 2011.
- [5] M. Nakayama, W. Nowak, H. Ishikawa, K. Asakawa, and Y. Ichibe, "Discovering irregular pupil light responses to chromatic stimuli using waveform shapes of pupillograms," *EURASIP J. in Bioinformatics and System Biology*, vol. #18, pp. 1–14, 2014.
- [6] W. Nowak, M. Nakayama, T. Kręcicki, E. Trypka, A. Andrzejak, and A. Hachoł, "Analysis for extracted features of pupil light reflex to chromatic stimuli in Alzheimer's patients," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 5, pp. 1–10, November 2019, e4.
- [7] W. Nowak, M. Nakayama, T. Kręcicki, and A. Hachoł, "Detection procedures for patients of Alzheimer's disease using waveform features of pupil light reflex in response to chromatic stimuli," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 6, pp. 1–11, December 2020, e6.
- [8] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "MINI-MENTAL STATE – a practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, pp. 189–198, 1975.
- [9] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment," *Journal of American Geriatrics Society*, vol. 53, pp. 695–699, 2005.
- [10] V. Peysakhovich, M. Causse, S. Scannella, and F. Dehais, "Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort," *International Journal of Psychophysiology*, vol. 97, pp. 30–37, 2015.
- [11] K. Ukai, "Pupil," in *SHIKAKU JYUHO SYORI HANDOBUKKA (Handbook of Visual Information Processing)*, Japan Society of Vision Science, Ed. Tokyo, Japan: Asakura Shoten, 2001.
- [12] Y. Ebisawa and M. Sugiura, "Influences of target and fixation point conditions on characteristics of visually guided voluntary saccade," *The Journal of the Institute of Image Information and Television Engineers*, vol. 52, no. 11, pp. 1730–1737, 1998.