

# BADANIE REGUŁ POSTRZEGANIA NATURALNEGO W CELU ICH WYKORZYSTANIA W INTELIGENTNYCH SYSTEMACH WIZYJNYCH

PIOTR AUGUSTYNIAK, ZBIGNIEW MIKRUT

*Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, {august zibi}@agh.edu.pl*

**Streszczenie.** Referat opisuje prace badawcze mające na celu określenie parametrów epizodów składowych procesu pobierania informacji wizyjnej przez system wzrokowy człowieka. System ten swą wysoką sprawność funkcjonalną zawdzięcza występującym naprzemiennie epizodom koncentracji uwagi i inteligentnego poszukiwania informacji. Istnieje zatem uzasadniona nadzieja, że użycie systemu wzrokowego człowieka jako wzorca dla rozwiązań użytych w systemach sztucznej inteligencji przyczyni się do istotnej poprawy ich parametrów. Opisano eksperymenty okoruchowe mające na celu określenie postrzeganego zakresu sceny podczas pojedynczego epizodu koncentracji uwagi. Dodatkowo dokonano oszacowania ilości informacji, która może zostać odtworzona na podstawie kontekstu, bez potrzeby ponownej zmiany punktu koncentracji uwagi. Otrzymane rezultaty dowodzą m. in., że gałka oczna jest słabym przyrządem optycznym o stosunkowo wąskim kącie widzenia. Rezultaty mogą być przydatne do dalszej analizy procesu postrzegania, a zwłaszcza czytania, oceny sprawności interpretacji obrazów i wreszcie w ocenie poprawności funkcjonowania układu okulomotorycznego.

## 1. WPROWADZENIE

Analiza procesu pobierania informacji przez system wzrokowy człowieka [6] jest niezwykle istotnym czynnikiem modelowania procesu pobierania informacji i stwarza nadzieję na zaimplementowanie rządzących nim reguł w inteligentnych systemach wizyjnych. Systemy te, działające wedle reguł sztucznej inteligencji zastępują coraz częściej człowieka w zadaniach obserwacyjno-interwencyjnych o szczególnej uciążliwości (np. ciągła analiza obrazów z kamer, przedstawiających ruch drogowy). Kolejnym krokiem jest uczynienie z nich funkcjonalnej repliki narządu wzrokowo-okoruchowego człowieka, wydaje się zatem naturalne uzupełnienie ich o algorytmy realizujące funkcje inteligentnego poszukiwania informacji wizyjnej. Ponieważ postrzeganie sceny realizowane jest przez człowieka podświadomie, a niekiedy z wykorzystaniem nawyków

okulomotorycznych, uzasadnione wydaje się badanie reguł postrzegania również bez angażowania świadomości obserwatora. Może to być dokonane przez akwizycję i analizę ścieżki wzrokowej, a następnie uogólnienie wyznaczonych parametrów do odpowiednich reguł.

Według badań okulistycznych i oftalmologicznych oko jest bardzo niedoskonałym przyrządem wizyjnym, a sprawność systemu wizyjnego człowieka jest osiągana dzięki naprzemiennym procesom pozyskiwania informacji, ich analizy i poszukiwania [2, 5].

Ponieważ zainteresowanie kolejnymi informacjami ze sceny jest uwarunkowane subiektywnie, ścieżka wzrokowa może być bardzo skomplikowana. Znacznym ułatwieniem może być serializacja informacji wizyjnej dokonywana przez użycie tekstu, którego odczytanie jest zdefiniowane przez następstwo liter od lewej do prawej i wierszy od wyższych do niższych.

Badania procesu postrzegania opisywane w niniejszym artykule miały na celu odpowiedź na następujące pytania:

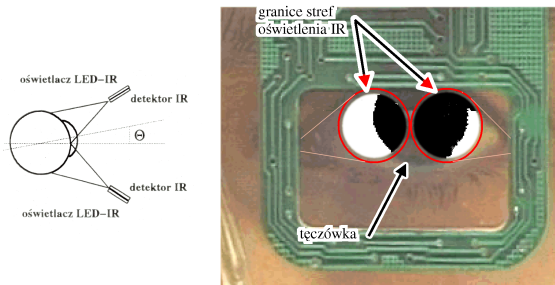
- jaką porcję sceny obserwator postrzega jednocześnie?
- jaka jest gęstość punktów koncentracji uwagi konieczna do prawidłowej percepcji sceny?

