POLITECHNIKA GDAŃSKA
Wydział Elektroniki, Telekomunikacji i Informatyki

Bartosz Kunka

System śledzenia punktu fiksacji wzroku jako narzędzie wspierające badania korelacji wzrokowo-słuchowych

Rozprawa doktorska

Promotor:
prof. dr hab. inż. Bożena Kostek,
prof. zw. Politechniki Gdańskiej
Wydział Elektroniki, Telekomunikacji i Informatyki
Politechnika Gdańska

Gdańsk 2011

Podziękowania

Autor pragnie złożyć serdeczne podziękowania Pani Promotor, prof. dr hab. inż. Bożenie Kostek za wszelkie uwagi i wskazówki udzielane podczas przygotowywania niniejszej rozprawy, Kierownikowi Katedry Systemów Multimedialnych, prof. dr hab. inż. Andrzejowi Czyżewskiemu za stworzenie i zapewnienie możliwości rozwoju zainteresowań naukowych, Kolegom z Katedry, a w szczególności koledze Rafałowi Rybackiemu za prace związane z oprogramowaniem systemu Cyber-Oko oraz żonie Agnieszce za dobre słowo i wyrozumiałość.

Praca została częściowo dofinansowana w ramach projektu Synat: „Utworzenie uniwersalnej, otwartej, repozytoryjnej platformy hostingowej i komunikacyjnej dla sieciowych zasobów wiedzy dla nauki, edukacji i otwartego społeczeństwa wiedzy" (umowa nr SP/I/1/77065/10).

Słownik pojęć

– znaczenie wybranych terminów w kontekście niniejszej rozprawy

Termin	Opis/wyjaśnienie
biofeedback	biologiczne sprzężenie zwrotne, polegające na pomiarze wybranego parametru fizjologicznego; w przypadku *biofeedbacku* oddechowego mierzy się rytm i długość wydechów; na podstawie zmierzonych wartości parametrów oddychania generuje się odpowiednie bodźce wzrokowo-słuchowe, które stymulują zmysły człowieka; celem *biofeedbacku* jest osiągnięcie stanu relaksacji, w którym efektywność umysłu wzrasta
Blu-ray 3D	format zapisu danych obrazu stereoskopowego (3D) na nośniku Blu-ray, każda składowa obrazu stereoskopowego jest zapisana w rozdzielczości 1080p
Cyber-Oko (ang. *Cyber-Eye*)	system śledzenia punktu fiksacji wzroku na monitorze komputera, opracowany w Katedrze Systemów Multimedialnych PG
dynamiczna mapa przejść (ang. *dynamic gaze plot*)	wizualizacja aktywności wzrokowej widza naniesiona na treść obrazu wizyjnego
kaszeta (ang. *letterbox*)	czarny margines w górnej i dolnej części ekranu, nadający wyświetlanemu obrazowi panoramiczności (zwiększenia szerokości kadru w stosunku do jego wysokości)
kątowa rozdzielczość słuchu (ang. *auditory localisation resolution*)	najmniejsza wartość kąta, dla której ludzki narząd słuchu wykrywa przesunięcie źródła dźwięku w płaszczyźnie poziomej
kontener AVI (ang. AVI *container*)	kontener danych audiowizualnych; może przechowywać dane wizyjne skomprymowane różnymi kodekami audio i wideo; w strukturze kontenera AVI wyróżnia się również indeks, przechowujący informacje o zawartych danych audiowizualnych
logatom (ang. *logatom*)	sylaba lub sztuczny wyraz służący do badania rozumienia mowy w audiometrii słownej; logatomy tworzone są po to, aby badany nie wspomagał się znajomością języka w procesie rozumienia mowy
mapa ciepła (ang. *heat map*)	wizualizacja aktywności wzrokowej widza, wyznaczonej przez system śledzenia wzroku w określonym przedziale czasu; mapę ciepła najczęściej stosuje się w celu zwizualizowania aktywności wzrokowej osoby badanej na stronie internetowej
metadane (ang. *movieDescription* – znacznik)	pojęcie związane z indeksacją treści wizyjnej obrazu; ogólne informacje o próbce wizyjno-fonicznej; reprezentowane przez znacznik *movieDescription* w strukturze XML

obszar (ang. *area* – znacznik)	pojęcie związane z indeksacją treści wizyjnej obrazu; określa wymiary obszaru zainteresowania i przechowuje etykietę obszaru, wyrażany w pikselach; reprezentowany przez znacznik *area* w strukturze XML
obszar zainteresowania (ang. *Region of Interest* – ROI, *Area of Interest* – AOI)	obszar/fragment obrazu charakterystyczny ze względu na treść lub zawartość, poddawaną analizie; w niniejszej rozprawie z obszarem zainteresowania (w materiale badawczym) związane jest źródło dźwięku, np. skrzypce, twarz bohatera
offset	pojęcie związane z algorytmem wyznaczania punktu fiksacji wzroku w przestrzeni; różnica między deklarowanymi a obliczonymi współrzędnymi punktu fiksacji
offset dla 'p+10' (ang. *'p+10' offset*)	pojęcie związane z algorytmem wyznaczania punktu fiksacji wzroku w przestrzeni; średni *offset* dla współrzędnej *Z* w płaszczyźnie 'p+10'
offset dla 'p-10' (ang. *'p-10' offset*)	pojęcie związane z algorytmem wyznaczania punktu fiksacji wzroku w przestrzeni; średni *offset* dla współrzędnej *Z* w płaszczyźnie 'p-10'
oprogramowanie otwarte (ang. *open source software*)	oprogramowanie, którego licencja pozwala na legalne i bezpłatne kopiowanie i dowolne modyfikowanie kodu źródłowego
paralaksa stereoskopowa (ang. *stereoscopic parallax*)	ledwie zauważalne różnice między obrazami stanowiącymi lewą i prawą składową obrazu stereoskopowego (3D); w zależności od rodzaju paralaksy (ujemna, zerowa, dodatnia) widz percypuje położenie obiektu w różnych odległościach od ekranu (odpowiednio: przed, na i za płaszczyzną ekranu)
percepcja odległości (ang. *distance perception*)	percepcja odległości, w której pozorne źródło dźwięku znajduje się od słuchacza
percepcja kierunku (ang. *directional perception*)	percepcja kierunku, z którego pozorne źródło dźwięku emituje falę dźwiękową
percepcja wielomodalna (ang. *multimodal perception, cross-modal perception*)	w przypadku jednoczesnej stymulacji zmysłu wzroku i słuchu – percepcja (odbiór) wrażenia wizyjno-fonicznego, różniącego się od złożenia wrażeń będących następstwem niezależnej stymulacji zmysłu wzroku i słuchu
płaszczyzna 'p+10' (ang. *'p+10' plane*)	pojęcie związane z algorytmem wyznaczania punktu fiksacji wzroku w przestrzeni; płaszczyzna równoległa do płaszczyzny ekranu, znajdująca się 10 cm za monitorem
płaszczyzna 'p0' (ang. *'p0' plane*)	pojęcie związane z algorytmem wyznaczania punktu fiksacji wzroku w przestrzeni; płaszczyzna ekranu monitora
płaszczyzna 'p-10' (ang. *'p-10' plane*)	pojęcie związane z algorytmem wyznaczania punktu fiksacji wzroku w przestrzeni; płaszczyzna równoległa do płaszczyzny ekranu, znajdująca się 10 cm przed monitorem

podstawka kamerowa, dublet kamerowy (ang. *camera rig*)	specjalna podstawka mocowana do statywu kamerowego umożliwiająca umieszczenie dwóch kamer bądź równolegle względem siebie, bądź prostopadle w celu rejestracji obrazu stereoskopowego; podstawka kamerowa wykorzystany podczas realizacji obrazu 3D w ramach niniejszej pracy wymaga równoległego ustawienia kamer względem siebie
pozorne źródło dźwięku (ang. *virtual sound source*)	w ogólności: pojęcie związane z wrażeniem dźwięku emitowanego przez źródło znajdujące się w panoramie stereofonicznej; źródło dźwięku, które znajduje się w miejscu, gdzie w rzeczywistości nie ma umiejscowionego źródła (np. między głośnikami)
próbka wizyjno-foniczna, bodziec wzrokowo-słuchowy (ang. *visual-auditory trial, bimodal stimulus*)	materiał zawierający treść wizyjną i foniczną, charakteryzującą się wystąpieniem jednego lub kilku bodźców stymulujących jednocześnie zmysł wzroku i słuchu
przedział czasowy (ang. *interval* – znacznik)	pojęcie związane z indeksacją treści wizyjnej obrazu; określa przedział czasu, w którym wybrany obszar zainteresowania występuje, wyrażany w milisekundach
pseudofon (ang. *pseudophone*)	instrument emitujący dźwięk w taki sposób, że jest on słyszany z innego kierunku niż w rzeczywistości dźwięk ten dochodzi
punkt fiksacji wzroku (ang. *fixation point, gaze point, point of regard* – PoR)	punkt na płaszczyźnie ekranu monitora, w którym osoba badana lub użytkownik skupia swój wzrok
punkt fiksacji wzroku w przestrzeni (ang. *3D gaze point, 3D fixation point*)	punkt w przestrzeni, w którym osoba badana lub użytkownik fiksuje wzrok, przy czym obiekt, na którym fiksowany jest wzrok nie jest obiektem fizycznym, a jedynie projekcją obiektu w technice stereoskopowej
QoE (ang. *Quality of Experience*)	subiektywna miara wskazująca na ocenę jakości usługi przez klienta; np. w systemach wirtualnej rzeczywistości *QoE* może wskazywać na jakość doświadczanych doznań, przekładającą się na jakość odzwierciedlenia rzeczywistości w świecie wirtualnym
QoMEX (ang. *Quality of Multimedia Experience*)	subiektywna miara jakości usługi w przypadku prezentacji, strumieniowania treści multimedialnej
rozdzielczość czasowa (ang. *temporal resolution*)	parametr określający częstotliwość wyznaczania punktu fiksacji wzroku w systemach śledzenia wzroku (wyrażona w Hz)

rozdzielczość przestrzenna (ang. *spatial resolution*)	parametr określający dokładność wyznaczania punktu fiksacji wzroku przez system śledzenia wzroku (wyrażona w stopniach), np. rozdzielczość przestrzenna systemu Tobii T60 wynosi 0,5°, co oznacza, że system wyznacza punkt fiksacji wzroku z dokładnością do ok. 5 mm
stereobaza (ang. *stereo sound basis*)	odległość pomiędzy środkami obiektywów aparatów fotograficznych lub kamer przeznaczonych do rejestracji obrazu stereoskopowego
technika stereoskopowa (stereoskopia) (ang. *stereoscopic technique, stereoscopy*)	technika obrazowania, która pozwala człowiekowi za pomocą specjalnych technik obserwacji percypować obraz przestrzenny, powstały w wyniku złożenia dwóch dwuwymiarowych obrazów widzianych z perspektywy lewego i prawego oka
tło akustyczne (ang. *atmosphere sound*)	w ogólności – dźwięki otoczenia; jedna z warstw ścieżki dźwiękowej stosowana w celu stworzenia atmosfery akustycznej odpowiedniej dla miejsca, w którym toczy się akcja filmu bądź słuchowiska
uwaga wzrokowa (ang. *visual attention*)	skupienie wzroku widza na prezentowanej mu treści wizyjnej; uwaga wzrokowa wskazuje na fiksowanie wzroku w zdefiniowanych obszarach zainteresowania
warstwa oprogramowania (ang. *software layer*)	zbiór aplikacji komputerowych pracujących w ramach określonego urządzenia lub systemu, np. w przypadku systemu CyberOko warstwę oprogramowania tworzy aplikacja serwerowa, w której zaimplementowane są algorytmy przetwarzania obrazu oraz różne aplikacje klienckie, o różnej funkcjonalności
warstwa sprzętowa (ang. *hardware layer*)	zbiór fizycznych elementów lub urządzeń tworzących system, np. w systemach śledzenia wzroku warstwę sprzętową stanowią: monitor, kamera, diody podczerwieni, jednostka obliczeniowa, sterownik zasilania diod, przewody
wpływ ściągający obrazu na percepcję dźwięku (ang. *image proximity effect, ventriloquism effect*)	zmiana percepcji kierunku źródła dźwięku, którego położenie nie pokrywa się ze związanym z nim bodźcem wzrokowym; wpływ ściągający występuje przy prezentacji bodźca wizyjno-fonicznego

Lista skrótów i oznaczeń

(w porządku alfabetycznym)

α	poziom istotności statystycznej, w niniejszej pracy przyjęto, że $\alpha=0,05$
BT	test podstawowy
c	zmienna określająca położenie bodźca wzrokowego na ekranie, związana ze współrzędną odciętą (x) środka ciężkości bodźca wzrokowego; wyrażona w [°]
CO	Cyber-Oko
d	wartość różnicy współrzędnych *x* punktów fiksacji oka lewego i prawego; wartość związana z paralaksą stereoskopową
df Błąd	parametr wyznaczany w ramach testu ANOVA; liczba stopni swobody wewnątrz grup
df Efekt	parametr wyznaczany w ramach testu ANOVA; liczba stopni swobody pomiędzy grupami
F	wartość testu F, związana z rozkładem F (Fischera-Snedeckora)
γ	kąt zbieżności gałek ocznych, zawarty pomiędzy liniami patrzenia lewego i prawego oka
H	wartość testu Kruskala-Wallisa
$\mathbf{H_0}$	hipoteza zerowa
$\mathbf{H_A}$	hipoteza alternatywna
MS Błąd	parametr wyznaczany w ramach testu ANOVA; błąd średniokwadratowy, zmienność wewnątrz grup
MS Efekt	parametr wyznaczany w ramach testu ANOVA; efekt średniokwadratowy, zmienność pomiędzy grupami
N	w statystyce: liczba obserwacji
N ważnych	w statystyce: liczba ważnych obserwacji
p	poziom *p* – poziom krytyczny testu, jest miarą dowodów przeciwko hipotezie zerowej $\mathbf{H_0}$,
PCD	metoda środków źrenic (wyznaczająca punkt fiksacji wzroku w przestrzeni)
PoR	punkt fiksacji wzroku

R Spearman	wartość współczynnika korelacji rang Spearmana; zakres wartości współczynnika korelacji: [-1, +1]
ROI	obszar zainteresowania
SS Błąd	parametr wyznaczany w ramach testu ANOVA; suma kwadratów odchyleń wewnątrz grup, prawdziwy błąd losowy; inaczej: wariancja niewyjaśniona przez eksperyment (wariancja błędu)
SS Efekt	parametr wyznaczany w ramach testu ANOVA; suma kwadratów odchyleń pomiędzy grupami; inaczej: wariancja wyjaśniona przez eksperyment (wariancja kontrolowana)
UW (lub *a*)	uwaga wzrokowa widza wyrażona w [%]; określa względne skupienie wzroku badanego na bodźcu wzrokowym w stosunku do czasu, w którym badany fiksował wzrok na prezentowanej mu treści wizyjnej
V	wpływ ściągający wyrażony w [°]
w	zmienna określająca szerokość bodźca wzrokowego prezentowanego na ekranie monitora wyrażona w [°]
W	wartość statystyki Shapiro-Wilka, testu badającego rozkład normalny badanej zmiennej

Spis treści

1 Wprowadzenie

Natura zjawisk zachodzących w centralnym układzie nerwowym człowieka od lat intryguje ludzi zajmujących się różnymi dziedzinami nauki. Na szczególne uznanie z pewnością zasługują osiągnięcia neurofizjologów, które przyczyniły się do poszerzenia wiedzy w tym zakresie. Jednakże pomimo licznych badań, prowadzonych w różnych jednostkach naukowo-badawczych, w dalszym ciągu znaczna część procesów zachodzących w mózgu, w tym zjawisk związanych z jednoczesną percepcją bodźców wzrokowych i słuchowych, pozostaje nieznana. Wiadomo, iż jednoczesny odbiór bodźca wzrokowego i słuchowego prowadzi do percepcji kompleksowego wrażenia wizyjno-fonicznego, które różni się od złożenia wrażeń, będących następstwem niezależnej stymulacji zmysłu wzroku i słuchu [26] [94]. Zatem odbiór spójnego bodźca wzrokowo-słuchowego w ogólności prowadzi do percepcji odmiennego wrażenia wizyjno-fonicznego. Związek pomiędzy treścią wizyjną i foniczną prezentowaną jednocześnie powoduje bądź ukrycie części informacji w uświadamianym wrażeniu, bądź prowadzi do jej przekłamania. Większość znanych korelacji wzrokowo-słuchowych wiąże się z wpływem obrazu na percepcję dźwięku. Wynika to bezpośrednio z faktu, że do ludzkiej świadomości trafia zdecydowanie więcej informacji dostarczanych przez zmysł wzroku. Wśród znanych zjawisk wynikających z korelacji wzrokowo-słuchowych wymienić należy efekt McGurka czy wpływ ściągający obrazu[s], związany ze zmianą percepcji położenia pozornego źródła dźwięku[s]. Temu ostatniemu poświęcono rozważania podjęte w niniejszej rozprawie.

Do niedawna większość badań prowadzonych w celu poznania zjawisk związanych z percepcją bodźców wzrokowych i słuchowych opierała się na subiektywnych ocenach osób, biorących w nich udział. Rozwój technik rejestracji i analizy fal mózgowych znacząco zwiększył możliwości z zakresu poznania procesów zachodzących w mózgu. Celem stosowania systemu śledzenia wzroku w badaniach korelacji wzrokowo-słuchowych jest dostarczenie informacji o uwadze wzrokowej[s] osoby badanej. Badanie aktywności elektrycznej centralnego układu nerwowego (analiza fal EEG) wskazuje na pracę określonych ośrodków mózgu, ale nie dostarcza informacji o tym, na których

[s] Termin, którego opis/wyjaśnienie zawarto w *Słowniku pojęć*

fragmentach prezentowanego obrazu osoba badana skupia wzrok. Dlatego niniejsza rozprawa podejmuje zagadnienie wykorzystania informacji o położeniu wzroku osoby badanej na prezentowanej w trakcie badania treści wizyjnej w kontekście wpływu obrazu na percepcję dźwięku.

Celem głównym niniejszej rozprawy jest, zatem, wykazanie słuszności prowadzenia badań korelacji wzrokowo-słuchowych z zastosowaniem systemu śledzenia wzroku, rejestrującego położenie punktów fiksacji wzroku[s] (ang. *fixation point, point of regard* – PoR) w trakcie trwania eksperymentu. Wykazanie powyższego celu ma bezpośredni związek ze stworzeniem stanowiska badawczego, przeznaczonego do prowadzenia badań nad wpływem obrazu na percepcję dźwięku z wykorzystaniem techniki śledzenia wzroku oraz opracowaniem materiału badawczego, zawierającego odpowiednio przygotowane próbki wizyjno-foniczne. Osiągnięcie celu głównego rozprawy wiąże się z udowodnieniem postawionych tez:

1. **Zastosowanie systemu śledzenia punktu fiksacji wzroku do prowadzenia badań nad wpływem obrazu na percepcję dźwięku prowadzi do obiektywizacji wyników.**
2. **Śledzenie punktu fiksacji wzroku umożliwia prowadzenie eksperymentów nad wpływem ściągającym obrazu przestrzennego na percepcję dźwięku.**

Ponadto, w ramach realizacji pracy pojawiły się dodatkowe cele cząstkowe rozprawy. Jednym z nich było wykazanie zależności **pomiędzy położeniem bodźca wzrokowego, na którym skupiony jest wzrok osoby badanej a wpływem ściągającym obrazu na percepcję dźwięku.** Dowiedzenie powyższej zależności z jednej strony potwierdza istnienie wpływu ściągającego opisanego w literaturze [4] [13] [14] [16] [18] [73] [74] [137] [144], z drugiej zaś strony, poszerza znaczenie bodźca wzrokowego w prezentowanym poczas eksperymentu materiale badawczym. Okazuje się bowiem, że wartość zaobserwowanego wpływu ściągającego bezpośrednio zależy od położenia bodźca wzrokowego (przykuwającego uwagę wzrokową widza) względem środka ekranu. Przy okazji udowadniania powyższego celu cząstkowego **zweryfiko-**

wano słuszność przyjętej metodyki prowadzenia eksperymentów, odnosząc się do wyników badań znanych z literatury. **Drugim celem cząstkowym** rozprawy było zbadanie wpływu ściągającego obrazu na percepcję dźwięku w przypadku wyświetlania treści wizyjnej na wyświetlaczach o różnej wielkości. Zbadano trzy rozmiary wyświetlacza: tzw. mały, średni i duży. W ramach przeprowadzonych eksperymentów udowodniono, że kąt widzenia obrazu obiektu (rozumiany jako szerokość obszaru zajmowanego przez obiekt, który jest widziany przez obserwatora) decyduje o wpływie ściągającym niezależnie od wielkości wyświetlanego obrazu. Zjawisko to w ramach niniejszej pracy nazywane jest umownie **„skalowalnością" wpływu ściągającego**.

Pomimo, iż wpływ bodźców wzrokowych na percepcję bodźców słuchowych, który wpisuje się w nurt badań korelacji wzrokowo-słuchowych, stał się tematem systematycznych badań już ponad sto lat temu, to temat ten jest ciągle aktualny. Wynika to ze zmian technologicznych, które zachodzą zarówno w sferze obrazu, jak i dźwięku. W kontekście oceny wpływu obrazu na percepcję dźwięku przeprowadzono wiele interesujących badań. W eksperymentach stosowano różne standardy systemów odsłuchowych – stereofonię dwu- i wielokanałową, różne sposoby wyświetlania obrazu – od monitorów kineskopowych o standardowym rozmiarze po wielkoformatowe wyświetlacze LCD oraz różną treść prezentowanych próbek testowych. Wyniki wspomnianych badań niezaprzeczalnie wskazują na istnienie tak zwanego wpływu ściągającego obrazu na percepcję dźwięku[s], który w literaturze anglojęzycznej nazywany jest „efektem bliskości obrazu" (ang. *image proximity effect*) [47] [73] [74] [103] [104] lub „efektem brzuchomówcy" (ang. *ventriloquism-effect*) [4] [12] [13] [14] [15] [16] [18] [20] [118] [137] [144]. Zjawisko to polega na tym, że osoba badana inaczej lokalizuje pozorne źródło dźwięku w panoramie stereofonicznej, gdy poddawany ocenie jest sam bodziec słuchowy i inaczej, gdy bodźcowi słuchowemu towarzyszy związany z nim bodziec wzrokowy. Wykazanie istnienia tego zjawiska w sposób bardziej obiektywny staje się możliwe dzięki śledzeniu wzroku osoby badanej i wyznaczaniu względnych wartości czasu skupienia wzroku w obszarze zainteresowania[s]. **Zatem najważniejsza różnica** w zaproponowanej przez autora rozprawy metodologii prowadzenia badań korelacji wzrokowo-słuchowych w porównaniu z badaniami prowadzonymi dotychczas, polega na wykorzystaniu informacji o skupieniu wzroku osoby badanej.

W ramach rozprawy zweryfikowano słuszność zaproponowanej metodologii prowadzenia badań z zakresu korelacji wzrokowo-słuchowych z wykorzystaniem systemu śledzenia wzroku. Udowodniono, iż **stosowanie systemu śledzenia wzroku w badaniu wpływu obrazu na percepcję dźwięku jest uzasadnione i prowadzi do obiektywizacji uzyskanych wyników**. Ponadto wykazano, że **śledzenie punktu fiksacji wzroku umożliwia prowadzenie eksperymentów nad wpływem ściągającym obrazu przestrzennego (3D) na percepcję dźwięku**. Dodatkowo, udowodniono istnienie silnej zależności pomiędzy wpływem ściągającym a położeniem bodźca wzrokowego w obrazie oraz potwierdzono słuszność stosowanej metodyki prowadzenia eksperymentów z zakresu korelacji wzrokowo-słuchowych. Wykazano również, że wpływ ściągający obrazu na percepcję dźwięku nie zależy istotnie od rozmiaru ekranu, na którym wyświetlany jest obraz wizyjny próbki wizyjno-fonicznej.

Przeprowadzone w ramach rozprawy badania poza aspektem naukowym mają również aspekt praktyczny. Warto zwrócić uwagę na fakt, że znajomość zjawisk wynikających z interakcji bodźców wzrokowych i słuchowych może wpłynąć na zmianę procesu przygotowywania utworów filmowych czy ogólnie pojętych treści multimedialnych. Treść wizyjno-foniczna uwzględniająca zjawisko tzw. percepcji wielomodalnej[s] może dostarczyć widzom dodatkowych wrażeń w porównaniu z treścią wizyjno-foniczną przygotowaną zgodnie z powszechnie znanymi zasadami.

W rozdziale 2. niniejszej rozprawy omówiono podstawowe zagadnienia percepcji widzenia i słyszenia. Opis zagadnień związanych z fizjologią narządu wzroku i słuchu stanowi punkt wyjścia do zrozumienia percepcji wielomodalnej, która jest wynikiem jednoczesnej stymulacji zmysłów, np. wzroku i słuchu. Zbadany w ramach rozprawy wpływ obrazu na percepcję dźwięku bezpośrednio wiąże się ze zjawiskiem percepcji wielomodalnej. Rozdział 3. zawiera przegląd badań w dziedzinie korelacji wzrokowo-słuchowych z uwzględnieniem kontekstu, w jakim badania te były i są prowadzone. Najważniejszym z punktu widzenia opisanych w niniejszej rozprawie badań jest kontekst lokalizacji pozornego źródła dźwięku. W rozdziale 4. scharakteryzowano zastosowaną w przeprowadzonych badaniach technikę śledzenia PoR[s] na ekranie komputera. W szczególności skoncentrowano się na opracowanym w Katedrze Systemów Multimedialnych systemie śledzenia punktu fiksacji wzroku, nazywanym Cyber-Oko[s].

Przedstawiono badania, w ramach których wyznaczono rozdzielczość przestrzenną systemu, wskazującą na dokładność wyznaczania kierunku patrzenia. Ponadto, w rozdziale 4. odniesiono się do technik śledzenia PoR w przestrzeni, które bezpośrednio wiążą się z badaniem korelacji wzrokowo-słuchowych z wykorzystaniem trójwymiarowego obrazu wizyjnego.

W rozdziałach 5. i 6. przedstawiono etapy zrealizowanej praktycznej części rozprawy. W rozdziale 5. opisano proces przygotowania materiału badawczego, przedstawiono poszczególne konfiguracje opracowanego stanowiska badawczego oraz zaprezentowano przebieg testów subiektywnych z wykorzystaniem systemu śledzenia wzroku. Wreszcie w ostatniej części rozdziału 5. scharakteryzowano wyniki otrzymane w wyniku wykonanych badań. W rozdziale 6. przeprowadzono analizę statystyczną wyników uzyskanych w ramach dwóch serii eksperymentu. Zestawiono informacje na temat istotności statystycznej zaobserwowanego wpływu ściągającego w poszczególnych próbkach wizyjno-fonicznych zarówno z konwencjonalnym obrazem 2D, jak i obrazem przestrzennym (3D). Zbadano również związek pomiędzy wpływem ściągającym a wybranymi obiektywnymi parametrami, związanymi z bodźcem wzrokowym. Badaniu poddano relację pomiędzy wpływem ściągającym a położeniem bodźca wzrokowego na ekranie, czasem skupienia wzroku na danym bodźcu, wyznaczonym przez system śledzenia punktu fiksacji wzroku oraz wielkością wyświetlanego obiektu. Ponadto, wykazano istnienie tzw. zjawiska skalowalności, zgodnie z którym wielkość wyświetlacza, na którym prezentowany jest materiał wizyjny, nie wpływa istotnie na zajwisko wpływu ściągającego obrazu na percepcję dźwięku. W rozdziale 6. przedstawiono również dowody postawionych w rozprawie tez.

W ostatnim rozdziale niniejszej rozprawy przedstawiono wnioski wynikające z przeprowadzonych badań oraz przytoczono dowody postawionych tez. W zakończeniu rozdziału zawarto propozycje kontynuacji prac badawczych oraz zaprezentowano schemat ideowy systemu przeznaczonego do prowadzenia dalszych zobiektywizowanych badań nad wpływem obrazu na percepcję dźwięku.

2 Wybrane aspekty percepcji widzenia i słyszenia

W niniejszym rozdziale przedstawiono wybrane zagadnienia percepcji obrazów i dźwięków. Skoncentrowano się na tych charakterystykach widzenia i słyszenia, które pozwolą zrozumieć, w jaki sposób obraz i dźwięk są percypowane w mózgu człowieka. Omówiono pokrótce elementy budowy anatomicznej oka i ucha ludzkiego oraz procesy zachodzące w mózgu, odpowiedzialne za percypowanie obrazów i dźwięków. Zagadnienia te są punktem wyjścia do zrozumienia tego, w jaki sposób oglądanie obrazu może wpłynąć na percepcję dźwięku (i odwrotnie) i tym samym stanowią podstawę przeprowadzonych badań.

