

# Rafał Linowiecki...

---

## GCAF. Platforma tworzenia aplikacji kontrolowanych wzrokiem – nowy sposób przygotowywania w pełni interaktywnych eksperymentów z użyciem okulografu

---

Lingwistyka Stosowana / Applied Linguistics / Angewandte Linguistik nr 20, 83-99

---

2016

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

**Rafał LINOWIECKI, Jacek MATULEWSKI,  
Agnieszka IGNACZEWSKA, Bibiana BAŁAJ,  
Joanna DRESZER, Magdalena KMIECIK, Włodzisław DUCH**  
Uniwersytet Mikołaja Kopernika w Toruniu

## **GCAF: narzędzie do tworzenia interaktywnych eksperymentów wykorzystujących okulograf i jego użycie w badaniach nabywania mowy przez niemowlęta<sup>1 2</sup>**

**Abstract: GCAF: the environment for designing fully interactive experiments involving eye tracking devices and its use for studying of speech acquisition by infants**

The aim of this paper is to present our computer software, the GCAF framework, which allow one to design and run applications using the eye trackers, among others the cognitive experiments for infants. In comparison to other tools for creating the experiments, the full interaction with user is possible. User can choose the object on screen activated by its gaze, which is most interesting for him, and therefore control the program activity. This framework and the domain language GIML were already used for designing of several experiments involving infants in Neurocognitive Laboratory located at the Interdisciplinary Centre for Modern Technologies in Nicolaus Copernicus University. One of its main objectives is to investigate the loss of the ability to differentiate speech sounds in native language before the end of the first year of life. In this project, we intend to design and verify the effectiveness of the interactive training, the aim of which is to influence this phenomenon.

### **Wstęp**

Celem artykułu jest prezentacja rozwijanego przez nas oprogramowania służącego do projektowania i uruchamiania aplikacji sterowanych wzrokiem. Składa się na nie platforma uruchamiania aplikacji GCAF (od. ang. *Gaze Controlled Application Framework*) oraz interpreter języka znaczników GIML (od ang. *Gaze Interaction Markup Language*). Jest to język służący do tworzenia aplikacji sterowanych wzrokiem, który był projektowany z myślą o osobach niebędących profesjonalnymi programistami. Oprogramowanie to przygotowaliśmy na potrzeby interaktywnych eksperymentów i treningów z udziałem niemowląt prowadzonych w Laboratorium Neurokognitywnym (LNK) na UMK w Toruniu. Badania te dotyczą procesu naby-

---

<sup>1</sup> Artykuł powstał na podstawie fragmentów pracy magisterskiej RL, której promotorem był JM.

<sup>2</sup> Badania są częścią projektu: *NeuroPerKog: rozwój słuchu i pamięci roboczej u niemowląt i dzieci* finansowanego z grantu Symfonia NCN (nr umowy UMO-2013/08/W/HS6/0033).

wania mowy przez niemowlęta między 6 a 12 miesiącem życia. Na tym przykładzie użycia naszego oprogramowania skupię się w tym artykule, ale może być ono używane także do innych celów, chociażby do szybkiego przygotowywania aplikacji ułatwiających komunikację z osobami, dla których spojrzenie jest jedynym sposobem porozumiewania się. Artykuł zawiera zarówno opis samych badań, co obrazuje typowy przypadek użycia GCAF i GIML, jak i garść informacji technicznych. Osoby zainteresowane jego testowaniem i używaniem prosimy o kontakt z autorem korespondencyjnym artykułu.

## 1. Badanie procesu nabywania mowy przez niemowlęta

Komunikacja wymaga wzajemnej interakcji uczestników biorących w niej udział, nawet jeżeli jednym z nich jest niemowlę. Niemowlęta bardzo szybko przestają być bowiem tylko biernymi odbiorcami, a stają się także nadawcami komunikatów, oczywiście na początku niewerbalnych, ale mimo to zwykle wyraźnie czytelnych dla rodziców (np. C. Leclère i in. 2014, N. Ramirez-Esparza i in. 2014, A.F.de C. Hamilton i in. 2016). Badania nad nabywaniem języka oraz percepcją dźwięków mowy w pierwszych latach życia mają fundamentalne znaczenia dla zrozumienia rozwoju ludzkiej mowy, jak i dla zidentyfikowania najważniejszych czynników kształtujących proces uczenia się w ogóle. Niemowlęta okresu przedwerbalnego są zdolne do różnicowania dźwięków mowy zarówno z języka natywnego, jak i języków obcych (P.K. Kuhl 2003). Do około dziewiątego miesiąca życia niemowlęta są w stanie rozróżniać wszystkie dźwięki z dowolnych języków, później stopniowo ta zdolność zanika. Dzieci w wieku od 8 do 12 miesięcy, w którym nabywają krytyczne dla swojego języka fonemy, często kierują swój wzrok na usta rozmówcy. Badania wykazują, że dzieci które w tym okresie częściej i dłużej przyglądają się ustom mówiącej do nich matki, w późniejszym okresie mają lepszą ekspresję języka, wyższy poziom socjalizacji i większy zasób słów.

Jedną z najważniejszych rzeczy, których ucą się niemowlęta jest kategoryzacja – zdolność do grupowania elementów posiadających wspólną cechę (percepcyjną lub funkcjonalną). U starszych dzieci i dorosłych wyniki kategoryzacji poznać można przez odpowiedzi werbalne lub określone zachowanie (np. wskazywanie). U niemowląt będących w fazie przedwerbalnej wykorzystywane są natomiast metody bazujące na zjawisku habituacji, warunkowaniu klasycznym i instrumentalnym. Badania prowadzone z udziałem niemowląt wymagają jednak dobrze określonej procedury i narzędzi, które wystarczająco długo utrzymują uwagę dziecka. Przy badaniach z udziałem niemowląt ważne jest, aby obiekt był atrakcyjny i przyciągający uwagę (np. postać z bajki lub zabawka); możemy manipulować jego wielkością, kształtem, kolorem oraz ruchem. Dobrym sposobem na zwrócenie uwagi dziecka jest wykorzystanie dźwięku, który towarzyszy poruszającemu się obiektowi.

Ważnym procesem w trakcie nauki mowy jest wykształcenie umiejętności rozróżniania fonemów mowy ojczystej, a jednocześnie dyskryminacja fonemów z języków obcych. Następuje to zwykle między 6 a 12 miesiącem życia. Możliwe jest jednak czasowe podtrzymanie wrażliwości na obce fonemy, które w przyszłości powinno zaowocować łatwiejszą nauką języków obcych u tych dzieci (P.K. Kuhl i

in. 2003, 2004, 2011, 2015). Eksperymenty oparte na tym założeniu przeprowadzane są aktualnie w naszym laboratorium.

