

# Iwona Mazur, Agnieszka Chmiel

---

## Odzwierciedlenie percepcji osób widzących w opisie dla osób niewidomych : badania okulograficzne nad audiodeskrypcją

---

Lingwistyka Stosowana / Applied Linguistics / Angewandte Linguistik nr 4,  
163-180

---

2011

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.

**Iwona MAZUR**

**Agnieszka CHMIEL**

Uniwersytet im. Adama Mickiewicza w Poznaniu

## **Odzwierciedlenie percepcji osób widzących w opisie dla osób niewidomych. Badania oku- lograficzne nad audiodeskrypcją<sup>1</sup>**

### **1. Wstęp**

Audiodeskrypcja (AD) umożliwia osobom niewidomym i słabowidzącym dostęp do mediów (audio)wizualnych. Na świecie (głównie w Stanach Zjednoczonych i Wielkiej Brytanii) ta forma przekładu audiowizualnego jest znana i stosowana od ponad dwudziestu lat. W Polsce osoby z dysfunkcją wzroku mogą z niej korzystać – na razie w bardzo ograniczonym zakresie – od roku 2006.

Wyzwania stojące przed audiodeskrypcją są liczne – musi ona dokonać selekcji opisywanych elementów, odpowiednio „ubrać je w słowa” i tak skondensować, aby opis mieścił się często w bardzo krótkich przerwach między dialogami (w przypadku filmów czy przedstawień teatralnych). W niniejszym artykule pragniemy skupić się na pierwszym z tych wyzwań, czyli na tym, które spośród licznych elementów obrazu filmowego należy opisać. Uważamy, że przy podejmowaniu takich decyzji pomocne mogą okazać się wyniki badań przy użyciu okulografu, dzięki któremu możliwe jest rejestrowanie ruchu gałek ocznych osób widzących.

W niniejszym artykule przedstawiamy wyniki badania przeprowadzonego w ramach pierwszego etapu projektu „Maria Antonina”, w którym wykorzystaliśmy m.in. metodologię okulograficzną. Swój wywód rozpoczynamy od przedstawienia głównych założeń audiodeskrypcji, następnie wprowadzamy podstawowe pojęcia w zakresie metodologii okulograficznej, po czym dokonujemy przeglądu badań okulograficznych istotnych z punktu widzenia naszych rozważań. Zasadniczą część artykułu stanowi opis badania oraz omówienie jego wyników. Ostatnią część artykułu stanowią wnioski.

---

<sup>1</sup> Praca naukowa finansowana ze środków na naukę w latach 2011-2012 jako projekt badawczy (N N104 054739).

## 2. Audiodeskrypcja jako intersemiotyczny przekład audiowizualny

Audiodeskrypcja polega na udostępnieniu treści (audio)wizualnych osobom niewidomym i słabowidzącym przez dołączenie dodatkowej ścieżki z opisem. Treści wizualne, które mogą być udostępniane osobom z dysfunkcją wzroku to nie tylko filmy (w kinie, na DVD, czy w telewizji), ale także spektakle teatralne i operowe, wydarzenia sportowe, czy ekspozycje w muzeach i galeriach (zob. np. B. Bencke 2004, P. Otero 2005, J. Snyder 2007).

Audiodeskrypcja opisuje istotne elementy wizualne: najbardziej charakterystyczne elementy obrazu czy rzeźby, a w przypadku dzieł filmowych czy teatralnych: kostiumy, scenografię, gesty, mimikę. Musi zrobić to w taki sposób, aby odbiorcy byli w stanie wyobrazić sobie opisywane elementy (sposób realizacji dzieła przez reżysera czy artystę). W przypadku filmów i spektakli niezwykle ważne jest zachowanie ciągłości narracyjnej tak, aby osoba niewidoma czy słabowidząca była w stanie podążać za akcją [por. J.-L. Kruger (2010), który twierdzi, że utrzymanie ciągłości narracyjnej jest rzeczą najwyższej wagi w AD i proponuje nawet zastąpienie określenia 'audiodeskrypcja' terminem 'audio-narracja' (AN)]. W związku z tym, iż audiodeskrypcja dokonuje swoistego „przekładu” obrazów na słowa, według klasyfikacji rodzajów tłumaczenia Jakobsona (1959) audiodeskrypcję traktuje się jako tłumaczenie intersemiotyczne, czyli przekład jednego kodu semiotycznego (wizualnego) na inny (werbalny) (por. J. Diaz Cintas 2007).

Sporym utrudnieniem w pracy audiodeskrypcyjisty są ograniczenia czasowe: opis musi się zmieścić w przerwach między dialogami i innymi istotnymi dźwiękami, co oznacza, że musi być on zwięzły i barwny, zawierający maksimum treści, a minimum słów. Bardzo często wybór tego, co jest istotne z punktu widzenia rozwoju akcji bądź doceniania warstwy wizualnej dzieła przez osoby niewidome jest niezwykle trudny – audiodeskrypcja ma wprawdzie do dyspozycji wytyczne [np. *ITC Guidance on Standards for Audio Description* (2000) dla języka angielskiego czy standardy opracowane przez Barbarę Szymańską i Tomasza Strzywińskiego (2010) dla języka polskiego], niemniej jednak każde opisywane dzieło jest inne, wymaga indywidualnego podejścia i w takim przypadku audiodeskrypcja polega głównie na własnej intuicji i subiektywnych odczuciach.

Uważamy, że w wyborze elementów istotnych dla zrozumienia (docenienia) dzieła zawierającego treści wizualne pomocne mogą okazać się badania okulograficzne, które umożliwiają lepsze poznanie percepcji wzrokowej osób bez dysfunkcji wzroku, a co za tym idzie, próbę jej odzwierciedlenia w AD. Zanim omówimy wyniki naszego badania, poniżej przedstawiamy krótki przegląd badań okulograficznych, ze szczególnym uwzględnieniem badań nad audiodeskrypcją.

## 3. Dotychczasowe badania okulograficzne

Okulograf (z ang. *eye-tracker*) śledzi ruch gałek ocznych, zapisując jednocześnie sakkady i fiksacje. Te pierwsze to podstawowe ruchy gałek ocznych. Między

sakkadami oko pozostaje względnie nieruchome przez ok. 200-300 ms i taki właśnie bezruch nazywamy fiksacją (K. Rayner 1998). Fiksacje i sakkady (jako regresje, czyli powrót do miejsca, w którym wcześniej nastąpiła fiksacja) uznawane są za miernik wysiłku kognitywnego wydatkowanego przy przetwarzaniu bodźców wizualnych (K. Rayner 1998). Można zatem uznać, że im liczniejsze i dłuższe fiksacje, tym bardziej intensywne przetwarzanie materiału wzrokowego.

Badania dotyczące postrzegania obrazów prowadzone są często w oderwaniu od audiodeskrypcyjnego nurtu przekładoznawczego. C.M. Masciocchi et al. (2009) zbadali postrzeganie naturalnych scen złożonych i odkryli, że badani dość konsekwentnie wybierają podobne punkty zainteresowania (*points of interest*) w danych scenach. Również D. Parkhurst et al. (2002) zauważyli umiarkowaną korelację między fiksacjami a tzw. istotnością obrazów (*saliency*).