2.1 Wybrane charakterystyki widzenia

Za proces widzenia odpowiedzialny jest narząd wzroku, w skład którego wchodzi oko i narządy dodatkowe. Oko obejmuje gałkę oczną oraz nerw wzrokowy, który łączy ją z mózgiem. Do narządów dodatkowych oka zalicza się: mięśnie gałki ocznej oraz powięzie oczodołu, powieki, spojówkę i narząd łzowy. Narząd wzroku mieści się w oczodole, pełniącym funkcję ochronną [24] [102].

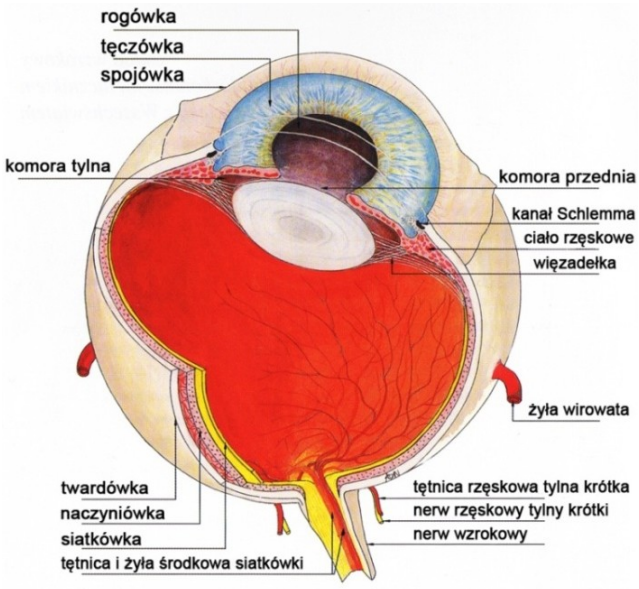
2.1.1 Fizjologia widzenia

Gałka oczna

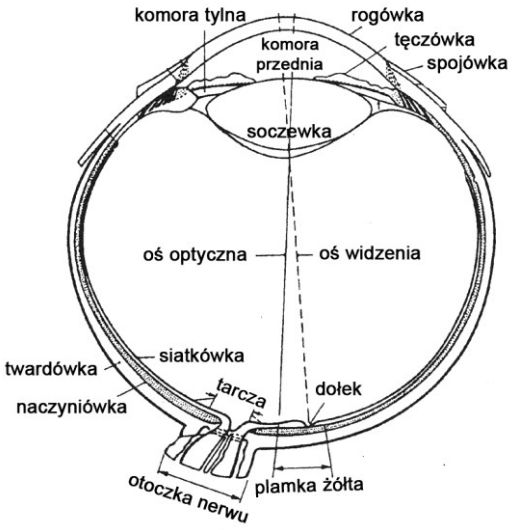
Gałka oczna zbudowana jest z trzech błon: błony zewnętrznej (włóknistej), składającej się z rogówki i twardówki; błony środkowej (naczyniowej), w skład której wchodzą: tęczówka, ciało rzęskowe, naczyniówka i źrenica oraz z błony wewnętrznej, inaczej zwanej siatkówką. Zadaniem gałki ocznej jest ogniskowanie obrazu świata zewnętrznego na siatkówce. Obraz ten jest rzeczywisty, pomniejszony, odwrócony i ostry. Droga wzrokowa przenosi obraz powstały na siatkówce do odpowiednich ośrodków kory mózgowej, gdzie jest odbierany. Anatomię gałki ocznej przedstawiono na rys. 2.1a.

Środki krzywizn rogówki i soczewki leżą na prostej zwanej osią optyczną oka. Oś optyczna nie pokrywa się jednak z osią widzenia, ponieważ miejsce najbardziej ostrego widzenia, plamka żółta, leży poza osią optyczną oka. Podczas obserwacji drobnych szczegółów oko samoczynnie ustawia się tak, aby obraz utworzył się na plamce żółtej. Obrót osi widzenia względem osi optycznej przyjmuje wartość około 5^{o}. Rys. 2.1b przedstawia schemat przekroju poprzecznego gałki ocznej z zaznaczoną osią optyczną i osią widzenia [24] [72].

a)



b)



Rys. 2.1 Gałka oczna: a) schemat budowy [102]; b) schemat przekroju poprzecznego [72]

Kształt gałki ocznej nie jest ściśle kulisty. Składa się ona z dwóch kulistych odcinków o różnym promieniu. Odcinek przedni, odpowiadający rogówce, stanowiący 1/6 powierzchni całej gałki, jest silniej wypukły. Promień jego krzywizny wynosi 7-8 mm. Odcinek tylny jest znacznie większy, a promień jego krzywizny wynosi 12 mm. W gałce ocznej należy odróżnić zawartość gałki i jej ścianę. Wiązka światła pada na siatkówkę, przenikając przez ścianę gałki w obrębie rogówki oraz trzy elementy położone wewnątrz gałki: ciecz wodnistą, soczewkę i ciało szkliste.

Ze względu na występujące w niniejszej rozprawie częste odniesienia do terminu *rogówka*, leżącej u podstaw działania systemu śledzenia punktu fiksacji wzroku, w tym podrozdziale skoncentrowano się na dokładniejszym opisie tej części gałki ocznej. Rogówka stanowi przezroczystą część przednią błony włóknistej. Przez rogówkę jest widoczna źrenica oraz tęczówka. Rogówka ma kształt elipsy, której średnica pozioma wynosi od 11 do 12 mm, a pionowa – od 10 do 11 mm. Rogówka jest częścią błony włóknistej. Jest mocna i odporna, przez co nadaje kształt gałce ocznej i jest jej narządem ochronnym. Głównym jednak zadaniem rogówki jest przepuszczanie i załamywanie promieni świetlnych. Zdolność refrakcyjna rogówki jest kilkakrotnie większa od zdolności załamywania światła przez soczewkę [24].

Ruchy gałek ocznych

Ruchy gałek ocznych spełniają dwa podstawowe zadania. Pierwszym z nich jest kompensacja ruchów głowy lub ruchów przedmiotów w polu widzenia w celu ustabilizowania obrazu na siatkówce. Drugą funkcją ruchów gałek ocznych jest ustawienie gałki ocznej względem obiektu, na którym fiksowany jest wzrok. Obraz obiektu jest wtedy rzutowany w obszarze plamki żółtej, czyli w obszarze o największej rozdzielczości widzenia związanej z największym zagęszczeniem czopków na siatkówce.

Ruchy gałek ocznych można ogólnie podzielić na szybkie i wolne, zgodnie z klasyfikacją zaproponowaną przez Roberta M. Steinmana [129]. Ruchy szybkie są to tak zwane sakkady. Mogą być dobrowolne (świadome) i osiągać prędkość szczytową do $550^{o}/s$. Sakkady umożliwiają rzutowanie obrazu oglądanego obiektu na obszar plamki żółtej. Innymi słowy, gałka oczna wykonuje ruch sakkadowy, gdy wzrok ma być

utkwiony na konkretnym obiekcie. Wolne ruchy gałki ocznej charakteryzują się stopniowym przyspieszaniem, a ich prędkość zawiera się w przedziale 3'/s (minut kątowych na sekundę) do ok. 200^{o}/s. Gałka oczna wykonuje wolny ruch w celu śledzenia obiektu poruszającego się w polu widzenia. Poza omówionymi powyżej ruchami gałka oczna znajduje się również w stanie względnego spoczynku – fiksacji. Należy zauważyć, że podczas normalnej aktywności człowieka, przez większość czasu oczy pozostają w stanie fiksacji. To właśnie w czasie fiksacji zachodzi pobieranie informacji wzrokowej z otoczenia. Czas trwania fiksacji jest zależny od sposobu przetwarzania informacji. Na ogół waha się w granicach od ok. 0,15s do 1,5s [129]. *Fiksacja* jest pojęciem, które często będzie powracać w niniejszej rozprawie przy okazji opisu systemu śledzenia punktu fiksacji wzroku (PoR), który wykorzystano w przeprowadzonych badaniach.

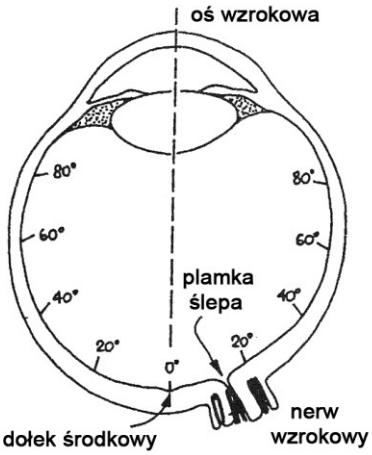
Proces widzenia

Za rzutowanie odwróconego obrazu oglądanej sceny na siatkówce oka odpowiedzialna jest rogówka i soczewka. Zjawisko powstawania obrazu w oku nie było znane aż do XVII w., kiedy poczyniono pierwsze teoretyczne badania i eksperymenty w celu poznania natury tego zjawiska. Proces widzenia wyjaśnił teoretycznie Kepler w 1604 r. [31].

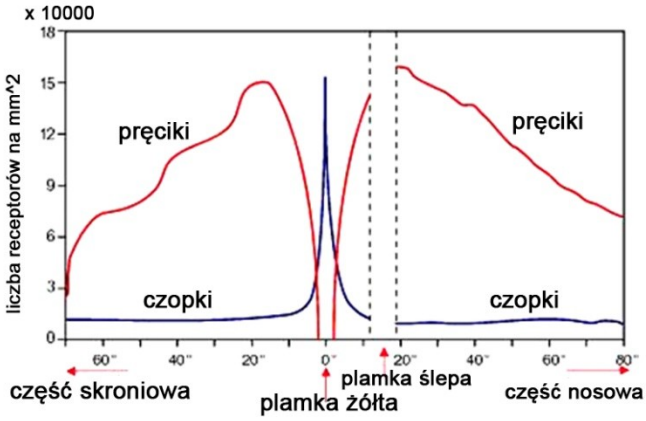
Współcześnie proces widzenia jest rozumiany jako proces elektrochemiczny. Gdy komórki wzrokowe pręcikowe (pręciki) lub czopkowe (czopki), wchodzące w skład warstwy światłoczułej siatkówki, zostają pobudzone światłem, chwilowo zmienia się kompozycja pigmentu. W uproszczeniu można powiedzieć, że zmiana ta powoduje powstanie bardzo małego prądu elektrycznego, który trafia do mózgu przez włókna nerwowe. Z pojedynczym włóknem nerwowym połączonych jest około sto pręcików, natomiast każdy czopek w dołku środkowym jest połączony z mózgiem indywidualnie. Czynnością pręcików jest przystosowanie oka do słabego oświetlenia i rozróżnianie zarysów przedmiotów (widzenie skotopowe), zaś czynnością czopków jest widzenie kształtu i barw przedmiotów w jasnym oświetleniu (widzenie fotopowe). Widzenie fotopowe (plamkowe) pozwala zatem na dokładne rozpoznanie szczegółów, kształtu i barwy – charakteryzuje się wysoką zdolnością rozdzielczą. Widzenie obwodem siat-

kówki daje orientację w przestrzeni [24] [72]. Na rys. 2.2 pokazano rozkład fotoreceptorów na powierzchni siatkówki.

a)



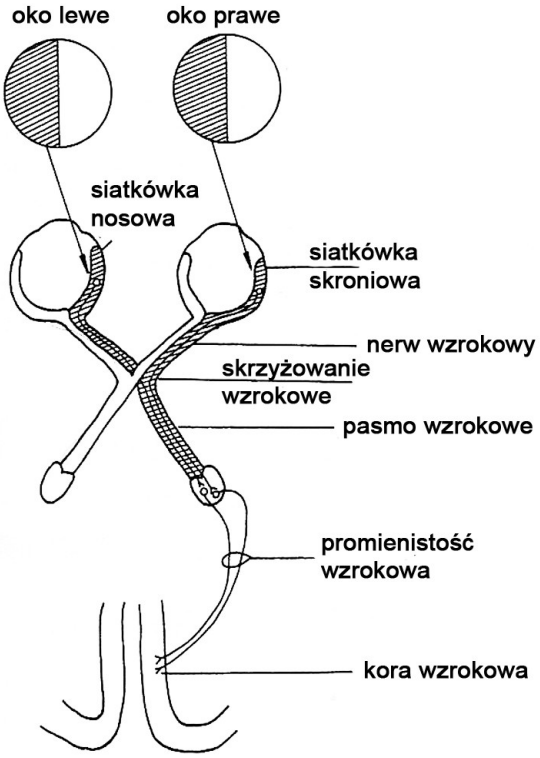
b)



Rys. 2.2 Rozmieszczenie czopków i pręcików w różnych obszarach siatkówki; a) przekrój gałki ocznej [72]; b) wizualizacja rozkładu czopków i pręcików na siatkówce [174]

Aksony komórek zwojowych siatkówki zbierają się w tarczy nerwu wzrokowego i tworzą nerw wzrokowy. W nim przekazywana jest impulsacja z siatkówki do ośrodków mózgowych. Nerw wzrokowy zawiera ok. 1 miliona aksonów komórek zwojowych, bezpośrednio odprowadzających impulsy z siatkówki. Nerwy wzrokowe lewego i prawego oka wchodzą do jamy czaszki i tworzą skrzyżowanie wzrokowe, w którym krzyżują się włókna z donosowych połówek siatkówki. Włókna z przyskroniowych połówek biegną dalej nieskrzyżowane. Włókna z przyśrodkowych części siatkówki krzyżują się tak, że prawe pasmo wzrokowe zawiera aksony pochodzące ze skroniowej – prawej (nieskrzyżowanej) połowy siatkówki prawego oka oraz z nosowej – prawej połowy siatkówki oka lewego. Reprezentacja w prawym paśmie wzrokowym zawiera więc ak-

sony z prawych połówek oka lewego i oka prawego. Analogiczna sytuacja ma miejsce w lewym paśmie wzrokowym. Większość włókien pasma wzrokowego kończy się w ciele kolankowatym bocznym, które stanowi jej główną stację przełącznikową. Wypustki ciała kolankowatego biegną w promienistości wzrokowej do płatu potylicznego, gdzie zlokalizowany jest ośrodek wzrokowy mózgu [72]. Schemat drogi wzrokowej przedstawiono na rys. 2.3.



Rys. 2.3 Schemat drogi wzrokowej [72]

2.1.2 Percepcja głębi

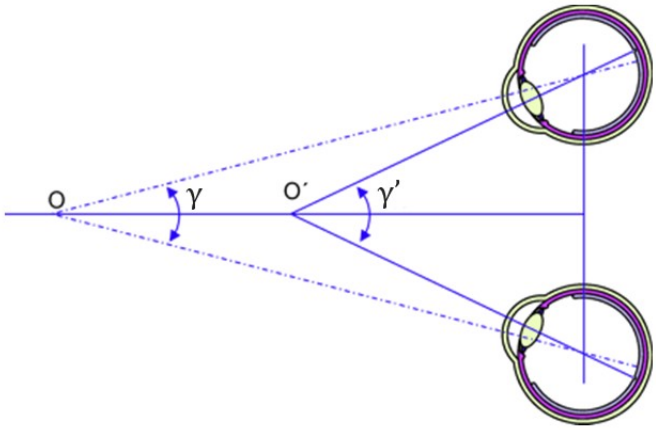
Z uwagi na fakt, iż w ramach rozprawy doktorskiej przeprowadzono badania związane z percepcją obrazu stereoskopowego, ta część podrozdziału została poświęcona podstawowym zagadnieniom z zakresu percepcji głębi stereoskopowej w postrzeganym obrazie.

Kierunki widzenia

Kierunek widzenia (inaczej: oś widzenia, promień kierunku Helmholtza) oznacza linię łączącą punkt obserwowanego obiektu z jego obrazem na siatkówce. Każde z oczu dostarcza mózgowi inny obraz tego samego obiektu, co wynika z faktu, iż inne są pozycje obu oczu. Choć mózg otrzymuje dwa różne obrazy postrzeganego świata (każdy z innej perspektywy), to w rzeczywistości osoby o prawidłowym wzroku widzą jeden obraz, a nie dwa. Wytłumaczenie tego zjawiska tkwi w teorii dotyczącej kierunków widzenia. Każde oko charakteryzuje się innym kierunkiem widzenia obiektu, na którym człowiek fiksuje wzrok. Jednakże mózg syntezuje oba te kierunki, dając poczucie jednego kierunku widzenia. Jest to tak zwana wspólna, subiektywna oś widzenia łącząca trzecie oko urojone (zlokalizowane u nasady nosa, stąd nazywane również cyklopowym) z punktem obserwowanego obiektu [19] [105].

Zbieżność oczu

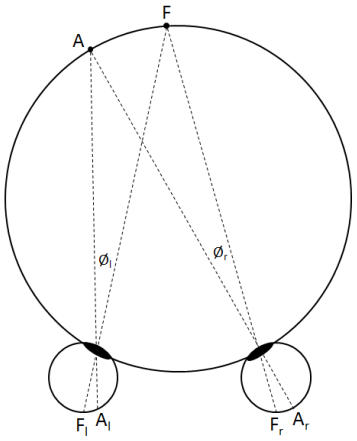
Zbieżność (ang. *convergence*) jest właściwością narządu wzroku, polegającą na obróceniu się gałek ocznych do środka w celu przecięcia osi widzenia każdego z oczu w punkcie, na którym jest fiksowany wzrok. Dostrajanie pozycji oka do pożądanego kąta zbieżności (γ) wykonywane jest przez mięśnie oka i trwa zwykle kilka dziesiątych sekundy. Kiedy obiekt, na którym fiksowany jest wzrok zbliża się do obserwatora, kąt zbieżności oczu wzrasta, jak pokazano na rys. 2.4 [102] [173]. Zagadnienie zbieżności wzroku w kontekście niniejszej rozprawy ma szczególne znaczenie w przypadku badania wpływu obrazu trójwymiarowego na percepcję słyszenia. Zakłada się, iż widz oglądając próbki filmowe 3D, fiksuje wzrok na „pozornych” obiektach znajdujących się przed i za płaszczyzną ekranu. Zatem wyznaczanie kąta zbieżności oczu pozwala określić, na które miejsce w przestrzeni patrzy osoba badana w danej chwili.



Rys. 2.4 Zbieżność oczu, gdzie kąt $\gamma' > \gamma$

Horopter

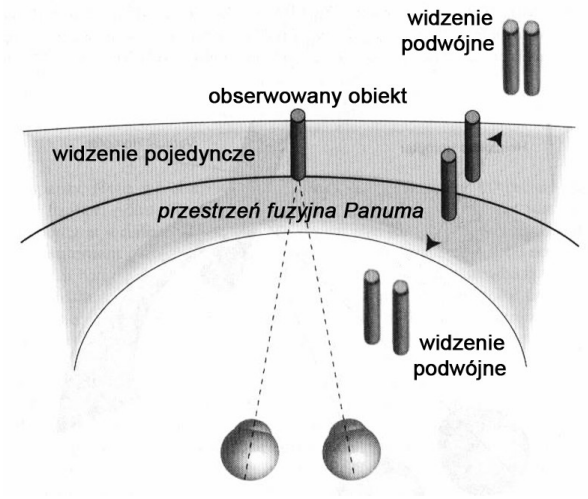
Termin „horopter" w wolnym tłumaczeniu oznacza „horyzont widzenia" (grec. *horos* – granica, *opter* – obserwator). Powszechnie obowiązująca definicja horoptera została zaproponowana przez Vietha w 1818 r. w jego dziele „Über die Rechtung der Augen" [136]. Zgodnie z tą definicją, horopter jest trójwymiarową krzywą, będącą zbiorem punktów w przestrzeni, z których światło pada na korespondujące miejsca dwóch siatkówek. Innymi słowy, horopter jest zbiorem punktów w przestrzeni wzrokowej, zapewniającym widzenie pojedyncze. Oczy obserwujące ten sam obiekt mają ten sam kierunek widzenia, czego efektem jest powstawanie jednakowych obrazów na siatkówkach obu oczu [161]. Rys. 2.5 wyjaśnia geometryczne zależności pomiędzy punktem fiksacji a korespondującymi z nim punktami na siatkówkach oczu. Punkt A na rys. 2.5 geometrycznie musi leżeć na okręgu przechodzącym przez punkt F, punkt fiksacji oraz środki soczewek obu oczu. Dla A_l i A_r kąty $\emptyset_l$ i $\emptyset_r$ muszą być jednakowe [161].



Rys. 2.5 Wizualizacja horoptera teoretycznego

Przestrzeń fuzyjna Panuma

Przestrzeń fuzyjna Panuma odnosi się do wąskiego pasma wokół horoptera, wewnątrz którego zachodzi fuzja (widzenie pojedyncze), dlatego też przestrzeń ta definiowana jest jako strefa widzenia stereoskopowego. Fuzja jest tym trudniejsza do uzyskania, im większa jest odległość punktu fiksacji od horoptera. Rozbieżności w postrzeganiu obiektu znajdującego się daleko od horoptera – poza przestrzenią fuzyjną, są na tyle duże, że uniemożliwiają fuzję, co prowadzi do widzenia podwójnego, jak zobrazowano na rys. 2.6. Przestrzeni fuzyjnej Panuma odpowiada powierzchnia na siatkówce zwana powierzchnią Panuma. Powierzchnie Panuma nie mają określonej wielkości i zależą od warunków stymulacji. Jednakże są one większe dla dużych obiektów, a węższe dla małych obiektów [130].



Rys. 2.6 Przestrzeń fuzyjna Panuma z zaznaczeniem zakresów widzenia pojedynczego i podwójnego [151]

Korespondencja siatkówkowa

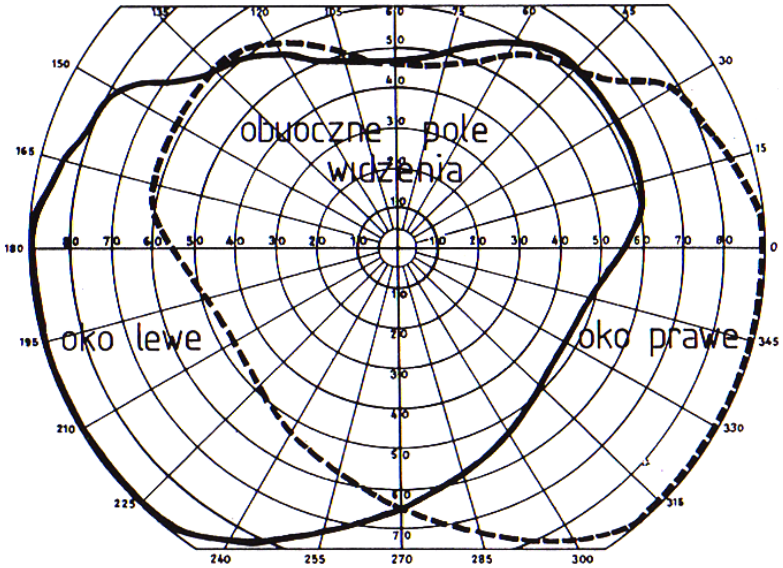
Jak wspomniano wcześniej, dwuoczna percepcja głębi bazuje na fakcie, że oczy odbierają dwa nieznacznie różniące się obrazy tej samej sceny. System wzrokowy musi zatem najpierw dopasować punkty dwóch obrazów siatkówkowych, które odpowiadają temu samemu punktowi w przestrzeni. Dla wszystkich n punktów obrazu istnieje n^2 możliwych konfiguracji, co znacznie utrudnia dopasowanie. Jednym z rozwiązań tego problemu jest istnienie punktów w dwóch siatkówkach, związanych ze sobą tak, że dla każdego punktu jednej siatkówki istnieje punkt na powierzchni drugiej siatkówki, który po stymulacji umożliwia widzenie w tym samym kierunku – punkty zdają się być nałożone na siebie w przestrzeni wizualnej. Takie pary nazywane są punktami korespondującymi i mają duże znaczenie w widzeniu dwuocznym [125].

Stereopsja

Widzenie stereoskopowe, czyli postrzeganie trzeciego wymiaru (lub inaczej percepcja przestrzeni) jest możliwe tylko wtedy, gdy narząd wzroku (w tym oko lewe i oko prawe) jest sprawny, a każde z oczu dostarcza do świadomości nieznieksztalcony obraz.

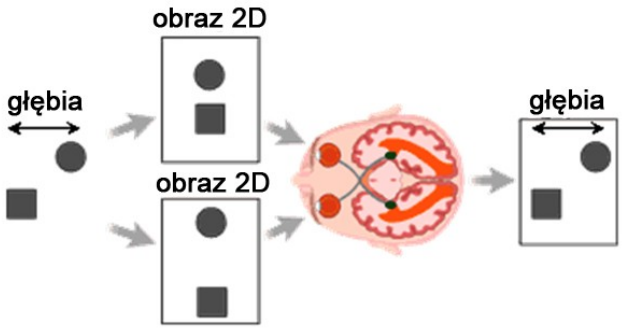
Prawidłowe pole widzenia każdego z oczu zawiera się w zakresie do 85^o-90^o od strony skroniowej, do 55^o-60^o od nosa, 45^o-55^o od góry oraz do 65^o-70^o od dołu. Obwodowe granice każdego z pola widzenia mogą być nieznacznie zmienione. Mają na to wpływ następujące czynniki: głębokość osadzenia oka w oczodole, wysoka nasada nosa, łuk brwiowy oraz opadająca powieka górna.

Pole widzenia, w którym oglądany fragment rzeczywistości postrzegany jest przestrzennie (obuoczne pole widzenia) powstaje w wyniku połączenia się pola oka prawego z polem oka lewego. W przybliżeniu obejmuje ono krąg o średnicy 60^o. Zakres poszczególnych pól widzenia przedstawiono na rys. 2.7. Prawidłowość pola widzenia pozwalającego percypować obraz przestrzenny zależy od ustawienia i ruchomości obu gałek ocznych [102].



Rys. 2.7 Granice obuocznego pola widzenia [102]

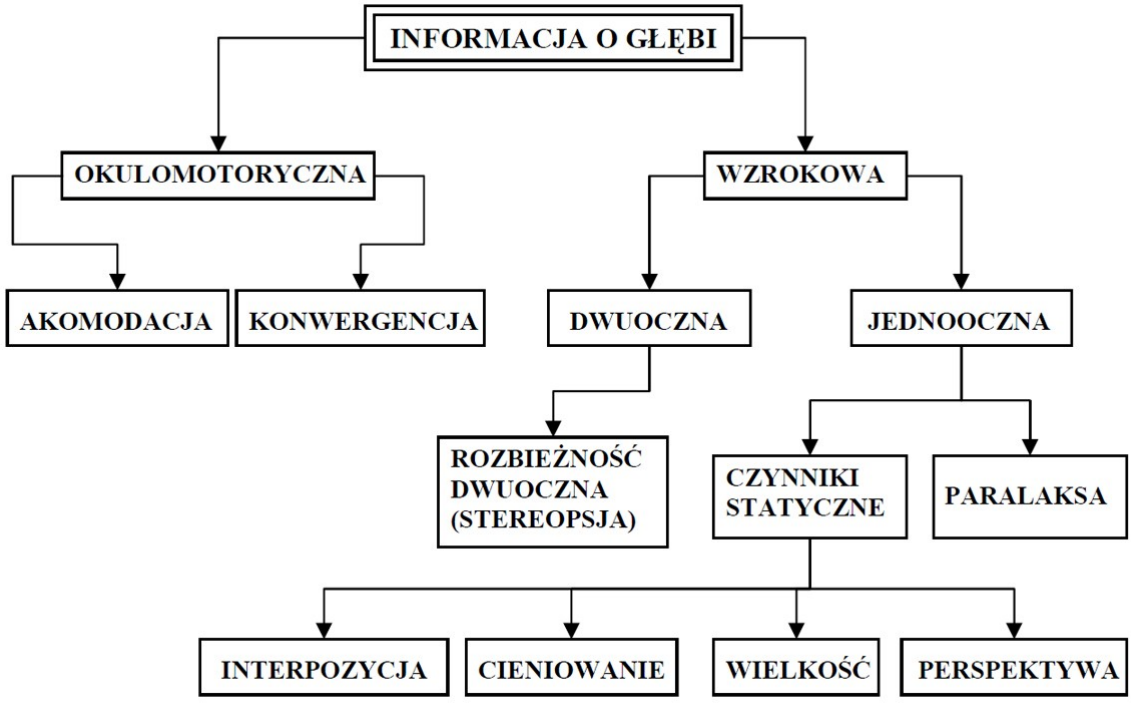
Stereopsja jest to proces percepcji głębi, będący skutkiem siatkówkowej rozbieżności, rozumianej jako różnica obrazów obserwowanego obiektu powstałych na siatkówkach oczu. Różnica ta wynika z faktu, że każde z oczu odbiera obraz z innej perspektywy. Dwa obrazy są ze sobą kojarzone dzięki procesowi korespondencji, który jest osiągany przez dopasowanie następujących cech obrazu: powierzchni teksturowych, krawędzi, rogów, segmentów [48]. Innymi słowy, stereopsja to zdolność fuzji obrazów, które stymulują w poziomie różne elementy siatkówek w obrębie przestrzeni fuzyjnej Panuma, czego wynikiem jest dwuoczne widzenie obiektu w trzech wymiarach. Proces percepcji obrazu trójwymiarowego przedstawiono na rys. 2.8.