### **1.1. Użycie okulo grafu w badaniach niemowląt**

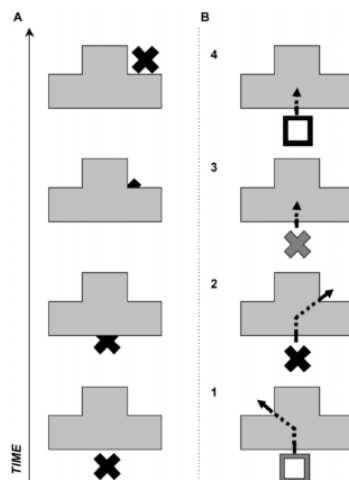
Bezpieczną i nieinwazyjną metodą często wykorzystywaną w badaniach niemowląt jest okulografia. Poetycko ujął to R.N. Aslin (2012) pisząc, że oczy dziecka są oknem do poznania jego rozwoju, szczególnie kiedy mowa właśnie o dzieciach okresu przedwerbalnego. Okulografia pozwala na śledzenie ruchów gałek ocznych, ścieżek spojrzeń, czy zmian wielkości źrenicy. Dzięki rejestracji reakcji wzrokowej niemowlęcia możemy w sposób obiektywny i dokładny ustalić co przyciąga jego wzrok i na jak długo. Zdolność niemowląt do sterowania przestrzenną uwagą selektywną może być mierzona za pomocą analizy fiksacji wzrokowych w określonych regionach zainteresowania. Przesunięcia uwagi wzrokowej są wykorzystywane również w badaniu różnicowania dźwięków mowy u niemowląt z zastosowaniem procedury wymagającej przewidywania. Zadania wymagające koordynacji oko-ręka pozwalają na szybsze i efektywniejsze zwrócenie uwagi, co może być znaczące w badaniach procesów poznawczych, w których zaangażowanie uwagi wzrokowej jest warunkiem koniecznym i etapem wstępnym dla dalszego przetwarzania danych.

### **1.2 Rola interakcji w procesie nabywania mowy**

W laboratoryjnych badaniach eksperymentalnych z udziałem niemowląt, rola dziecka jest często ograniczana do biernego odbiorcy bodźców prezentowanych przez prowadzącego badanie na ekranie komputera, po których zaprezentowaniu mierzona jest jego reakcja. Niemowlę nie ma w tej sytuacji żadnej kontroli nad procesem, w którym bierze udział, trudno więc mówić o nawiązaniu komunikacji, czy choćby kontaktu, które mają być przecież przedmiotem badania. Niewiele jest badań wprowadzających choćby prostą interakcję wzrokową do eksperymentów z udziałem niemowląt. Te, które są znane w literaturze umożliwiają uruchomienie kolejnej próby eksperymentu, np. gdy niemowlę spojrzy na postać której towarzyszy dźwięk (B. Albareda-Castellot i in. 2011). Eksperymenty Patrycji P.K. Kuhl i jej współpracowników (2003, 2004, 2011, 2015) pokazują jak bardzo interakcja jest ważna dla nabywania mowy u niemowląt. Bierna ekspozycja dziecka na język, bez jego aktywnego udziału w komunikacji, jest znacznie mniej skuteczna.

### **1.3. Wybrane procedury eksperymentalne w badaniach nabywania mowy przez niemowlęta**

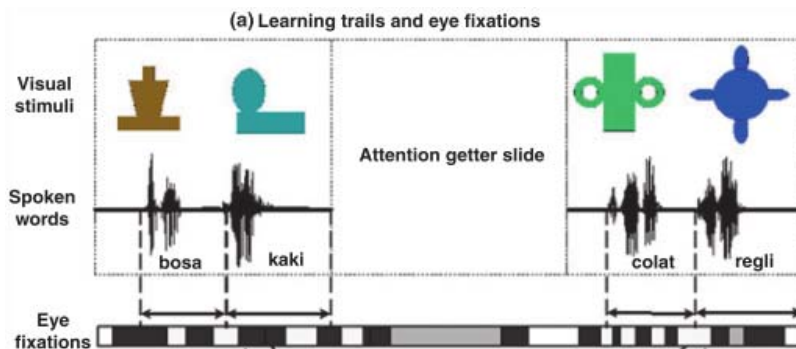
Stosowane w eksperymentach procedury opierają się zwykle na umiejętności przewidywania pozycji pojawiającego się na ekranie obiektu, która zależy od podanego wcześniej bodźca dźwiękowego (por. B. Albareda-Castellot 2011, J. Bjerva 2011). Przeprowadzono jednak również szereg innych eksperymentów, w których ten paradigmat został zmodyfikowany.



Rysunek 1. Procedura eksperymentu zaproponowana przez B. McMurray'a i R.N. Aslina (2004).

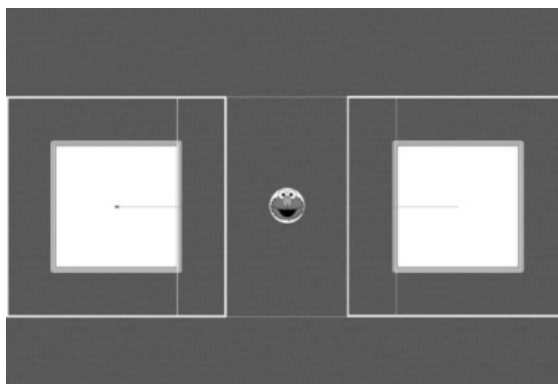
Jednym z przykładów są badania B. McMurray'a oraz R.N. Aslina (2004), którzy współpracując z sześciomiesięcznymi niemowlętami zaprojektowali serię czterech eksperymentów opartych na zasadzie wymuszonego wyboru z dwoma możliwościami (ang. *two alternative forced choice*, 2AFC). Pierwsze trzy sprawdzały kategoryzację koloru i kształtu. Początkowo stosowano procedurę, w której obiekty na ekranie ułożone były poziomo, z lewej lub prawej strony bodźca centralnego, którym był kwadrat, krzyżyk lub koło. Stopniowo jednak procedurę utrudniano. W pierwszym eksperymencie, w zależności od obiektu widocznego po środku (kwadrat lub krzyżyk), nagroda w postaci zdjęcia zwierzęcia pojawiała się po jego lewej, bądź prawej stronie. W drugim eksperymencie jako obiekt warunkujący pojawienie się nagrody wprowadzono koła z pionowymi i poziomymi kreskami. Trzeci etap wprowadzał przesłonę w kształcie odwróconej litery T (rysunek 1), natomiast kształt obiektu startowego, w tym wypadku ponownie kwadrat bądź krzyżyk, warunkował stronę jego pojawienia się. Eksperyment czwarty wprowadzał dodatkowo kategoryzację dźwiękową. Badania nie wykazały preferencji dla koloru; kwadrat jako kształt wzbudzał więcej zainteresowania. Dzieci szybko uczyły się samej procedury i wybierały poprawnie zarówno w badaniach prezentujących pojawianie się obiektu linearnie, jak i tych, które tę linearność zaburzały. Wiedza ta dała podstawę do stworzenia zasadniczego czwartego eksperymentu, w którym wykorzystywano przesłonę w kształcie odwróconej litery T (rysunek 1). W chwili początkowej u podstawy pojawiał się bodziec wzrokowy (koło), który po aktywacji wzrokiem zniknął za przesłoną i w zależności od tego, jaki dźwięk został odtworzony, pojawiał się po pół sekundzie z lewej lub z prawej strony przesłony. Bodźcami dźwiękowymi były angielskie słowa *lamb* i *teak*. Były one powtarzane trzykrotnie, w momencie zniknięcia obiektu za przesłoną. Całość badania wykazała znaczącą preferencję strony lewej dla słowa *lamb*.