W dziedzinie językoznawstwa większość dotychczasowych badań okulograficznych dotyczyła procesu czytania (zob. np. K. Rayner 1998). W ostatnim czasie metodologia okulograficzna zyskuje coraz większą popularność w badaniach nad procesem tłumaczenia (np. F. Alves et al. 2010, N. Pavlović i K. Jensen 2009, S. Sharmin et al. 2008, G.M. Shreve et al. 2010, A. L. Jakobsen i K. Jensen 2009). Z uwagi na fakt, iż badania te nie są istotne dla rozważań zawartych w niniejszym artykule, nie omawiamy ich tutaj bardziej szczegółowo [ich przeglądu dokonujemy w innym naszym artykule – A. Chmiel i I. Mazur (w przygotowaniu)].

Od niedawna metodologia okulograficzna zaczęła być stosowana w badaniach nad audiodeskrypcją. Badania te są na tyle nowe, że w momencie pisania tego artykułu większość opracowań na ich temat była „w przygotowaniu” lub „w druku”. Z uwagi na ich istotność dla poruszanych tutaj zagadnień, poniżej krótko je opisujemy.

J-L. Kruger (w druku) zastosował metodę okulograficzną do badania kognitywnego odbioru filmów przez osoby widzące. Wyniki jego badania wydają się sugerować, że przy tworzeniu audiodeskrypcji i audionarracji większe znaczenie ma istotność narracyjna (*narrative saliency*) niż istotność wizualna (*visual saliency*). Oznacza to, że uwzględnienie w opisie elementów peryferyjnych pod względem wizualnym, a istotnych z narratologicznego punktu widzenia, może okazać się ważniejsze dla spójnego opisu, aniżeli uwzględnienie elementów istotnych wizualnie, a mniej ważnych dla narracji filmu. Co za tym idzie, w przypadku gdy opis wszystkich elementów danej sceny jest niemożliwy, audiodeskryptor powinien w pierwszej kolejności opisać elementy istotne pod względem narracyjnym.

Z kolei E. di Giovanni (2011) przygotowała dwie audiodeskrypcje do wybranych fragmentów filmu – pierwszy opis był standardową audiodeskrypcją, natomiast drugi odzwierciedlał wyniki uzyskane przez badaczkę w badaniu okulograficznym przeprowadzonym na grupie osób widzących. Innymi słowy, zawierał on tzw. „priorytety wizualne”, czyli elementy, na które badani zwracali najczęściej uwagę. Opis ten odzwierciedlał także kolejność, w jakiej badani patrzyli na te elementy. Następnie jedna z dwóch tak przygotowanych wersji AD została zaprezentowana dwóm grupom osób niewidomych w celu ustalenia, czy AD odzwierciedlająca „priorytety wizualne” osób widzących będzie bardziej zrozumiała dla osób niewidomych. Okazało się, że osoby, którym zapre-

zentowano klip ze standardową audiodeskrypcją, były nieco zdezorientowane sekwencją prezentowanych zdarzeń. Takiego problemu nie zgłosiły osoby, które obejrzały klip z AD opracowaną przy wykorzystaniu danych okulograficznych. Powyższe wyniki sugerują, że dane okulograficzne mogą być skutecznie wykorzystywane w tworzeniu audiodeskrypcji. W swoim badaniu A. Vilaró et al. (w druku) skoncentrowali się na wpływie ścieżki dźwiękowej w tekście audiowizualnym na percepcję i rozumienie takiego tekstu. Przygotowali oni cztery różne ścieżki dźwiękowe do tej samej sceny, a następnie zarejestrowali przy pomocy okulografu ruchy gałek ocznych uczestników, którym pokazano klip z różnymi ścieżkami dźwiękowymi. Dane okulograficzne uzupełniono dodatkowo danymi z kwestionariusza. Okazało się, że przy różnych ścieżkach dźwiękowych występowały różnice w percepcji wzrokowej – badani kierowali wzrok na inne obszary danego obrazu. Co więcej, pomimo iż badanym pokazywano dokładnie ten sam klip w czterech wersjach, niektórzy z nich – jak się okazało w kwestionariuszu – mieli co do tego wątpliwości. Oznacza to, że ścieżka dźwiękowa ma wpływ zarówno na percepcję, jak i rozumienie tekstów audiowizualnych, a co za tym idzie, powinna być brana pod uwagę przy opracowywaniu audiodeskrypcji. Autorzy wnioskuje, że audiodeskrypcjonarzy oraz badaczy AD mogą traktować dane okulograficzne pochodzące od osób widzących jako wyznacznik tego, co należy w danej scenie opisać (przy założeniu, że różnice występujące pomiędzy różnymi grupami widzów nie będą istotne). Proponują oni także dalsze badania wykorzystujące fragmenty filmów z AD i bez niej – jeśli okaże się, że wyniki uzyskane dla wersji z AD znacząco różnią się od tych uzyskanych dla klipów bez AD, to może to oznaczać, iż opis koncentruje się na niewłaściwych elementach obrazu i odciąga uwagę widzów od tego, co w obrazie istotne.

I. Krejtz et al. (w przygotowaniu) jako pierwsi zbadali przy wykorzystaniu okulografu przydatność AD dla osób widzących. Czterdziestu czterech 8-9-latków podzielono na dwie grupy: jednej grupie pokazano fragment animowanego filmu edukacyjnego z AD, natomiast drugiej ten sam fragment bez AD. Okazało się, że badani w grupie pierwszej mieli m.in. więcej fiksacji na opisywanych elementach. Oznacza to, że AD jest czynnikiem sterującym uwagą dzieci i w związku z tym mogłaby być wykorzystana jako dodatkowa pomoc naukowa przy wprowadzaniu nowych pojęć. Twierdzą oni, że audiodeskrypcja może być skutecznym narzędziem sterującym uwagą dzieci widzących, dzięki któremu można skupić uwagę dzieci na wybranych elementach obrazu, a przez to ułatwić im przyswajanie nowych informacji (np. naukę specjalistycznego słownictwa).

#### **4. Projekt Maria Antonina**

**1. Cel.** Zgodnie z jedną z podstawowych zasad audiodeskrypcji opisywane wydarzenia powinny być istotne dla rozwoju akcji lub zrozumienia treści filmu lub spektaklu. Niemniej jednak audiodeskrypcjonarz kieruje się często w tym zakresie swoim subiektywnym odczuciem. W związku z tym proponowane opisy są często

przedmiotem krytyki ze strony osób niewidomych i niedowidzących, a podstawowe zarzuty to zbyt szczegółowy opis wypełniający wszystkie przerwy między dialogami, co bardzo szybko powoduje zmęczenie i 'wyłączenie się' widza lub, wręcz przeciwnie, zbyt skąpy opis, który uniemożliwia pełne zrozumienie filmu.

Dzięki możliwościom oferowanym przez metodologię okulograficzną możliwym stało się zbadanie, na jakie elementy wizualne zwracają uwagę osoby widzące podczas oglądania filmu. Przedmiotem naszego zainteresowania były m.in. sceny bardzo bogate pod względem wizualnym, czyli takie, w których jest wiele elementów, które mogą potencjalnie przyciągnąć uwagę widza. Film „Maria Antonina” (2006) w reżyserii Sofii Coppoli wydawał się być bardzo dobrym materiałem do przeprowadzenia tego rodzaju badań – w filmie tym mamy pełną przepychu scenografię, a kostiumy są niejako kolejną postacią w filmie.

Celem projektu jest zatem zbadanie percepcji osób widzących, to jest stwierdzenie, na jakie elementy wizualne zwracają oni uwagę oglądając film, a następnie uwzględnienie tych elementów w opisie dla osób niewidomych. Aby osiągnąć powyższy cel, podzieliliśmy projekt na dwa etapy. Celem etapu pierwszego jest zbadanie percepcji wybranych fragmentów filmu, natomiast celem etapu drugiego jest poznanie preferencji niewidomych względem audiodeskrypcji opracowanej w „tradycyjny” sposób oraz audiodeskrypcji odzwierciedlającej percepcję osób widzących. W niniejszym artykule omawiamy wyniki badań przeprowadzonych w ramach etapu pierwszego.