Rys. 2.8 Proces powstawania obrazu przestrzennego [152]

Najbardziej znane techniki obrazowania stereoskopowego, stosowane również obecnie w fotografii, kinie, telewizji i grach komputerowych, są oparte właśnie na zjawisku fuzji. Wystarczy wymienić chociażby technikę anaglifową [168], polaryzacyjną [153] czy migawkową [180]. Wszystkie te techniki charakteryzuje wspólna cecha: są oparte na filtrze, decydującym o tym, która składowa obrazu stereoskopowego trafia do konkretnego oka. W technice anaglifowej występuje filtr koloru (widz zakłada dwubarwne okulary: z czerwonym i cyjanowym szkiełkiem), w technice polaryzacyjnej – filtr polaryzacyjny (współcześnie najczęściej stosowaną polaryzacją jest polaryzacja kołowa: prawoskrętna i lewoskrętna), natomiast w technice migawkowej występuje filtr czasu – w danym momencie obraz jest dostarczany tylko do jednego oka. W ramach niniejszej rozprawy doktorskiej przeprowadzono badania z wykorzystaniem (między innymi) materiału wideo 3D, wyświetlanego w technice anaglifowej i polaryzacyjnej.

Warto w tym miejscu zaznaczyć, że nie tylko stereopsja pozwala informację o przestrzenności otaczającego go świata. Na rys. 2.9 zaprezentowano podział źródeł informacji o głębi, na podstawie których człowiek pozyskuje informację o przestrzeni [164].



Rys. 2.9 Główne źródła informacji o głębi [164]

Dwoma najważniejszymi źródłami informacji są: stereopsja (wymagająca współdziałania obu oczu) oraz czynniki jednooczne, które pozwalają postrzegać głębię przy użyciu jednego oka. Warto wspomnieć również o tym, że głębię można percypować dzięki ruchom głowy, oczu, a także dzięki zmianie paralaksy [33]. W niniejszym podrozdziale (2.1.2) skoncentrowano się na zagadnieniach związanych z percepcją dwuoczną, ponieważ w przeprowadzonych w ramach niniejszej rozprawy eksperymentach, zaprezentowano badanym próbki wideo, przygotowane w technice stereoskopowej[s].

2.2 Wybrane charakterystyki słyszenia

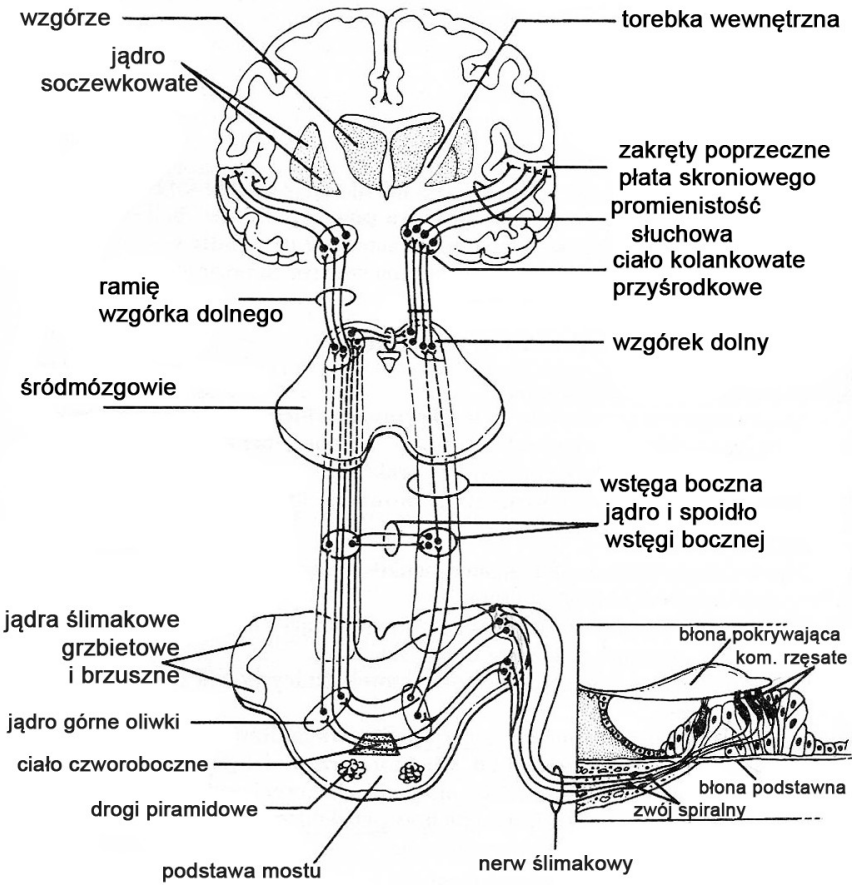
Po zaprezentowaniu charakterystyk widzenia w kontekście badań korelacji wzrokowo-słuchowych, w dalszej części niniejszego rozdziału zostaną omówione wybrane zagadnienia percepcji dźwięku, stanowiące podstawę merytoryczną rozważań na temat wzajemnego wpływu dźwięku i obrazu. Rozważania te ograniczą się do prezentacji drogi słuchowej i charakterystyki percepcji przestrzenności dźwięku.

2.2.1 Fizjologia słyszenia

Droga słuchowa

Przyjmuje się, że droga słuchowa ma swój początek w narządzie Cortiego, czyli właściwym narządzie słuchu, znajdującym się na błonie podstawnej [72], zaś kończy się w korze słuchowej mózgu. Nerw ślimakowy łączy się z neuronami dwóch jąder ślimakowych, grzbietowego i brzusznego, w moście (most – część mózgowia). Z jąder ślimakowych droga słuchowa biegnie w głąb mostu, gdzie łączy się z neuronami jądra brzusznego i grzbietowego ciała czworobocznego. Aksony pochodzące z neuronów ciała czworobocznego biegną po tej samej stronie albo przechodzą na stronę przeciwną, łącząc się później we wspólną drogę zwaną wstęgą boczną. Wstęga boczna biegnie przez śródmózgowie i kończy się synapsami na neuronach wzgórka dolnego. Następnie impulsy słuchowe przewodzone są do ciał kolankowatych przyśrodkowych. Drogi i ośrodki przewodzące impulsację słuchową do kory mózgowej zobrazowano na rys. 2.10 [43] [72] [181].

Impulsy słuchowe wychodzące z ciała kolankowatego przyśrodkowego są przewodzone przez torebkę wewnętrzną i promienistość słuchową do pierwszorzędowego pola słuchowego w korze mózgowej w obrębie zakrętów skroniowych poprzecznych (Heschla) [72].



Rys. 2.10 Uproszczony schemat drogi słuchowej [43]

Korowy ośrodek słuchu w mózgu

W korze słuchowej można wyróżnić pierwszorzędowe pole słuchowe oraz obszar reprezentacji drugorzędowej i trzeciorzędowej. Zakres częstotliwości, na który reaguje kora słuchowa, jest znacznie węższy niż ten, na który reagują neurony drogi słuchowej. Oznacza to, że pasmo przewodzonych częstotliwości ulega zwężeniu podczas przepływu impulsów z receptorów do kory słuchowej.

Pierwszorzędowe pole słuchowe odbiera i analizuje impulsy słuchowe wywołane dźwiękami o różnej częstotliwości, natężeniu, czasie trwania i kierunku pochodzenia.

Drugorzędowa reprezentacja słuchowa różnicuje i porównuje wysokości i natężenia poszczególnych tonów [72].

2.2.2 Percepcja przestrzenności dźwięku

Dźwięki otoczenia generowane przez poszczególne źródła docierają do peryferyjnych warstw narządu słuchowego jako złożone pole dźwiękowe, nie zaś jako niezależne bodźce dźwiękowe. Wielu badaczy uważa, że procesy określania i determinacji źródeł dźwięków stanowią najważniejszy aspekt percepcji dźwięku. Złożone pole dźwiękowe w pierwszej kolejności podlega kodowaniu wynikającemu z geometrii głowy i małżowiny usznej. Zakodowany dwuwymiarowy (czasowo-częstotliwościowy) opis fizyczny zjawisk akustycznych zachodzących w otoczeniu przybiera formę pewnej reprezentacji impulsacji słuchowej, percypowanej przez człowieka jako odpowiednie wrażenie słuchowe [22] [28].

Dwuuszna lokalizacja źródła dźwięku

Fala akustyczna docierająca do uszu słuchacza podlega odbiciom i ugięciom, wynikającym z geometrii głowy. Te zjawiska fizyczne decydują o powstaniu międzyusznej różnicy czasów (ang. *Interaural Time Difference* – ITD) oraz międzyusznej różnicy natężeń (ang. *Interaural Intensity Difference* – IID) – znanych międzyusznych czynników kodujących informację o lokalizacji źródła dźwięku. Międzyuszna różnica czasów – ITD, występuje ze względu na fakt, że impulsy nerwowe z narządu Cortiego ucha znajdującego się bliżej źródła dźwięku, docierają do kory słuchowej wcześniej niż po stronie przeciwnej. Natomiast główną przyczyną powstawania IID jest lokalizacja uszu po przeciwnych stronach głowy. Sygnał dochodzący do jednego ucha może być stłumiony w stosunku do sygnału odbieranego przez drugie ucho. W tym przypadku mówi się o występowaniu zjawiska cienia akustycznego [98].

Parametr ITD można wyznaczyć zgodnie z formułą 2.1, przy założeniu, że głowa ma kształt idealnej kuli. ITD stosuje się dla niskich częstotliwości.

$$ITD = \frac{3a}{c} \cdot \sin(\theta) \tag{2.1}$$

gdzie: a – promień głowy, c – prędkość dźwięku, θ – kąt określający położenie źródła (przy czym kąt 0^o oznacza kierunek na wprost głowy słuchacza)

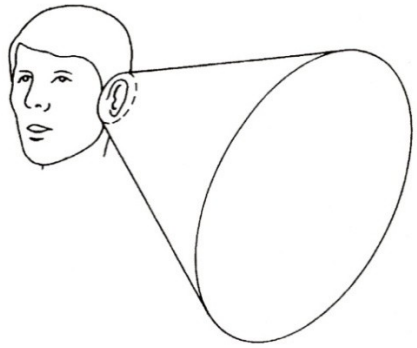
Dla częstotliwości wysokich i średnich dominującym czynnikiem służącym do określenia kierunku dźwięku w płaszczyźnie poziomej jest parametr IID. Można go wyrazić wzorem 2.2.

$$IID \;\; dB \;\; = 20\log(P_L) - 20\log(P_R) \tag{2.2}$$

gdzie: P_L i P_R oznaczają wartości ciśnienia akustycznego odpowiednio w lewym i prawym uchu.

Warto zaznaczyć, że w paśmie 1500 do 3000 Hz częstotliwość jest zbyt wysoka, aby ITD mogło być czynnikiem dominującym, z drugiej zaś strony – jest zbyt niska, żeby zapewnić odpowiednie wartości IID. Z tego powodu we wspomnianym zakresie częstotliwości lokalizacja źródła dźwięku jest trudniejsza i mniej efektywna [98].

Należy dodać, że w przestrzeni otaczającej słuchacza istnieje wiele punktów, w których wartości ITD i IID są równe. Punkty te leżą na powierzchni tak zwanego stożka niepewności, jak pokazano na rys. 2.11. Takie same wartości różnicy dróg i czasów dojścia dźwięku do lewego i prawego ucha występują dla źródeł położonych na podstawie tego stożka. Jednym ze sposobów zlokalizowania źródła dźwięku leżącego na stożku niepewności jest wykonanie ruchu głową.



Rys. 2.11 Stożek niepewności dla ustalonej międzyusznej różnicy czasów (przy założeniu, że głowa jest sferą) [98]

Poza czynnikami międzyusznej różnicy czasów i natężeń, informacja o kierunku źródła dźwięku jest również zakodowana w różnicy faz sygnałów (niskie i średnie częstotliwości). W przypadku, gdy źródło dźwięku znajduje się dokładnie w płaszczyźnie symetrii głowy, różnica faz sygnałów jest równa zero. Zmiana kierunku słyszenia jest odczuwalna, gdy źródło dźwięku przesunie się zaledwie o 3 stopnie w lewo lub prawo od tej płaszczyzny. Detekcja kierunku na podstawie różnicy faz jest bardziej precyzyjna dla kierunków zbliżonych do płaszczyzny symetrii głowy, aniżeli dla kierunków bocznych [55] [72] [121] [181].

W procesie wyznaczania kierunku źródła dźwięku znaczenie mają nie tylko czynniki, w których zakodowana jest informacja o położeniu źródła, ale także przetwarzanie informacji na wyższych piętrach drogi słuchowej. Badania przeprowadzone przez Zimmera i Macaluso [147] z wykorzystaniem fMRI wykazały, że lokalizacja na podstawie ITD jest możliwa tylko wtedy, gdy wejściowe sygnały akustyczne dla lewego i prawego ucha mają podobne profile czasowo-częstotliwościowe. Duża koherencja dwuuszna zapewnia wysoką skuteczność lokalizacji źródła dźwięku. Zaobserwowano, że w przypadku dużej koherencji aktywność kory słuchowej jest większa niż w przypadku mniejszej koherencji dwuusznej.

Warto w tym miejscu wspomnieć również o tak zwanej „charakterystyce przenoszenia głowy” (ang. *Head Related Transfer Function* – HRTF), która określa wpływ m.in. małżowiny usznej i kształtu głowy na rozkład poziomów w funkcji częstotliwości dla różnych położeń źródła fali akustycznej [145] [148]. Funkcja HRTF zawiera wszystkie informacje na temat sferycznego środowiska dźwiękowego (we wszystkich kierunkach), w tym również wartości parametrów ITD i IID. Funkcja HRTF umożliwia regulowanie położenia pozornych źródeł dźwięku w wielokanałowej panoramie stereofonicznej oraz modelowanie dźwięku w słuchawkach stereofonicznych w celu jego wirtualizacji. Ostatni wątek badawczy jest intensywnie rozwijany w Katedrze Akustyki Politechniki Wrocławskiej [36] [108] [109] [110].

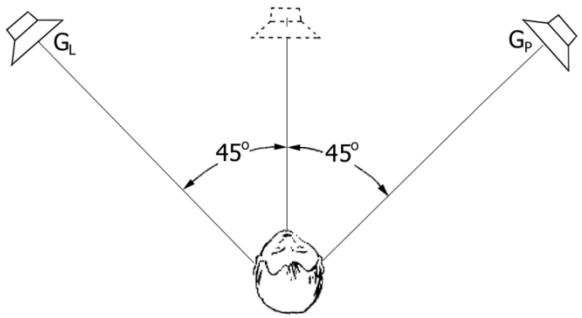
Rozdzielczość czasowa słuchu

Rozdzielczość czasowa słuchu jest zdecydowanie wyższa niż rozdzielczość czasowa wzroku. Znaczna część elementów wchodzących w skład narządu słuchu jest

wyspecjalizowana w pomiarze czasu. Jednak główny wpływ na wysoką rozdzielczość czasową słuchu ma budowa narządu odbierającego dźwięki. W przypadku narządu wzroku, światło padające na siatkówkę oka jest przekształcane na impulsy elektryczne w stosunkowo powolnym procesie elektrochemicznym, zachodzącym w czopkach i pręcikach. W przypadku narządu słuchu natomiast dźwięk jest transformowany na impulsy nerwowe poprzez szybką drogę mechaniczną i bioelektryczną. Można zatem wysnuć wniosek, że w procesie percepcji bodźców wzrokowo-słuchowych narząd wzroku jest odpowiedzialny za lokalizację bodźca w przestrzeni, zaś narząd słuchu – za lokalizację bodźca w czasie. O funkcjonalności poszczególnych zmysłów w kontekście percepcji wspomniano w podrozdziale 2.3.

Zjawisko Haasa

Rozdzielczość narządu słuchu ma bezpośredni związek ze zjawiskiem Haasa, zgodnie z którym opóźnienie fali docierającej do obserwatora ma znacznie większy wpływ na lokalizację pozornego źródła dźwięku (ang. *virtual sound source*) niż różnica poziomów dźwięków emitowanych przez źródła oddalone od siebie w przestrzeni. W przypadku dwóch identycznych źródeł, emitujących dźwięk o tym samym natężeniu, odbiorca lokalizuje pozorne źródło dźwięku dokładnie pośrodku tych źródeł (dźwięki ze źródeł G_L i G_P dochodzą do uszu jednocześnie). Przypadek taki przedstawiono na rys. 2.12.



Rys. 2.12 Lokalizacja pozornego źródła dźwięku w przypadku, gdy źródła G_L i G_P emitują dźwięk o tym samym natężeniu jednocześnie

Zwiększanie opóźnienia jednego z sygnałów powyżej 1 ms powoduje stopniowe przesunięcie percypowanego położenia pozornego źródła dźwięku w stronę głośnika emitującego sygnał bezpośredni. Zatem zgodnie z teorią Haasa w rozproszonym polu

akustycznym o lokalizacji pozornego źródła dźwięku decyduje pierwsza fala dźwiękowa docierająca do słuchacza [21] [50].

2.3 Percepcja wielomodalna

Celem opisu fizjologii widzenia i słyszenia zawartego w podrozdziałach 2.1 i 2.2 było przybliżenie zagadnień związanych z procesem percepcji bodźców wzrokowych i słuchowych dostarczanych przez zmysły wzroku i słuchu. Należy jednak zwrócić uwagę na fakt, że opisywane w niniejszej rozprawie badania odnoszą się do zjawisk związanych z tak zwaną percepcją wielomodalną (ang. *multimodal perception, cross-modal perception*), a więc do zjawisk, które są wynikiem jednoczesnej stymulacji zmysłów wzroku i słuchu. W związku z powyższym niniejszy podrozdział został poświęcony opisowi percepcji wielomodalnej, w szczególności w kontekście wpływu obrazu na percepcję dźwięku. Przykładem percepcji wielomodalnej jest tak zwany efekt McGurka [94]. McGurk i McDonald przeprowadzili eksperyment, w wyniku którego udowodnili, że przy dostarczaniu zmysłom wzroku i słuchu bodźców o tym samym czasie trwania, ale przenoszących inne informacje, człowiek percypuje bodziec słuchowy, który jest niejako syntezą informacji zawartej w bodźcu wzrokowym i słuchowym. Przeprowadzony przez nich eksperyment polegał na tym, że badani oglądali twarz człowieka wypowiadającego sylaby „ga-ga”, natomiast w ścieżce dźwiękowej prezentowano, synchronicznie do obrazu, sylaby „ba-ba”. Badani w większości zgodnie stwierdzali, że w prezentowanej im próbce wizyjno-fonicznej słyszą sylaby „da-da” [93] [94]. Warto w tym miejscu przywołać również stwierdzenie polskiej psycholog, profesor Bogdanowicz. W jednej ze swoich książek napisała, że „stymulowanie jednych narządów zmysłu powoduje różnorodne zmiany w funkcjonowaniu innych, np. pod wpływem bodźców dźwiękowych zmieniają się progi wrażliwości i czułości wzrokowej, a określone oświetlenie sprawia, że dźwięki są lepiej słyszane i wydają się głośniejsze” [26].

W ciągu minionych dziesięcioleci przeprowadzono wiele badań, których wyniki wykazały, że informacja dostarczana przez zmysł wzroku, przy jednoczesnej stymulacji innych zmysłów (w szczególności słuchu i dotyku), jest dominująca. Badania przeprowadzone przez autora niniejszej rozprawy nawiązują do znanego zjawiska wpływu ob-

razu na percypowany kierunek dźwięku. Można zauważyć, że zjawisko wpływu obrazu na lokalizację pozornego źródła dźwięku znacznie częściej podlega badaniu niż wpływ bodźca dźwiękowego na lokalizację źródła dźwięku w obrazie. Wydaje się, że jednym z głównych powodów tej asymetrii jest fakt, że informacja o położeniu źródła dźwięku zawarta w bodźcu wzrokowym jest dokładniejsza [118]. Jednakże w ogólności, percepcja różnych bodźców zależy od typu analizowanej cechy bodźca i rodzaju aktywności. Zgodnie z hipotezą zaproponowaną przez Welcha i Warrena, rozbieżności pomiędzy różnymi bodźcami są rozstrzygane w zależności od tego, który z nich zawiera dokładniejszą i bardziej pożyteczną informację w kontekście danej sytuacji [140]. Na przykład zmysł wzroku zazwyczaj dominuje nad zmysłem słuchu w przypadku określania odległości, czyli w analizie położenia źródeł dźwięku w przestrzeni. Słuch natomiast dominuje nad wzrokiem w sytuacjach, gdy wymagana jest duża rozdzielczość czasowa [140].

Wpływ ściągający obrazu na percepcję dźwięku

W niniejszej rozprawie zbadany został wpływ kierunku patrzenia na lokalizację pozornego źródła dźwięku w panoramie stereofonicznej. W kontekście psychofizjologicznym zjawisko to wiąże się bezpośrednio z wpływem bodźca wzrokowego na percepcję bodźca słuchowego. Eksperymenty potwierdzające istnienie tego zjawiska przeprowadzane są w dwóch etapach. W pierwszym etapie badanym prezentowany jest jedynie bodziec słuchowy, zwany również próbką dźwiękową (ang. *unimodal stimulus*, *auditory-only trial*). W drugim zaś etapie następuje prezentacja bodźca wzrokowo-słuchowego, nazywanego próbką wizyjno-foniczną[s] (ang. *bimodal stimulus*, *visual-auditory trial*). Jednocześnie zakłada się, że próbka dźwiękowa wykorzystana w pierwszym i drugim etapie badania jest dokładnie taka sama. W obu etapach zadaniem badanych jest wskazanie położenia pozornego źródła dźwięku, związanego z prezentowanym bodźcem w panoramie stereofonicznej. Zauważono, że badani inaczej percypują położenie pozornego źródła dźwięku w przypadku, gdy prezentowana jest im tylko próbka dźwiękowa i inaczej, gdy próbce dźwiękowej towarzyszy obraz z nią związany. Zjawisko przesunięcia pozornego źródła dźwięku w kierunku bodźca wzrokowego nazywane jest powszechnie „wpływem ściągającym obrazu na percepcję dźwięku[s]". W literaturze anglojęzycznej „wpływ ściągający" jest określany jako „efekt zbliżenia obrazu" (ang.

image proximity effect) [47] [73] [74] [103] [104] lub „efekt brzuchomówstwa” (ang. *ventriloquism effect*) [4] [12] [13] [14] [15] [16] [18] [20] [118] [137] [144]. W związku ze stosunkowo często używanym w literaturze anglojęzycznej terminem “ventriloquism” autor podjął rozważanie nad słusznością stosowania tego terminu w języku polskim. Przypatrując się etymologii słowa „*ventriloquism*” (z łac. ‘*venter*’ + ‘*loqui*’ – „brzuch” + „mówić”) można wykazać, iż termin ten nie jest zgodny z definicją wpływu ściągającego obrazu na percepcję dźwięku[!]. „Brzuchomówstwo” sugeruje umiejscowienie źródła dźwięków w brzuchu, co może wprowadzać w błąd osoby niezaznajomione z tym pojęciem. W związku z powyższym autor sugeruje posługiwanie się znacznie częściej stosowanym określeniem „wpływu ściągającego obrazu na percepcję dźwięku”, które nie budzi wątpliwości związanych z rodzajem źródła dźwięku, ponieważ może nim być dowolny dźwięk, w tym również sygnał mowy.

Odpowiednią ilustracją zjawiska wpływu ściągającego obrazu na percepcję dźwięku może być eksperyment przeprowadzony przez Witkina w 1952 roku [141]. Badani biorący udział w eksperymencie wskazywali położenie pozornego źródła dźwięku (w tym przypadku głosu spikera telewizyjnego) w przestrzeni. Eksperyment składał się z dwóch etapów – w pierwszym badani widzieli twarz spikera i słyszeli jego głos, zaś w drugim słyszeli tylko jego głos. Wyniki eksperymentu wskazały, że pierwszym przypadku badani lokalizowali głos spikera jako „dochodzący ze środka”. Natomiast w przypadku, gdy mieli zamknięte oczy, słyszeli jego głos jako dochodzący z lewej lub prawej strony. Eksperyment przeprowadzony przez Witkina opisany został dokładniej w podrozdziale 4.2 niniejszej rozprawy. Wpływ ściągający obrazu na percepcję dźwięku występuje również w przypadku bodźców wzrokowo-słuchowych podczas oglądania innego rodzaju treści wizyjno-fonicznych, np. filmów [128]. Znaczna część próbek wizyjno-fonicznych, wykorzystana w przeprowadzonych w ramach niniejszej rozprawy badaniach, stanowi fragmenty filmów zrealizowanych w technice stereoskopowej (3D). Warto zaznaczyć, że wpływ kierunku patrzenia na percepcję położenia pozornego źródła dźwięku został udowodniony naukowo i jego istnienie jest bezsprzeczne [106] [111].

[!] W rozważaniach nad etymologią i znaczeniem słowa “*ventriloquism*” autor otrzymał wsparcie doc. dra inż. Gustawa Budzyńskiego, emerytowanego pracownika naukowego Katedry Systemów Multimedialnych Politechniki Gdańskiej.

Następstwo przesunięcia percepcji kierunku dźwięku

W literaturze opisano również bardzo interesujący aspekt wpływu ściągającego obrazu na percepcję dźwięku. Zauważono, że zmiana w percepcji kierunku dźwięku może trwać po zakończeniu projekcji próbki wizyjno-fonicznej (ang. *ventriloquism aftereffect*) [89] [112] [116] [142]. Jeżeli przez okres kilkudziesięciu minut osobie badanej prezentowana jest próbka wizyjno-foniczna, w której przesunięcie pozornego źródła dźwięku pomiędzy bodźcem wzrokowym i słuchowym jest stałe, a następnie badana jest lokalizacja pozornego źródła dźwięku bez towarzyszenia obrazu, to osoba badana przez pewien czas wskazuje położenie pozornego źródła dźwięku tak, jakby prezentowany był jej bodziec wzrokowo-słuchowy. Oznacza to, że percepcja dźwięku zostaje w pewien sposób zmieniona w wyniku stymulacji bodźcem wzrokowym. Zjawisko to może wystąpić, gdy pozorne źródło dźwięku w bodźcu wzrokowym jest nieznacznie przesunięte względem pozornego źródła dźwięku w bodźcu słuchowym, a próbka wizyjno-foniczna jest percypowana z tego samego miejsca [116] lub gdy przesunięcie pozornego źródła dźwięku jest stosunkowo duże [89]. Nie zbadano dokładnie, jak długo może trwać to zjawisko. Przyjmuje się jednak, że „przesunięcie percepcji pozornego źródła dźwięku" trwa do czasu, kiedy osoba badana wróci do „normalnego" otoczenia, a bodźce słuchowe i wzrokowe będą zawierały spójną informację o położeniu pozornego źródła dźwięku. Warto zauważyć, że „następstwo przesunięcia percepcji kierunku dźwięku" ma zupełnie inny charakter w porównaniu z pozostałymi zjawiskami, które są wynikiem adaptacji centralnego układu nerwowego [54] [135]. Zjawisko nazywane w literaturze anglojęzycznej „*ventriloquism aftereffect*" może trwać kilkadziesiąt minut, a przesunięcie percypowanego kierunku dźwięku jest zgodne z kierunkiem przesunięcia w prezentowanym bodźcu adaptującym [45] [117].

W niniejszym rozdziale opisane zostały charakterystyki widzenia i słyszenia, istotne w kontekście zrozumienia procesów zachodzących podczas percepcji obrazów i dźwięków oraz zjawisko percepcji wielomodalnej bezpośrednio związane z badaniami przeprowadzonymi przez autora niniejszej rozprawy.