Również R.N. Aslin i B. McMurray (2005) sprawdzali wrażliwość niemowląt na różnice w obrębie jednej kategorii głosek, czyli spółgłoski dźwięcznej i jej bezdźwięcznego odpowiednika. W tym badaniu niemowlętom podawano 80 reprezentacji różnych słów zaczynających się na /b/ lub /p/. Porównywali zarówno różnicowanie tych liter między sobą, jak i różnicowanie długości czasu odsunięcia dźwięczności (ang. *voice-onset time*, VOT) w obrębie każdej z nich. Jak pokazują wyniki, niemowlęta bardzo dobrze różnicują zarówno /b/ i /p/, jak i długie /b\*/ i krótsze /b/. Co więcej, badania te wykazały również znaczącą preferencję dla słów zawierających /b\*/.



Rysunek 2. Trening realizowany przez C. Yu i L. B. Smith (2010).

C. Yu i L.B. Smith (2010) za pomocą reakcji wzrokowych sprawdzali jak szybko dzieci w wieku 14 miesięcy uczą się słów „na bieżąco”. Wykorzystana do tego została omówiona wyżej procedura 2AFC (rysunek 2). Początkowo dzieci przechodziły część treningową, podczas której prezentowano na ekranie komputera figurę geometryczną i 3-krotnie powtarzano dopasowane do niej pseudosłowo („kaki”, „bosa”, „manu”, „regli”, „golat” i „basser”). Celem sesji treningowej było stworzenie skojarzeń między obiektami wizualnymi i każdym z sześciu pseudosłów. Część właściwa badania polegała na przedstawianiu dwóch rysunków jednocześnie, które pojawiały się z opóźnieniem w stosunku do prezentowanego bodźca dźwiękowego. Tym sposobem sprawdzano, czy podczas sesji treningowej poprawnie zapamiętano pary. Eksperyment ten potwierdził, że dzieci są w stanie nauczyć się rozróżniać pseudosłowa, jednak umiejętność ta jest silnie zależna od częstotliwości prezentowania materiału.



Rysunek 3. Procedura zaproponowana przez M. Shukłę et al. (2011).

Kolejnym przykładem badań umiejętności językowych niemowląt są eksperymenty przeprowadzone przez M. Shukłę i współpracowników (2011). Początkowa procedura, podobnie jak u R.N. Aslin (2004), opierała się na wykorzystaniu przesłony w kształcie odwróconej litery T. Została jednak zmieniona na dwa obszary z przemieszczającym się obiektem centralnym z rysunkiem postaci z Ulicy Sezamkowej (Elmo lub Ciasteczkowy Potwór) (rysunek 3). Podawanymi dźwiękami były sylaby /ta/ i /ga/ w różnych reprezentacjach głosowych. Badania te wykazały, że niemowlęta są w stanie pracować multimodalnie różnicując zarówno bodźce dźwiękowe, jak i prezentowany bodziec wzrokowy. Niemowlęta w trakcie kolejnych prób z coraz większą poprawnością przewidywały obszar, na którym pojawi się rysunek, zarówno na podstawie wyświetlanej postaci, jak i biorąc pod uwagę bodziec dźwiękowy. Co ciekawe M. Shukla nie korzystał z tradycyjnych programów do projektowania eksperymentów psychologicznych, ale podobnie jak my stworzył własną platformę do badań. Platforma ta nazywa się SMART-T (od ang. *System of Monitoring Anticipations in Real Time with the Tobii*); jak wskazuje jej nazwa wykorzystuje eyetracker firmy Tobii, część informatyczna została przygotowana w środowisku MATLAB z użyciem modułu *Talk2Tobii*.

## 2. Platforma GCAF i język GIML

Celem pełniejszego wykorzystania interakcji wzrokowej u niemowląt stworzyliśmy narzędzie – program komputerowy, który pozwala przygotowywać eksperymenty, w których samo niemowlę inicjuje wzrokiem działanie poszczególnych elementów aplikacji i ma możliwość wyboru spośród kilku bodźców tego, który je najbardziej zainteresuje. Wchodzi zatem w interakcję z programem, który zmienia się w zależności od decyzji podejmowanych przez dziecko. Równocześnie okulograf rejestruje dane dotyczące miejsca spojrzenia. Jak już wspomnieliśmy wyżej, oprogramowanie to nazwaliśmy GCAF (od ang. *Gaze Controlled Application Framework*).

Pozostała część tego podrozdziału przeznaczona jest przede wszystkim dla osób, które zainteresowane możliwościami platformy GCAF i języka GIML, chcą poznać szczegóły techniczne i być może spróbować samodzielnie stworzyć eksperymenty korzystające z tych narzędzi. Jest to siłą rzeczy bardzo wstępny opis, daleki od kom-

pletności, ale mający przekonać czytelników, że samodzielne tworzenie przez nich w pełni interaktywnych eksperymentów jest jak najbardziej możliwe. Dostępny jest również tutorial ułatwiający naukę języka GIML.

## 2.1. Platforma GCAF

Powtórzmy: GCAF jest platformą, w której korzystając z nowego języka znaczników GIML można tworzyć aplikacje sterowane wzrokiem. Platforma ta obsługuje w tej chwili okulometry firmy SMI, ale zaawansowane są już prace nad wdrożeniem otwartego protokołu *Open Eye-Gaze Interface* (C. Hennessey i in. 2010), które wykorzystywany jest m.in. w tańszych urządzeniach firm Mirametrix i Tobi.

Odpowiednio przygotowane pliki XML z kodem GIML są analizowane przez platformę GCAF i na ich podstawie budowany jest zestaw scen pokazywanych użytkownikowi. Scena może zawierać wiele obszarów pokazujących tekst, rysunki lub filmy. Użytkownik patrząc na poszczególne obszary może sterować działaniem aplikacji, w tym: włączaniem, wyłączaniem i aktywacją poszczególnych obszarów, uruchamianiem animacji i nawigacją pomiędzy scenami.