Oczywiście, jesteśmy świadome ograniczeń niniejszego projektu. Nie zakładamy, że badania te pozwolą nam dogłębnie poznać zasady rządzące ludzką percepcją ruchomych treści wizualnych, która jest procesem niezwykle złożonym. Nie liczymy też na to, że dzięki wynikom naszych badań uda nam się opracować wytyczne do tworzenia idealnej audiodeskrypcji-protezy, wiernie naśladującej funkcje pełnione przez zdrowy narząd wzroku. Niemniej jednak mamy nadzieję, że badania te przyczynią się do lepszego poznania ludzkiej percepcji wzrokowej, a co za tym idzie, w perspektywie długofalowej, do próby opracowania zasad tworzenia AD, przynajmniej w jakimś stopniu odzwierciedlającej taką percepcję.

**(2) Metody badawcze (*study design*).** W projekcie Maria Antonina połączono kilka metod badawczych. Podstawową jest śledzenie ruchów gałek ocznych, czyli okulografia, wzbogacona o kwestionariusze i badanie recepcji audiodeskrypcji przez osoby niewidome i słabowidzące. W niniejszym artykule omawiamy wyniki pierwszej fazy projektu, czyli badania okulograficznego i kwestionariuszowego. Faza pierwsza projektu replikuje wspomniane wcześniej badania C.M. Masciocchiego et al. i D. Parkhurst et al. w ich ogólnym zamyśle, ale z zastosowaniem klipów wideo, a nie statycznych obrazów. Jeśli osiągnięte zostaną podobne wyniki i dowody na niewielką różnorodność wykazywaną przez badanych w wyborze punktów zainteresowania, tak spójne i obiektywne wyniki można będzie łatwo odzwierciedlić w audiodeskrypcji.

**(3) Uczestnicy.** W badaniu wzięło udział 18 osób widzących bez wady wzroku lub z jej korekcją. Byli to studenci filologii angielskiej na poziomie magisterskim, łącznie 15 kobiet i 3 mężczyzn. Ich średnia wieku wyniosła 23,5 roku.



**(4) Bodźce.** Do badania wybrano cztery klipy z filmu pt. „Maria Antonina”. W pierwszym klipie, na potrzeby badania zatytułowanym „Komnata”, Maria Antonina po raz pierwszy ogląda swoje komnaty. Witana przez służbę i damy dworu kolejno przechodzi przez bogato zdobione salony pełne barokowych złoceń, kryształowych żyrandoli i wyściełanych aksamitem mebli. Klip trwa minutę i 40 sekund, brak w nim dialogów. Kamera prowadzi widza poprzez kolejne pomieszczenia, czasami ukazując zbliżenie na główną bohaterkę lub detale wystroju.

Drugi klip trwa dokładnie minutę. Zatytułowany został „Noc poślubna”. W sypialni król Ludwik XV, dostojnicy królewscy, damy dworu i kardynał otaczają łożem małżeńskie Marii Antoniny i przyszłego króla Francji, Ludwika XVI. Nowożeńcy kładą się na łożu i przyjmują błogosławieństwo od kardynała i króla. Scena obfituje w statyczne ujęcia (szeroki plan obejmujący łożo i zgromadzonych ludzi) przeplatane zbliżeniami nowożeńców, kardynała i króla. Jedyne wypowiedziane kwestie to krótkie błogosławieństwa.

W trzecim klipie zatytułowanym „Śniadanie” nowożeńcy siedzą przy suto zastawionym stole. Na bogato zdobionych i złotych naczyniach różne potrawy. Na środku stołu imponujący piętrowy półmisek z warzywami i piramidą zielonych szparagów. Wokół nowożeńców ponownie mnóstwo służby i dworzan. W tle orkiestra. Ludwik XVI spożywa przystawkę, a Maria Antonina popija przejrzysty płyn z kryształowego kieliszka. Podobnie jak w poprzednim klipie, ta scena również zawiera wiele ujęć statycznych, pokazujących Ludwika i Antoninę siedzących obok siebie, a między nimi wspomniany półmisek z warzywami. Antonina próbuje nawiązać rozmowę z mężem, lecz uzyskuje prawie monosylabicznie i niezbyt zachęcające odpowiedzi na swoje pytania. Klip trwa minutę i 50 sekund.

Czwarty klip o długości 45 sekund na potrzeby badania zatytułowano „Buty i ciastka”. Jest to fragment filmu nagrany w stylistyce teledysku. Dynamiczna muzyka (rockowy utwór „I Want Candy”) towarzyszy równie dynamicznemu montażowi ukazującemu feerię barw, wyszukane wzory dworskich pantofelków (zaprojektowane przez Manolo Blahnika, odważnie zestawiające rokokowy przesyt ze współczesnym stylem *glamour*), zapierające dech w piersiach aksamity i satyny, błyszczące kolie, przepięknie zaaranżowane półmiski z eleganckimi pralinkami. Widz może również zobaczyć, jak Maria Antonina opycha się łakociami, opija szampanem i podziwia tkaniny. W ujęciu, na którym widzimy stopę bohaterki w nowych trzewikach, przez około dwie sekundy na drugim planie widoczne są inne pary butów (pasujące stylistycznie do epoki) oraz ... para współczesnych niebieskich tenisówek. Sama reżyserka nigdy nie przyznała otwarcie, czy trampki w filmie to żartobliwy zabieg, czy przeoczenie rekwizytora. Ujęcie jest o tyle ciekawe w niniejszym badaniu, że mózg ludzki zupełnie nie spodziewa się takiego obuwia w filmie i widzowie często po prostu nie zauważali tenisówek w tym fragmencie (P. Orero – kontakt osobisty). Pozostaje zatem pytanie, czy taki element scenografii powinien zostać opisany w audiodeskrypcji.

Sceny wybrano tak, aby nie dominowały w nich dialogi. Wykorzystano wersję oryginalną filmu, aby nie wprowadzać dodatkowego czynnika w formie tłumaczenia – tłumaczenie w formie napisów zdecydowanie zniekształciłoby wyniki bada-

nia okulograficznego, ponieważ badani kierowaliby swój wzrok nie tylko na obraz, ale także na napisy. Nie zdecydowaliśmy się także na wykorzystanie tłumaczenia w formie lektorskiej – dodatkowa ścieżka mogłaby również wpłynąć na przetwarzanie bodźców wizualnych (por. A. Vilaró et al. w druku). Ponieważ uczestnikami badania byli studenci anglistyki, nie mieli oni problemu ze zrozumieniem niedługich kwestii z rzadka wypowiedzianych przez bohaterów.

**(5) Aparatura.** Do badań wykorzystano okulograf Tobii T60. Zastosowana rozdzielczość ekranu to 1280x1024, a promień fiksacji to 35 ppi. Są to ustawienia domyślne w oprogramowaniu Tobii Studio, które z kolei wykorzystano do przetwarzania, częściowej analizy i wizualizacji danych.