W kolejnym rozdziale w stosunkowo szerokim zakresie przedstawiono badania korelacji wzrokowo-słuchowych.

3 Przegląd badań w dziedzinie korelacji wzrokowo-słuchowych

Wpływ obrazu na percepcję dźwięku jest zjawiskiem znanym naukowcom od wielu lat. Historia badań nad korelacjami wzrokowo-słuchowymi sięga końca XIX wieku. Prekursorem badań w tej dziedzinie był Stratton, którego eksperymenty opisał Meares w swoim referacie [95]. W eksperymencie przeprowadzonym w 1896 roku Stratton wykazał, że bodźce wzrokowe mają istotny wpływ na lokalizację źródła dźwięku w przestrzeni. Obecnie można lepiej zrozumieć proces percepcji bodźców słuchowych i wzrokowych na przykład dzięki obrazowaniu fMRI (ang. *Functional Magnetic Resonance Imaging*) [139] czy pozytonowej, emisyjnej tomografii komputerowej (ang. *Positron Emission Tomography* – PET). Jednak w dalszym ciągu zjawiska zachodzące w korze słuchowej i wzrokowej pozostają nie do końca poznane. Warto odnieść się do wyników eksperymentu naukowców, którzy postanowili wyznaczyć przepływność informacji „transmitowanej” przez poszczególne zmysły. Podano przybliżoną liczbę komórek nerwowych związanych z każdym zmysłem. Następnie, znając przepływność pojedynczej komórki, oszacowali „szerokość pasma” poszczególnych zmysłów. Wyniki uzyskane w tym eksperymencie zestawiono w tab. 3.1 [150].

Tab. 3.1 Przepływność zmysłów

zmysł	**przepływność [b/s]**
wzrok	10 000 000
słuch	100 000
dotyk	1 000 000
smak	1 000
zapach	100 000
w sumie:	**11 201 000**

Jak wynika z powyższego zestawienia, stosunek ilości informacji przetwarzanej przez zmysł wzroku i słuchu wynosi 100:1. Co prawda, nie wszystkie informacje dostarczane przez zmysły są przez człowieka świadomie percypowane. W wyniku skom-

plikowanych badań oceniono, iż człowiek odbiera jedynie ok. 40 bit/s z ponad dostarczanych 11 milionów bitów na sekundę. W kontekście percepcji treści wizyjno-fonicznej wydaje się, że to podświadomość odgrywa istotną rolę. Mając na uwadze powyżej przedstawione dane, można zauważyć, że w procesie percepcji obrazu i dźwięku muszą brać udział informacje przechowywane w pamięci. Jednak do ich wydobycia konieczny jest odpowiedni bodziec, który zainicjalizuje proces przetwarzania. W filmie czy ogólnie pojętym materiale wizyjno-fonicznym, przyjmuje się, że takim bodźcem jest dźwięk. Za przykład wyjaśniający to zjawisko może posłużyć scena przedstawiająca słabo oświetlone pomieszczenie z towarzyszącym dźwiękiem, np. krzykiem. U większości widzów taka scena wyzwoli poczucie niepokoju, a nawet strachu [150].

W prowadzonych w ramach niniejszej rozprawy badaniach założono, iż osoby badane lokalizują źródło dźwięku w panoramie stereofonicznej nie tylko w oparciu o informacje, których są świadome, ale również, a może przede wszystkim – w oparciu o informacje, które są przetwarzane przez mózg poza ich świadomością. Przykładem tego zjawiska może być opisany w podrozdziale 2.3 efekt McGurka. Znaczenie faktu, w jaki sposób bodźce słuchowe i wzrokowe wpływają na siebie, oraz w jaki sposób ta interakcja przekłada się na percepcję informacji zawartych w tych bodźcach jest dość istotne. Dzięki bliższemu poznaniu natury tego zjawiska możliwe stanie się przygotowanie treści wizyjno-fonicznej w taki sposób, aby zakres bodźców percypowanych przez widza oglądającego film był szerszy w porównaniu z filmem zrealizowanym według dotychczas znanych reguł. W tym kontekście badania korelacji wzrokowo-słuchowych wpisują się w nurt prac mających na celu poznanie czynników wpływających na wielkość parametrów QoE^s (ang. *Quality of Experience*) oraz $QoMEX^s$ (ang. *Quality of Multimedia Experience*), które określają jakość prezentowanej treści multimedialnej w ocenie widza bądź użytkownika systemu [46]. Warto zauważyć, że bliższe poznanie wzajemnych zależności pomiędzy dźwiękiem i obrazem (wpływające na wielkość QoMEX) może mieć również istotne znaczenie w przypadku projektów takich jak na przykład Laboratorium Zanurzonej Wizualizacji Przestrzennej, powstające obecnie w Politechnice Gdańskiej [179].

Możliwe jest udowodnienie, że człowiek koncentrując wzrok na wybranym fragmencie obrazu wyświetlanego na ekranie, lokalizuje źródło dźwięku w panoramie ste-

reofonicznej, sugerując się jego umiejscowieniem w obrazie. W takim właśnie kontekście rozumiane są w niniejszej rozprawie korelacje wzrokowo-słuchowe. W badaniach podjęto również **aspekt skalowalności wpływu ściągającego**, polegający na zbadaniu wpływu obrazu na percepcję dźwięku w przypadku prezentacji treści wizyjnej w obszarze wyświetlania o różnej wielkości. Ocenie badanych poddano próbki testowe w różnych konfiguracjach, z zastosowaniem obrazu dwuwymiarowego oraz obrazu trójwymiarowego z towarzyszeniem dwukanałowego dźwięku stereofonicznego.

W niniejszym rozdziale dokonano przeglądu wyników badań korelacji wzrokowo-słuchowych opisanych w literaturze [1] [9] [10] [11] [17] [29] [32] [35] [47] [56] [57] [66] [70] [71] [86] [87] [88] [90] [91] [92] [95] [100] [101] [113] [114] [115] [122] [123] [131] [141] [175] i dodatkowo przeprowadzonych w Katedrze Systemów Multimedialnych Politechniki Gdańskiej [39] [73] [74] [79] [80] [81] [82] [103] [104].

Badania korelacji wzrokowo-słuchowych były i w dalszym ciągu są prowadzone w różnych obszarach nauki. Autor niniejszej rozprawy zaproponował klasyfikację badań korelacji wzrokowo-słuchowych w zależności od kontekstu ich przeprowadzania. W pierwszym podejściu korelacje wzrokowo-słuchowe można rozumieć jako dopasowanie bodźca słuchowego do bodźca wzrokowego w dziedzinie czasu, a zatem można je badać w kontekście synchronizacji dźwięku i obrazu [1] [56] [87] [91] [126]. Drugie podejście, chronologicznie najstarsze, jest związane z wpływem bodźca wzrokowego na percepcję bodźca słuchowego w kontekście lokalizacji źródła dźwięku [10] [11] [17] [29] [47] [66] [70] [71] [88] [90] [92] [95] [101] [141] [113] [122] [123] [175]. Korelacje wzrokowo-słuchowe można badać również w kontekście wpływu prezentowanej widzowi treści wizyjno-fonicznej na odbierane przez niego wrażenia (w kontekście dziedziny *Quality of Experience*) [9] [35] [57] [131]. Wreszcie, badania korelacji wzrokowo-słuchowych wpisują się w rozwój metod kompresji obrazu wideo, opartych na informacji zawartej w sygnale dźwiękowym [32] [86] [100] [114] [115]. Chronologicznie ostatnie podejście jest najmłodsze. W związku z powyższym, opracowane w ramach pierwszych eksperymentów metody kodowania obrazu nie są jeszcze powszechnie stosowane.

W ramach niniejszej rozprawy przeprowadzono badania korelacji wzrokowo-słuchowych w kontekście wpływu obrazu na lokalizację źródła dźwięku w panoramie stereofonicznej. Zaproponowana przez autora rozprawy metoda prowadzenia badań, polegająca na zastosowaniu systemu śledzenia punktu fiksacji wzroku, powinna prowadzić do obiektywizacji uzyskanych wyników.

3.1 Kontekst synchronizacji dźwięku i obrazu

W latach 40. XX wieku w *Bell Laboratories* w Stanach Zjednoczonych przeprowadzono badania w wyniku, których wyznaczono wartości progów czasowych powyżej i poniżej których odczuwalne jest przesunięcie w czasie i obrazu. Zauważono, że widz percypuje rozsynchronizowanie pomiędzy dźwiękiem i obrazem, gdy dźwięk wyprzedza obraz o więcej niż 35 ms, oraz gdy występuje po nim o więcej niż 100 ms [165]. W ciągu kolejnych lat wielu naukowców przeprowadziło podobne badania. Konkluzje wynikające z przeprowadzonych przez nich badań można sprecyzować następująco [87]:

- percypowane progi desynchronizacji dźwięku i obrazu ściśle zależą od osoby badanej,
- progi desynchronizacji są asymetryczne – rozsynchronizowanie jest łatwiej percypowane, gdy dźwięk wyprzedza obraz,
- zakładając najbardziej kontrolowane warunki, zjawisko rozsynchronizowania dźwięku i obrazu jest percypowane, gdy dźwięk wyprzedza obraz o 75 ms i gdy jest opóźniony w stosunku do obrazu o 90 ms.

W 1998 r. organizacja standaryzująca ITU (ang. *International Telecommunication Union*) opracowała normę ITU-R[1] BT[2].1359, która określiła wartości progowe czasu opóźnień dźwięku i obrazu przy transmisji. Oszacowano, że przesunięcie w czasie ścieżki dźwiękowej w stosunku do ścieżki obrazu jest wykrywalne dla +45 ms (gdy dźwięk wyprzedza obraz) i -125 ms (gdy dźwięk jest opóźniony względem obrazu) [62].

[1] ITU-R – ang. *International Telecommunication Union – Radiocommunication Sector*
[2] BT – ang. *Broadcasting Service (Television)*

Z kolei inna organizacja – ATSC[3] opracowała normę IS-191, która określa, że dźwięk nie powinien wyprzedzać obrazu o więcej niż 15 ms i nie powinien być opóźniony w stosunku do niego o więcej niż 45 ms [59].

Przesunięcie dźwięku w stosunku do obrazu można nie tylko mierzyć, ale również modyfikować. Wielu badaczy podjęło się opracowania inteligentnych metod analizy obrazu i dźwięku, mających na celu kompensację odczuwalnej desynchronizacji wizji i fonii. Najczęściej przedmiotem badania był obraz twarzy mówiącego bohatera. Liu i Sato opracowali metodę dopasowywania w dziedzinie czasu ścieżki dźwiękowej i wizyjnej analizując ruchy ust bohatera [91]. Podobne badania przeprowadził ze swoimi współpracownikami Abel. Zaproponowana przez niego metoda zwiększenia współczynnika korelacji wzrokowo-słuchowej, rozumianego jako stopień zsynchronizowania dźwięku i obrazu, była oparta na detekcji samogłosek w sygnale mowy oraz śledzeniu obszaru ust w obrazie [1]. Inne badania, przeprowadzone przez Holliera potwierdziły, iż prawdopodobieństwo poprawnego wykrycia błędu synchronizacji dźwięku i obrazu zależy od treści zawartej w próbce wizyjno-fonicznej, a dokładniej od typu bodźca wzrokowo-słuchowego. Na przykład, bardziej prawdopodobne jest wykrycie desynchronizacji dźwięku i obrazu w przypadku próbki prezentującej spadający długopis czy topór uderzający o podłogę, niż w przypadku nagrania osoby mówiącej [56].

W następnym podrozdziale dokonano przeglądu badań nad korelacjami wzrokowo-słuchowymi w kontekście wpływu bodźca wzrokowego na lokalizację pozornego źródła dźwięku w panoramie stereofonicznej.

3.2 Kontekst lokalizacji źródła dźwięku

Badania przeprowadzone w ramach niniejszej rozprawy odnoszą się w głównej mierze do opisanego w tym podrozdziale podejścia. Badając wpływ obrazu na lokalizację pozornego źródła dźwięku, szczególnie w wielokanałowej panoramie stereofonicznej, należy pamiętać o kątowej rozdzielczości ludzkiego słuchu w płaszczyźnie horyzontalnej. Słuch człowieka jest bowiem bardziej czuły na dźwięk i dochodzące z kierunku przedniego niż na dźwięk z kierunków bocznych. Warto zaznaczyć, iż kierunek

[3] ATSC – ang. *Advanced Television Systems Committee*

przedni rozumie się jako obszar zawarty w przedziale kątowym (-59^o, $+59^o$), kierunki boczne: (-60^o, -119^o) oraz ($+60^o$, 119^o), natomiast kierunek tylny – to obszar w przedziale (-120^o, $+120^o$). Rozdzielczość ludzkiego słuchu można wyznaczyć w oparciu o metodę najmniejszego przesunięcia źródła dźwięku. Najmniejsze słyszalne przesunięcie wyznacza rozdzielczość kątową ludzkiego słuchu[s]. Często wartość tego przesunięcia określa się mianem minimalnego kąta słyszenia (ang. *Minimum Audible Angle* – MAA) [99]. Mills, Hartman i Grantham wykazali w swoich pracach, że optymalne warunki odsłuchowe, biorąc pod uwagę parametr MAA, są spełnione, gdy źródło dźwięku jest umieszczone bezpośrednio przed słuchaczem [49] [53] [97]. Przy spełnieniu tych warunków możliwe jest wykrycie przesunięcia źródła dźwięku nawet o 1^o. Blauert zbadał, że rozdzielczość przestrzenna słuchu w kierunkach bocznych może być od trzech do dziesięciu razy mniejsza w porównaniu z rozdzielczością słuchową w kierunku przednim i w przybliżeniu dwa razy słabsza w porównaniu do kierunku tylnego [23]. Spostrzeżenia te są istotne z perspektywy przeprowadzenia analizy i interpretacji wyników badań przeprowadzonych w ramach niniejszej pracy.

Prekursorem badań korelacji wzrokowo-słuchowych, rozumianych jako wpływ bodźca wzrokowego na percepcję kierunku dźwięku związanego z tym bodźcem, był Stratton [95]. W jego pierwszym eksperymencie badani lokalizowali źródła dźwięku na obrazie odwróconym w płaszczyźnie pionowej. Wyniki tego badania wskazały, że wrażenia słuchowe pochodzące od źródeł dźwięku znajdujących się w polu widzenia obserwatora były lokalizowane zgodnie z ich położeniem w obrazie, czyli odwrotnie w stosunku do ich rzeczywistego położenia [95]. Potwierdzeniem badań Strattona stały się obserwacje Klemma [66]. W przeprowadzonym przez niego eksperymencie badani percypowali dźwięk przyporządkowanych do głośników młoteczków. Głośniki były umieszczone odpowiednio po lewej i prawej stronie badanego. Wskazania lokalizacji źródeł dźwięku w przypadku, gdy badani mieli zamknięte oczy, nie pokrywały się ze wskazaniami, gdy badani mogli obserwować ruch młoteczków. Eksperyment ten pozwolił Klemmowi sformułować tzw. prawo komplikacji przestrzennej, głoszące, że pobudzenie zmysłów różnymi bodźcami powoduje „zlewanie" się wrażeń odbieranych przez człowieka [66]. W roku 1941 Thomas udowodnił, że bodźce wizualne nie muszą być ściśle związane z towarzyszącymi im bodźcami akustycznymi. W swoich badaniach

Thomas wykorzystał światło lampy i dźwięk dzwonka [134]. Wspomniany w podrozdziale 2.3 eksperyment przeprowadzony przez Witkina miał na celu zbadanie wpływu obrazu twarzy spikera na lokalizację jego głosu w panoramie stereofonicznej. W badaniu wykorzystano tzw. pseudofon[s]. Wrażenie zmiany położenia głosu w płaszczyźnie horyzontalnej uzyskano dzięki zastosowaniu dwóch mosiężnych teleskopowych rur. Długość rur mogła się zmieniać, dzięki czemu dźwięk dochodził do obu uszu z różnymi opóźnieniami. Badani przebywali w komorze dźwiękoszczelnej, natomiast obraz twarzy spikera przesyłano do pomieszczenia badawczego przez okno znajdujące się za ich plecami. Eksperyment Witkina podzielony był na dwa etapy. W pierwszym etapie badani mieli otwarte oczy, zatem percypowali zarówno bodźce wzrokowe, jak i słuchowe. W drugim etapie natomiast badanym zaprezentowano ścieżkę dźwiękową i poproszono ich o zamknięcie oczu. Badani, patrząc na twarz spikera w lustrze, postawionym ok. 0,5 m przed twarzą, lokalizowali jego głos jako słyszany „ze środka”, na wprost. Inaczej natomiast ci sami badani lokalizowali położenie źródła dźwięku w panoramie stereofonicznej w sytuacji, gdy mieli zamknięte oczy. Wówczas głos spikera lokalizowali zgodnie z rzeczywistością – jako dochodzący odpowiednio z lewej i prawej strony spoza obszaru zajmowanego przez odbicie obrazu w lustrze [141]. Badania Witkina dowiodły, że bodźce wzrokowe mają bardzo duży wpływ na postrzeganie kierunku dźwięku. W latach 50. Held ponownie zwrócił szczególną uwagę na ten aspekt w badaniu korelacji wzrokowo-słuchowych [10]. Podobnie jak Witkin, Held również wykorzystał pseudofon. Badani mieli nałożone słuchawki, na które wysyłano sygnały z dwóch mikrofonów, które ustawiono po przeciwległych stronach głowy. Held założył, że źródłem dźwięku będą krótkie impulsy elektryczne. Impulsy emitował głośnik ustawiony na linii wzroku osoby badanej. Wyniki tego eksperymentu potwierdziły, że w przypadku, gdy badani mieli otwarte oczy, nie odczuwali różnicy w położeniu źródła sygnału (głośnika) w bazie stereofonicznej w stosunku do sygnału słyszanego w słuchawkach. Natomiast w sytuacji, gdy badani mieli zamknięte oczy, dźwięk odbierali z kierunków przesuniętych w stosunku do prawidłowych o kąt rozsunięcia mikrofonów względem osi uszu [10].

Ze względu na podjęty w niniejszej rozprawie wątek badania wpływu obrazu trójwymiarowego na percepcję dźwięku nie można pominąć badań przeprowadzonych

przez Gardnera w latach 60. [47]. Postanowił on bowiem jako pierwszy zbadać percepcję głębi dźwięku w kontekście korelacji wzrokowo-słuchowych. Na linii wzroku osoby badanej ustawił pięć głośników, jeden za drugim tak, że pierwszy całkowicie zasłaniał cztery pozostałe. Badany słuchał dźwięku odtwarzanego przez losowo wybrany głośnik. Wyniki eksperymentu wyraźnie wskazały, iż położenie głośnika w płaszczyźnie przód-tył nie miało większego znaczenia, ponieważ badani percypowali dźwięk emitowany przez różne głośniki tak, jakby pochodził on z głośnika, na który patrzyli [47]. Można zatem uznać, że badanie Gardnera zapoczątkowało wątek badań korelacji wzrokowo-słuchowych, związany z lokalizowaniem źródła dźwięku w płaszczyźnie przód-tył.

W latach 70. Blauert poszerzył zakres badań nad wpływem obrazu spikera na percepcję kierunku jego głosu [11]. Przeprowadził dwa eksperymenty związane z wpływem ekspozycji obrazu telewizyjnego na lokalizację źródła dźwięku w płaszczyźnie horyzontalnej i wertykalnej. Dźwiękiem testowym w obu badaniach były logatomy[s] wypowiadane przez spikera. Badani znajdowali się w odległości 7 m od monitora. W pierwszym eksperymencie wykorzystane zostały dwa monitory – jeden powyżej linii wzroku badanego, drugi poniżej. Na wysokości wzroku (na osi 0^{o}) znajdowało się pięć głośników rozmieszczonych w linii jeden obok drugiego, oddzielających jednocześnie oba monitory. W doświadczeniu pierwszym podawano dźwięk w losowej kolejności na każdy z pięciu głośników we wszystkich trzech poniższych konfiguracjach [11]:

- przy włączonym monitorze dolnym i wyłączonym górnym,
- przy włączonym monitorze górnym i wyłączonym dolnym,
- przy wyłączonych monitorach.

W każdej z wymienionych konfiguracji nagrane logatomy podawano pięciokrotnie na każdy z głośników. Zadaniem badanych było zaklasyfikowanie kierunku percypowanego dźwięku do jednej z dwóch kategorii: „lewy" lub „prawy". W drugim eksperymencie wykorzystano jeden monitor, umieszczony na linii wzroku badanego oraz trzy głośniki jeden nad drugim powyżej monitora i trzy głośniki jeden pod drugim poniżej monitora. Zadaniem badanych w tym doświadczeniu było lokalizowanie dźwięku w płaszczyźnie wertykalnej, polegające na przyporządkowaniu kierunku dźwięku do jednej z dwóch kategorii: „góra" lub „dół". Tę część badania przeprowadzono dla dwóch konfiguracji:

- przy włączonym monitorze,
- przy wyłączonym monitorze.

Analiza statystyczna wyników uzyskanych dla obu eksperymentów nie wykazała istotnych różnic przy porównaniu ocen badanych w konfiguracji z obrazem i bez obrazu. W badaniu Blauerta nie stwierdzono zatem znaczącego wpływu telewizyjnego obrazu spikera na zmianę percepcji kierunku jego głosu. Niemniej jednak interpretacja wyników pozwoliła wysnuć ważny wniosek dotyczący przygotowania treści materiału wizyjno-fonicznego poddanego badaniu. Okazało się bowiem, że wpływ obrazu na percepcję dźwięku silnie zależy od tego, w jakim stopniu treść obrazu stymuluje uwagę widza. Jest to ważne spostrzeżenie, które należy brać pod uwagę przy przygotowywaniu materiałów do badań korelacji wzrokowo-słuchowych.

Interesujące badania nad wpływem obrazu na percepcję dźwięku na przełomie lat 70. i 80. minionego stulecia prowadzili Radeau i Bertelson [17] [113]. Wykazali oni nie tylko istnienie wpływu ściągającego, ale również badali naturę tego zjawiska w zależności od prezentowanych bodźców oraz od wpływu adaptacji (przyzwyczajenia się) badanego na mierzalny wpływ ściągający. W swoich eksperymentach stymulowali badanych między innymi bodźcem wzrokowym w postaci światła modulowanego dźwiękiem [17] [113]. Współcześnie wykorzystuje się technikę jednoczesnej stymulacji bodźcem świetlnym i skorelowanym z nim dźwiękiem w celu zwiększenia wydajności umysłu w procesie uczenia się. W ramach dygresji warto wspomnieć o dostępnym na rynku urządzeniu o nazwie SITA. SITA przetwarza oddech człowieka na bodźce świetlne i dźwiękowe. Człowiek, odbierając te bodźce, podświadomie reguluje i wyrównuje częstotliwość i głębokość własnego oddechu. W ten sposób powstaje tzw. biofeedback[s], czyli sprzężenie zwrotne pomiędzy oddechem a wrażeniami wzrokowymi i słuchowymi, co sprzyja synchronizacji półkul mózgowych. Ta z kolei wpływa na kilkukrotny wzrost zdolności mózgu do percepcji i zapamiętywania informacji.

W latach 80. XX wieku przeprowadzono wiele badań nad wzajemnym wpływem obrazu i dźwięku na potrzeby telewizji. Sakamoto i Brook są uważani za prekursorów badań w tej dziedzinie [29] [122] [123]. Istotny wkład w badania korelacji wzrokowo-słuchowych w kontekście zastosowań telewizyjnych wniosła również japońska sieć telewizyjna NHK. W laboratoriach sieci NHK Komiyama i Nakabayashi [71] badali

wpływ obrazu wielkoformatowego na percepcję dźwięku w płaszczyźnie pionowej. W badaniach tych wykorzystano płaski ekran o przekątnej 72", stanowiący wówczas rozwiązanie wysoce zaawansowane technologicznie. Badania Komiyamy i Nakabayashiego dowiodły, że istnieje silny wpływ usytuowania obrazu na lokalizację dźwięku w kierunku pionowym. Wnioski z tych badań znajdują zastosowanie w realizacjach telewizyjnych wykorzystujących monitory wielkoformatowe [71]. W następnych eksperymentach Komiyama i Nakabayashi podjęli się zbadania wpływu ściągającego w zależności od płci osoby badanej. Badania wykazały, iż płeć badanego determinuje jego ocenę lokalizacji źródła dźwięku w panoramie stereofonicznej. Zdecydowanie większy wpływ ściągający występował, gdy mężczyźnie prezentowano obraz spikerki, a kobiecie obraz spikera, niż w przypadku, gdy badany i spiker byli tej samej płci [70]. Można zatem założyć, iż zjawisko wpływu ściągającego obrazu na percepcję dźwięku ma podstawy psychologiczne i w pewnej mierze jest wynikiem indywidualnych uwarunkowań i upodobań.

W 2003 roku Nakayama przeprowadził interesujący eksperyment z wykorzystaniem obrazu stereoskopowego (3D) i dźwięku przestrzennego. Celem badania było zdefiniowane wymagań dla szerokopasmowej telewizji 3D. Nakayama zaproponował metodę sterowania obrazem dźwięku przestrzennego emitowanego przez macierz głośników (ang. *loudspeakers array*), która dowolnie mogła sterować położeniem obrazu dźwiękowego. Przeprowadzono dwa rodzaje testów subiektywnych: badających percepcję odległości[s] (ang. *distance perception*) oraz percepcję kierunku[s] (ang. *directional perception*). Takie podejście do oszacowania przez badanego odległości i kierunku pozornego źródła dźwięku zostało również zastosowane w przeprowadzonych w ramach rozprawy eksperymentach. Nakayama udowodnił, że na percypowany przestrzenny obraz dźwiękowy bezpośrednio wpływa prezentowany obraz wideo 3D [101]. Z kolei Majdak wraz ze współpracownikami zbadał lokalizowanie pozornego źródła dźwięku w środowisku wirtualnym [92].

Zaznaczyć należy, iż także w Katedrze Systemów Multimedialnych na przestrzeni ostatniej dekady prowadzone były badania z zakresu korelacji wzrokowo-słuchowych. W głównej mierze koncentrowały się one na badaniu wpływu ściągającego obrazu na percepcję dźwięku. W jednej z serii przeprowadzonych eksperymentów

do analizy uzyskanych wyników wykorzystano algorytm genetyczny, co w kontekście badań prowadzonych przez inne ośrodki naukowe stanowiło innowacyjne podejście [73] [74] [103] [104].

We wszystkich przedstawionych powyżej badaniach, wpływ bodźca wzrokowego na lokalizację pozornego źródła dźwięku, nie uwzględniał kierunku, w którym badany koncentrował swój wzrok. W literaturze można znaleźć publikacje Lewalda, w których opisana została metodyka przeprowadzania badań korelacji wzrokowo-słuchowych z jednoczesną analizą kierunku patrzenia osoba badanej. Na potrzeby eksperymentów Lewald skonstruował stanowisko badawcze, w skład którego wchodziło 9 głośników, leżących na okręgu o promieniu 3,35 m (jeden na wprost osoby badanej, cztery po lewej i cztery po prawej stronie, przy czym każdy głośnik był od siebie odseparowany o $2{,}75^{o}$) oraz 5 diod leżących na okręgu o promieniu 1,3 m, gdzie każda dioda była od siebie oddalona o wartość kąta $22{,}5^{o}$. Założono, że diody leżące na linii wzroku badanego, a jednocześnie w obszarze odsłuchowym nie wpływały na odbiór dźwięków emitowanych przez głośniki. Informacja o kierunku patrzenia była bezpośrednio związana z położeniem diody LED w płaszczyźnie horyzontalnej. Badani zostali poinstruowani, aby koncentrowali wzrok na diodzie, która w danym momencie emitowała światło. Wyniki badania Lewalda wykazały, iż przesunięcie położenia percypowanego źródła dźwięku w stosunku do rzeczywistego źródła dźwięku nie zawsze było takie samo. Część badanych oceniła, iż słyszała dźwięk w kierunku przeciwnym do kierunku patrzenia. Z ocen pozostałych uczestników eksperymentów wynikało, że pozorne źródło dźwięku przesuwało się w kierunku, w którym badani koncentrowali wzrok [88] [90].