Platforma GCAF jeszcze nie w pełni implementuje projekt języka GIML. Obecnie trwają prace nad dodaniem kontrolki służącej do wyświetlania tekstu z automatycznym podziałem na zdania, wyrazy i litery, co ułatwi przygotowywanie eksperymentów badających czytanie, oraz kontrolki umożliwiającej wprowadzanie tekstu. Planowane są cztery rozwiązania, klasyczna klawiatura QWERTY ze stałym czasem zatwierdzenia dla każdej wprowadzanej litery (ang. *dwel-time*) i alternatywnie z płynną metodą wprowadzania (ang. *swipe*), a także dwa nowatorskie rozwiązania zaproponowane przez A. Hackaufa/ M. Urbina (2007) oraz J.O. Wobbrocka i in. (2008).

W kontekście projektowania badań psychologicznych, GIML i GCAF wyróżnia pełna interaktywność tworzonych z ich pomocą eksperymentów. Możliwa jest dowolna ścieżka przebiegu eksperymentu – to badany wybiera kolejność aktywowanych bodźców i tym samym związane z ich aktywacją działania programu. Istnieje jednak również możliwość narzucenia kolejności aktywowanych obszarów lub ograniczenia ich dostępności w ustalonych przez projektanta aplikacji warunkach. To pozwala na zaprojektowanie eksperymentów, których przebieg może być kontrolowany w stopniu, który ustala projektujący badanie.

## 2.2. Gramatyka języka GIML

Język GIML jest językiem deklaratywnym, opisującym efekt, który chcemy uzyskać (tj. wygląd i zachowanie aplikacji), a nie algorytm, który do takiego efektu prowadzi. To czyni go łatwiejszym w użyciu dla osób, które nie są profesjonalnymi programistami. Jak wspomniano, wygląd aplikacji opisywany jest znacznikami. Tworzenie aplikacji polega zatem na edycji plików XML, które zawierają odpowiednią strukturę znaczników określonych standardem języka GIML. Rozważaliśmy przygotowanie specjalnego edytora, który pozwalałby na projektowanie interfejsu za pomocą myszy, ale porzuciliśmy ten pomysł ze względu na liczbę opcji, jakie musia-



by ten edytor udostępniać użytkownikowi – rozmiary menu przytłoczyłyby każdego użytkownika tworząc trudny do pokonania tzw. próg wejścia. Zwróćmy uwagę, że bardzo podobna sytuacja jest w przypadku języków HTML i XAML: pomimo, że dostępne są wizualne edytory, w których posługujemy się przede wszystkim myszą, znacznie bardziej popularne są zwykłe edytory kodu wyposażone w mechanizm podpowiadania składni. Również i my planujemy stworzyć moduł podpowiadający znaczniki i atrybuty GIML w ramach mechanizmu *IntelliSense* dostępnego w darmowym środowisku *Visual Studio Community*. Ponadto programowanie w GIML ułatwia walidator wbudowany w platformę GCAF, który wskazuje błędne znaczniki w pliku. Zostanie on również zintegrowany z Visual Studio.

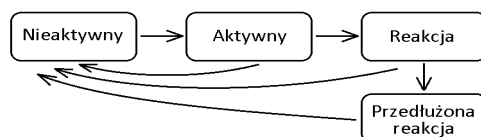
Łatwość używania języka GIML bierze się również z tego, że jest on stosunkowo prosty – składa się z zaledwie kilku znaczników. Ogromna jest natomiast liczba atrybutów, jakimi możemy kontrolować wygląd i zachowanie kontrolki. Jednak początkujący użytkownik, aby móc pisać pierwsze aplikacje musi znać tylko kilka z nich.

### 2.3. Sceny

Listing 1. Szablon interfejsu opisany w GIML (z lewej w języku polskim, z prawej w języku angielskim).

<pre>&lt;sceny nazwaScenyDomyślnej="0" ...&gt;   &lt;scena nazwa="0" ...&gt;     &lt;obszar ...&gt;       &lt;aktywacja /&gt;       &lt;reakcja /&gt;     &lt;/obszar&gt;     ...   &lt;/scena&gt;   ... &lt;/sceny&gt;</pre>	<pre>&lt;scenes nameOfDefaultScene="0" ...&gt;   &lt;scene name="0" ...&gt;     &lt;region ...&gt;       &lt;activation /&gt;       &lt;reaction /&gt;     &lt;/region&gt;     ...   &lt;/scene&gt;   ... &lt;/scenes&gt;</pre>
---	---

Struktura interfejsu aplikacji GIML odzwierciedlona jest w hierarchii znaczników (listing 1). Interfejs aplikacji GIML zorganizowany jest w sceny; każda scena zawiera obszary, które mogą znajdować się w różnych stanach przełączanych spojrzeniem (por. rysunek 4). Podstawowym stanem obszaru jest *nieaktywny*. Spojrzenie użytkownika, ale nie dłuższe niż jedna sekunda (ten parametr można modyfikować), przełącza obszar w stan aktywny. Natomiast spojrzenie dłuższe niż sekunda przełącza go w stan *reakcji*. Przeniesienie spojrzenia na inny fragment sceny powoduje powrót obszaru do stanu nieaktywnego. Możliwe jest niezależne ustalenie wyglądu i zachowania obszaru w każdym z tych trzech stanów, włączając w to zmianę tekstu, rysunku, dźwięku, filmu, animacji czy akcji, jakie obszar może wyzwać.



Rysunek 4. Diagram stanów obszaru. Stan przedłużona reakcja dotyczy sytuacji, w której obszar wykonuje dłuższą czynność aktywowaną przez zmianę stanu np. płynna zmiana pozycji obszaru.

Jedynym obowiązkowym atrybutem znacznika scena jest jej nazwa. Więcej obowiązkowych atrybutów ma znacznik obszar, w którym poza nazwą musimy określić jego kształt, rozmiar i położenie na ekranie. Poza tym dla każdego stanu osobno można określić wyświetlany tekst, czcionkę, kolor i grubość ramki, rysunek, dźwięk, film, obszary włączane i wyłączane przy wejściu lub wyjściu obszaru do danego stanu oraz scenę, do której nawigujemy. Można również określić animację obszaru lub samej jego zawartości. Ważnym atrybutem obszaru jest też treść komunikatu zapisywanego w danych rejestrowanych przez eyetracker przy zmianie stanu obszaru – jego użycie znakomicie ułatwia późniejsze analizy danych okولوجraficznych. Opisy stanów aktywnego i reakcji umieszczone są w osobnych pod-znacznikach aktywacja i reakcja zagnieżdżonych w znaczniku obszar.