**(6) Procedura.** Uczestnicy badania siadali przed monitorem okulografu w zalecanej odległości ok. 60 cm. Kalibracji okulografu dokonywano poprzez poproszenie uczestników o śledzenie wzrokiem kolorowej kuli poruszającej się po białym ekranie. W przypadku kilku uczestników kalibrację należało powtórzyć. Następnie uczestnicy otrzymywali krótką instrukcję – mieli obejrzeć fragmenty filmu i odpowiedzieć na kilka pytań po ich obejrzeniu. Uczestnicy oglądali zatem cztery klipy, ruch ich gałek ocznych rejestrowany był przez okulograf, po każdym z klipów badani odpowiadali na pytania zadawane na podstawie kwestionariusza. Cała sesja eksperymentalna trwała ok. 20 min.

**(7) Rodzaj danych.** W trakcie eksperymentu zbierano dane dwojakiego rodzaju – dane okulograficzne i dane kwestionariuszowe. Te ostatnie zostaną omówione w części dotyczącej wyników, natomiast te pierwsze krótko definiujemy poniżej.

W punkcie 3. wyjaśniliśmy podstawowe pojęcia w okulografii, tj. fiksacje i sakiady. Konieczne jest wyjaśnienie jeszcze jednego pojęcia używanego przy analizie danych w programie Tobii Studio. AOI, czyli z angielskiego *Area of Interest* (obszar zainteresowania), to pole ekranu o dowolnym kształcie i wielkości, dla którego analizujemy dane o fiksacjach. Może być to konkretny element obrazu, np. twarz, napis lub powierzchnia całego ekranu. W analizowanych klipach wytyczono AOI wokół konkretnych elementów sceny (dokładny opis w punkcie dotyczącym wyników).

W opisywanym badaniu wykorzystano takie zmienne zależne, jak: długość obserwacji, liczba fiksacji i odsetek uczestników. Długość obserwacji to całkowity czas liczony w sekundach, kiedy badany spojrzął na AOI, zaczynając od fiksacji na AOI, a kończąc na fiksacji poza AOI. Liczba fiksacji to całkowita liczba fiksacji na danym AOI, a odsetek uczestników to procent uczestników, którzy mieli przynajmniej jedną fiksację na AOI. Wszystkie te dane standardowo obliczane są w programie Tobii Studio.

**(8) Wyniki.** Pierwszy klip „Komnata” przedstawia bogato urządzone wnętrza. Problemem dla audiodeskryptora w takim przypadku jest wybór elementów do opisu, ponieważ w krótkim czasie słowami trudno oddać całość obrazu. Audiodeskryptor musi zatem podjąć decyzję, czy i w jakim stopniu opisywać scenerię ogólnie, jakie szczegóły uwzględnić w opisie i w jakiej kolejności, jeśli wiele elementów pojawia się na ekranie jednocześnie. Jedno ujęcie w tym fragmencie filmu zdominował olbrzymi kryształowy żyrandol, który jest zaledwie jednym z elementów pełnego przepychu wnętrza.



W części kwestionariuszowej skupiliśmy się zatem na tym aspekcie opisu i zadałyśmy pytanie o elementy wystroju wnętrza, które szczególnie przyciągnęły uwagę uczestnika badania. Jedynie 30% uczestników wskazało na kryształowy żyrandol, większość odpowiedzi była ogólna. Respondenci wskazywali takie cechy, jak: dużo złota, kwiatowe wzory, bogate wnętrza, przepych, ornamenty. Takie wyniki niosą za sobą pewne implikacje dla audiodeskrypcji. Ponieważ uczestnicy opisywali scenę raczej holistycznie, bez nadmiaru szczegółów, opis audiodeskrypcyjny (aby odzwierciedlić percepcję osób widzących) powinien również syntetycznie przekazać wydźwięk sceny i ewentualnie zilustrować go kilkoma szczegółami (żyrandol, baldachim, złożony ptak nad łóżkiem).

W brytyjskiej audiodeskrypcji do tego fragmentu „Marii Antoniny” zastosowano podobne rozwiązanie. Najpierw w opisie pojawia się ogólne stwierdzenie (In a magnificent circular chamber lit by a huge crystal chandelier... [We wspaniałej okrągłej komnacie oświetlonej wielkim kryształowym żyrandolem...]). Następnie pojawia się wiele szczegółów na temat wystroju i służby (np. wysokie okna, niebieskie zasłony, krzesła obite błękitnym brokatem, lokaje przenoszący kwietniki i niebieskie pudła, kłaniające się damy dworu), co wskazuje na zbytnią dokładność angielskiego opisu.

Sugerowane powyżej podejście jest również zgodne z polskimi standardami audiodeskrypcji, wg których: „[t]worzenie audiodeskrypcji wymaga wnikliwej obserwacji opisywanego obrazu oraz uporządkowania cech według pewnej logicznej kolejności. Aby opis zachował przejrzystość, zwracamy uwagę tylko na te cechy, które są wartościowe i niezbędne do odzwierciedlenia celu i charakteru konkretnej produkcji.” (B. Szymańska i T. Strzymiński 2010: 23-24).

W tym przypadku opis wybranych szczegółów wnętrza przyczynia się do oddania owego charakteru produkcji, która zachwyca widza rokokowym przepychem scenarii i kostiumów.

Kolejny analizowany klip zatytułowany jest „Noc poślubna”. Przedstawia on scenę zbiorową i jest to scena raczej statyczna – praktycznie nie ma tu ruchu kamery, co więcej bohaterowie również nie przemieszczają się. W ujęciu tym widzimy parę nowożeńców w łóżu, obok którego stoją król i kardynał udzielający małżonkom błogosławieństwa. Na drugim planie liczni członkowie dworskiej świty. Dylemat, przed jakim staje audiodeskrytor w takim przypadku, to kogo i jak szczegółowo w takiej sytuacji opisać: czy tylko postaci pierwszoplanowe, czy może także te z planu drugiego.

W angielskiej audiodeskrypcji opis sceny jest bardzo szczegółowy, choć sama scena wprowadzona jest dość ogólnie. (Courtiers line the wall in the royal bed-chamber facing the bed. (...) On the other side of the bed a lady-in-waiting eases off Marie Antoinette's robe. [Dworzanie ustawieni wzdłuż ściany w królewskiej sypialni twarzą do łóżka. (...) Po drugiej stronie łóżka pokojówka ściąga szatę z ramion Marii Antoniny]). W dalszej części opisu audiodeskrytor podaje pozycje ciała nowożeńców i koncentruje się na ubiorze obojga bohaterów. Wspomina też kardynała i błogosławieństwo króla.

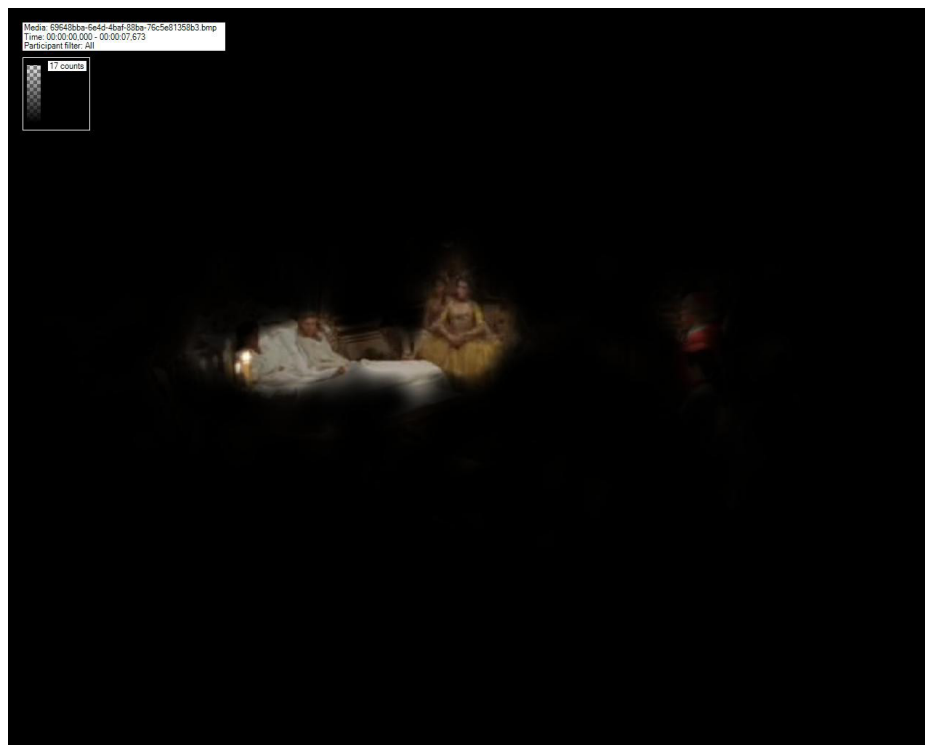
Uzyskane w badaniu okulograficznym wyniki pokazują, że uczestnicy najczęściej skupiali wzrok na parze nowożeńców oraz na dwóch damach dworu stojących