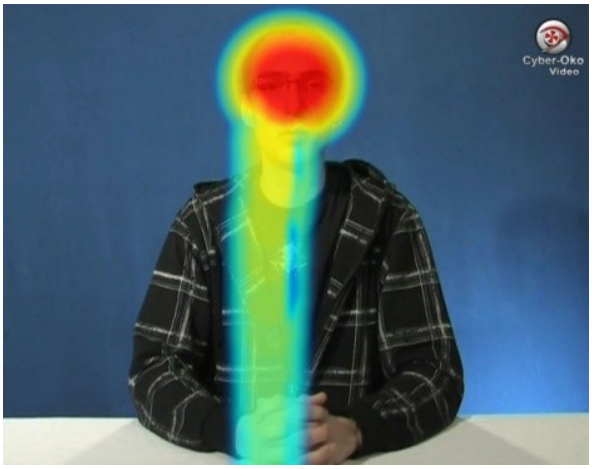
Należy zaznaczyć, że system śledzenia wzroku był już wykorzystywany w badaniu korelacji wzrokowo-słuchowych [119]. Niemniej jednak podejście zaproponowane przez autora rozprawy różni się od założeń przyjętych przez Rordena, opisującego wykorzystanie technik śledzenia wzroku w swoich eksperymentach [119]. Założenia badań prowadzonych przez autora różnią się w odniesieniu do badań Rordena w dwóch najważniejszych kwestiach. Po pierwsze, w zastosowanym systemie śledzenia wzroku – autor wykorzystał dwa bezkontaktowe systemy śledzenia wzroku, zaś Rorden – nagłowny system (ang. *head-mounted eye tracking systems*), montowany na głowie osoby badanej. Po drugie, Rorden przeprowadził badania, w których bodźcem wzrokowym by-

ło światło emitowane przez diody LED. Autor rozprawy natomiast wykorzystał w swoich eksperymentach materiał badawczy oparty przede wszystkim na fragmentach rzeczywistych filmów, prezentowanych w konwencjonalny sposób (film 2D) oraz w technice anaglifowej (film 3D). Wobec powyżej przedstawionych różnic można uznać, iż zaproponowana przez autora rozprawy metodologia prowadzenia badań korelacji wzrokowo-słuchowych w kontekście lokalizacji pozornego źródła dźwięku jest oryginalna i nowatorska.

Warto wspomnieć ponadto, iż autor rozprawy prowadził badania korelacji wzrokowo-słuchowych z wykorzystaniem systemu śledzenia wzroku na wcześniejszym etapie prac nad niniejszą rozprawą [79] [80] [81]. W eksperymencie badającym wpływ kierunku patrzenia na percepcję dźwięku w płaszczyźnie horyzontalnej posłużono się systemem śledzenia punktu fiksacji wzroku Cyber-Oko [80]. Koncepcję badania oparto na założeniach eksperymentu Witkina [141], o którym już wspomniano w niniejszym podrozdziale. W badaniu wykorzystano próbki, zawierające obraz twarzy spikera, znajdującego się w centrum kadru. Zawartość obrazu nie ulegała zmianie, jednak zmieniało się położenie pozornego źródła dźwięku w bazie stereofonicznej. Badani wypełniali formularz ankiety wskazując kierunek, z którego dochodził percypowany dźwięk: lewy, lewy-środek, środek, prawy-środek, prawy. Eksperyment składał się z dwóch etapów, w każdym z nich badani oceniali położenie źródła dźwięku w panoramie stereofonicznej. W wyniku eksperymentu wygenerowane zostały tzw. „dynamiczne mapy przejść"[s] (ang. *dynamic gaze plot*) naniesione na prezentowany materiał wizyjny. Dynamiczna mapa przejść stanowi wizualizację aktywności wzrokowej widza podczas projekcji próbki. Zatem odzwierciedla ona skupienie wzroku widza w czasie rzeczywistym. Warto odróżnić dynamiczną mapę przejść od tzw. mapy ciepła[s] (ang. *heat map*), która wskazuje na fragmenty prezentowanego obrazu, na których osoba badana najczęściej koncentrowała wzrok. Mapa ciepła, w odróżnieniu od dynamicznej mapy przejść, jest generowana za określony przedział czasu.

W pierwszym etapie badanym prezentowana była tylko ścieżka dźwiękowa próbki filmowej, w drugim zaś – ścieżka dźwiękowa z towarzyszeniem obrazu wizyjnego. Dodatkowo, w drugim etapie badania w tle pracował system śledzenia punktu fiksacji wzroku, rejestrujący położenie wzroku widza na konkretnych obszarach prezentowa-

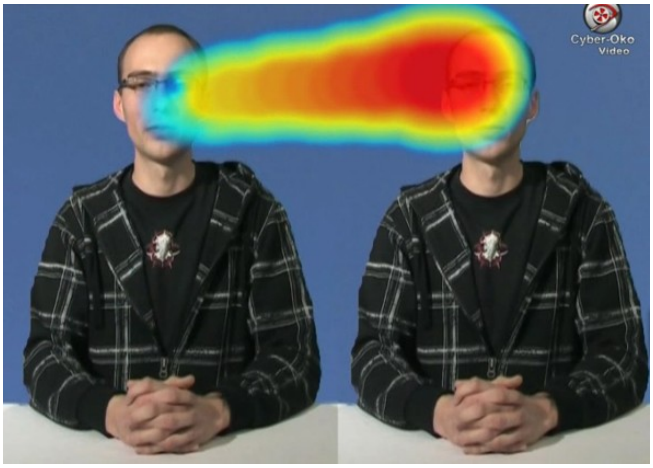
nego obrazu. Przeprowadzony eksperyment wykazał, iż badani percypowali położenie pozornego źródła dźwięku tak, jakby znajdowało się ono pomiędzy rzeczywistym źródłem dźwięku a PoR. Badania potwierdziły zatem istnienie wpływu ściągającego obrazu na percepcję dźwięku. Przykładową klatkę filmu z naniesioną dynamiczną mapą przejść przedstawiono na rys. 3.1 Dokładniejszy opis przeprowadzonego eksperymentu znajduje się w artykule zatytułowanym „Exploiting audio-visual correlation by means of gaze tracking” [80].



Rys. 3.1 Klatka filmu będącego wynikiem badania; naniesiona dynamiczna mapa przejść oznacza położenie wzroku badanego

W kolejnym etapie autor rozprawy rozwinął opisane powyżej badanie, zmieniając treść prezentowanego materiału wizyjno-fonicznego. Przygotowano dwa rodzaje próbek wizyjnych: obraz z podwojoną postacią spikera – próbki przygotowano na etapie postprodukcji, umieszczając skopiowaną postać spikera obok postaci oryginalnej – oraz próbki przedstawiające typową rozmowę dwóch osób. Eksperymenty, w których wykorzystano oba zestawy próbek, wykonano zgodnie z metodyką badania, które przeprowadzono wcześniej z wykorzystaniem obrazu spikera znajdującego się w centrum kadru. Przygotowane materiały wizyjno-foniczne różniły się między sobą jedynie lokalizacją głosu spikera lub rozmówcy w panoramie stereofonicznej. Okazało się, że w większości przypadków badani koncentrowali wzrok na twarzy spikera/rozmówcy znajdującego się po tej stronie kadru, która bezpośrednio korespondowała z kierunkiem dochodzenia dźwięku. Wygenerowane przez system śledzenia wzroku dynamiczne mapy potwierdzają to zjawisko w sposób obiektywny. Na rys. 3.2 pokazano przykładową klatkę wynikowego filmu z naniesioną na obraz dynamiczną mapą przejść.

Przedstawione powyżej badania zostały opisane w artykule pt. „Objective audio-visual correlation methodology" [81].



Rys. 3.2 Klatka filmu przedstawiającego podwojoną postać spikera; naniesiona dynamiczna mapa przejść oznacza położenie wzroku badanego

W opisie badań korelacji wzrokowo-słuchowych w kontekście lokalizowania źródła dźwięku w obrazie należy odnieść się również do wstępnych wyników projektu badawczego DIEM z 2010 roku [175]. Naukowcy postanowili zbadać aktywność wzrokową widzów podczas oglądania filmów za pomocą systemu śledzenia wzroku. Jednak zaproponowane przez nich podejście znacząco różniło się od tego, w jaki sposób w swoich badaniach techniki śledzenia wzroku wykorzystał autor rozprawy. Ich badania bowiem koncentrowały się wokół wyjaśnienia procesu percepcji obrazu przez człowieka oraz tego, jakie czynniki w procesie widzenia odgrywają istotną rolę. W badaniach nad percepcją obrazu naukowcy wykorzystali system śledzenia wzroku.

Niedawno pojawiła się praca, w której autorzy również wykorzystali technikę śledzenia punktu fiksacji wzroku w badaniu wpływu kierunku patrzenia na lokalizację pozornego źródła dźwięku [124]. Warto jednak zaznaczyć, iż pierwsze wyniki badań w tej tematyce autor opublikował już w 2009 roku [79] [80] [81]. Można stwierdzić, że wykorzystanie systemu śledzenia wzroku w badaniach korelacji wzrokowo-słuchowych jest obecnie szeroko rozwijane.

3.3 Kontekst odbioru treści wizyjno-fonicznej przez widza

W ciągu ostatnich kilkunastu lat silnie rozwinęły się badania korelacji wzrokowo-słuchowych prowadzone w kontekście wzajemnego wpływu bodźców wzrokowych i słuchowych na percepcję obrazu i dźwięku. W pierwszej kolejności należy wspomnieć o badaniach przeprowadzonych przez Becha i współpracowników [9], którzy badali wzajemne oddziaływanie bodźców wzrokowo-słuchowych w systemie kina domowego przy zastosowaniu wyświetlaczy o różnym rozmiarze. Wyniki eksperymentów wykazały, iż rozmiar ekranu wpływa zarówno na odbiór dźwięku, jak i obrazu, a także na ogólną ocenę przestrzenności. Znaleziono logarytmiczną relację pomiędzy rozmiarem ekranu a odbiorem prezentowanego materiału wizyjno-fonicznego [9]. Wyniki uzyskane przez Becha i jego współpracowników korespondują z kontekstem badań przeprowadzonych w ramach niniejszej rozprawy. Jednym z celów cząstkowych rozprawy było zbadanie czy wpływ ściągający obrazu na percepcję dźwięku jest skalowalny, czyli, że występuje niezależnie od wielkości wyświetlacza, na który patrzy widz.

Kolejne badania korelacji wzrokowo-słuchowych w kontekście interakcji obrazu na percepcję dźwięku przestrzennego prowadzili Hollier i Voelcker [57]. Postanowili potwierdzić eksperymentalnie hipotezę, że jakość obrazu ma wpływ na postrzeganie jakości dźwięku oraz zbadać wpływ interakcji dźwięku i obrazu na ogólną ocenę prezentowanej treści wizyjno-fonicznej. Badania potwierdziły istnienie wpływu degradacji jakości jednego z bodźców na ocenę drugiego [57]. Z kolei, wyniki badań przeprowadzonych przez Davisa wykazały, że dźwięk dodany do obrazu znacząco zwiększa poczucie uczestnictwa w wirtualnej rzeczywistości. Badani percypowali większą głębię i objętość pomieszczeń prezentowanych w wirtualnym świecie, gdy były one wypełnione dźwiękami. Eksperymenty przeprowadzone przez Davisa wykazały również, że wirtualny świat wydawał się badanym bardziej realistyczny, gdy obrazowi towarzyszyły dźwięki. Ponadto, Davis wykazał, że na odbieraną jakość obrazu (związaną z jakością wyświetlacza), pozytywnie wpływa obecność dźwięku [35]. Gdyby uznać obserwacje Davisa za słuszne, zastosowanie dźwięku mogłoby kompensować gorszą jakość środowiska wirtualnego. Interesujący jest również fakt, iż badani uznali, że jakość dźwięku (niska czy wysoka) nie miała wpływu na to, jak ostatecznie oceniano jakość obrazu. Do innych wniosków doszli na podstawie swoich badań Storms i Zyda [131]. Zauważyli

bowiem, że istnieje ścisły związek pomiędzy jakością dźwięku a percepcją wizyjno-foniczną. Wyniki przeprowadzonych badań wykazały, że dźwięk wysokiej jakości towarzyszący obrazowi wysokiej jakości wpływa na to, że obraz jest postrzegany jako lepszej jakości w porównaniu z tym samym obrazem bez towarzyszenia dźwięku. Nie wykryto wpływu dźwięku niskiej jakości na percepcję obrazu. Zauważono jednak, że taki dźwięk w połączeniu z obrazem wysokiej jakości, jest odbierany jako gorszej jakości w porównaniu z jego oceną bez obecności obrazu [131]. Inni badacze, Beerends i de Calowe [10], zauważyli, że jakość obrazu (nie zaś jakość dźwięku) była decydującym czynnikiem w ocenie całego materiału wizyjno-fonicznego. Wyniki tych badań zostały również potwierdzone wynikami eksperymentów przeprowadzonych w ramach prac dyplomowych w Katedrze Systemów Multimedialnych [5] [132]. Można więc wysnuć wniosek, że przy produkcjach filmowych czy telewizyjnych główny nacisk powinien być położony na poprawną realizację obrazu. Z takiego założenia wychodzi większość polskich producentów filmowych. W kontekście przytoczonych powyżej wyników badań można uznać, że takie podejście jest jak najbardziej uzasadnione, zwłaszcza w sytuacji szukania oszczędności przez producenta.

Jak wspomniano, badania korelacji wzrokowo-słuchowych w kontekście odbioru treści wizyjno-fonicznej przez widza były prowadzone również w Katedrze Systemów Multimedialnych. Wczesne prace naukowo-badawcze związane z tą tematyką były prowadzone w Zakładzie Inżynierii Dźwieku (dawna nazwa Katedry Systemów Multimedialnych) przez Hajdukiewicza. Wystarczy wspomnieć publikację w tej tematyce z roku 1985 [51] czy rozprawę doktorską z 1987 r., której tematem były „Współzależności realizacji dźwięku i obrazu w studyjnej technice telewizyjnej" [52]. Ponadto, interesujące eksperymenty w tym kontekście badań korelacji wzrokowo-słuchowych przeprowadzili w ramach swoich prac dyplomowych Florek i Szczuko [44], Szymków i inni [132], Andrzejewski [5] oraz Dybski i Napierski [39]. Florek i Szczuko badali korelacje wzrokowo-słuchowe w kontekście zgodności treści wizyjnej z treścią foniczną [44]. Szymków i inni badali m.in. wpływ trzeciego wymiaru w dziedzinie wizji i fonii na wrażenia odbierane przez widza [132]. Andrzejewski badał wzajemny wpływ kompresji materiału wizyjnego i fonicznego na odbieraną przez widza jakość prezentowanej treści [5]. Zaś Dybski i Napierski badali wpływ ściągający obrazu na percepcję dźwięku w

zależności od rodzaju treści wizyjno-fonicznej dla różnych grup badanych [39]. Należy przy tym wspomnieć, że autor rozprawy wspierał prace prowadzone w tej tematyce jako konsultant od roku 2008.

Autor rozprawy w pierwszych badaniach nad wpływem obrazu na percepcję dźwięku przeprowadził eksperyment, który wykazał, iż uwaga wzrokowa widza w większości przypadków pozostaje skupiona, gdy w prezentowanej mu próbce wizyjno-fonicznej dźwięk i obraz są do siebie dopasowane (np. silnie pogłosowy dźwięk we wnętrzu kościoła). Zbadanie tego zjawiska było możliwe dzięki zastosowaniu systemu śledzenia punktu fiksacji wzroku. Ponadto, w ramach wspomnianych badań, sprawdzono czy uwaga wzrokowa widza pozostawała niezmienna podczas badania, a prezentowana widzowi próbka wizyjno-foniczna była wystarczająco interesująca. Dzięki temu możliwe stało się zweryfikowanie czy uzyskane wyniki eksperymentów można uważać za wiarygodne [79] [82]. We wspomnianych badaniach autor analizował między innymi ścieżkę patrzenia (ang. *scan path*) osoby badanej, podobnie jak Augustyniak i Tadeusiewicz analizowali aktywność wzrokową badanych, którym prezentowano wykresy EKG [6] [7].

3.4 Kontekst kompresji obrazu wizyjnego

Metody kompresji obrazu wizyjnego oparte na korelacjach wzrokowo-słuchowych zaczęto rozwijać stosunkowo niedawno. Jednak mimo to, powstało już w tej tematyce wiele prac badawczych i naukowych. W 1996 roku Rao i Chen zaproponowali metodę kompresji wideo bazującą na informacji zawartej w sygnale dźwiękowym [114] [115]. Podstawowe założenie opracowanej przez nich metody opierało się na przewidywaniu treści wideo na podstawie dźwięku zsynchronizowanego z obrazem. Dlatego Rao i Chen posłużyli się w swoich badaniach materiałem wizyjno-fonicznym przedstawiającym twarz osoby mówiącej. Założenie to znacznie uprościło algorytm kodowania, który zakładał m.in., że gdy w sygnale mowy występują spółgłoski ‘p’, ‘b’, czy ‘m’, to można przewidzieć zamknięcie ust w obrazie. Analogicznie, gdy wystąpią samogłoski – można przewidzieć, że usta będą otwarte [114] [115].

W 1998 roku ci sami badacze wykazali, że zintegrowane przetwarzanie dźwięku i obrazu (ang. *joint audio-video processing*) znacznie poprawia jego efektywność oraz zapewnia większą wiarygodność w porównaniu z wynikami niezależnego przetwarzania dźwięku i obrazu [32]. Rao i Chen wskazali obszary praktycznego wykorzystania takiego podejścia. W swoim artykule zaproponowali następujące zastosowania zintegrowanego przetwarzania dźwięku i obrazu: synchronizacja dźwięku z obrazem na podstawie ruchu ust (w telefonii – wideorozmowy, systemy wideokonferencyjne), kompresja wideo czy bimodalne rozpoznawanie osób (na podstawie obrazu twarzy i sygnału mowy) [32].

Interakcje pomiędzy sygnałem dźwięku i obrazu były i w dalszym ciągu są badane przez wielu naukowców. W 2002 roku Mujal i Kirlin [100] zaproponowali wykorzystanie wzajemnych zależności dźwięku i obrazu w celu opracowania nowej metody kompresji w systemach o niskiej przepływności (ang. *low-bit rate encoding systems*). Uzyskane przez nich wyniki okazały się satysfakcjonujące, co potwierdziło słuszność idei zintegrowanego przetwarzania sygnału dźwięku i obrazu [100].

W ostatnio opublikowanych badaniach, Lee zaproponował metodę kompresji obrazu wideo opierającą się na detekcji źródła dźwięku w kodowanym obrazie [86]. Dzięki takiemu podejściu obraz może być kodowany nierównomiernie, w zależności od lokalizacji źródła dźwięku. Obszar, w którym znajduje się źródło dźwięku oraz jego otoczenie jest kodowane z mniejszym stopniem kompresji w porównaniu z obszarami oddalonymi od tego źródła. Należy jednak zaznaczyć, że zaproponowana metoda kodowania obrazu wizyjnego zakłada, iż źródło dźwięku znajdujące się w obrazie przyciąga uwagę wzrokową widza. Można przypuszczać, że to założenie nie zawsze jest spełnione, zwłaszcza, że w obrazie często występują poruszające się obiekty nie będące źródłem dźwięku, skutecznie przyciągające wzrok widza [86]. Warto wspomnieć, iż wątek kompresji obrazu znalazł się również w obszarze badań autora rozprawy. Zostały przeprowadzone eksperymenty w dziedzinie QoE, a dokładniej QoMEX [138] i kontekście wykorzystania systemu śledzenia punktu fiksacji wzroku w badaniach nad kompresją obrazu wizyjnego [75].

Podsumowując przegląd badań przeprowadzonych i aktualnie prowadzonych w dziedzinie korelacji wzrokowo-słuchowych, należy stwierdzić, iż bodźce wzrokowe i

słuchowe wzajemnie na siebie wpływają. Obraz i dźwięk przenoszą informacje w nich zawarte, jednak mogą być one inaczej percypowane ze względu na fakt ich jednoczesnego wystąpienia (percepcja wielomodalna). Współcześnie korelacje wzrokowo-słuchowe obejmują bardzo szeroki obszar badawczy: od synchronizacji dźwięku i obrazu, poprzez wzajemny wpływ obrazu i dźwięku na percepcję aż po kompresję obrazu wizyjnego. Autor niniejszej rozprawy skoncentrował się na badaniu korelacji wzrokowo-słuchowych w kontekście wpływu obrazu na percepcję dźwięku, a precyzyjniej – w kontekście wpływu bodźca wzrokowego (wpływu kierunku patrzenia) na lokalizację pozornego źródła dźwięku w panoramie stereofonicznej. W przeprowadzonych eksperymentach autor odniósł się do badań: Witkina (badającego wpływ ściągający obrazu twarzy spikera na kierunek słyszenia) [141], Gardnera (którego badania wykazały istnienie wpływu ściągającego obrazu w płaszczyźnie przód-tył) [47], Becha (badającego zależność między rozmiarem ekranu a percepcją bodźców wzrokowo-słuchowych) [9] oraz Lewalda (badającego percepcję dźwięku w zależności od kierunku patrzenia) [88]. Dalsza część pracy została poświęcona badaniom, które zostały przeprowadzone w ramach niniejszej rozprawy.

W następnym rozdziale (rozdz. 4) scharakteryzowano systemy śledzenia punktu fiksacji wzroku, w szczególności podano charakterystyki systemów wykorzystanych w badaniach korelacji wzrokowo-słuchowych. Szczególną uwagę poświęcono systemowi Cyber-Oko, opracowanemu w Katedrze Systemów Multimedialnych. Opis tego typu systemów jest niezbędny dla podkreślenia faktu, że tego typu interfejsy mogą pełnić istotną rolę w obiektywizacji wyników badań korelacji wzrokowo-słuchowych.

4 System śledzenia punktu fiksacji wzroku

Jak wspomniano w zakończeniu rozdziału trzeciego, w badaniach przeprowadzonych w ramach niniejszej rozprawy wykorzystano system, a właściwie dwa systemy śledzenia punktu fiksacji wzroku (ang. *eye gaze tracking system*). W przeprowadzonych eksperymentach zastosowano system komercyjny Tobii T60 oraz opracowany w Katedrze Systemów Multimedialnych system Cyber-Oko. Należy zaznaczyć, iż opracowanie i skonstruowanie Cyber-Oka nie stanowi przedmiotu przeprowadzonych badań, choć w aspekcie analizy aktywności wzrokowej widza oglądającego film trójwymiarowy, badania te wspomaga. Autor rozprawy jest jednym z współtwórców tego systemu i jednocześnie osobą współpracującą nad rozwijaniem jego funkcjonalności.

Cyber-Oko, podobnie jak wszystkie dostępne na rynku systemy tego typu, umożliwia wyznaczanie PoR na płaszczyźnie monitora, na który patrzy użytkownik. Warto zwrócić uwagę na fakt, iż Cyber-Oko jest systemem bezkontaktowym (ang. *non-contact eye gaze tracking system*). Oznacza to, że kamera rejestrująca ruchy gałek ocznych jest umiejscowiona w pobliżu monitora komputerowego, nie zaś na specjalnym diademie, zakładanym na głowę, jak w przypadku systemów nagłownych (ang. *head-mounted eye tracking systems*). Interesujące badania nad rozwojem nagłownego systemu śledzenia wzroku prowadzą naukowcy z Katedry Inżynierii Biomedycznej Politechniki Gdańskiej [67] [68]. W przypadku eksperymentów, w ramach których badany jest wpływ obrazu na percepcję dźwięku uzasadnione jest stosowanie bezkontaktowego systemu śledzenia punktu fiksacji wzroku, ponieważ widz jest poddany badaniu w praktycznie rzeczywistych warunkach. Bezkontaktowy interfejs, śledzący położenie wzroku na ekranie, w mniejszym stopniu rozprasza i dekoncentruje widza. Cyber-Oko pracuje w czasie rzeczywistym, jednakże jego zastosowanie w badaniu wzajemnego wpływu dźwięku i obrazu w ujęciu percepcyjnym nakłada na ten system zdecydowanie większe wymagania techniczne. Pierwszy z dwóch parametrów, które należy brać pod uwagę w kontekście wspomnianych wyżej badań, jest związany z liczbą wyznaczanych w jednostce czasu punktów fiksacji wzroku (PoR). W nomenklaturze technicznej parametr ten jest nazywany rozdzielczością czasową[s] systemu (ang. *temporal resolution*). Drugi parametr określa natomiast zdolność rozróżniania przez system obszarów ekranu, na które patrzy użyt-

kownik. Przyjęło się nazywanie tego parametru rozdzielczością przestrzenną[s] (ang. *spatial resolution*). Więcej szczegółowych informacji odnoszących się do tych dwóch parametrów zamieszczono w podrozdziale 4.3 niniejszej rozprawy. Rozdzielczość czasowa i przestrzenna stanowią zatem istotne parametry systemu śledzenia punktu fiksacji wzroku w badaniu korelacji wzrokowo-słuchowych.

4.1 Przegląd systemów śledzenia punktu fiksacji wzroku

4.1.1 Rozwój technik śledzenia wzroku

Historia badań nad wyznaczaniem kierunku patrzenia sięga roku 1879. Wówczas Javal zauważył, że czytanie tekstu składa się z serii krótkich przerw (fiksacji), podczas których następuje pobieranie informacji oraz z szybkich ruchów pomiędzy tymi przerwami [58]. Pierwsze urządzenie umożliwiające śledzenie wzroku skonstruował Huey w 1898 roku. Zastosował on specjalny rodzaj soczewki kontaktowej z otworem na źrenicę. Soczewka była połączona z aluminiowym wskaźnikiem, poruszającym się wraz z ruchem gałki ocznej. Dzięki temu możliwe stało się zweryfikowanie tego, na których słowach osoba badana zatrzymywała wzrok podczas czytania. Pierwszy nieinwazyjny system śledzenia wzroku został skonstruowany przez Buswella [30] [162]. Wykorzystał on promienie światła, które odbijały się od powierzchni oka, a następnie były rejestrowane na taśmie filmowej. W 1947 roku Fitts użył kamer filmowych do zarejestrowania obrazów oczu pilota podczas lądowania. W swoim eksperymencie skoncentrował się nad tym, w jaki sposób pilot korzysta z urządzeń znajdujących się w zasięgu jego wzroku w kokpicie samolotu. Eksperyment ten uważa się za pierwsze badanie użyteczności z wykorzystaniem techniki śledzenia wzroku [38] [42] [63]. W latach 50. XX wieku Yarbus odkrył związek pomiędzy fiksowaniem wzroku a procesem myślenia:

> *„Zapisy ruchów gałek ocznych wskazują na to, że uwaga osoby patrzącej na obraz zazwyczaj jest skupiona jedynie na jego elementach… Ruchy oczu odzwierciedlają proces myślenia człowieka…”* [143].

Od lat 70. XX wieku prace nad rozwojem technik śledzenia wzroku były ściśle związane z mocą obliczeniową komputerów. W latach 80. wzrost wydajności kompute-

rów przyczynił się do rozwoju systemów śledzenia punktu fiksacji wzroku. Poza tym, dzięki operacjom przetwarzania obrazu wizyjnego (ang. *video-based processing*), pobieranego z kamery podłączonej do komputera i śledzącej twarz użytkownika, możliwa stała się interakcja człowieka z komputerem (ang. *Human-Computer Interaction*). Pierwszy taki system został zaprezentowany przez Bolta w 1981 r. [27]. W tym czasie również po raz pierwszy system śledzenia wzroku został użyty przez osobę niepełnosprawną. Od lat 90. minionego stulecia zainteresowanie, a przede wszystkim stosowanie opisywanych w niniejszym rozdziale systemów, stale wzrasta. Obecnie systemy śledzenia punktu fiksacji wzroku znajdują zastosowanie w przemyśle rozrywkowym, życiu codziennym [38] [63], a także w interesujących badaniach naukowych, m.in. z zakresu medycyny [6] [7].

4.1.2 Przegląd komercyjnych systemów śledzenia punktu fiksacji wzroku

Jak wspomniano we wprowadzeniu rozdziału czwartego, Cyber-Oko jest interfejsem bezkontaktowym. W niniejszym podrozdziale scharakteryzowano zatem jedynie interfejsy umożliwiające bezkontaktową interakcję użytkownika z komputerem. Obecnie czołowym producentem systemów śledzenia punktu fiksacji wzroku jest szwedzka firma Tobii, oferująca bogatą ofertę interfejsów dedykowanych do różnych zastosowań. Wszystkie systemy śledzenia wzroku produkowane przez firmę Tobii charakteryzują się tym, że pracują w zakresie promieniowania podczerwonego (ang. *infrared* (IR) *radiation*). Zalety pracy systemu w tym przedziale długości fali omówiono w podrozdziale 4.3.1 niniejszej rozprawy. Jak wspomniano na początku niniejszego rozdziału w przeprowadzonych w ramach rozprawy eksperymentach zastosowano jeden z modeli systemów śledzenia wzroku firmy Tobii – model T60. W tab. 4.1 zestawiono parametry techniczne wykorzystanego w badaniach interfejsu Tobii T60 oraz innych modeli interfejsów wzrokowych firmy Tobii: T120, T60 XL i TX300 [177].