## 2. MATERIAŁY I METODY

Badania mające na celu odpowiedź na wyżej sformułowane pytania przeprowadzono metodą dwóch eksperymentów wizualnych (ang. *visual task*) z użyciem aparatury rejestrującej ścieżkę okoruchową (ang. *eyetracker*) Ober-2. Podczas realizacji zadań scena postrzegana przez obserwatora jest wcześniej precyzyjnie przygotowana, a analiza trajektorii ruchu gałek ocznych jest dokonywana w kontekście zawartości sceny. W charakterze obserwatorów występowały wolontariusze wyłonieni z grup studenckich (8 osób, 5 kobiet, 3 mężczyźni, średnia wieku 22 lata).

Do rejestracji położenia gałki ocznej wykorzystano urządzenie Ober-2 [4], które mierzy moc wiązki promieniowania podczerwonego odbitego od gałki ocznej. Oświetlacze oraz detektory promieniowania podczer-

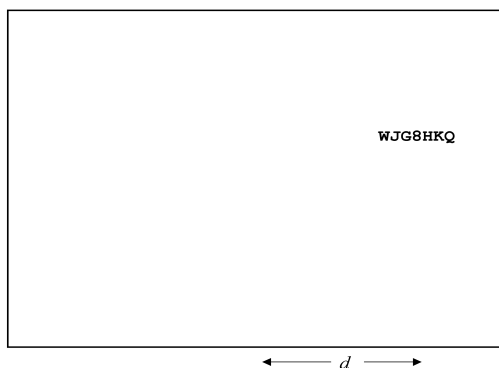
wonego umieszczone są w goglach zakładanych przez obserwatora (rys. 1), skąd sygnał elektryczny pobierany jest do systemu OBER2 a po przetworzeniu na postać cyfrową (12 bitów, 500 Hz) dostępny na łączu komunikacji szeregowej w postaci ciągu czwórek liczb, będących reprezentacją pozycji centrum każdego z oczu w osi pionowej i poziomej. Poprawna interpretacja danych pochodzących z urządzenia jest możliwa dzięki ustaleniu pozycji obserwatora względem prezentowanej sceny (monitora komputerowego) a także kalibracji urządzenia, polegającej na obserwacji prostokąta o standardowych wymiarach każdorazowo przed i po pokazie.



Rys. 1. Fizyczna zasada działania urządzenia do rejestracji sygnału okoruchowego Ober-2.

## 2.1 Pokazy ciągów znaków

Pierwszy z eksperymentów polegał na obserwacji pojawiających się w dowolnych miejscach sceny ciągów znaków (0-9,A-Z - por. rys. 2). Eksperyment ten miał na celu zbadanie zależności poprawności rozpoznania ciągu znaków od długości tego ciągu, a także stwierdzenie, jaka długość ciągu jest jeszcze obserwowalna za pomocą jednego epizodu koncentracji uwagi [1, 3]. Zawiera on zarówno aspekt geometryczny (dłuższy ciąg zajmuje większy fragment sceny) jak i informacyjny (dłuższy ciąg zawiera większą ilość informacji). Ciągi znaków były prezentowane w ograniczonym czasie dobieranym z przedziału od 0,1 do 1 [s] tak, aby obserwator nie zdążył wielokrotnie skanować sceny, co było kontrolowane za pomocą systemu okoruchowego Ober-2.



Rys. 2. Przykład ciągu znaków wyświetlanego na ekranie,  $d$  jest współrzędną poziomą względem środka ekranu.