obok łóżka. W znacznie mniejszym stopniu badani kierowali wzrok na kardynała, a właściwie jedynie jego głowę. Poniższe wizualizacje z Tobii Studio - mapa cieplna (*heat map*) i odwrócona mapa cieplna (*gaze opacity*) w sposób obrazowy przedstawiają uzyskane wyniki. W przypadku mapy cieplnej, im ciemniejszy kolor, tym więcej fiksacji na danym obszarze zainteresowania, natomiast w przypadku odwróconej mapy cieplnej – im więcej fiksacji, tym mniej zaciemnienia na obrazie.



Ryc. 1 „Noc poślubna” – mapa cieplna.

Otrzymane dla niniejszego klipu wyniki sugerują, iż w przypadku sceny zbiorowej zwracamy uwagę na wybrane postaci – zarówno te, które są kluczowe dla danej sceny (małżonkowie, kardynał), jak i te, które odgrywają w niej jedynie role drugoplanowe (dwie damy dworu). Być może fiksacje na damach dworu były związane z ich umiejscowieniem obok łóżka, na które najczęściej padał wzrok badanych. Na podstawie powyższych wyników można wstępnie założyć, iż jeśli audiodeskrypcja miałaby odzwierciedlać percepcję osób widzących, to w przypadku scen zbiorowych należałoby zasygnalizować, że jest to scena zbiorowa, w sposób syntetyczny wymieniając osoby biorące w niej udział, a następnie bardziej szczegółowo opisać postaci istotne dla danej sceny.



Ryc. 2 „Noc poślubna” – odwrócona mapa cieplna.

Trzeci analizowany klip, „Śniadanie”, w pewnym sensie łączy w sobie cechy dwóch poprzednich scen – jest to z jednej strony scena zbiorowa, a z drugiej jest ona bardzo bogata pod względem wizualnym. Audiodeskryptor musi dokonać selekcji opisywanych elementów oraz podjąć decyzję, jak je opisać. W naszej analizie skupiliśmy się na dwóch elementach – na przezroczystym płynie, jaki pije Maria Antonina do śniadania, oraz na suto zastawionym stole, a w szczególności na centralnie umieszczonym piętrowym półmisku z warzywami i piramidą zielonych szparagów.

Wyniki dla pierwszego aspektu (przezroczysty płyn) miały nam pokazać, na ile uczestnicy będą interpretować to, co zobaczyli. Można założyć, że Maria Antonina najprawdopodobniej piła do śniadania wodę, niemniej jednak mógł to być też szampan. Zgodnie z obowiązującymi standardami AD opis powinien być możliwie obiektywny, a audiodeskryptor powinien unikać subiektywnej oceny czy interpretacji opisywanych zdarzeń. Jedna z trzech złotych zasad audiodeskrypcji brzmi: „Nie interpretuj, nie przedstawiaj swoich wniosków, opinii ani motywów opisywanych postaci” (B. Szymańska i T. Strzymiński 2010: 22).

Należy jednak podkreślić, iż kwestia interpretacji w AD jest w tej chwili przedmiotem zagorzałej dyskusji wśród praktyków i teoretyków tej formy przekładu audiowizualnego – część z nich twierdzi, że pewien stopień interpretacji w opisie jest nieunikniony, choćby ze względu na bardziej skrótowy charakter takiego opisu, a co za tym idzie jego lepsze zrozumienie przez docelowych odbiorców, czy też z uwa-

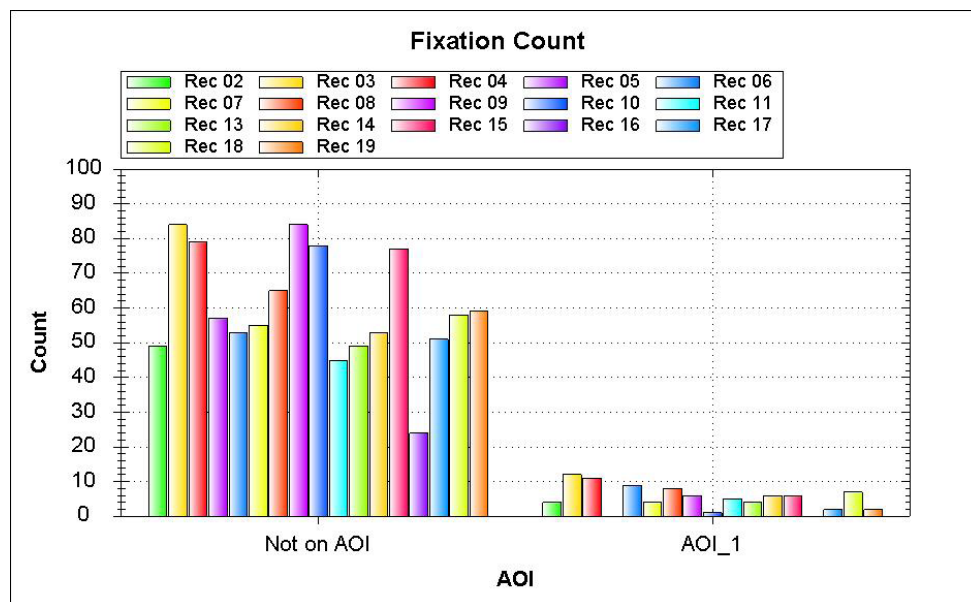
gi na fakt, iż audiodeskrytor dokonując opisu, patrzy na treści wizualne przez pryzmat swoich doświadczeń, wiedzy czy odczuć (podsumowanie debaty nt. interpretacji w audiodeskrypcji – zobacz I. Mazur i A. Chmiel w druku).

Jeżeli chodzi o pierwszy analizowany aspekt, to w kwestionariuszu po badaniu okulograficznym zapytałyśmy uczestników, co piła Maria Antonina do śniadania. Zdecydowana większość (16 na 18 badanych) odpowiedziała, że wodę, jedna osoba, że wino lub szampana, natomiast jedna osoba odpowiedziała, że nie pamięta. Oznacza to, że oglądając film, większość z nas czyni pewne założenia na temat tego, co widzimy. Oczywiście przykład z napojem pitym do śniadania wydaje się być trywialny z punktu widzenia audiodeskrypcji tej konkretnej sceny – prawdopodobnie dla większości odbiorców AD nie będzie miało znaczenia, czy napój będzie opisany jako woda, szampan czy przezroczysty płyn, jako że pozostanie to bez wpływu na rozwój akcji czy zrozumienie sceny. Niemniej jednak, uzyskane wyniki dla tego aspektu pokazują, że umysł ludzki czyni pewne założenia na temat oglądanych treści i poddaje je interpretacji. Autor angielskiej audiodeskrypcji poczynił właśnie takie założenie i umieścił „kieliszek wody” w opisie tej sceny.