Warto zwrócić uwagę na bardzo wysoką precyzję tego systemu w wyznaczaniu punktu fiksacji na ekranie monitora. Rozdzielczość przestrzenna o wartości $0{,}5^{o}$ oznacza, że gdy użytkownik siedzi w odległości 60 cm od monitora, system śledzenia wzroku zwraca współrzędne punktu fiksacji z dokładnością do 5,2 mm (w płaszczyźnie po-

ziomej i pionowej). Tak wysoka precyzja pozwala na zastosowanie tych systemów na przykład w badaniach użyteczności stron internetowych, gdzie wymagana jest możliwie największa dokładność odwzorowania ścieżki patrzenia na badanej stronie internetowej.

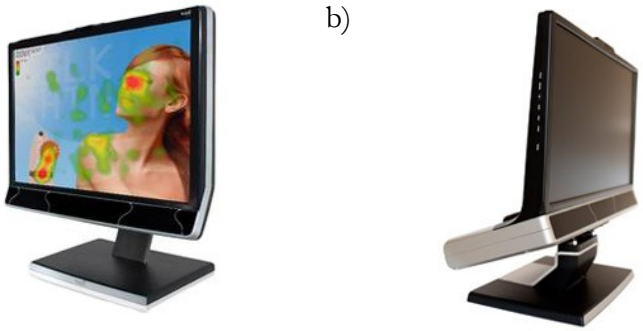
Tab. 4.1 Zestawienie parametrów technicznych różnych modeli systemów śledzenia wzroku firmy Tobii [177]

	T60	T120	T60 XL	TX 300
rozdzielczość przestrzenna	**0,5°**	0,5°	0,5°	0,4°
rozdzielczość czasowa	**60 Hz**	120 Hz	60 Hz	300 Hz
przekątna i proporcje ekranu	**17”, 5:4**	17”, 5:4	24”, 16:10	23”, 16:9
maks. rozdzielczość ekranu	**1280x1024**	1280x1024	1920x1200	1920x1080

Na rys. 4.1 zaprezentowano część sprzętową systemu Tobii T60 wykorzystanego w przeprowadzonym w ramach rozprawy eksperymencie, zaś na rys. 4.2 przedstawiono pozostałe interfejsy firmy Tobii, których parametry wyspecyfikowano w tab. 4.1. Łatwo zauważyć, że systemy śledzenia wzroku firmy Tobii do złudzenia przypominają zwyczajne monitory LCD. Część interfejsu odpowiedzialna za rejestrację obrazu i emisję światła w zakresie fal podczerwonych jest zintegrowana z monitorem, stąd trudno odróżnić go od zwykłego monitora komputerowego.

Rys. 4.1 System śledzenia punktu fiksacji wzroku Tobii T60 [177]

a) b)



Rys. 4.2 Systemy śledzenia punktu fiksacji wzroku firmy Tobii: a) model T60 XL, b) model TX300 [177]

Dodatkowo warto również zaznaczyć, iż system Tobii TX300 poprzez rozłączenie ekranu, który nie jest połączony na stałe z pozostałą częścią urządzenia, umożliwia badanie PoR nie tylko na płaszczyźnie ekranu monitora, ale również na obiektach rzeczywistych czy na ekranie projektora. Podobną funkcjonalność posiada również system Cyber-Oko. Śledzenie wzroku osoby badanej na ekranie projektora jest możliwe dzięki zastosowaniu specjalnej ramki imitującej ekran monitora, przedstawionej na rys. 4.9b.

Bezpośrednio po wprowadzeniu na rynek modelu Tobii TX300 firma SMI zaprezentowała w 2010 r. nowy model interfejsu wzrokowego – SMI RED500. Dokładność wyznaczania PoR w porównaniu z systemem Tobii TX300 okazała się większa. SMI RED500 charakteryzuje się rozdzielczością przestrzenną większą niż $0{,}4^{o}$. Na większą uwagę zasługuje jednak wzrost rozdzielczości czasowej RED500 w porównaniu z najnowszym modelem interfejsu wzrokowego firmy Tobii. Informacja o położeniu PoR jest bowiem odświeżana co 1/500 s, czyli co 2 ms. Oznacza to, że system umożliwia wykrycie i śledzenie ruchów sakkadowych gałek ocznych. Podobnie jak interfejs Tobii TX300, „gaze tracker" firmy SMI pozwala na śledzenie kierunku patrzenia również na ekranie projekcyjnym [171]. Na rys. 4.3 przedstawiono dwie konfiguracje użytkowe systemu SMI RED500: z monitorem i wolnostojącą – do badania PoR na obiektach rzeczywistych i na ekranie projektora.

Rys. 4.3 System śledzenia punktu fiksacji SMI RED500 [171]

Interesujące rozwiązanie zaproponowała amerykańska firma LC Technologies, wprowadzając na rynek system Eyefollower 2.0, przedstawiony na rys. 4.4. Interfejs ten zapewnia bardzo wysoką dokładność wyznaczania punktu fiksacji na ekranie monitora – rozdzielczość przestrzenna wynosi 0,45^{o}. Wysoką dokładność wyznaczania PoR uzyskano dzięki zastosowaniu czterech kamer, z czego jedna śledzi położenie głowy, a trzy pozostałe, wyposażone w teleobiektywy, automatycznie ustawiają ostrość na śledzonym oku, przechwytując obraz wysokiej rozdzielczości. Rozdzielczość czasowa również jest bardzo wysoka i wynosi 120Hz. W kontekście przeprowadzonych w ramach rozprawy badań, wspomnieć należy o tym, że Eyefollower 2.0 zwraca również przestrzenne współrzędne punktów fiksacji, zarówno dla rzeczywistych obiektów (nie wykorzystując monitora), jak i wirtualnych obiektów 3D wyświetlanych na monitorze stereoskopowym [158].

Rys. 4.4 System śledzenia punktu fiksacji wzroku Eyefollower 2.0 [158]

Zaprezentowane powyżej interfejsy wzrokowe cechuje wysoka dokładność i jakość wykończenia. Z pewnością są to istotne czynniki wpływające na stosunkowo wysoką cenę tych rozwiązań. Jedną z firm produkujących systemy śledzenia wzroku o

nieznacznie słabszych parametrach jest Eye Response Technologies, będąca obecnie częścią amerykańskiej DynaVox Mayer-Johnson. Firma ta wypromowała interfejs o nazwie ERICA również często stosowany w badaniach użyteczności stron internetowych. ERICA umożliwia przeprowadzenie badania z rozdzielczością przestrzeną 0,5^{o} i rozdzielczością czasową 60Hz [38] [155]. System ERICA przedstawiono na rys. 4.5.

Rys. 4.5 System śledzenia punktu fiksacji wzroku [156]

Na rynku dostępnych jest jeszcze wiele innych bezkontaktowych interfejsów wzrokowych różnych producentów. Ze względu na tematykę niniejszej pracy zrezygnowano jednak z dokładnej charakterystyki tych interfejsów, ograniczając się jedynie do podania nazw niektórych z nich. Poniżej wymieniono nazwy wybranych interfejsów, charakteryzujących się podobnymi parametrami technicznymi:

- system Intelligaze niemieckiej firmy Alea Technologies [149],
- system SeeTech PRO niemieckiej firmy Humanelektronik [170],
- system EyeTech TM3 amerykańskiej firmy EyeTech Digital Systems [159],
- system EMR-AT VOXER japońskiej firmy NAC Technologies [167].

Wśród bezkontaktowych systemów śledzenia wzroku na szczególną uwagę zasługuje prototypowe rozwiązanie opracowane w wyniku współpracy firm Lenovo i Tobii. W dniu 01.03.2011 na targach CeBIT w Hannoverze Lenovo zaprezentowało niekomercyjną wersję laptopa z wbudowanym systemem śledzenia wzroku. Wersja prototypowa prezentuje możliwości, które zostaną wprowadzone na rynek konsumencki w postaci produktu w przeciągu dwóch lat. Obecna wersja laptopa Lenovo umożliwia wyznaczanie punktu fiksacji z rozdzielczością przestrzenną 0,5^{o} 30-40 razy w ciągu sekundy. Rys. 4.6 przedstawia prototypową wersję laptopa sterowanego wzrokiem [178].

Rys. 4.6 Prototypowy laptop firmy Lenovo z wbudowanym systemem śledzenia wzroku [178]

Wszystkie przedstawione powyżej interfejsy wzrokowe są systemami komercyjnymi. Warto jednak zauważyć, że istnieją zaawansowane prace badawczo-naukowe poświęcone rozwijaniu otwartego oprogramowania[s] (ang. *open source software*), umożliwiającego efektywne śledzenie wzroku użytkownika komputera. Największe osiągnięcia w tej dziedzinie należy przypisać europejskiemu projektowi COGAIN (Communication by Gaze Interaction), którego celem było rozwijanie nowych technologii śledzenia wzroku i udoskonalanie już istniejących na potrzeby komunikacji z osobami niepełnosprawnymi fizycznie [154]. Grupa naukowców z Uniwersytetu Technicznego w Kopenhadze związana z projektem COGAIN udostępniła darmową aplikację ITU Gaze Tracker [160], będącą niskobudżetową alternatywą dla komercyjnych interfejsów wzrokowych. Innymi przykładami otwartego oprogramowania pozwalającego na wyznaczanie punktu fiksacji wzroku w czasie rzeczywistym są między innymi aplikacje: Opengazer [163] i openEyes [176]. Choć istnieje jeszcze kilka innych aplikacji otwartych tego typu, zdecydowano się je pominąć ze względu na tematykę niniejszej rozprawy. Warto zauważyć, iż rozwój projektów, w ramach których opracowywane jest oprogramowanie otwarte ma pozytywny wpływ na rozpowszechnienie się niskobudżetowych wersji interfejsów wzrokowych [2] [3]. To z kolei powoduje wzrost zainteresowania systemami śledzenia wzroku w pracach naukowo-badawczych, jak również w zastosowaniach użytkowych.

4.2 Założenia

Jak wspomniano we wprowadzeniu, celem badań prowadzonych w ramach niniejszej rozprawy jest wykazanie słuszności stosowania systemu śledzenia punktu fiksacji wzroku w eksperymentach z zakresu korelacji wzrokowo-słuchowych. W celu udowodnienia postawionych w rozprawie tez wykorzystano system Cyber-Oko, jak również komercyjny system Tobii T60. Poprzez wyznaczanie współrzędnych punktów fiksacji wzroku system dostarcza obiektywnej informacji o tym, na których fragmentach wyświetlanego obrazu użytkownik skupia uwagę wzrokową. Istotnymi parametrami technicznymi, które należało brać pod uwagę w kontekście przeprowadzonych badań były: rozdzielczość przestrzenna i rozdzielczość czasowa systemu. Dokładność wyznaczania PoR w czasie opisano w podrozdziale 4.3.2.

Udowodnienie tez rozprawy wiąże się bezpośrednio z udowodnieniem istnienia wpływu ściągającego obrazu na percepcję dźwięku w sposób zobiektywizowany. Dla przypomnienia wpływ ściągający określa sytuację, w której widz inaczej lokalizuje źródło dźwięku w dwu- lub wielokanałowej bazie stereofonicznej, gdy prezentowana jest mu tylko ścieżka dźwiękowa w porównaniu z sytuacją, gdy ścieżce dźwiękowej towarzyszy wyświetlany obraz. Zastosowanie Cyber-Oka w eksperymentach badających efekt ściągający polegało na zarejestrowaniu współrzędnych punktów fiksacji na oglądanym obrazie wideo. Rozkład uwagi wzrokowej zsynchronizowany z obrazem wideo posłużył do sprecyzowania wniosków na temat wpływu obrazu na percepcję dźwięku. Zapis położenia punktów fiksacji wzroku jest obiektywnym źródłem informacji o tym, które elementy wyświetlanego obrazu przyciągnęły uwagę widza. W związku z powyższym założono, że obiektywna informacja o położeniu wzroku na ekranie monitora wraz z subiektywnym wskazaniem lokalizacji źródła dźwięku w bazie stereofonicznej, będą stanowić podstawę do udowodnienia istnienia tak zwanego wpływu ściągającego.

Dane o położeniu punktów fiksacji wzroku w czasie trwania wizyjno-fonicznej próbki testowej są zapisywane do pliku w formacie XML. Na podstawie informacji zawartych w pliku XML generowana jest dynamiczna mapa przejść, wizualizująca uwagę wzrokową widza w czasie trwania danej próbki testowej. Dane zawarte w pliku XML mogą być przetwarzane w dowolny sposób. Założono, że do interpretacji wyników wykorzystane zostaną dane zapisane w strukturze XML. Wizualizacja wyników w

postaci dynamicznej mapy przejść stanowi jedynie dodatkową funkcjonalność aplikacji Cyber-Oka, która z kolei może posłużyć do weryfikacji otrzymanych wyników.

Założono, iż w ramach rozprawy zbadany zostanie wpływ trójwymiarowego obrazu wizyjnego na percepcję dźwięku. W związku z powyższym należało opracować metodykę śledzenia punktu fiksacji wzroku badanego w przestrzeni. Metody zbadane w ramach pracy zostały opisane w podrozdziałach 4.4.1 i 4.4.2. Niemniej jednak, ze względu na ograniczoną rozdzielczość przestrzenną i czasową systemu Cyber-Oko oraz wysokie wymagania zaproponowanych metod, niemożliwe okazało się ich bezpośrednie zastosowanie w praktyce. Zaproponowano więc inne podejście śledzenia aktywności wzrokowej widza na prezentowanym mu obrazie trójwymiarowym. Podejście to zostało opisane w podrozdziale 4.4.3 i opiera się na indeksacji treści obrazu wizyjnego. Założono, że skupienie wzroku badanego na obrazie 3D będzie analizowane nie na podstawie PoR wyznaczonego w przestrzeni, lecz na podstawie informacji o położeniu wzroku na zdefiniowanych wcześniej obszarach zainteresowania w określonych przedziałach czasu, tak zwanych interwałach.

Zgodnie z założeniem materiałem badawczym przeprowadzonych badań były przede wszystkim fragmenty znanych utworów filmowych, zrealizowanych w ostatnich latach jako filmy stereoskopowe (potocznie nazywane filmami 3D). Dodatkowo, w badaniach wykorzystano także trójwymiarowe próbki z nagrania koncertu, przygotowane przez autora rozprawy. Cyber-Oko umożliwia wczytywanie próbek z konwencjonalnym, dwuwymiarowym obrazem wizyjnym oraz filmów 3D z dźwiękiem wielokanałowym. Obraz 3D przygotowano i wyświetlano w technice anaglifowej. Zatem separacja lewej i prawej składowej obrazu stereoskopowego odbywała się przez filtr koloru (czerwony na lewym oku i cyjanowy – na prawym). Zaletą techniki anaglifowej jest możliwość wyświetlania obrazu zarówno na monitorze komputerowym, jak i na ekranie z wykorzystaniem projektora multimedialnego bez znaczącego pogorszenia postrzeganego efektu 3D.

Podsumowując, autor niniejszej rozprawy założył, iż parametry techniczne systemu Cyber-Oko umożliwiają przeprowadzenie badań z zakresu korelacji wzrokowo-słuchowych na dwuwymiarowym i trójwymiarowym obrazie wizyjnym. Ponadto założono, że na obiektywną miarę efektu ściągającego składają się subiektywne wskazania

położenia pozornego źródła dźwięku w panoramie stereofonicznej, jak również informacja o położeniu wzroku na określonych obszarach zainteresowania, stanowiącymi fragmenty obrazu prezentowanego na ekranie.

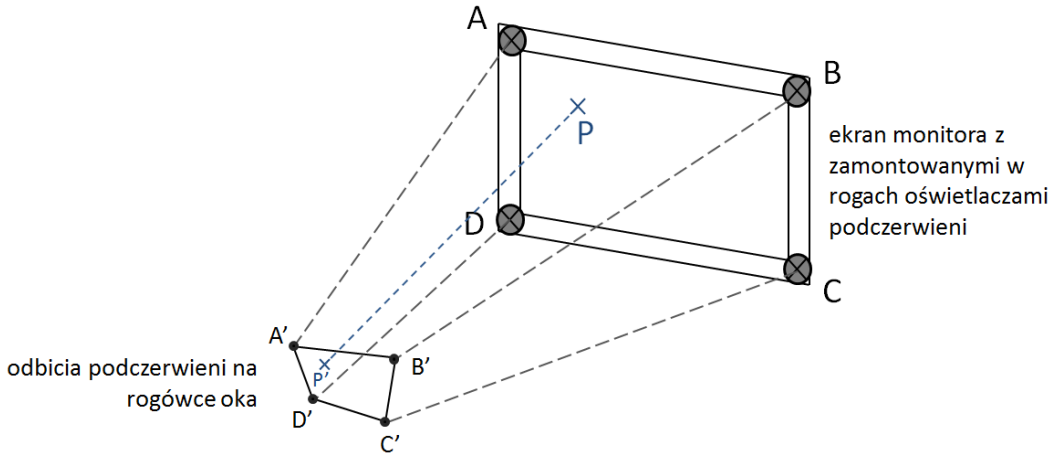
4.3 Charakterystyki opracowanego systemu

System Cyber-Oko ma istotne znaczenie dla przedstawionych w rozprawie badań, w związku z czym niniejszy rozdział został poświęcony jego opisowi. Warto na samym początku zaznaczyć, iż rozdzielczość przestrzenna Cyber-Oka pierwotnie miała wynosić 3x3 [34]. Jednak w miarę postępów prac rozwojowych jego rozdzielczość została znacząco poprawiona. W celu jej dokładnego wyznaczenia, zostały przeprowadzone badania z udziałem studentów Wydziału Elektroniki, Telekomunikacji i Informatyki Politechniki Gdańskiej. Metodykę wyznaczenia rozdzielczości przestrzennej przedstawiono w podrozdziale 4.3.2. Rozdzielczość czasową Cyber-Oka zdeterminowała kamera, która pozwalała na akwizycję pięciu klatek obrazu na sekundę. Dodać należy, że wszystkie pobrane w ciągu sekundy klatki były przetwarzane przez system, zatem jego rozdzielczość czasowa wynosiła 5Hz. Oznacza to, że informacja o położeniu PoR na ekranie monitora była odświeżana co 200 ms.

Koncepcja wyznaczania punktu fiksacji wzroku

Przed przejściem do właściwego opisu systemu konieczne jest zrozumienie zasady jego funkcjonowania w kontekście wyznaczania punktu fiksacji wzroku na ekranie monitora. Współrzędne PoR są wyznaczane na podstawie informacji o środku źrenicy oka oraz konfiguracji odbić promieniowania podczerwonego, powstałych na rogówce oka. Odbicia promieniowania podczerwonego, nazywane również pierwszym obrazem Purkinjego lub glintami, nie zmieniają swojego położenia, gdy gałka oczna się porusza. Dzięki temu stanowią one punkty odniesienia dla środka źrenicy, który w opracowanym systemie jest punktem charakterystycznym, zmieniającym swoje położenie wraz z ruchem gałki ocznej. Wyznaczenie punktu fiksacji wzroku użytkownika pracującego z systemem polega na pomiarze odległości środka źrenicy od każdego z czterech odbić. Relacja pomiędzy odbiciami na rogówce oka a środkiem źrenicy odzwierciedla relacje

geometryczne pomiędzy oświetlaczami podczerwieni zamontowanymi w rogach ekranu a punktem fiksacji wzroku. Relację tę przedstawiono na rys. 4.7, gdzie A, B, C i D oznaczają cztery oświetlacze w rogach ekranu, P – punkt fiksacji wzroku (PoR), A’, B’, C’, i D’ – odbicia podczerwieni odpowiadające kolejnym oświetlaczom, a punkt P’ – środek źrenicy.



Rys. 4.7 Metodyka wyznaczania punktu fiksacji wzroku na ekranie monitora

Należy zaznaczyć, że współrzędne środka źrenicy i czterech odbić wyrażone są we współrzędnych obrazu, pobieranego z kamery (1600x1200 pikseli), zaś współrzędne PoR są wyrażone w jednostkach z zakresu wartości (0; 100). Punkt o współrzędnych (0; 0) znajduje się w lewym górnym rogu ekranu.

W następnym podrozdziale scharakteryzowano system Cyber-Oko, opisując jego komponenty sprzętowe oraz zaimplementowane algorytmy przetwarzania obrazu.

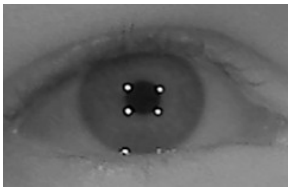
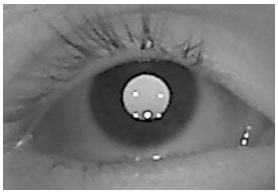
4.3.1 Specyfikacja systemu Cyber-Oko

Cyber-Oko jest systemem składającym się z dwóch warstw – warstwy sprzętowej[s] i warstwy oprogramowania[s]. Na część sprzętową Cyber-Oka składają się: jednostka obliczeniowa – średniej klasy komputer klasy PC, monitor, zmodyfikowana kamera USB, diody podczerwieni (ang. *infrared* – IR *diodes*), sterownik zasilania diod oraz przewody. Poniżej scharakteryzowano bardziej szczegółowo każdą z warstw systemu.

Konfiguracja sprzętowa systemu Cyber-Oko

Podstawowym wymogiem, jakie musi spełniać monitor współpracujący z Cyber-Okiem jest stosunek długości boków ekranu: 5:4. System współpracuje również z monitorami o innych proporcjach, w tym z monitorami panoramicznymi, ale optymalna rozdzielczość wyświetlanego obrazu wynosi 1280x1024. Monitor z diodami zamontowanymi w rogach ekranu przedstawiono na rys. 4.9. W badaniach przeprowadzonych w ramach niniejszej rozprawy system Cyber-Oko współpracował z monitorem 19" marki *Hyundai*. Ponadto, w badaniach wykorzystano również specjalną ramkę, na której rogach umieszczono diody emitujące światło podczerwone. Zastosowanie ramki imitującej ekran monitora pozwoliło na przeprowadzenie badań korelacji wzrokowo-słuchowych przy prezentacji obrazu wizyjnego na ekranie projektora. Przy spełnieniu odpowiednich warunków możliwe stało się śledzenie położenia wzroku na obrazie wizyjnym prezentowanym na ekranie o znacznie większym rozmiarze niż dedykowany ekran systemu Cyber-Oko. Poza monitorem z zamontowanymi w rogach ekranu diodami podczerwieni istotnym komponentem systemu Cyber-Oko jest kamera internetowa. Do pracy z systemem wybrano dwa modele kamer USB firmy *Logitech* – *QuickCam Pro9000* (kamera statyczna) i *QuickCam Sphere AF* (kamera z ruchomą głowicą). W przypadku konfiguracji systemu z kamerą ruchomą możliwe jest śledzenie aktywności użytkownika. Dzięki temu w trakcie pracy z systemem użytkownik może czuć się swobodnie, poruszając się w określonym obszarze. W celu przystosowania kamery do pracy z Cyber-Okiem konieczne było wykonanie kilku modyfikacji związanych przede wszystkim ze zwiększeniem czułości kamery w zakresie fal podczerwonych. W pierwszej kolejności usunięto przylegający do matrycy kamery filtr podczerwieni. W następnym kroku wybrano obiektyw, zapewniający odpowiednie pole widzenia kamery, czyli obszar analizowanego obrazu. Po przeprowadzeniu badań okazało się, że optymalna długość ogniskowej obiektywu wynosi 12 mm. Następnie dodano filtr pasmowy, przepuszczający światło w zakresie podczerwieni (ang. IR-*pass filter*). Ostatnia modyfikacja kamery USB, wykorzystywanej w opracowanym systemie, polegała na zamontowaniu wokół obiektywu kamery grupy diod IR, odpowiedzialnych za powstawanie tzw. efektu jasnej źrenicy (ang. *bright-eye effect*). Przyjęto, że Cyber-Oko może pracować w dwóch trybach – z ciemną źrenicą (gdy diody IR wokół obiektywu są wyłączone) i z jasną źre-

nicą. Tryb pracy jest wybierany automatycznie na początku pracy z systemem i zależy od koloru oczu użytkownika korzystającego z systemu oraz od panujących warunków oświetlenia. Na rys. 4.8 przedstawiono obrazy oczu dla każdego z dwóch trybów pracy systemu – z efektem ciemnej i jasnej źrenicy.

a)  b) 

Rys. 4.8 Obrazy oczu analizowane przez system Cyber-Oko: a) z efektem ciemnej źrenicy, b) z efektem jasnej źrenicy

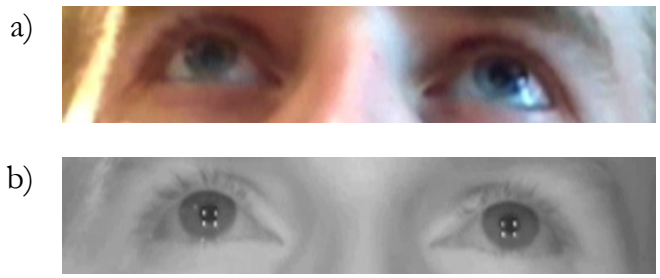
Umiejscowienie kamery w opracowanym systemie śledzenia punktu fiksacji wzroku jest takie samo, jak w przypadku wszystkich wspomnianych wcześniej komercyjnych systemów tego typu. Kamera znajduje się bezpośrednio pod monitorem, dokładnie na pionowej osi symetrii monitora. Takie położenie jest uzasadnione ze względu na optymalne warunki akwizycji obrazu oka, zwłaszcza gdy użytkownik fiksuje wzrok w dolnej części ekranu monitora. Rys. 4.9 prezentuje warstwę sprzętową Cyber-Oka w dwóch konfiguracjach: z monitorem i ruchomą kamerą oraz z ramką imitującą ekran monitora i kamerą statyczną.

a)  b) 

Rys. 4.9 Cyber-Oko: a) w konfiguracji z dedykowanym monitorem i ruchomą kamerą, b) w konfiguracji z ramką imitującą ekran monitora i kamerą statyczną

Promieniowanie podczerwone

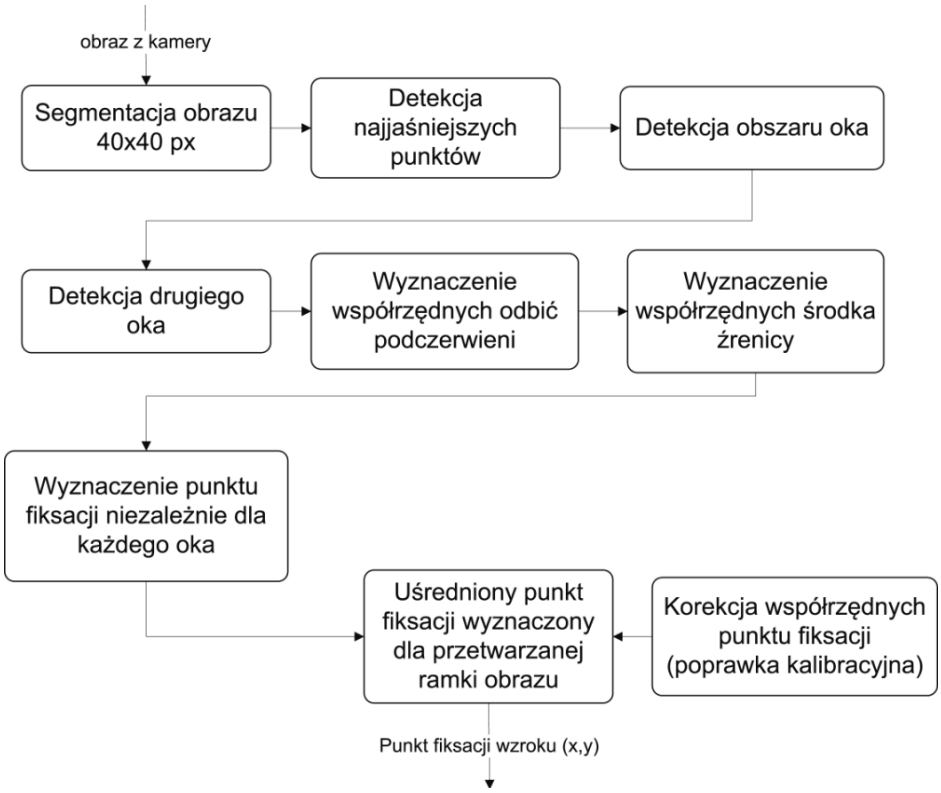
Jak wspomniano wcześniej, system Cyber-Oko wykorzystuje promieniowanie podczerwone, podobnie jak większość komercyjnych systemów śledzenia wzroku. W związku z tym, że promieniowanie podczerwone znajduje się poza oknem optycznym, czyli zakresem światła widzialnego, nie wpływa na komfort pracy użytkownika korzystającego z systemu. Stosowanie promieniowania podczerwonego w systemach tego typu jest uzasadnione, ponieważ znacząco wspiera proces przetwarzania obrazu. Pierwszą z dwóch istotnych zalet jest wzrost kontrastu pomiędzy źrenicą i tęczówką oka. Odpowiedni kontrast zapewnia prawidłową detekcję konturu źrenicy analizowanego oka. Prawidłowa detekcja konturu źrenicy jest istotna ze względu na dokładność wyznaczenia punktu fiksacji wzroku, która bezpośrednio zależy od precyzyjnego wyznaczenia współrzędnych środka źrenicy. Drugą zaletą stosowania promieniowania podczerwonego w systemie Cyber-Oko oraz w systemach śledzenia punktu fiksacji wzroku w ogólności, jest powstawanie czterech charakterystycznych odbić na rogówce oka, o których wspomniano wcześniej [77]. Dla zobrazowania różnic pomiędzy standardową kamerą i kamerą pracującą w zakresie fal podczerwonych w systemie Cyber-Oko na rys. 4.10 porównano obrazy z obu tych kamer.