Wyróżnikiem GIML jest to, że atrybuty mogą określać nie tylko wygląd, ale również zachowanie obiektów na ekranie, a także ich reakcję na działania użytkownika, w tym przede wszystkim na spojrzenie. Odpowiadają za to atrybuty znacznika obszar. Przykład takiego znacznika widoczny jest w listingu 2 (na listingu wytłuszczone są atrybuty obowiązkowe). W tym przykładowym znaczniku w obszarze wyświetlany jest jedynie tekst, inny w każdym stanie. Zmienia się również kolor czcionki. Jak widać z listingu oznacza to, że atrybuty tekst i kolor czcionki powinny być nadpisywane w pod-znacznikach aktywacja i reakcja. Najważniejsze jest jednak to, że w stanie *reakcja* określona jest akcja polegająca na zmianie bieżącej sceny na inną. Inne możliwe akcje to narysowanie ramki ściągającej uwagę, czy przesunięcie obszaru w inne miejsce.

Listing 2. Prosty przykład znacznika Obszar – najbardziej rozbudowanego znacznika w języku GIML

```

<obszar nazwa="PrzykładowyObszar"
kształt="prostokąt" rozmiarX="400" rozmiarY="50"
położenieŚrodkaX="300" położenieŚrodkaY="450"

  tekst="Dłuższe spojrzenie na obszar zmienia scenę"
  czcionka="Times" rozmiarCzcionki="20"
  kolorCzcionki="Black" >

  <aktywacja tekst="Nastąpiła aktywacja"
    kolorCzcionki="Navy" />

  <reakcja typAkcji="PrzejscieDoSceny"
    nazwaDocelowejSceny="innaScena"
    tekst="Nastąpiła reakcja"
    kolorCzcionki="Blue" />

</obszar>

```

Obszary są domyślnie włączone, ale można także zadeklarować obszar, który jest początkowo wyłączony, a włączany jest dopiero po zajściu jakiegoś zdarzenia. To daje możliwość dowolnego zmieniania zawartości sceny w wyniku interakcji programu z użytkownikiem.

#### 2.4. Rysunki, dźwięki i filmy. Szablon pliku GIML

Jak wspomniano każdy z obszarów może mieć przypisany obraz, dźwięk lub film. W praktyce obiekty te często są wykorzystywane wielokrotnie w różnych obszarach i na różnych scenach. Aby uniknąć redundantnego alokowania pamięci, wprowadziliśmy abstrakcyjną warstwę zasobów (ang. *assets*), czyli zajmujących pamięć obiektów wykorzystywanych w programie. Dla przykładu każdy wykorzystywany rysunek musi być najpierw zadeklarowany w znaczniku `rysunek`, gdzie nadawana jest mu unikalna nazwa (por. listing 3). Obszary odwołują się do zasobów korzystając właśnie z tych nazw. To pozwala nie tylko uniknąć nadmiarowego obciążenia pamięci (ważne szczególnie w przypadku filmów), ale także umożliwia centralne zarządzanie zasobami oraz dodatkową kontrolę nad nimi (np. ustalenie liczby powtórzeń lub poziomu głośności).

Listing 3. Szablon pliku GIML pokazujący sposób wykorzystywania zasobów

```
<?xml version="1.0" encoding="UTF-8"?>
<ustawienia katalog="D:\GIML\Zasoby" kodJęzyka="pl">

  <rysunki>
    <rysunek nazwa="elmo" ścieżka="rysunki\elmo.png" />
    ...
  </rysunki>
  <dźwięki>
    <dźwięk ... />
  </dźwięki>
  <filmy>
    <film nazwa="ciastko" ścieżka="filmy\ciastko.avi"
      liczbaPowtórzeń="2" głośność="0.3" />
  </filmy>

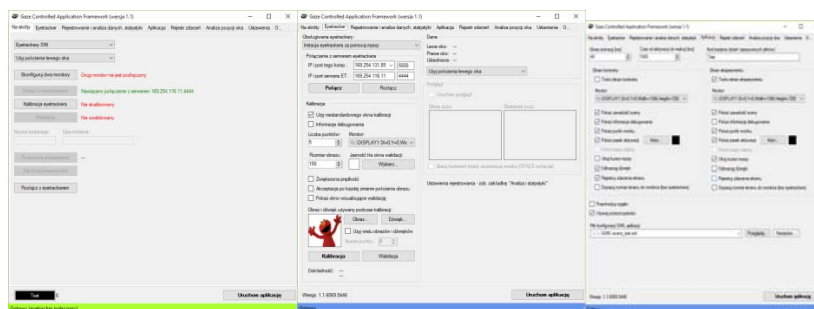
  <sceny nazwaScenyDomyślnej="ulicaSezamkowa">
    <scena nazwa="ulicaSezamkowa">
      <obszar nazwa="Obszar wyświetlający rysunek" ...
        nazwaRysunku="elmo" />
      <obszar nazwa="Obszar wyświetlający film" ...
        nazwaRysunku="ciastko" nazwaDźwięku="ciastko" />
    </scena>
  </sceny>
</ustawienia>
```

Listing 3 pokazuje ogólny szablon pliku GIML. Tzw. korzeniem drzewa XML jest w nim znacznik `ustawienia`. Nieobowiązkowymi atrybutami tego znacznika są: ścieżka katalogu, w którym znajdują się zasoby oraz kod języka używanego w bieżącym pliku (na razie może to być język polski lub angielski). Wewnątrz tego znacznika mogą znajdować się cztery znaczniki: `rysunki`, `dźwięki`, `filmy` i

sceny. W każdym z nich może znajdować się tylko jeden rodzaj znacznika, odpowiednio: rysunek, dźwięk, film i scena. Ten ostatni został już pobieżnie omówiony. Pozostałe trzy są od niego znacznie prostsze. W każdym z nich obowiązkowe są tylko dwa atrybuty: nazwa zasobu oraz ścieżka do pliku, w jakim ów zasób się znajduje. Możliwe jest jednak także np. dopasowanie głośności dźwięku lub ustalenie liczby jego powtórzeń.

Powyższy opis GIML jest daleko niekompletny. Jego zadaniem jest bowiem tylko przekonanie Czytelnika, że tworzenie aplikacji z użyciem języka GIML jest podobne do przygotowywania stron HTML i można się go szybko nauczyć nawet nie będąc informatykiem. Aby to potwierdzić planujemy przeprowadzenie testów tzw. metodą zaznajamiania (ang. *coaching method*) J. Nielsena (1993), w której moderator pomaga użytkownikowi w nauce języka, odpowiadając (co jest nietypowe w testach użyteczności) na wszelkie pytania. Analiza pytań zadawanych przez osoby uczące się GIML pozwoli na ocenę użyteczności języka i ewentualne wykrycie miejsc, które są trudne do opanowania.

## 2.5. Uruchamianie aplikacji GIML. Dane zbierane przez platformę GCAF



Rysunek 5. Okno konfiguracji platformy GCAF; od lewej zakładka „Na skróty”, zakładka konfiguracji okulografu i zakładka uruchamiania aplikacji GIML.