Drugi analizowany w tej scenie aspekt to piętrowy półmisek z warzywami. Poniższa mapa cieplna oraz wykres ilustrują liczbę fiksacji na tym obszarze zainteresowania (AOI\_1).



Ryc. 3 „Śniadanie” – mapa cieplna.



Wykres 1. „Śniadanie” – liczba fiksacji na AOI\_1 i poza AOI.

Jak wynika z powyższych wizualizacji, większość fiksacji występowała poza AOI\_1 – badani skupiali wzrok głównie na twarzach księcia i Marii Antoniny. Mimo iż półmisek umieszczony jest centralnie i zajmuje znaczną część analizowanego ujęcia, jedynie dwie osoby miały ponad 10 fiksacji na tym obszarze zainteresowania. Analiza statystyczna przy użyciu testu t-Studenta wykazała, że różnica między średnimi liczbami fiksacji na obszarze zainteresowania i poza tym obszarem jest istotna –  $t(16)=14,59$ ,  $p<0,000$ . W kwestionariuszu po badaniu na pytanie, jakie warzywa znajdowały się na stole (np. jako część dekoracji), zaledwie 39% badanych odpowiedziało, że zielone szparagi, 17% badanych zauważyło na stole „coś zielonego”. Z kolei na pytanie, czy zauważyli jakieś inne jedzenie na stole, 28% uczestników, odpowiedziało, że tak i wymieniło takie artykuły spożywcze, jak: pomidory, truskawki, owoce, ser (nie do końca zgodne ze stanem faktycznym).

Powyższe wyniki potwierdzają wyniki uzyskane dla klipu pierwszego („Komnata”) – w scenach bogatych pod względem wizualnym widzowie wydają się „omiatać wzrokiem” daną scenę, nie zwracając uwagi na detale, które ją budują. Jeśli te wyniki miałyby mieć przełożenie na AD, wówczas wystarczyłoby wspomnieć o suto zastawionym stole (ewentualnie wymienić kilka znajdujących się na nim potraw). Naturalnie ta propozycja musi zostać zweryfikowana w badaniu odbioru wśród osób z dysfunkcją wzroku.

Co dziwne, w angielskiej audiodeskrypcji, która zazwyczaj jest bardzo szczegółowa, stół zostaje opisany jedynie jako „zastawiony jedzeniem”, a opis koncentruje się na poszczególnych osobach i ich czynnościach (podawanie dań przez służbę, oglądanie pary królewskiej przez przechadzających się gości).

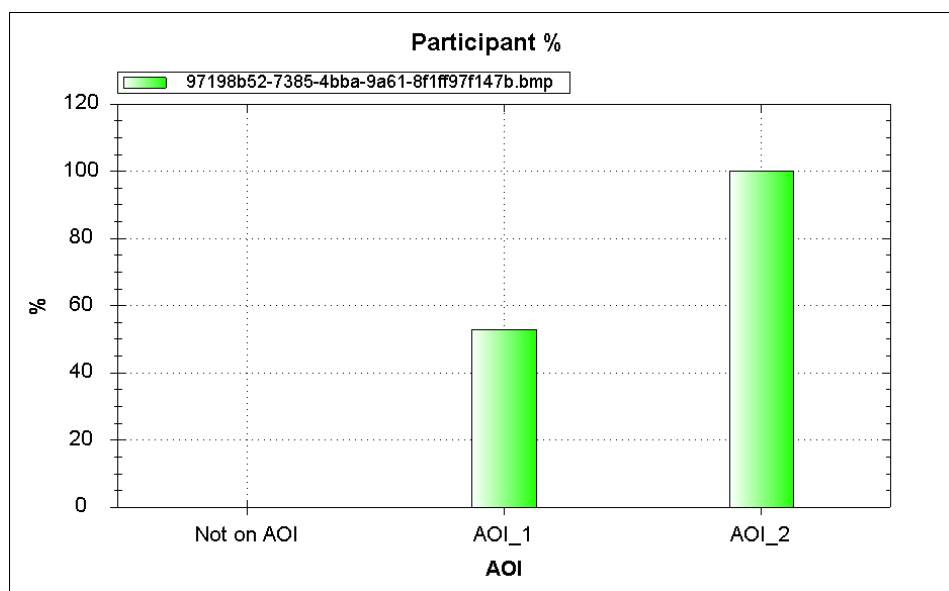


Analizując kolejny klip zatytułowany „Buty i ciastka”, byliśmy ciekawe, czy badani zauważą element zupełnie niepasujący do prezentowanej sceny i epoki, mianowicie współczesne niebieskie tenisówki. W kwestionariuszu przeprowadzonym po badaniu okulograficznym okazało się, że jedynie dwie osoby zauważyły trampki. Co ciekawe, analiza danych okulograficznych pokazała, że aż połowa uczestników miała fiksacje na tym obszarze zainteresowania (AOI\_1), choć było ich niewiele (jedna, dwie lub trzy w przypadku jednej osoby). Różnica w liczbie fiksacji na tym obszarze i poza nim jest statystycznie istotna –  $t(16)=7,33$ ,  $p<0,000$ . Czas obserwacji dla tego obszaru wahał się od 0,15ms do 1s (patrz wykres 4 poniżej). Wyniki te pokazują, iż sam fakt patrzenia na dany obiekt nie oznacza, że go zauważymy. Ponownie porównanie między średnią długością obserwacji na obszarze zainteresowania i poza nim okazuje się statystycznie istotne –  $t(16)=15,01$ ,  $p<0,000$ .

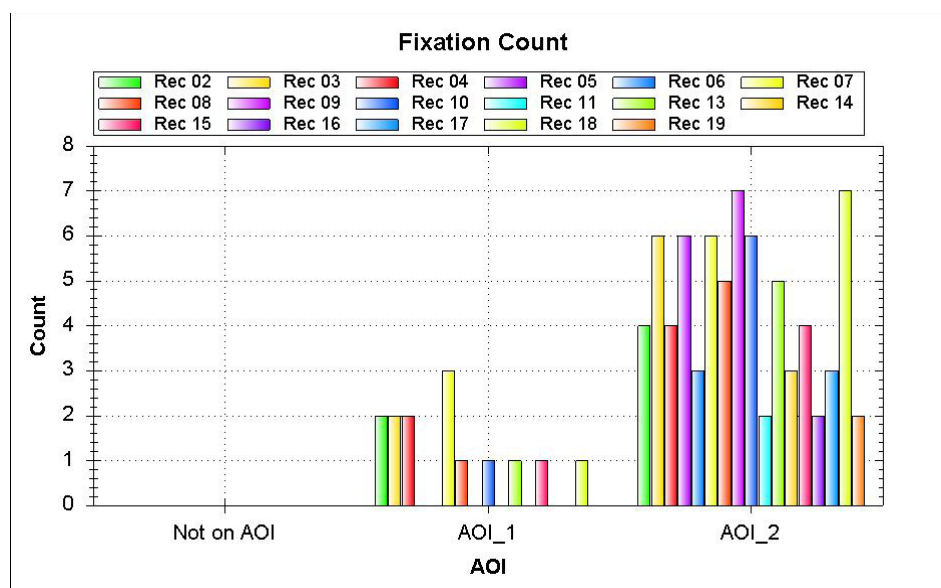


Ryc. 4 „Buty i ciastka” – mapa cieplna.