Rys. 4.10 Porównanie obrazu ze standardowej kamery i kamery wykorzystywanej przez system Cyber-Oko: a) standardowa kamera, b) kamera Cyber-Oka

Warstwa oprogramowania

Na warstwę oprogramowania systemu Cyber-Oko w głównej mierze składają się algorytmy przetwarzania obrazu. Proces przetwarzania obrazu ma na celu ekstrakcję charakterystycznych punktów w obrazie, czyli odbić podczerwieni i środka źrenicy, w celu obliczenia współrzędnych PoR na ekranie. Na rys. 4.11 przedstawiono schemat algorytmu wyznaczania PoR.



Rys. 4.11 Schemat algorytmu wyznaczania PoR przez system Cyber-Oko

W pierwszym etapie dokonywana jest segmentacja pobranej przez kamerę ramki obrazu. Operacja segmentacji polega na podzieleniu całego obszaru ramki (1600x1200 pikseli) na segmenty o wymiarach 40x40 pikseli. W każdym z segmentów wyznaczane są współrzędne najjaśniejszych punktów. Jeżeli konfiguracja najjaśniejszych punktów w wybranym segmencie tworzy kształt czworokąta, którego kąty wewnętrzne są kątami ostrymi, a długości krawędzi spełniają określone warunki, wówczas zakłada się, że w tym segmencie wykryto pierwsze oko. Obszar, w którym poszukiwane jest drugie oko, zostaje ograniczony do segmentów znajdujących się po lewej i prawej stronie segmentu, w którym znaleziono pierwsze oko. Następnie wyznaczane są współrzędne każdego z odbić podczerwieni na podstawie obrazu obu oczu. Współrzędne poszczególnych odbić są wyznaczane w oparciu o środek ciężkości (ang. *center of gravity*) obszarów reprezentujących poszczególne odbicia. W kolejnym etapie w obrazie oka poszukiwany jest obszar źrenicy. Kształt źrenicy jest znajdowany na podstawie dopasowania elipsy (ang. *ellipse fitting*). Współrzędne środka elipsy aproksymującej kształt źrenicy są uważane za współrzędne środka źrenicy. Na podstawie współrzędnych odbić podczerwieni oraz środka źrenicy wyznaczany jest punkt fiksacji niezależnie dla lewego i prawego

oka. W ostatnim etapie uwzględniana jest tzw. poprawka kalibracyjna, uzyskiwana w drodze kalibracji systemu, przeprowadzanej w pierwszej fazie pracy z Cyber-Okiem. Proces kalibracji ma na celu wyznaczenie wektora przesunięcia pomiędzy oczekiwanym (rzeczywistym) a obliczonym PoR. Podczas kalibracji użytkownik wpatruje się w dziewięć punktów wyświetlanych kolejno na ekranie. Na całej powierzchni ekranu może być wyznaczonych kilka wektorów przesunięcia, ponieważ różnice we współrzędnych punktów oczekiwanych i obliczonych nie są równomierne w obszarze ekranu. Współrzędne PoR są zatem obliczane na podstawie współrzędnych punktów fiksacji lewego i prawego oka oraz wartości przesunięcia uzyskanej w procesie kalibracji [76].

4.3.2 Pomiary charakterystyk opracowanego systemu

W niniejszym podrozdziale dokonano opisu charakterystyk istotnych w kontekście użytkowania opracowanego systemu. W pierwszej kolejności przedstawiono obliczenia inżynierskie pokazujące, że natężenie emitowanego promieniowania podczerwonego (promieniowania IR) spełnia określone normy bezpieczeństwa. Aspekt ten jest niezwykle istotny ze względu na stosowanie systemu w praktyce, nie tylko w warunkach laboratoryjnych. Drugą charakterystyką systemu opisaną w niniejszym rozdziale jest rozdzielczość przestrzenna Cyber-Oka. Rozdzielczość przestrzenna bezpośrednio wskazuje na dokładność systemu.

Pomiar natężenia promieniowania IR

System Cyber-Oko pracuje w zakresie fal podczerwieni, przy czym źródła promieniowania podczerwonego emitują światło o długości fali 860 nm. Jak wspomniano wcześniej system może pracować w dwóch trybach i w przypadku każdego z nich inna liczba diod IR emituje światło. W przypadku pracy systemu w trybie ciemnej źrenicy aktywnych jest osiem diod w każdym rogu ekranu, czyli w sumie trzydzieści dwie diody. Gdy system pracuje w trybie jasnej źrenicy aktywnych jest sześć diod zamontowanych na osi kamery oraz po cztery diody umieszczone w każdym z rogów ekranu, czyli łącznie dwadzieścia dwie diody. Zgodnie z normą EN 60825-1:2005 zastosowane w

systemie oświetlacze podczerwieni można zakwalifikować do jednej z dwóch poniżej przedstawionych klas laserów:

- lasery klasy 1 – lasery bezpieczne dla ludzkiego wzroku, umożliwiają spoglądanie na wiązkę przez przyrządy optyczne;
- lasery klasy 1M – lasery, dla których spoglądanie w wiązkę optyczną poprzez przyrządy optyczne jest zabronione [40].

Zakres bezpiecznego natężenia promieniowania dla długości fali 860nm oszacowano na podstawie modelu źródła promieniowania, dla którego przyjęto dwie wartości średnicy: 1 mm i 3 mm. Średnica równa 1 mm, zgodnie z normą 60825-1:2005 część 12 [40], jest typową wartością średnicy pojedynczej diody podczerwieni. Norma ta zakłada, iż sumaryczne natężenie promieniowane jest przez jedną diodę umieszczoną w osi oka. W przypadku gdy średnica źródła promieniowania przyjmuje wartość 3 mm podobnie zakłada się, że sumaryczne natężenie promieniowane jest przez jedną diodę umieszczoną w osi oka. Przyjęto, że długość średnicy równa 3 mm bardziej odpowiada oświetlaczowi złożonemu z kilku diod. W związku z powyższym dalszą analizę przeprowadzono dla modelu źródła promieniowania, którego średnica jest równa 3 mm. Dla tego modelu obliczono wartości graniczne natężenia promieniowania dla laserów klasy 1 i 1M. Uzyskane wyniki zestawiono w tab. 4.2 [78].

Tab. 4.2 Obliczone wartości maksymalnych dopuszczalnych dawek emisji promieniowania

Parametry źródła	MDE[4] dla klasy 1	MDE dla klasy 1M
Źródło o średnicy 3 mm (kąt źródła α=30mrad)	3,42 W/sr	3,42 W/sr

Parametry pracy diod podczerwieni zostały dobrane w taki sposób, aby natężenie prądu przez nie przepływającego nie przekraczało wartości 50mA (taka jest maksymalna wartość dopuszczalnego natężenia prądu, jaki może płynąć przez wybrane diody) przy napięciu do 5,25V (maksymalna wartość napięcia na pojedynczym złączu huba USB). Przy założeniu, że napięcie zasilania dostarczanego przez port USB jest równe 5V, wartość natężenia prądu płynącego przez każdą z diod jest równa ok. 45mA. Spe-

[4] MDE – maksymalna dopuszczalna dawka emisji promieniowania

cyfikacja techniczna wykorzystanych w Cyber-Oku diód[5] wskazuje, że dla prądu o natężeniu 50mA pojedyncza dioda wypromieniowuje od 50 do 100mW/sr (rozrzut technologiczny). Na potrzeby obliczeń założono najbardziej niekorzystny wariant, zgodnie z którym na osi oka umieszczona jest pojedyncza dioda o mocy pozwalającej na emisję promieniowania podczerwonego odpowiadającego 22 (jasna źrenica) oraz 32 (ciemna źrenica) diodom. W takim wariancie wynikowe wartości natężenia promieniowania wynoszą odpowiednio: 2,2W/sr oraz 3,2W/sr, jak zestawiono w tab. 4.3 [78].

Tab. 4.3 Wyznaczone wartości natężenia promieniowania dla modelu źródła o średnicy 3 mm

Tryb pracy systemu	Natężenie promieniowania
Jasna źrenica (22 diody)	2,2 W/sr
Ciemna źrenica (32 diody)	3,2 W/sr

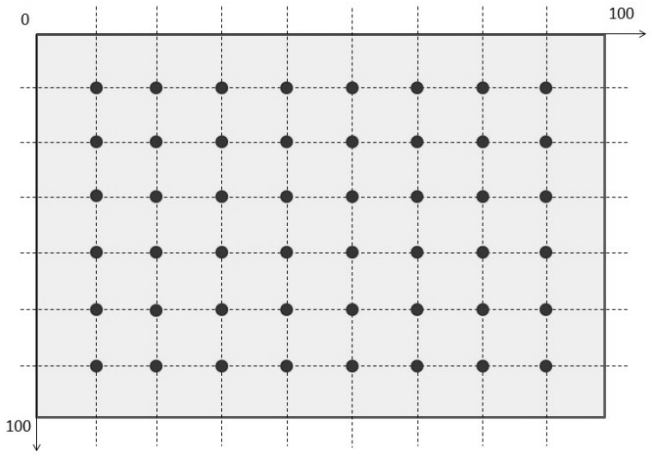
Porównanie wartości granicznych zestawionych w tab. 4.2 z wartościami wyznaczonymi dla systemu Cyber-Oko, przedstawionymi w tab. 4.3, pozwala wywnioskować, że system ten jest bezpieczny dla ludzkiego wzroku. W najbardziej niekorzystnym wariancie maksymalna dopuszczalna dawka emisji promieniowania nie zostaje przekroczona.

Rozdzielczość przestrzenna

W celu wyznaczenia rozdzielczości przestrzennej systemu Cyber-Oko wykonano badania testowe z udziałem studentów Wydziału Elektroniki, Telekomunikacji i Informatyki. W badaniach wzięło udział 163 studentów, którzy wcześniej nie obsługiwali tego typu systemu. W pierwszym etapie przeprowadzono proces kalibracji systemu, w tym kalibrację przestrzenną opisaną w podrozdziale 4.3.1. Badanie rozdzielczości przestrzennej Cyber-Oka w większości przypadków odbywało się w stabilnych warunkach, to znaczy przy zachowaniu stałej odległości pomiędzy głową badanego a monitorem. Wyniki badania, podczas którego badani nie zachowali stałej odległości zostały odrzucone i nie wzięto ich pod uwagę przy wyznaczaniu rozdzielczości przestrzennej syste-

[5] w systemie wykorzystano diody L-53SF6C

mu. Optymalna odległość badanego od ekranu została zdeterminowana przez konfigurację sprzętową Cyber-Oka i wynikała z zastosowanej ogniskowej obiektywu. Ogniskowa obiektywu równa 12 mm zapewniła uzyskanie ostrego obrazu oczu badanego przy zachowaniu odległości równej 60 cm. Warto również zaznaczyć, że odległość ta odpowiada warunkom komfortowej pracy użytkownika komputera. W badaniu wykorzystano monitory 19”, których szerokość i wysokość wynosiły odpowiednio 38 i 30 cm. Jak wspomniano wcześniej, współrzędne punktów fiksacji są zapisywane przez system CO zgodnie z założeniem, że początek układu współrzędnych znajduje się w lewym górnym rogu ekranu, jak pokazano na rys. 4.12. Każda badana osoba skupiała wzrok kolejno na 48 punktach testowych rozmieszczonych w sześciu wierszach po osiem punktów. Czas fiksowania wzroku na każdym z punktów wynosił 10s, zatem dla każdego punktu testowego zebrano 50 punktów fiksacji. Na rys. 4.12 przedstawiono równomierne rozłożenie punktów testowych na ekranie monitora.



Rys. 4.12 Punkty testowe wykorzystane w badaniu rozdzielczości przestrzennej Cyber-Oka

W związku z faktem, że nie wszyscy badani spełnili wymagania dotyczące dyscypliny przeprowadzenia eksperymentu, wyniki 48 badanych zostały odrzucone. Stabilne utrzymanie głowy w jednej pozycji podczas kalibracji i w trakcie eksperymentu nie było zbyt komfortowe, jednak zapewniało spełnienie optymalnych warunków dla przeprowadzanego badania. Ostatecznie do wyznaczenia rozdzielczości przestrzennej sytemu wzięto pod uwagę wyniki 115 badanych. Wyniki części badanych z tej grupy wskazywały na to, że nie zachowali oni w pełni wymaganej dyscypliny trzymania głowy w jednej pozycji, ale mimo to wzięto je pod uwagę przy wyznaczaniu rozdzielczości przestrzen-

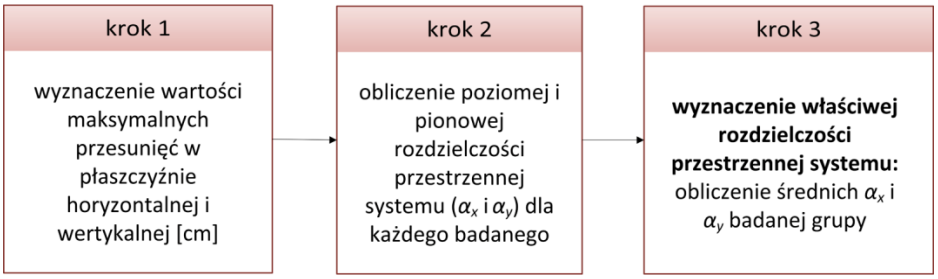
nej. Oszacowana w ten sposób dokładność systemu jest mniejsza, ale dzięki temu odpowiada rzeczywistej dokładności systemu podczas pracy w praktycznych warunkach. Przyjęto, że rozdzielczość przestrzenna systemu CO w płaszczyźnie horyzontalnej (a_x) i wertykalnej (a_y) ekranu zostanie wyznaczona w trzech krokach, zgodnie z algorytmem przedstawionym na rys. 4.13. Wartości rozdzielczości przestrzennych w drugim kroku algorytmu obliczono zgodnie z formułami 4.1 i 4.2.

$$\alpha_x = arctg \frac{x_{max}}{d} \tag{4.1}$$

gdzie: x_{max} – maksymalne przesunięcie punktu fiksacji wyznaczonego przez Cyber-Oko względem punktu wzorcowego w płaszczyźnie poziomej; d – odległość badanego od płaszczyzny ekranu monitora

$$\alpha_y = arctg \frac{y_{max}}{d} \tag{4.2}$$

gdzie: y_{max} – maksymalne przesunięcie punktu fiksacji wyznaczonego przez Cyber-Oko względem punktu wzorcowego w płaszczyźnie pionowej; d – odległość badanego od płaszczyzny ekranu monitora



Rys. 4.13 Algorytm wyznaczania rozdzielczości przestrzennej systemu Cyber-Oko

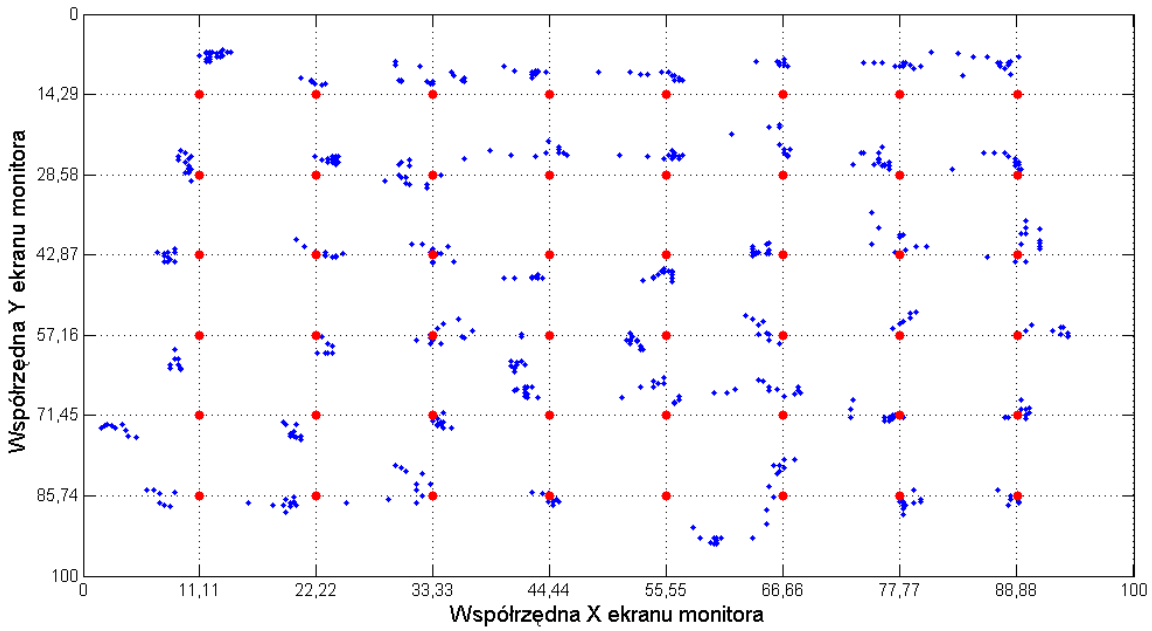
W tab. 4.4 zestawiono maksymalne wartości przesunięć oraz wyznaczone na ich podstawie wartości rozdzielczości przestrzennej w płaszczyźnie poziomej i pionowej. Maksymalne przesunięcia wybrano spośród wyników całej grupy testowej z uwzględnieniem wszystkich 48 punktów testowych. Dodatkowo w tabeli zamieszczono minimalne wartości przesunięcia pomiędzy punktem wyznaczonym przez system CO a punktem wzorcowym. W ten sposób można zauważyć, jak duży wpływ na dokładność systemu ma stabilne trzymanie głowy w jednej ustalonej pozycji.

Tab. 4.4 Zestawienie wyników badania rozdzielczości przestrzennej Cyber-Oka

	wartość przesunięcia [cm]		**rozdzielczość przestrzenna [°]**	
	x	y	α_x	α_y
min	2,08	2,67	1,99	2,55
max	5,8	5,12	5,52	4,88
średnia	3,48	3,54	**3,32**	**3,38**

Jak wynika z wartości zamieszczonych w powyższej tabeli, rozdzielczość przestrzenna systemu Cyber-Oko w płaszczyźnie horyzontalnej wynosi 3,32°, zaś w płaszczyźnie wertykalnej: 3,38°. Oznacza to, że dla przeciętnego użytkownika siedzącego w odległości 60 cm od płaszczyzny ekranu, system będzie efektywnie rozróżniał obszary o szerokości 3,48 cm i wysokości 3,54 cm. Warto w tym miejscu zaznaczyć, że dokładność systemu Cyber-Oko w konfiguracji z ramką imitującą ekran monitora i wyświetlaniu obrazu na ekranie projektora nie ulega pogorszeniu. Zmienia się natomiast rozmiar minimalnego rozróżnialnego obszaru, którego szerokość wynosi 16,53 cm, zaś wysokość 16,83 cm.

W celu graficznego przedstawienia dokładności systemu na rys. 4.14 przedstawiono rozłożenie punktów fiksacji wzroku wyznaczonych przez system Cyber-Oko dla losowo wybranego uczestnika badania. Czerwone punkty oznaczają punkty testowe (wzorcowe), na których badany skupiał wzrok. Z wartości zamieszczonych w tab. 4.4 i z wykresu przedstawionego na rys. 4.14 wynika, że system Cyber-Oko charakteryzuje się większą dokładnością wyznaczania PoR w płaszczyźnie horyzontalnej. Na podstawie rozkładu wyznaczonych punktów fiksacji można stwierdzić, że rozdzielczość przestrzenna w pionie maleje wraz z oddalaniem się punktu fiksacji od osi symetrii ekranu. Biorąc pod uwagę ten fakt, można przypuszczać, że ograniczenie pionowej rozdzielczości przestrzennej systemu wynika przede wszystkim z krzywizny gałki ocznej oraz pozycji kamery rejestrującej obraz oczu osoby badanej.



Rys. 4.14 Wizualizacja przykładowych wyników badania rozdzielczości przestrzennej Cyber-Oka

Warto jednak zauważyć, iż w kontekście przeprowadzonych w rozprawie badań, nie jest wymagana wysoka dokładność wyznaczania PoR. Zgodnie z założeniem, opisanym w podrozdziale 4.4.3, w interpretacji wyników istotna jest informacja o skupieniu wzroku badanego w zdefiniowanym obszarze obrazu, nie zaś precyzyjne wartości współrzędnych punktów fiksacji oka lewego i prawego. Poza tym, założono, że w trakcie eksperymentu badani powinni opierać głowę na specjalnej podstawce, której zadaniem było utrzymanie położenia głowy w stabilnej odległości od ekranu oraz na stałej wysokości względem powierzchni stołu. Oba założenia zdecydowanie zmniejszyły wymagania dotyczące rozdzielczości przestrzennej wykorzystanego systemu CO i umożliwiły jego zastosowanie w badaniach korelacji wzrokowo-słuchowych.

4.4 Badanie położenia wzroku w trójwymiarowym obrazie wizyjnym

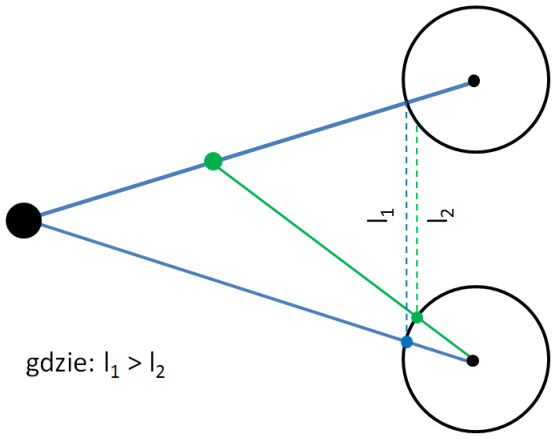
Ze względu na przeprowadzanie badań korelacji wzrokowo-słuchowych z wykorzystaniem trójwymiarowego materiału wizyjnego powstała konieczność śledzenia

wzroku osoby badanej fiksującej wzrok na trójwymiarowych elementach prezentowanego obrazu. Autor niniejszej rozprawy przeprowadził badania, których celem było opracowanie metodyki wyznaczania PoR w przestrzeni na podstawie danych uzyskanych z systemu śledzenia wzroku. W pierwszej kolejności zbadano metodę opisywaną obszernie w literaturze. Metoda ta polegała na wyznaczaniu punktu fiksacji w trzecim wymiarze na podstawie zmian odległości pomiędzy środkami źrenic. Po przeprowadzeniu eksperymentu, którego celem była wstępna weryfikacja efektywności tejże metody, autor zaproponował wyznaczanie punktu fiksacji w przestrzeni na podstawie zmian odległości pomiędzy punktami fiksacji wyznaczonymi niezależnie dla lewego i prawego oka. Przeprowadzono odpowiednie eksperymenty, w wyniku których zweryfikowano możliwość zastosowania tej metody w praktyce. Zauważono, że wyznaczanie współrzędnych punktów fiksacji oka lewego i prawego nie jest wystarczająco precyzyjne i zastosowanie Cyber-Oka w celu wyznaczania położenia wzroku w przestrzeni nie jest w praktyce możliwe. W związku z powyższym ostatecznie zaproponowano inne podejście śledzenia wzroku skupionego na obiektach wyświetlanego obrazu trójwymiarowego. Podejście to polegało na indeksacji treści obrazu wizyjnego i ostatecznie zostało wykorzystane w przeprowadzonych badaniach.

4.4.1 Metoda środków źrenic

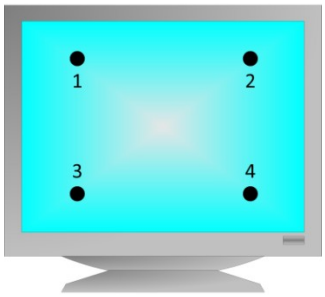
Metoda środków źrenic (ang. *Pupil Center Distance* – PCD) opiera się na zjawisku zbieżności gałek ocznych. Podczas fiksowania wzroku na obiektach znajdujących się coraz bliżej obserwatora, zwiększa się kąt zbieżności oczu. Zauważono, że wraz ze wzrostem kąta zbieżności maleje odległość pomiędzy środkami źrenic obu oczu. Ideę wyznaczania punktu fiksacji w oparciu o metodę PCD przedstawiono na rys. 4.15.

Koreańscy naukowcy, Ki i Kwon [64] [65] [84] [85], którzy przeprowadzili opisywany eksperyment, wykazali, że skuteczność wyznaczania PoR w przestrzeni w oparciu o metodę PCD jest większa niż 92,5%. Jednocześnie umieścili w swojej publikacji tabelę, zgodnie z którą różnica odległości środków źrenic w przypadku, gdy osoba badana fiksuje wzrok na obiekcie oddalonym o 500 mm i 2600 mm wynosi zaledwie 1,2 mm [64].



Rys. 4.15 Zasada wyznaczania punktu fiksacji wzroku w przestrzeni w oparciu o metodę PCD

Biorąc pod uwagę ograniczoną rozdzielczość przestrzenną systemu śledzenia wzroku, który został wykorzystany w opisywanym eksperymencie, należy przypuszczać, że wyznaczanie punktu fiksacji wzroku w przestrzeni w oparciu o analizę zmian odległości pomiędzy środkami źrenic nie jest optymalnym podejściem. Ponadto, warto zwrócić uwagę na fakt, że wartość kąta zbieżności oczu zawiera informację o skupieniu wzroku jedynie w płaszczyźnie horyzontalnej. Na tej podstawie bardzo trudno oszacować kierunek patrzenia (góra, dół) widza, nawet wówczas, gdy poprawnie została wyznaczona odległość obiektu, na którym osoba badana skupiała wzrok. Uwzględniając powyżej opisane ograniczenia i trudności wynikające z implementacji tej metody, zdecydowano się na jej odrzucenie. Niemniej jednak, autor rozprawy przeprowadził prosty eksperyment, mający na celu ostateczne zweryfikowanie efektywności metody PCD w kontekście zaimplementowania jej w systemie śledzenia wzroku CO. Eksperyment polegał na zebraniu próbek z obrazem oczu osoby skupiającej wzrok na obiektach, znajdujących się w różnych odległościach od badanego. Badanie przeprowadzono dla trzech odległości: 70 cm (10 cm za płaszczyzną ekranu), 60 cm (na płaszczyźnie ekranu) i 50 cm (10 cm przed płaszczyzną ekranu). W przypadku każdej odległości badany skupiał wzrok na czterech punktach, zgodnie z wzorcem przedstawionym na rys. 4.16.



Rys. 4.16 Rozmieszczenie punktów testowych na płaszczyźnie ekranu

W celu zarejestrowania obrazu oczu w trakcie fiksacji wzroku na punktach testowych „za płaszczyzną ekranu” zastosowano specjalną ramkę imitującą ekran CO przedstawioną na rys. 4.9b. Dla zebranych próbek obrazu wyznaczono z wysoką dokładnością odległości pomiędzy środkami źrenic. Rozdzielczość pobranych ramek obrazu wynosiła 1600x1200 pikseli, ponieważ obrazy rejestrowała kamera wykorzystywana w CO. W eksperymencie udział wzięły trzy osoby. Niemniej jednak z wartości, uzyskanych nawet dla tak małej grupy badanych można wysnuć wniosek, że wyznaczanie PoR w przestrzeni w oparciu o zaimplementowaną w systemie CO metodę PCD nie jest w praktyce możliwe. W tab. 4.5 zestawiono wyniki przeprowadzonego eksperymentu. Wartości w tabeli określają zmierzoną liczbę pikseli. Analiza tych wartości pozwala na sprecyzowanie następujących wniosków. Po pierwsze, porównując średnie wartości średnich odległości środków źrenic, wyznaczone niezależnie dla każdej z trzech głębi (wartości te zaznaczono w tabeli wytłuszczoną czcionką), można zauważyć, że są one bardzo do siebie zbliżone.