Po uruchomieniu platformy pojawia się okno (rysunek 5), które umożliwia konfigurację jej działania, połączenie z okulometrem, wybór, weryfikację poprawności i uruchomienie pliku GIML oraz selekcję danych, które mają być zapisywane w trakcie działania aplikacji. Oprócz zbierania i rejestrowania danych, program może analizować je „w locie” wykrywając podstawowe zdarzenia: fiksacje, sakady i mruknięcia. Możliwa jest również rejestracja przebiegu eksperymentu na filmie. Po uruchomieniu, aplikacja GIML zajmuje cały ekran (na pozostałych ekranach może być wyświetlana „zasłona”, aby ich zawartość nie zwracała uwagi badanego). Aplikację tę możemy zakończyć z poziomu okna konfiguracyjnego (jeżeli korzystamy z wielu ekranów) lub naciskając klawisz *Escape* na klawiaturze. Przykłady aplikacji GIML używane w eksperymentach nad rozwojem mowy u niemowląt omówione zostały w części 4.

Platforma GCAF zapisuje kilka rodzajów plików z danymi zebranymi w trakcie działania aplikacji GIML. Są to m.in: sekwencja zrzutów ekranu lub film z przebie-

gu badania, położenia spojrzenia przekazywane przez eyetracker, statystyki regionów zainteresowania, tzw. logi relacjonujące przebieg działania aplikacji, w tym zmiany stanów obszarów i przełączanie scen. Wszystkie tworzone pliki to pliki tekstowe (CSV), ich ścieżki ustalane są z poziomu okna konfiguracyjnego. Wraz z uruchomieniem aplikacji platforma może również włączyć rejestrację spojrzenia na serwerze eyetrackera.

### 3. Czy użytkownicy mogą sami tworzyć aplikacje korzystające z eyetrackera?

Specjalistyczne oprogramowanie tworzone na zamówienie dostarczane jest zwykle w postaci pakietu zawierającego gotowy do uruchomienia (skompilowany) program oraz instrukcję obsługi. Kod źródłowy najczęściej pozostaje własnością firmy tworzącej oprogramowanie. To oznacza, że klient otrzymuje program, który posiada z góry ustalone funkcjonalności zgodne z zadeklarowanymi wcześniej wymaganiami, a zmiana lub rozbudowa programu wymaga złożenia nowego zamówienia w tej samej firmie. W taki prosty sposób firmy wytwarzające oprogramowanie „wiążą” swoich klientów. Taki scenariusz jest tym bardziej prawdopodobny, jeżeli zamawiane oprogramowanie jest wysoce specjalizowane. Tak będzie np. w sytuacji, gdy oprogramowanie jest przygotowywane dla konkretnej osoby niepełnosprawnej i jest dostosowane do jej możliwości ruchowych lub jeżeli aplikacja ma korzystać z nietypowych urządzeń, choćby eyetrackera. Jest to sytuacja szczególnie niekorzystna w przypadku osób niepełnosprawnych, których sytuacja materialna jest często trudna. Ale również w przypadku oprogramowania tworzonego na potrzeby badań naukowych wysoka cena ogranicza możliwość swobodnego próbowania różnych rozwiązań i szukania najlepszego z nich. Ponieważ sam proces wytwarzania oprogramowania jest dość kosztowny, w praktyce wymusza zawężenie realizowanego zamówienia do nieinnovacyjnego rozwiązania, którego skuteczność można jednak potwierdzić przynajmniej referencjami do literatury przedmiotu.

Aby tę sytuację zmienić w obu tych dziedzinach, w których ważne jest użycie okulometrów, a więc w przypadku aplikacji projektowanych dla osób niepełnosprawnych i aplikacji wykorzystywanych do badań z zakresu psychologii, tworzymy oprogramowanie, które pozwoli osobom, które nie są profesjonalnymi programistami na przygotowywanie stosunkowo rozbudowanych aplikacji, w których możliwa będzie pełna interakcja z programem. Osoby te to np. psycholog lub nawet student psychologii projektujący eksperyment z użyciem okulometru lub ktoś z otoczenia osoby niepełnosprawnej, kto chce przygotować aplikację umożliwiającą komunikację z chorym.

Twierdzenie, że języka GIML i platformy GCAF mogą używać osoby nie będące programistami, a mimo to tworzyć z ich pomocą w pełni użyteczne aplikacje wykorzystujące okulometry, wydaje się trudne do obrony. Oznacza to bowiem, że platforma i język muszą dawać maksymalnie dużą swobodę użytkownikowi, który projektuje aplikację, ale jednocześnie projektowanie to musi być łatwe. Na szczęście są przykłady pokazujące, że te dwa z pozoru sprzeczne warunki są możliwe do pogodzenia. Najlepiej znanym jest język HTML (od ang. *Hyper-Text Markup Language*), który został zaprojektowany do wymiany informacji przez fizyków pracujących

w CERN, a obecnie używany jest jako podstawowy język Internetu służący do opisu zawartości stron WWW. W podstawowej wersji HTML może być w ciągu jednego dnia opanowany przez każdego gimnazjalistę. Naszą główną inspiracją podczas projektowania GIML nie był jednak HTML, a język XAML (ang. *eXtensible Application Markup Language*) firmy Microsoft, który służy do projektowania wyglądu interfejsu aplikacji, a częściowo również sposobu ich zachowania.

Wzorując się na HTML i XAML, język GIML również uczyniliśmy językiem znaczników (ang. *markup language*). Mówiąc bardziej precyzyjnie, wykorzystaliśmy ogólny schemat języka znaczników XML, w którym zdefiniowaliśmy grupę znaczników pozwalających na opisanie poszczególnych elementów interfejsu aplikacji oraz atrybuty określające zachowania tych elementów. Warto wspomnieć, że jednym z założeń GIML jest jego wielojęzyczność – obecnie każdy znacznik i atrybut GIML ma wersję polską i angielską. W najbliższych planach jest przygotowanie wersji niemieckiej i francuskiej.