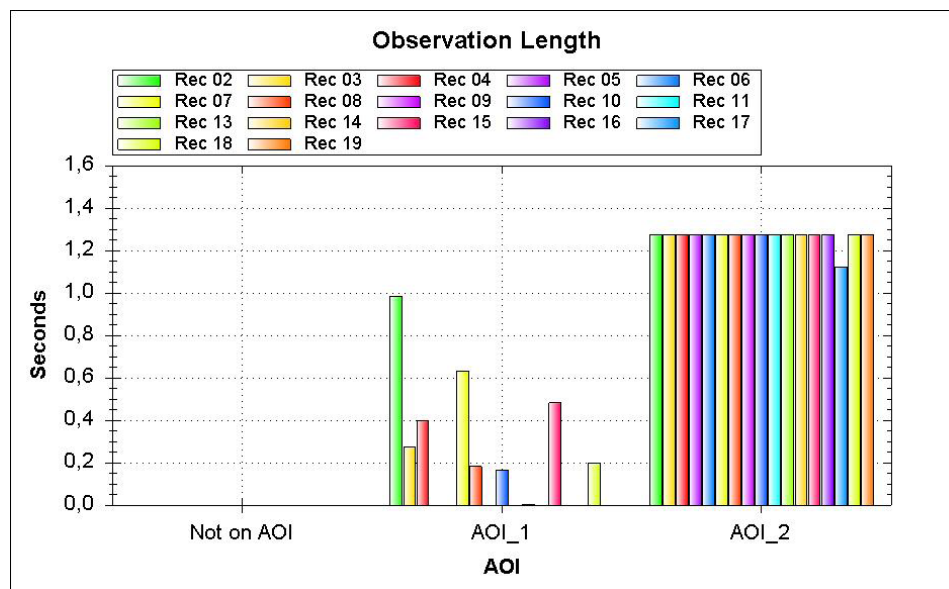




Wykres 2. Odsetek badanych z fiksacjami w dwóch obszarach zainteresowania (AOI\_1 – tenisówki, AOI\_2 – cały ekran) w klipie „Buty i ciastka”.



Wykres 3. Liczba fiksacji na dwóch obszarach zainteresowania (AOI\_1 – tenisówki, AOI\_2 – cały ekran) w klipie „Buty i ciastka”.



Wykres 4. Czas obserwacji dla dwóch obszarów zainteresowania (AOI\_1 – tenisówki, AOI\_2 – cały ekran) w klipie „Buty i ciastka”.

Mając na uwadze powyższe wyniki, należy sobie zadać pytanie, czy audiodeskrytor opisując tę scenę, powinien wspomnieć o tenisówkach. Po pierwsze, element ten wydaje się być nieistotny dla rozwoju akcji bądź zrozumienia sceny. Poza tym większość widzów nie zauważa tenisówek. Z drugiej strony, nie wiemy, czy nie był to celowy zabieg reżyserski, w końcu rockowa ścieżka dźwiękowa też nie pasuje do prezentowanej epoki. Należy podkreślić, że tenisówki były przedmiotem debaty na forach internetowych – jeśli ten element nie zostanie opisany, użytkownicy AD są z takiej debaty wyłączeni. Ponadto, jak zauważa J. Snyder (2007), audiodeskrytor powinien odznaczać się ponadprzeciętną zdolnością obserwacji, być uważnym obserwatorem (*an active seer*) i dostrzegać elementy, które mogą umknąć uwadze przeciętnego widza. W związku z powyższym trudno jest wyciągnąć z powyższych wyników jednoznaczne wnioski dla AD – audiodeskrytor, bazując na swoim zmyśle obserwacji, wiedzy filmoznawczej, a także biorąc pod uwagę ograniczenia narzucone przez audiodeskryptowany materiał (dialogi, ścieżka dźwiękowa itp.), każdorazowo musi podjąć decyzję, czy dany element opisać czy też nie.

W angielskiej audiodeskrypcji napotykamy na opis, który najpierw syntetycznie interpretuje całą sekwencję, a następnie koncentruje się na niektórych elementach:

Time for a mega shopping spree without ever having to leave home. Shoes, fans, bolts of fabric and dresses are presented to Marie Antoinette, Comtesse de Polignac and Princesse de Lamballe while they eat chocolate and sip champagne. Delicious plates of food come thick and fast along with glove samples and card games. All washed down with endless supplies of champagne. They gamble with pale pink chips and there is still more champagne.

Czas na szaleństwo zakupowe bez wychodzenia z domu. Buty, wachlarze, bele materiału i suknie prezentowane są Marii Antoninie, Księżnej de Polignac i Księżniczce de Lamballe, które jedzą czekoladę i popijają szampana. Półmiski smacznych przekąsek serwowane są jedno po drugim wraz z przykładami rękawiczek i gier karcianych. Wszystko pośród lejącego się w nieskończoność szampana. Bohaterki używają bladuróżowych żetonów do gry, a szampana wciąż przybywa.

Autor tego opisu zupełnie zignorował niepasujące do epoki trampki i całkowicie zrezygnował z dokładniejszego opisu butów.

## 5. Wnioski

Przygotowanie dobrej audiodeskrypcji do filmu jest niezwykle trudne – po pierwsze powinna zawierać ona wszystkie elementy, które są niezbędne do zrozumienia fabuły, po drugie powinna umożliwić odbiorcom wyobrazenie sobie warstwy wizualnej dzieła, i po trzecie powinna być na tyle zwięzła, by nie nadwyręzać koncentracji widza, pozwalając mu „na oddech” i na wsłuchanie się w ścieżkę dźwiękową filmu. Wybierając elementy do opisu, audiodeskryptor bazuje przede wszystkim na własnym zmyśle obserwacji, subiektywnych odczuciach i intuicji.

Metoda okulograficzna umożliwia odejście od takiego subiektywnego podejścia w stronę większego obiektywizmu w identyfikacji elementów obrazu, na które zwracają uwagę osoby widzące oglądające dany film. Oczywiście metodologia ta nie jest pozbawiona wad. Przede wszystkim należy pamiętać, że wiele rzeczy dostrzegamy tzw. kątem oka (*widzenie peryferyjne*) – skupiając się na jednym elemencie obrazu, jesteśmy w stanie niejako „omieść wzrokiem” całość ekranu, nawet jeśli nie dostrzegamy wszystkich szczegółów, jesteśmy w stanie zarejestrować ogólny charakter, klimat danej sceny. Poza tym – jak pokazują uzyskane przez nas dane – sam fakt patrzenia nie oznacza, że dany obiekt zauważamy. Co więcej, możemy patrzeć i zauważać wybrane elementy, a mimo to nie zrozumieć fabuły czy przesłania filmu.

Aby nieco wyeliminować powyższe ograniczenia, uważamy, iż w przypadku badań nad audiodeskrypcją metody okulograficzne powinny być stosowane w połączeniu z innymi metodami. Na przykład dane okulograficzne powinny być poddawane dodatkowej weryfikacji w formie wywiadu czy kwestionariusza. Ponadto, uzyskane w ten sposób wyniki powinny zostać poddane kolejnej weryfikacji – tym razem przez osoby niewidome i słabowidzące – np. w ramach badań odbioru (*reception studies*).