Wstępna analiza 465 pokazów pozwoliła na ustalenie reguł interpretacji ścieżek wzrokowych oraz zjawisk zakłócających zależność liczby epizodów koncentracji

uwagi od geometrycznego rozmiaru prezentowanej informacji:

- zbyt krótki czas pokazu powoduje, że obserwator nie zdąży przeanalizować zauważonej sekwencji znaków i rozpocząć poszukiwania brakującej części (w takich przypadkach otrzymywano jeden punkt koncentracji uwagi, a informacja zawarta w dalszej części ciągu znaków była zgubiona);
- zbyt długi czas pokazu powoduje, że obserwator po całkowitym rozpoznaniu prezentowanego ciągu znaków zaczyna poszukiwać innych interesujących elementów sceny i ścieżka wzrokowa reprezentuje również dodatkowe punkty koncentracji uwagi.

## 2.2 Rekonstrukcja tekstu zniekształconego

Drugi eksperyment polegał na czytaniu tekstu zniekształconego poprzez losową zamianę liter w obrębie wyrazów. Prezentowany tekst zawierał narastającą liczbę przestawień (rys. 3) zmuszających obserwatora do interpolacji zniekształconej informacji, a przy przekroczeniu określonego progu - do poszukiwania brakującej informacji w kontekście sąsiednich wyrazów. Progowa wartość odsetka przestawionych liter była możliwa do określenia przy pomocy generatora zniekształconego tekstu. Eksperyment umożliwiał rejestrację sygnału okoruchowego, a następnie stwierdzenie, w którym wierszu ścieżka wzrokowa obserwatora traciła charakter jednokierunkowo-skokowy, typowy dla czytania płynnego, i zawierała przeskoki i powroty kolejnych punktów koncentracji uwagi, związane z próbami interpretacji zniekształconej informacji w szerszym kontekście.

Tak jak uniersawny jest sam garnitur, tak i dtodaki do niego mogą być bardzo różne. Kolzsue doaberimy więc, zaeźnlie od natosrju czy ozakji, kowloroe lub bñiae, gidakie, w deltikana katrkę czy modne w tym senzoie paski. Ponobdie krwaaty - doasmowpujey je koyorysltcznie do klozusi i garitunru, kieujarc się wñsnaym wycyziuem. Podorzapdkowac nlaeży się tylko żelzanej ziasiadze dobroru butów: czarne władakmy do uranbia szagreo, cienombłeitnegko i grawatonego, do garnirtuów w pozoístaych barawch - tkylo brwazoe.

Rys. 3. Przykład tekstu zniekształconego. Podczas eksperymentu tekst był wyświetlany linijka po linijce.

## 3. REZULTATY

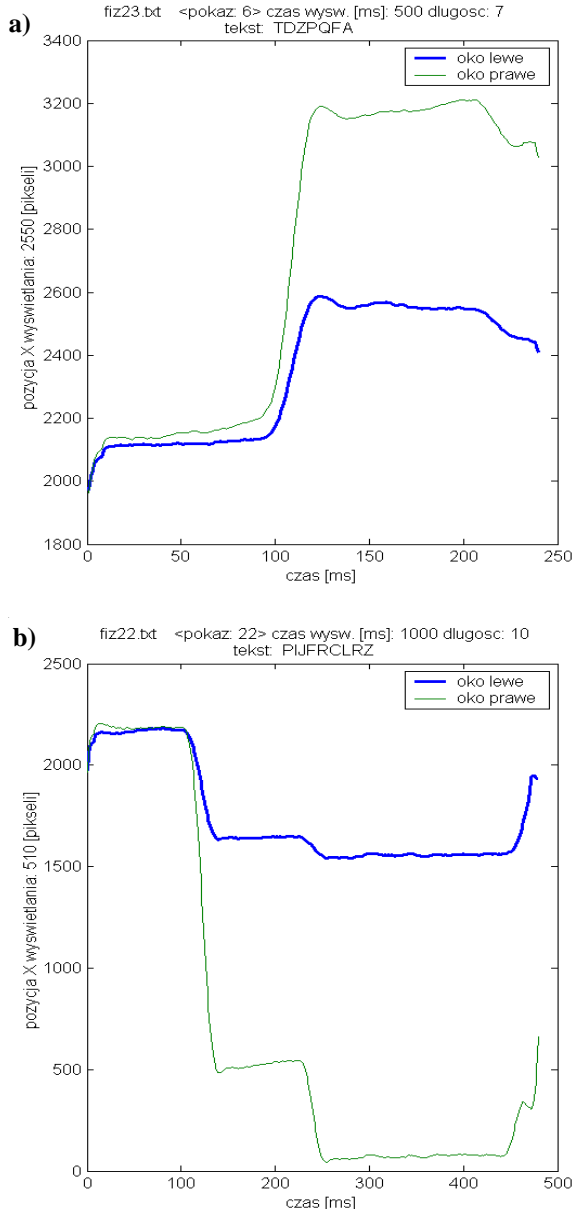
Trudności z opracowaniem w pełni automatycznego algorytmu interpretującego ścieżkę wzrokową i uwzględniającego powyższe ograniczenia były przyczyną manualnej interpretacji ścieżek wzrokowych. Ze zbioru wyników eksperymentów, opisanych w punkcie 2.1, usunięto ścieżki wzrokowe nie zawierające precyzyjnie zlokalizowanych punktów koncentracji uwagi, a także te, dla których czas pokazu był mniejszy niż 500 [ms]. Przyjęty czas uznano za wystarczający do dwukrotnej koncentracji uwagi, jeśli wymaga tego proces postrzegania. W zbiorze pozostałych 245 obserwacji otrzymano wysoką korelację ilości punktów koncentracji uwagi z długością prezentowanego ciągu znaków ( $r$  Pearsona = 0.5196) oraz niewielkie zależności od czasu prezentacji ( $r$  = 0.1948) i odległości prezentowanego ciągu od centrum sceny ( $r$  = 0.1627). Żadna ścieżka wzrokowa nie zawierała trzech punktów koncentracji uwagi. Prawdopodobieństwo wystąpienia dwóch punk-