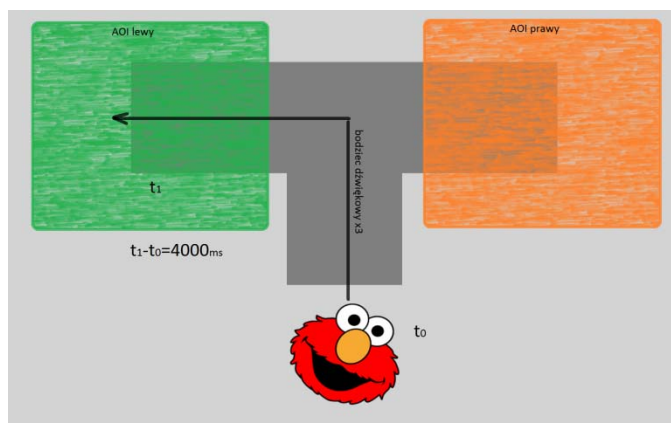
#### **4. Przykłady użycia platformy GCAF w eksperymentach badających nabywanie języka u niemowląt**

Platforma GCAF została użyta w badaniach realizowanych w LNK w ramach projektu *NeuroPerKog: rozwój słuchu i pamięci roboczej u niemowląt i dzieci*. Jednym z jego celów jest zbadanie utraty zdolności do różnicowania dźwięków mowy nie występujących w języku natywnym przed ukończeniem pierwszego roku życia z równoczesnym wykształcaniem specjalizacji w rozpoznawaniu dźwięków mowy języka ojczystego. W ramach tego projektu zamierzamy zaprojektować i zweryfikować skuteczność interaktywnego treningu, którego celem jest opóźnienie tego zjawiska. W szczególności chcemy sprawdzić, czy obecnie powszechnie przyjmowana hipoteza o konieczności społecznej interakcji opiekunów z niemowlakami jest słuszna, czy też do nabywania zdolności rozróżniania fonemów różnych języków wystarczy interakcja z komputerowym systemem sterowanym przez reakcje dziecka.

Opisany w części 2 język znaczników GIML używany jest w platformie GCAF do opisu interfejsu i działania aplikacji, które mogą być zarówno testami diagnostycznymi, jak i treningami. Za pomocą języka GIML przygotowaliśmy dwie metody typu 2AFC, w których niemowlę przewiduje pozycję obiektu pojawiającego się na ekranie na podstawie wcześniej słyszanego bodźca dźwiękowego. W naszym przypadku bodźcami są pseudosłowa z par różniących się tylko jednym fonemem.

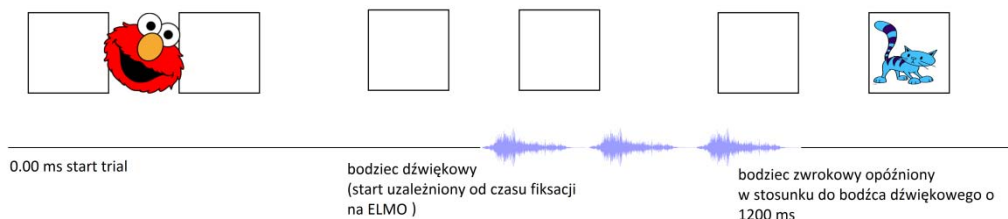
W przypadku pierwszej metody diagnostycznej używanej w prowadzonych w naszym laboratorium badaniach, wzorujemy się na procedurze eksperymentu zrealizowanego przez B. Albareda-Castellot i in. (2011). Dziecko widzi rysunek postaci Elmo u dołu ekranu, który ściąga jego uwagę. Spojrzenie na rysunek uruchamia odtwarzanie pseudosłowa, a jednocześnie rysunek chowa się za przesłoną w kształcie litery „T”. Po 4 sekundach rysunek wyłania się zza lewego lub prawego ramienia przesłony; strona zależy od prezentowanego bodźca dźwiękowego. Pomiar miejsca spojrzenia dziecka (zob. obszary zainteresowania zaznaczone na rysunku 6) wykonywany jest zanim rysunek Elmo wyłoni się zza przesłony, a więc pokazuje przewidywaną przez dziecko stronę, co z kolei pozwala na ocenę poprawności rozróżniania

przez dziecko pseudosłów. Cały eksperyment obejmuje 24 sceny, w tym 18 scen z bodźcem dźwiękowym i 4 sceny ściągające uwagę niemowlęcia (animowany obiekt po środku ekranu, który znika gdy spojrzenie dziecka skupi się na nim przez 300 ms).

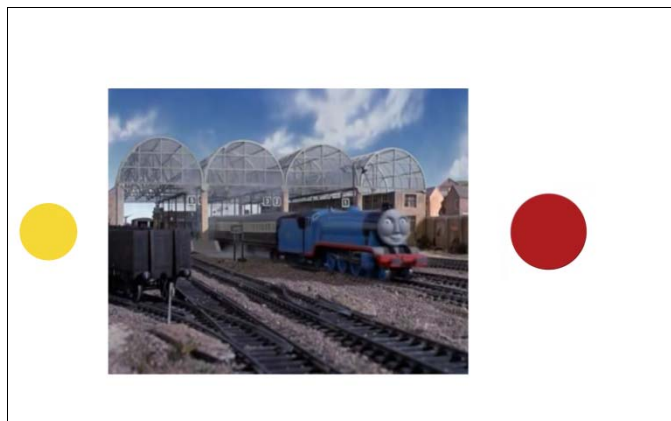


Rysunek 6. Implementacja metody B. Albareda-Castellot i in. (2011) w GIML.

Metoda druga wzorowana jest na opisanym wyżej eksperymencie J. Bjerva (2011). Jej koncepcja jest podobna do tej z metody przedstawionej powyżej. W tym przypadku na ekranie stale widoczne są dwa obszary, w których mogą pokazywać się rysunki (rysunek 7). W pierwszym etapie rysunek pojawia się jednak między obszarami – jego zadaniem jest ściągnięcie uwagi dziecka. Następnie niemowlę słyszy bodziec dźwiękowy, a po 1200 ms pojawia się rysunek. Analogicznie jak w pierwszej metodzie, mierzona jest pozycja spojrzenia przed pojawieniem się rysunku w jednym z obszarów. W przypadku tej metody prezentowane jest 36 scen, z których 18 to sceny zgodne z powyższym opisem, a pozostałe mają zadanie przyciągnąć uwagę dziecka. Ponadto co czwarta scena z bodźcem dźwiękowym pozbawiona jest bodźca wizualnego, aby uzyskać jak najlepsze potwierdzenie dla umiejętności różnicowania dźwięków mowy, poprzez efekt oczekiwania na obiekt wizualny w odpowiednim dla dźwięku obszarze.



Rysunek 7. Zmodyfikowany schemat metody J. Bjerva (2011).



Rysunek 8. Trening, w którym dziecko może „przewijać” bajki spoglądając na czerwoną kropkę.

Część treningowa badań również obejmowała dwa eksperymenty przygotowane za pomocą języka GIML i uruchamiane na platformie GCAF. W pierwszym wykorzystaliśmy schemat zaproponowany przez Q. Wang (2012). W oryginale eksperyment ten polegał na prezentowaniu serii zdjęć zwierząt, które powoli zmieniały się przenikając z jednego w drugi. Proces ten można było jednak znacznie przyspieszyć, jeżeli badany skupił wzrok na czerwonej kropce widocznej z prawej strony rysunku. Kropka pełniła zatem rolę przełącznika. Zmodyfikowaliśmy ten eksperyment projektując na jego bazie trening, w którym zdjęcia zwierząt zastąpiono fragmentami bajek (filmami) z francuskim dubbingiem. Dzięki kropce dziecko miało możliwość wyboru: czy ogląda aktualnie odtwarzany fragment bajki do końca, czy spoglądając na kropkę przełącza go na kolejny (rysunek 8).