Uzyskane w naszym badaniu wyniki pokazują, że badani najczęściej skupiają wzrok na twarzach (tam fiksacje były najbardziej zbieżne). W scenach, w których było bardzo dużo detali, fiksacje były bardziej zróżnicowane. Co więcej, w przypadku takich scen niewielu uczestników zapamiętywało szczegóły – pamiętali oni raczej ogólny klimat czy charakter danej sceny. Ponadto badani mieli ten-

dencję do interpretowania tego, co widzieli na ekranie. Bazując na tych wynikach, można by zaproponować, iż AD odzwierciedlająca percepcję osób widzących powinna być mniej szczegółowa, oddająca bardziej charakter danej sceny niż opisująca szczegółowo wybrane detale. Mogłaby również zawierać interpretację tego, co zostało przedstawione na ekranie. Oczywiście propozycje tego typu należałoby skonsultować z docelowymi odbiorcami audiodeskrypcji, co planujemy uczynić w drugim etapie projektu.

Jednocześnie pragniemy podkreślić, iż zakres tego badania jest ograniczony. Nie zakładamy, że uzyskane przez nas wyniki pozwolą odpowiedzieć na pytanie, jak stworzyć doskonałą audiodeskrypcję. Niemniej jednak, mamy nadzieję, że projekt ten przyczyni się do nieco lepszego poznania ludzkiej percepcji wzrokowej, a co za tym idzie, w połączeniu z innymi podobnymi badaniami, w perspektywie długoterminowej, będzie stanowił wkład w udoskonalenie tej swoistej protezy dla osób z dysfunkcją wzroku tak, aby spełniała ona funkcje zbliżone do niezaburzonej percepcji wzrokowej.

#### BIBLIOGRAFIA

- ALVES F., PAGANO A. S., DA SILVA I. L. (2010), *Investing processing effort in the unfolding of the translation process: Methodological issues in empirical-experimental research using eye-tracking data*. Referat wygłoszony na 5<sup>th</sup> International Maastricht-Lódź Duo Colloquium on Translation and Meaning, Lódź.
- BENECKE B. (2004), *Audio Description*, (w:) *Meta* 49 (1), 78-80.
- CHMIEL A., MAZUR I. (w przygotowaniu), *Eyetracking sight translation performed by trainee interpreters*.
- DI GIOVANNI E. (2011), *Visual and narrative priorities of the blind and non-blind: Eye-tracking and audio description*. Referat wygłoszony na konferencji III Advanced Research Seminar on Audio Description ARSAD, Barcelona.
- DIAZ CINTAS J. (2007), *Accessibility and/in translation training*. Referat wygłoszony na 5. Kongresie EST „Why Translation Studies matters?”, Lublana.
- ITC Guidance on Standards for Audio Description* (2000), [http://www.ofcom.org.uk/static/archive/itc/uploads/ITC\\_Guidance\\_On\\_Standards\\_for\\_Audio\\_Description.doc](http://www.ofcom.org.uk/static/archive/itc/uploads/ITC_Guidance_On_Standards_for_Audio_Description.doc), 26.08.2011.
- JAKOBSEN A. L., JENSEN K. (2009), *Eye Movement Behaviour Across Four Different Types of Reading Task*, (w:) Göpferich S., Jakobsen A.L., Mees I.M. (red.), *Looking at eyes - Eye Tracking Studies of Reading and Translation Processing. Copenhagen Studies in Language 36*, Copenhagen, Samfundslitteratur, 103-124.
- JAKOBSON R. (1959/2000). *On linguistic aspects of translation*, (w:) *On Translation* (red.), Cambridge, Harvard University Press, 232-239.
- KREJTZ I., SZARKOWSKA A., KREJTZ K., WALCZAK A., DUCHOWSKI A. (w przygotowaniu), *Audio description as an aural guide of children's visual attention: Evidence from an eye-tracking study*.
- KRUGER J-L. (2010), *Audio narration: Re-narrativising film*, (w:) *Perspectives: Studies in Translatology*, 3(2010), 231-249.
- KRUGER J-L. (w druku), *Making meaning in AVT: Eye tracking and viewer construction of narrative*, (w:) *Perspectives: Studies in Translatology*.

- MASCIOCCHI C. M., MIHALAS S., PARKHURST D., NIEBUR E. (2006), *Everyone knows what is interesting: Salient locations which should be fixated*, (w:) Journal of Vision, 9 (11): 25, 1–22.
- MAZUR I., CHMIEL A. (w druku), *Audio description made to measure: Reflections on interpretation in AD based on the Pear Tree Project data*, (w:) Carroll M., Orero P., Remael A. (red.), *Media for All 3*, Amsterdam/New York, Rodopi, [brak numerów stron].
- ORERO P. (2005), *Audio Description: Professional Recognition, Practice and Standards in Spain*, (w:) Translation Watch Quarterly 1(2005), 7-18.
- PARKHURST D., LAW K., NIEBUR E. (2002), *Modelling the role of salience in the allocation of visual selective attention*, (w:) Vision Research, 42, 107–123.
- PAVLOVIĆ N., JENSEN K. (2009), *Eye tracking translation directionality*, [http://isg.urv.es/publicity/isg/publications/trp\\_2\\_2009/chapters/jensenpavlovic.pdf](http://isg.urv.es/publicity/isg/publications/trp_2_2009/chapters/jensenpavlovic.pdf)
- RAYNER K. (1998), *Eye movements in reading and information processing: 20 years of research*, (w:) Psychological Bulletin 124 (3), 372-422.
- SHARMIN S., ŠPAKOV O., RÄIHÄ K.-J., JAKOBSEN A. L. (2008), *Effects of time pressure and text complexity on translators' fixations*, (w:) ETRA '08 Proceedings of the 2008 Symposium on Eye Tracking Research & Applications, 123-126.
- SHREVE G.M., I. LACRUZ, E. ANGELONE (2010), *Cognitive effort, syntactic disruption, and visual interference in a sight translation task*, (w:) G.M. Shreve, E. Angelone (red.), *In Translation and Cognition* (red.), Amsterdam/Philadelphia, John Benjamins Publishing Company, 63–84.
- SNYDER J. (2007), *Audio Description: The Visual Made Verbal*, (w:) The International Journal of the Arts in Society, t. 2, [http://www.audiodescribe.com/about/articles/ad\\_international\\_journal\\_07.pdf](http://www.audiodescribe.com/about/articles/ad_international_journal_07.pdf), 26.08.2011.
- SZYMAŃSKA B., STRZYMIŃSKI T. (2010), *Standardy tworzenia audiodeskrypcji do produkcji audiowizualnych*, <http://www.audiodeskrypcja.org.pl/index.php/standardy-tworzenia-audiodeskrypcji/do-produkcji-audiowizualnych>.
- VILARÓ A., DUCHOWSKI A., ORERO P., GRINDINGER T.J., TETREAULT S., DI GIOVANNI E. (w druku), *How Sound is The Pear Tree? Testing the Effect of Varying Audio Stimuli on Visual Attention Distribution*, (w:) Perspectives: Studies in Translatology.

### **Reflecting perception of sighted audiences in description for the blind. Eye-tracking research in audio description**

Professional audio describers often face a dilemma what to include in their descriptions, how to order described elements and what to omit due to time constraints. It is thus useful to examine how particular film scenes are perceived by sighted viewers and to what extent audio description can influence that perception. This paper is a work-in-progress report on the first stage of an eye-tracking study. The participants watched clips from “Marie Antoinette” by Sofia Coppola and responded to questions afterwards. The data analysis focused on the number of fixations and the total observation time in selected Areas of Interest and gaze sequences in selected scenes. The results can contribute to further development of audio description standards by showing audio describers how the perception of sighted viewers should be reflected in the description in order to facilitate scene comprehension for the blind and visually impaired.