tów koncentracji uwagi w zależności od długości tekstu przedstawia tabela 1.

Tab.1. Prawdopodobieństwo postrzegania ciągu znaków za pomocą dwóch epizodów koncentracji uwagi.

długość ciągu znaków	kąt widzenia [°]	prawdopodobieństwo wystąpienia dwóch epizodów koncentracji uwagi
4	1.77	0.0889
5	2.22	0.3462
6	2.66	0.4000
7	3.11	0.6207
8	3.55	0.6471
9	3.99	0.7407
10	4.44	0.8026

Przykładowe wykresy składowych poziomych ścieżek wzrokowych przy ograniczonym czasowo rozpoznawaniu ciągu znaków przedstawia rysunek 4.



Rys. 4. Składowe poziome ścieżki wzrokowych dla: a) pojedynczego epizodu koncentracji uwagi b) dwóch epizodów koncentracji uwagi.

W drugim eksperymencie (por. pkt 2.2) parametry ścieżki wzrokowej były opisywane przez różnice prędkości i kierunku w osi poziomej. Graniczny odsetek przestawionych liter wyraża maksymalny interwał, w którym obserwatorowi łatwiej jest interpolować zniekształconą informację niż poszukiwać jej z użyciem systemu wizyjnego. Wybór ten jest oczywiście dokonywany podświadomie i zależy od cech osobniczych.

Spośród 288 pokazów linii tekstu zniekształconego zawierających średnio  $74.7 \pm 9.2$  znaków zidentyfikowano 62 sytuacje powrotu (przesunięcia w lewo) punktu koncentracji uwagi. Przyjęto następujące kryteria:

- obie gałki oczne muszą wykonać ruch powrotny,
- ruch powrotny występuje po rozpoczęciu skanowania linii tekstu i przed zakończeniem (ruchy powrotne styczne do granic tekstu nie były brane pod uwagę),
- długość skoku powrotnego wyrażono jako ułamek całkowitej długości ścieżki wzrokowej w płaszczyźnie poziomej, a zatem do długości tekstu.

Na rysunku 5 przedstawiono przykładowe wykresy składowych poziomych ścieżek wzrokowych, zarejestrowanych podczas obserwacji jednej linijki tekstu.