Drugi trening to interaktywna bajka sterowana przez dziecko spojrzeniem (rysunek 9). Wykorzystano w niej rysunki i dźwięki (w języku francuskim) pochodzące z bajki *Świnka Peppa*. W każdej scenie widoczne są trzy elementy, które po spojrzeniu (stan aktywacji obszaru, por. rysunek 4 poniżej) zaczynają się obracać. W przypadku dłuższego spojrzenia (stan reakcji) cała scena poza wybranym obszarem jest zasłonięta półprzezroczystą przesłoną, a postać z obszaru zaczyna „opowiadać”. W tym treningu dziecko samo decyduje, spoglądając na elementy widoczne na scenie, która z postaci zaczyna się poruszać i „mówić”. W oparciu o literaturę (por. P.K. Kuhl 2011, P. Tomalski i in. 2013, A.F. de C. Hamilton 2016) oczekujemy, że możliwość interakcji dziecka z bajką spowoduje, że wyniki treningu będą lepsze, niż w przypadku treningu biernego. Projekt interaktywnej bajki jest obecnie głównym poligonem testowym platformy GCAF.





Rysunek 9. Interaktywna bajka. Z lewej scena, w której pozostały dwa z trzech aktywnych elementów, z prawej – scena podczas aktywacji dolnego elementu.

## Bibliografia

- Albareda-Castellot, B./ F. Pons/ N. Sebastián-Gallés (2011), *The acquisition of phonetic categories in bilingual infants: new data from an anticipatory eye movement paradigm: Acquisition of phonetic categories in bilingual infants*, (w:) *Developmental Science*, 14 (2), 395–401.
- Aslin, R.N. (2012), *Infant Eye: A Window on Cognitive Development*, (w:) *Infancy*, 17 (1), 126–140.
- Bénédicte de Boysson-Bardies B./ M.B. Debevoise (2001), *How Language Comes to Children: From Birth to Two Years* (Bradford Books). Cambridge/ London.
- Bjerva, J./ E. Marklund/ J. Engdahl/ F. Lacerda (2011), *Anticipatory Looking in Infants and Adults*, (w:) *Proceedings of EyeTrackBehavior*, 2011, 1–4.
- Conboy, B.T./ R. Brooks/ A.N. Meltzoff/ P.K. Kuhl (2015), *Social Interaction in Infants' Learning of Second-Language Phonetics: An Exploration of Brain–Behavior Relations*, (w:) *Developmental Neuropsychology*, 40 (4), 216–229.
- Gallaway, C./ B.J. Richard (red.) (1994), *Input and Interaction in Language Acquisition*. Cambridge.
- Hamilton A.F.de C./ V. Southgate/ E.L. Hill (2016), *The development of action cognition*, (w:) A.K. Engel/ K.J. Friston/ D. Kragic (red.), *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science*. Cambridge, 35–47.
- Hamilton A.F.de C. (2016), *Gazing at me: the importance of social meaning in understanding direct gaze cues*, (w:) *Philosophical Transactions of the Royal Society B*, 371, 20150077.
- Hennessey C./ A.T. Duchowski (2010), *An open source eye-gaze interface: expanding the adoption of eye-gaze in everyday applications*, (w:) C.H. Morimoto/ H. Istance (red.) *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. Austix, TX, 81–84
- Huckauf, A./ M. Urbina (2007), *Gazing with pEYE: new concepts in eye typing*, (w:) C. Wallrafen, V. Sundstedt (red.) *APGV 2007: Proceedings of the 4th symposium on Applied perception in graphics and visualization*. Tübingen, 141–141.
- Jusczyk, P.W. (2000), *The Discovery of Spoken Language (Language, speech, and communication)*. Cambridge/ London.

- Kuhl, P.K./ F.M. Tsao/ H.M. Liu (2003), *Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning*, (w:) PNAS, 100, 9096–9101.
- Kuhl, P.K. (2004), *Early language acquisition: cracking the speech code*, (w:) Nature Reviews Neuroscience, 5, 831–843.
- Kuhl, P.K. (2011), *Early Language Learning and Literacy: Neuroscience Implications for Education*. *Mind, Brain and Education*, (w:) The Official Journal of the International Mind, Brain, and Education Society, 5 (3), 128–142.
- Leclere, C./ S. Viaux/ M. Avril/ C. Achard/ M. Chetouani/ S. Missonnier et al. (2014), *Why Synchrony Matters during Mother-Child Interactions: A Systematic Review*, (w:) PLoS ONE, 9 (12), 1–12.
- McMurray, B./ R.N. Aslin (2005), *Infants are sensitive to within-category variation in speech perception*, (w:) Cognition, 95, B15–B26.
- McMurray, B./ R.N. Aslin (2004), *Anticipatory Eye Movements Reveal Infants. Auditory nad Visual Categories*, (w:) Infancy, 6 (2), 203–229.
- Nielsen, J. (1993), *Usability Engineering*. Cambridge.
- Ramirez-Esparza, N./ A. Garcia-Sierra/ P.K. Kuhl (2014), *Look who's talking: Speech style and social context in language input are linked to concurrent and future speech development*, (w:) Developmental Science, 17, 880–891.
- Shukla, M./ J. Wen/ K.S. White/ R.N. Aslin (2011), *SMART-T: A system for novel fully automated anticipatory eye-tracking paradigms*, (w:) Behav Res Methods, 43 (2), 384–398.
- Tomalski P./ D.G. Moore/ H. Ribeiro/ E. Axelsson/ E. Murphy/ A. Karmiloff-Smith/ M.H. Johnson/ E. Kushnerenko (2013), *Socio-economic status and functional brain development – associations in early infancy*, (w:) Developmental Science, 16 (5), 676–687.
- Wang, Q./ J. Bolhuis/ C.A. Rothkopf/ T. Kolling/ M. Knopf/ J. Triesch (2012), *Infants in Control: Rapid Anticipation of Action Outcomes in a Gaze-Contingent Paradigm*, (w:) PLoS ONE, 7 (2), 1–14.
- Wobbrock, J.O./ J. Rubinstein/ M.W. Sawyer/ A.T. Duchowski (2008), *Longitudinal evaluation of discrete consecutive gaze gestures for text entry*, (w:) J.J. Rääiä, A.T. Duchowski (red.), ETRA 2008: Proceedings of the 2006 Symposium on Eye tracking Research & Applications. New York, 11–18.
- Yu, C./ L.B. Smith (2010), *What you learn is what you see: Using eye movements to study infant cross-situational word learning*, (w:) Developmental Science, 2010, 1–16.