Długość ruchu powrotnego stanowiła średnio 10.95% długości ścieżki wzrokowej ( $\pm 6.44\%$ ). Zatem biorąc pod uwagę, że ścieżka wzrokowa zawiera średnio 12.8 punktów koncentracji uwagi na odczytywanym tekście sformułowano następujące wnioski:

- niepowodzenie interpolacji tekstu powoduje powrót koncentracji uwagi do znanego fragmentu,
- średnia ilość liter przypadająca na punkt koncentracji uwagi wynosi 5.8359 i jest nieco mniejsza niż w przypadku pokazów izolowanych ciągów znaków.

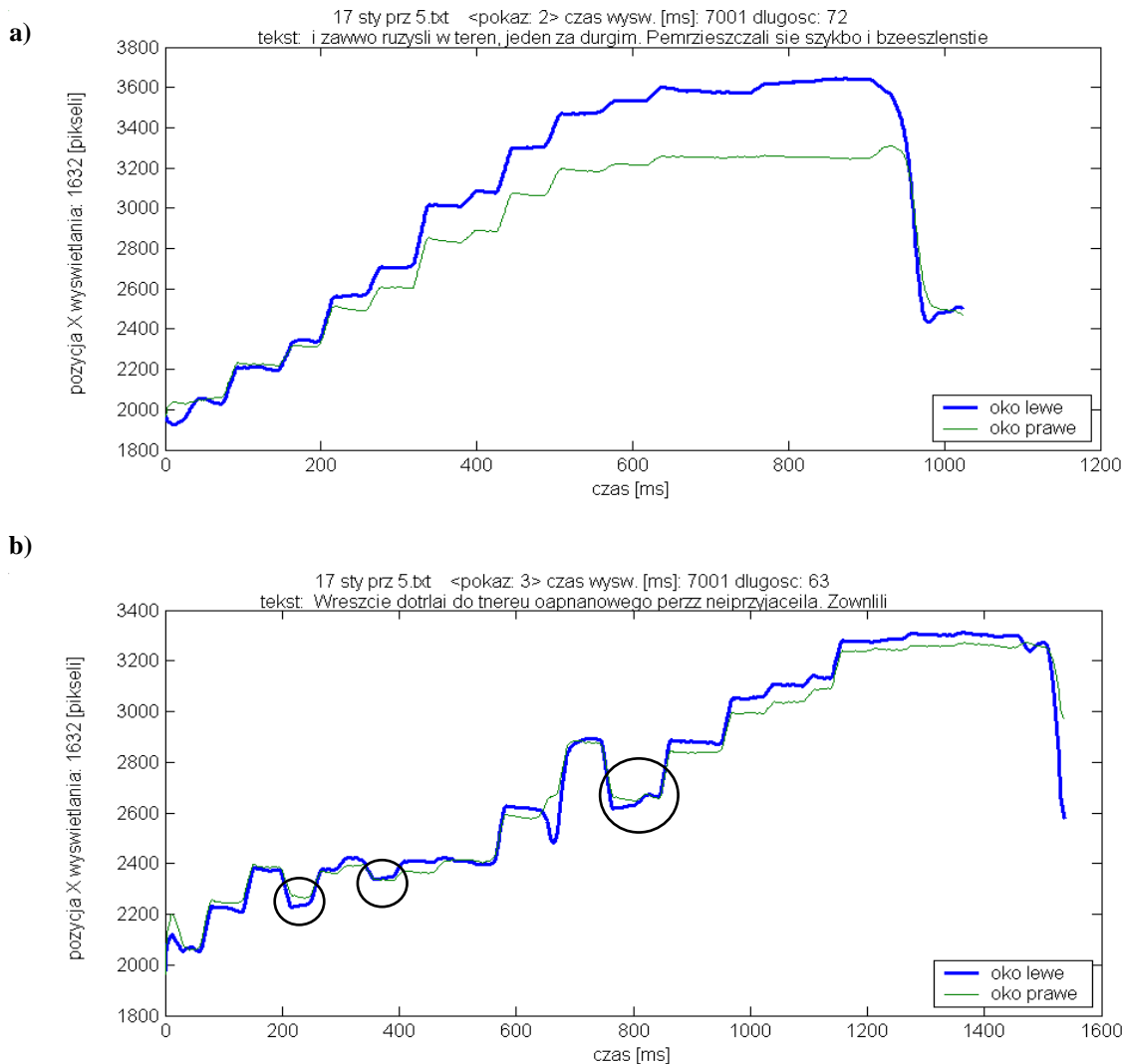
## 4. DYSKUSJA

Przeprowadzone eksperymenty i analiza otrzymanych ścieżek wzrokowych pozwoliły na udzielenie odpowiedzi na pytania, sformułowane we wprowadzeniu:

- porcja tekstu widziana podczas pojedynczego epizodu koncentracji uwagi jest większa niż skok gałki ocznej – średnia odległość punktów koncentracji uwagi stanowi zatem dolne oszacowanie porcji informacji pobieranej ze sceny,
- porcja tekstu możliwa do interpolacji bez angażowania systemu okulomotorycznego do poszukiwań kontekstowych jest mniejsza niż średnia odległość pomiędzy punktami koncentracji uwagi podczas powrotu.

Przeprowadzone eksperymenty wykazały także silną zależność parametrów ścieżki wzrokowej od cech osobniczych. Niektórzy obserwatorzy postrzegali znacznie większe porcje tekstu podczas pojedynczego epizodu koncentracji uwagi.

Przeprowadzono także próby głośnego czytania zniekształconego tekstu, co miało na celu bieżącą weryfikację poprawności interpolacji zniekształconej informacji. Okazało się jednak, że głośna wymowa powoduje znaczną niestabilność pozycji głowy i w konsekwencji zniekształcenie rejestrowanej ścieżki wzrokowej.



Rys. 5. Składowe poziome ścieżki wzrokowych: a) ścieżka wzrokowa o charakterze jednokierunkowo-skokowym b) ścieżka wzrokowa z powrotami koncentracji uwagi (sytuacje powrotu oznaczono okręgami).

Zastosowanie wyników analizy pobierania informacji wizyjnej przez człowieka otwiera możliwości znacznego rozszerzenia pojęcia interaktywności mediów informacyjnych. Jako przykłady pomysłów wartych dalszego rozwijania podać można:

- hipertekst rozwijany w zależności od koncentracji uwagi czytelnika na wybranym wyrazie,
- tekst inteligentny, w którym treść kolejnych wyświetlanych wierszy jest uzależniona od parametrów ścieżki wzrokowej czytelnika,
- interaktywny pokaz wizyjny (informacyjny, edukacyjny lub rozrywkowy), w którym celowa modyfikacja treści kieruje uwagę obserwatora.

## PODZIĘKOWANIA

Opisywane prace badawcze były finansowane przez Akademię Górniczo-Hutniczą w Krakowie, nr umowy: 11.11.120.249. Autorzy wyrażają swe podziękowania studentom: Łukaszowi Chornikowskiemu, Łukaszowi Czajczykowi, Przemysławowi Koprowskiemu oraz Franciszkowi Korcie za zaangażowanie i kreatywność.

## LITERATURA

- [1] Augustyniak P., Mikrut Z.: Lokalizacja istotnych cech obrazu metodą śledzenia ścieżki wzrokowej, *XIII Krajowa Konferencja Biocybernetyka i Inżynieria Biomedyczna*, 931-936, Gdańsk 2003.
- [2] Boccignone G.: An Information-theoretic Approach to Active Vision, *11th International Conference on Image Analysis and Processing ICIAP*, 2001.
- [3] Mikrut Z., Augustyniak P.: Estimation of Execution Time for Tasks of Objects Counting and Localization Using the Ober2 Device, *IFMBE Proc. 2*: 144-145, Reykjavik 2002.
- [4] Ober J., Hajda J., Loska J., Jamicki M.: Application of Eye Movement Measuring System OBER2 to Medicine and Technology, *SPIE, Infrared technology and applications* 3061(1):327-333, Orlando, USA, 1996.
- [5] Pelz J. B., Canosa R.: Oculomotor behavior and perceptual strategies in complex tasks, *Vision Research* 41:3587-3596, 2001.
- [6] Yarbus A. F.: *Eye Movements and Vision*, New York, Plenum Press, 1967.