Ponadto, w wyniku porównania wartości średnich wyznaczonych dla odległości 60 i 50 cm uzyskuje się różnicę równą 2,75 piksela. Porównując różnice wartości średnich wyznaczonych dla każdego badanego niezależnie, największą wartość otrzymuje się dla badanego o numerze 2 przy porównaniu wartości odległości środków źrenic dla 60 i 50 cm. Różnica ta wynosi zaledwie 5 pikseli. Zatem efektywne wyznaczanie punktu fiksacji w przestrzeni zakłada, że system śledzenia wzroku byłby w stanie rozróżnić dwa punkty z dokładnością większą niż 1,5 mm, tak więc jego rozdzielczość przestrzenna powinna wynosić ok. $0{,}14^{o}$.

Tab. 4.5 Zestawienie wartości odległości pomiędzy środkami źrenic [px] dla różnych wartości głębi

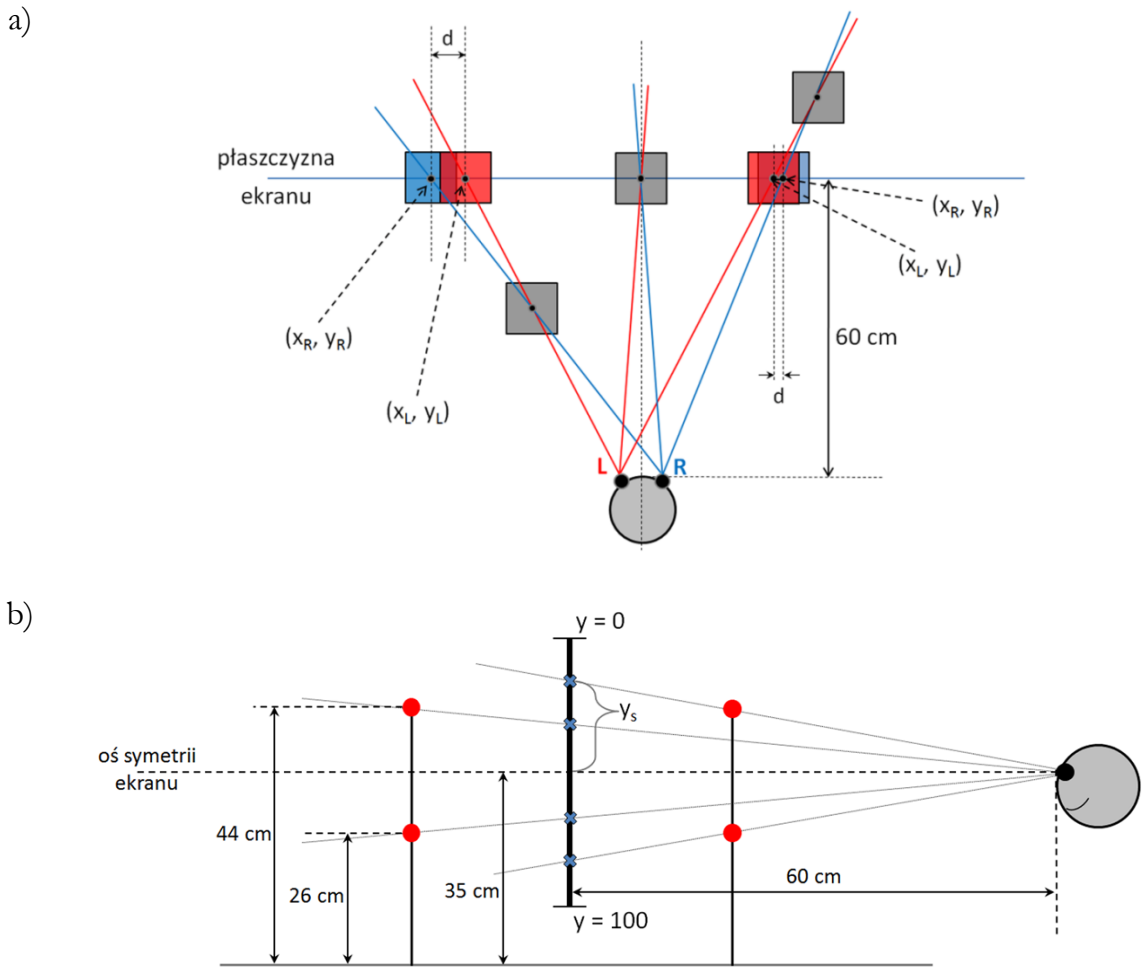
nr badanego	1	2	3		
	odległość = 70 cm			średnia	σ
1	525	531	546	534	10,82
2	523	530	546	533	11,79
3	525	529	542	532	8,89
4	524	529	543	532	9,85
średnia	524,25	529,75	544,25	**532,75**	
	odległość = 60 cm				
1	520	529	550	533	15,40
2	520	527	551	532,67	16,26
3	521	527	548	532	14,18
4	520	529	550	533	15,39
średnia	520,25	528	549,75	**532,67**	
	odległość = 50 cm				
1	517	524	543	528	13,45
2	518	522	545	528,33	14,57
3	519	524	550	531	16,64
4	519	522	556	532,33	20,55
średnia	518,25	523	548,5	**529,92**	

Kolejny istotny wniosek, jaki nasuwa się podczas analizy wartości zestawionych w tab. 4.5 wynika ze stosunkowo dużych różnic pomiędzy wartościami średnimi odległości pomiędzy środkami źrenic wyznaczonymi dla kolejnych badanych. Na przykład, gdy badani skupiali wzrok na obiekcie oddalonym od nich o 70 cm, średnia wartość odległości pomiędzy środkami źrenic dla badanego nr 2 wynosiła 529,75 pikseli, natomiast dla badanego nr 3 odpowiednio: 544,25 pikseli. Tak duża rozbieżność wartości średnich pomiędzy różnymi badanymi oraz potencjalnymi użytkownikami systemu oznacza konieczność kalibracji wykonywanej dla każdego użytkownika niezależnie. Ponadto, o konieczności przeprowadzania procedury kalibracyjnej świadczą również stosunkowo duże wartości odchylenia standardowego, którego maksymalną wartość jest równa 20,55. Przedstawiona powyżej analiza efektywności metody PCD została wykonana dla systemu śledzenia punktu fiksacji wzroku CO i nie podważa wyników uzyskanych przez Ki i Kwona, opisanych w ich publikacjach [64] [65] [84] [85]. Biorąc pod uwagę wyżej sprecyzowane wnioski, jak również rozdzielczość przestrzenną systemu CO

uznano, że wyznaczanie punktu fiksacji w 3D z wykorzystaniem tego systemu nie jest w praktyce możliwe.

4.4.2 Metoda paralaksy stereoskopowej

W związku z tym, że efektywność metody PCD, znanej z literatury, okazała się zdecydowanie niewystarczająca, autor rozprawy zaproponował metodę wyznaczania PoR w przestrzeni opartą na analizie współrzędnych punktów fiksacji, wyznaczonych niezależnie dla oka lewego i prawego. Idea tej metody opiera się na zjawisku paralaksy stereoskopowej[s] (ang. *stereoscopic parallax*), dlatego też zaproponowaną metodę nazywa się „metodą paralaksy stereoskopowej”. W wektorze parametrów wyznaczanych podczas pracy przez system CO znajdują się między innymi współrzędne punktów fiksacji wyznaczone niezależnie dla oka lewego i prawego. Wykorzystanie informacji o położeniu tych punktów na płaszczyźnie ekranu może posłużyć do wyznaczenia współrzędnych punktu fiksacji w przestrzeni. Istotnym założeniem tej metody jest to, aby osoba badana trzymała głowę w jednej pozycji względem płaszczyzny ekranu. Punkt fiksacji wzroku w przestrzeni[s] (ang. *3D gaze point*) jest wyznaczany na podstawie zależności geometrycznych. Oszacowanie współrzędnych (x, y, z) punktu fiksacji składa się z dwóch etapów: w pierwszym wyznaczany jest kierunek patrzenia w płaszczyźnie horyzontalnej, w drugim zaś – kierunek skupienia wzroku w płaszczyźnie wertykalnej. Na kierunek patrzenia w płaszczyźnie horyzontalnej bezpośrednio wpływa odległość od ekranu obiektu, na którym osoba badana skupia wzrok, zaś na kierunek patrzenia w płaszczyźnie wertykalnej bezpośredni wpływ ma umiejscowienie obiektu względem środka ekranu w płaszczyźnie góra-dół. Ideę wyznaczania PoR w przestrzeni w oparciu o metodę paralaksy stereoskopowej przedstawiono na rys. 4.17.



Rys. 4.17 Wyznaczanie punktu fiksacji wzroku w przestrzeni w oparciu o metodę paralaksy stereoskopowej: a) w płaszczyźnie horyzontalnej (rzut z góry), b) w płaszczyźnie wertykalnej (rzut z boku)

Jak wspomniano powyżej, system CO zwraca współrzędne punktu fiksacji wzroku niezależnie dla oka lewego i prawego, które oznaczono na rys. 4.17 odpowiednio jako (x_L, y_L) i (x_R, y_R). W zależności od tego, w jakiej odległości od widza (lub od ekranu) znajduje się obiekt, na którym fiksuje on swój wzrok, wartość różnicy współrzędnych x punktów fiksacji oka lewego i prawego (d) zmienia się zgodnie z formułą 4.3.

Warto zaznaczyć, że rys. 4.17a może stanowić również graficzną ilustrację powstawania efektu 3D w zależności od przesunięcia względem siebie lewej i prawej składowej obrazu stereoskopowego. Przyjęta nazwa omawianej metody wynika z tego właśnie podobieństwa.

$$d = x_R - x_L \tag{4.3}$$

gdzie: x_R – współrzędna odcięta punktu fiksacji wzroku wyznaczonego dla oka prawego; x_L – współrzędna odcięta punktu fiksacji wzroku wyznaczonego dla oka lewego

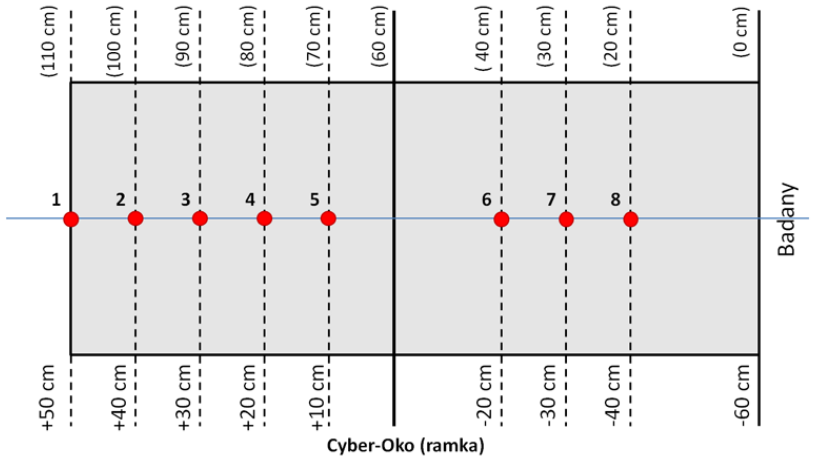
W ogólności przyjmuje się dwa zakresy wartości parametru *d*: *d>0* oraz *d<0*. Gdy *d>0*, punkt fiksacji oka lewego znajduje się po lewej stronie względem punktu fiksacji oka prawego w płaszczyźnie ekranu monitora. Wówczas jest to przypadek tzw. paralaksy pozytywnej, co w przypadku wyznaczania punktu fiksacji w przestrzeni oznacza, że widz skupia swój wzrok „za płaszczyzną” ekranu. Gdy parametr *d* przyjmuje wartości mniejsze od 0, występuje przypadek paralaksy negatywnej, ponieważ punkt fiksacji lewego oka znajduje się względem punktu fiksacji oka prawego po prawej stronie. W tej sytuacji przyjmuje się, że widz skupia swój wzrok „przed płaszczyzną” ekranu. Parametr *d* może jednak przyjmować również wartość równą 0, co oznacza, że widz skupia swój wzrok dokładnie na płaszczyźnie ekranu, na którym wyświetlany jest obraz. Wówczas współrzędne punktu fiksacji oka lewego są równe współrzędnym punktu fiksacji oka prawego.

W celu zweryfikowania efektywności zaproponowanej „metody paralaksy stereoskopowej” przeprowadzono dwa eksperymenty. W każdym z nich zadbano o stabilne położenie głowy osoby badanej, zarówno w płaszczyźnie przód-tył, jak i góra-dół. W tym celu zastosowano specjalną podstawkę, która zapewniała umieszczenie głowy na wysokości 35 cm od powierzchni stołu (wysokość ta odpowiadała środkowi ekranu) oraz w odległości 60 cm od zestawu Cyber-Oka. Ponadto wykorzystano ramkę imitującą ekran, o której wspomniano już przy opisie eksperymentu badającego efektywność metody środków źrenic. W rogach ramki umieszczone były diody emitujące światło w zakresie podczerwieni. Zastosowanie ramki pozwoliło osobom badanym skupiać wzrok na obiekcie fizycznym (metalowym pręcie), nie zaś na obiektach wirtualnych (powstałych w wyniku projekcji obrazu stereoskopowego). To z kolei umożliwiło przeprowadzenie badania w stabilnych, powtarzalnych warunkach. Na podstawie uzyskanych wyników wykonano symulacje, dzięki którym możliwe stało się wyznaczenie

zakresu głębi, w którym obliczanie punktu fiksacji w przestrzeni (w idealnych warunkach) jest optymalne. Konfigurację stanowiska badawczego przedstawiono na rys. 4.18.

Rys. 4.18 Konfiguracja stanowiska badawczego wykorzystującego ramkę imitującą ekran monitora Cyber-Oka

W pierwszym z eksperymentów zbadano przydatność metody, wyznaczając wartości parametru *d* w przypadku, gdy osoba badana skupiała wzrok na obiekcie umieszczonym w różnych odległościach od ekranu. Obiekt, na którym badany fiksował wzrok znajdował się na stałej wysokości, 35 cm od powierzchni stołu i pokrywał się dokładnie ze środkiem ramki imitującej ekran. Zbadano fiksowanie wzroku w ośmiu różnych odległościach od widza: 110 cm (50 cm za ramką imitującą ekran CO), 100 cm (40 cm za ramką), 90 cm (30 cm za ramką), 80 cm (20 cm za ramką), 70 cm (10 cm za ramką), 40 cm (20 cm przed ramką), 30 cm (30 cm przed ramką) i 20 cm (40 cm przed ramką). Ze względu na obecność kamery Cyber-Oka umieszczoną pod ramką imitującą ekran nie zbadano współrzędnych punktów fiksacji oka lewego i prawego przy skupianiu wzroku na obiekcie w odległości 50 cm od widza. Osoba badana była proszona o patrzenie na każdy z kolejnych punktów przez 20s. Oznacza to, że dla każdego punktu zebrano po 100 wektorów parametrów zapisywanych przez system CO. Rozłożenie punktów testowych, wykorzystanych w pierwszym eksperymencie przedstawiono na rys. 4.19. Eksperyment ten przeprowadzono z udziałem sześciu osób, pracowników Katedry Systemów Multimedialnych PG, którzy charakteryzowali się takim samym rozstawem oczu, wynoszącym 65 mm. Średnie wartości parametru *d* (różnicy pomiędzy współrzędną x_R i x_L), uzyskane w wyniku eksperymentu, zestawiono w tab. 4.6.



Rys. 4.19 Rozłożenie punktów testowych przed i za płaszczyzną ramki imitującej ekran Cyber-Oka w pierwszym eksperymencie (rzut z góry)

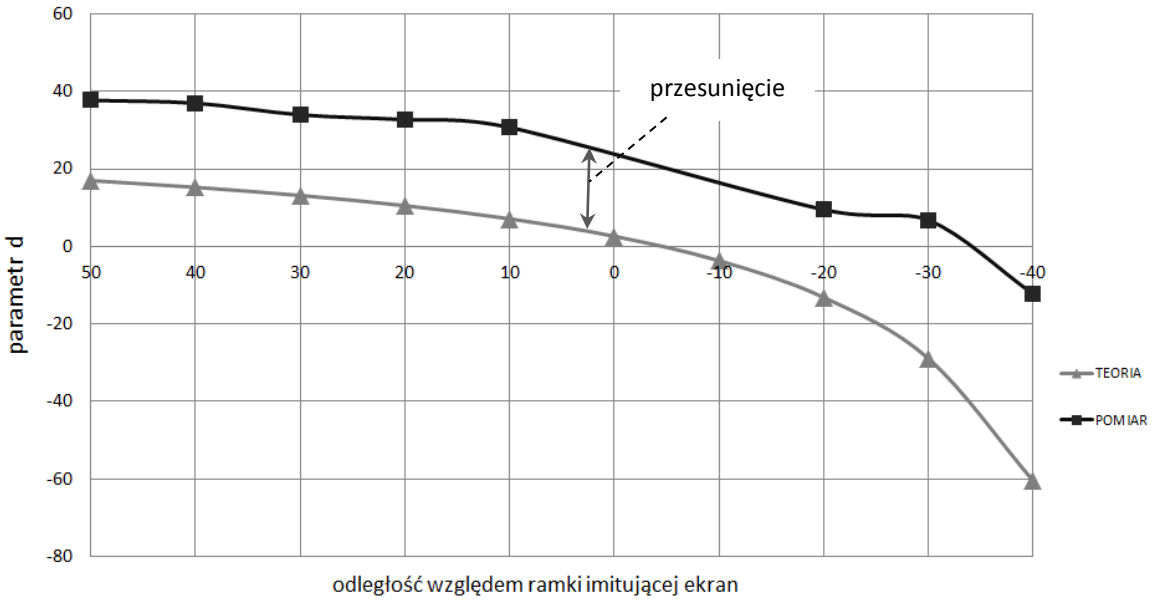
Należy pamiętać, że współrzędne PoR są wyznaczane przez system Cyber-Oko w jednostkach 1/100, przy założeniu, że całkowita szerokość ekranu wynosi 100 jednostek. Dodatkowo, na podstawie zależności geometrycznych wyznaczono teoretyczne wartości parametru *d* (oznaczone w tab. 4.6 jako d_t) dla kolejnych odległości. Zostały one zestawione w ostatniej kolumnie tab. 4.6.

Tab. 4.6 Zestawienie pomierzonych i teoretycznych wartości parametru *d* i $\measuredangle d$ dla różnych odległości

odległość	x_L	x_R	d	$\lvert\Delta d\rvert$	d_t	$\lvert\Delta d_t\rvert$
50	32,35	70,04	37,69	-	16,99	-
40	32,27	69,23	36,95	0,74	15,26	1,72
30	33,70	67,66	33,96	2,99	13,16	2,11
20	34,55	67,26	32,69	1,27	10,53	2,63
10	34,74	65,44	30,69	2,00	7,14	3,38
0	brak pomiaru	brak pomiaru	-	-	2,63	4,51
-10	brak pomiaru	brak pomiaru	-	-	-3,68	6,32
-20	43,76	53,26	9,49	-	-13,16	9,47
-30	49,45	56,23	6,78	2,71	-28,95	15,79
-40	58,19	45,87	-12,32	19,11	-60,53	31,58

gdzie: $|\measuredangle d|$ - różnica parametru *d* między kolejnymi odległościami, $|\measuredangle d_t|$ - różnica parametru *d* wyznaczonego teoretycznie między kolejnymi odległościami

Rys. 4.20 stanowi wizualizację wartości, jakie dla kolejnych odległości przyjmuje parametr *d*. Znacznikiem w kształcie „trójkąta" zaznaczono wartości obliczone na podstawie zależności geometrycznych (dane teoretyczne), zaś znacznikiem w kształcie „kwadratu" – wartości średnie wszystkich wyników uzyskanych w przeprowadzonym eksperymencie (dane pomiarowe). Porównując wartości teoretyczne z wartościami pomierzonymi, można stwierdzić, że w obu przypadkach zmiany parametru *d* mają dokładnie ten sam charakter. Warto zwrócić uwagę na stałe przesunięcie wartości uzyskanych w drodze pomiaru z wartościami teoretycznymi. Zauważono, że wielkość tego przesunięcia zmienia się dla każdego badanego. Oznacza to, że w procesie wyznaczania PoR w przestrzeni w oparciu o metodę paralaksy stereoskopowej konieczne jest wykonanie kalibracji w celu określenia wartości przesunięcia niezależnie dla każdego badanego.



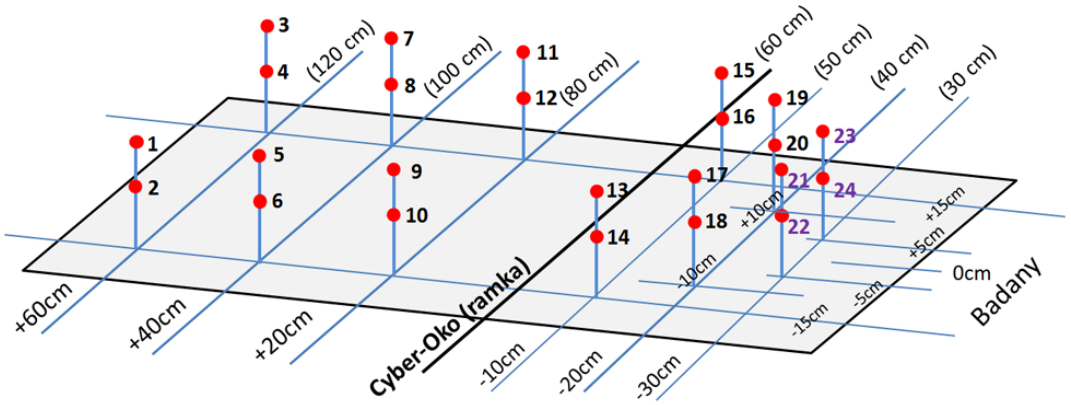
Rys. 4.20 Porównanie teoretycznych i pomierzonych wartości parametru *d* w funkcji odległości (eksperyment pierwszy)

Analizując wartości Δd, obliczone na podstawie pomierzonych wartości parametru *d*, można stwierdzić, że system CO byłby w stanie efektywnie rozróżniać położenie obiektu w zakresie od 30 do 40 cm przed monitorem, czyli od 30 do 20 cm przed widzem. Wymagana minimalna dokładność systemu w tym zakresie powinna zapewnić rozróżnianie obszarów o szerokości mniejszej niż 7,2 cm, zatem rozdzielczość przestrzenna CO w tym zakresie byłaby wystarczająca. W zakresie od 10 cm za ekranem do 10 cm przed ekranem, czyli w najbardziej interesującym obszarze z uwagi na wyświe-

tlany obraz trójwymiarowy, $\measuredangle d$ przyjmuje w przybliżeniu wartość równą 5 jednostek. Dla tak niewielkich różnic system śledzenia wzroku powinien zapewnić poprawne rozróżnianie obszarów o szerokości mniejszej niż 2 cm. W związku z powyższym, zastosowanie systemu CO do śledzenia aktywności wzrokowej widza oglądającego trójwymiarowy film nie jest w praktyce możliwe. Niemniej jednak, w celu dokładniejszego zbadania efektywności Cyber-Oka w kontekście wyznaczania położenia punktu fiksacji wzroku w przestrzeni, przeprowadzono drugi eksperyment. Eksperyment ten uwzględniał również zmiany wartości współrzędnej *y* punktu fiksacji oka lewego i prawego, zgodnie z metodyką przedstawioną na rys. 4.17b.

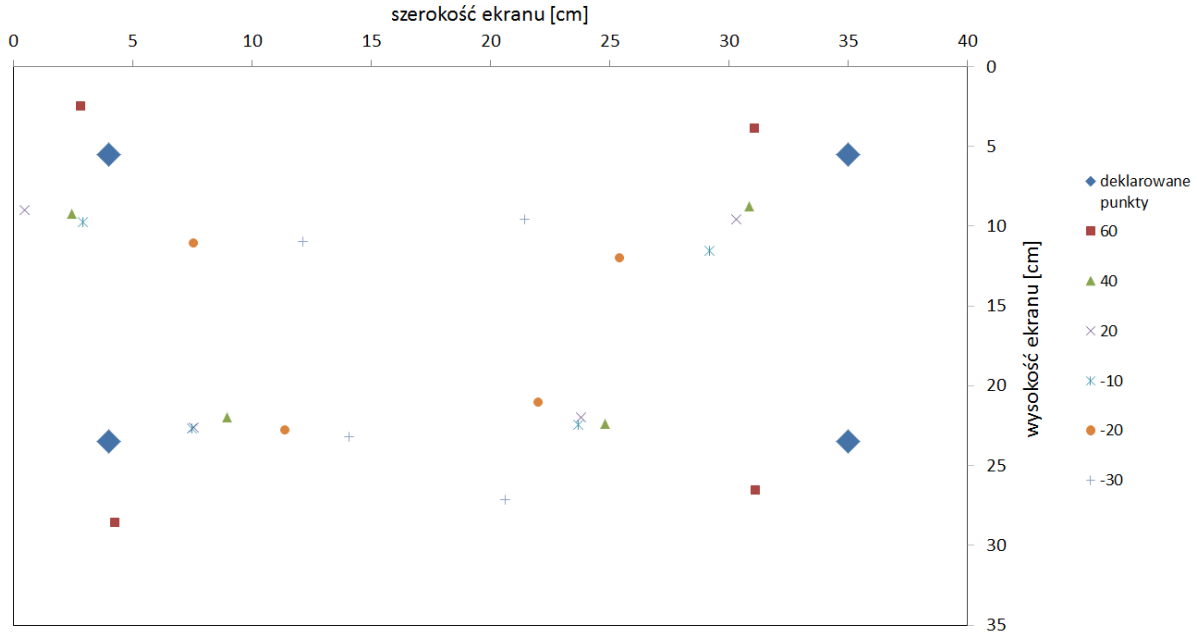
W drugim eksperymencie badani byli poproszeni o fiksowanie wzroku kolejno na 24 obiektach, przy czym obiekt znajdował się na dwóch wysokościach: 26 i 44 cm, czyli ±9 cm od środka ekranu, znajdującego się na wysokości 35 cm. Korzystając z wniosków pierwszego eksperymentu, punkty testowe przed ekranem umieszczono w mniejszych odległościach, ponieważ zmiany wartości parametru *d* w tym obszarze były zdecydowanie większe niż w obszarze za ekranem. Badanie współrzędnych punktów fiksacji oka prawego i lewego zbadano dla następujących odległości od widza: 120 cm (60 cm za ramką), 100 cm (40 cm za ramką), 80 cm (20 cm za ramką), 50 cm (10 cm przed ramką), 40 cm (20 cm przed ramką) i 30 cm (30 cm przed ramką). Ponadto, na każdej odległości położenie PoR badano w czterech miejscach. Umieszczono po dwa punkty testowe oddalone od środka ekranu o 15 cm z każdej strony. W przypadku punktów testowych znajdujących się w odległości 40 i 30 cm od widza, symetryczne rozłożenie punktów testowych względem środka ekranu zmniejszono (odpowiednio) do wartości: 10 i 5 cm. Rozłożenie punktów, wykorzystanych w drugim eksperymencie przedstawiono na rys. 4.21.

W wyniku przeprowadzonego eksperymentu zebrano współrzędne punktów fiksacji oka lewego i prawego sześciu badanych, którzy skupiali wzrok na obiektach znajdujących się w różnych odległościach od ekranu Cyber-Oka. Uzyskane w ten sposób wyniki zostały przetworzone i na ich podstawie obliczono współrzędne *x* i *y* punktów testowych (znajdujących się w różnych odległościach w przestrzeni), które zrzutowano na płaszczyznę ekranu Cyber-Oka.



Rys. 4.21 Rozłożenie punktów testowych przed i za płaszczyzną ramki imitującej ekran Cyber-Oka w drugim eksperymencie

Rys. 4.22 przedstawia uśrednione dla sześciu badanych współrzędne kolejnych punktów testowych. Znaczniki w kształcie rombu, widoczne na wykresie, oznaczają położenie punktów testowych w odległości odniesienia, czyli na płaszczyźnie ekranu.



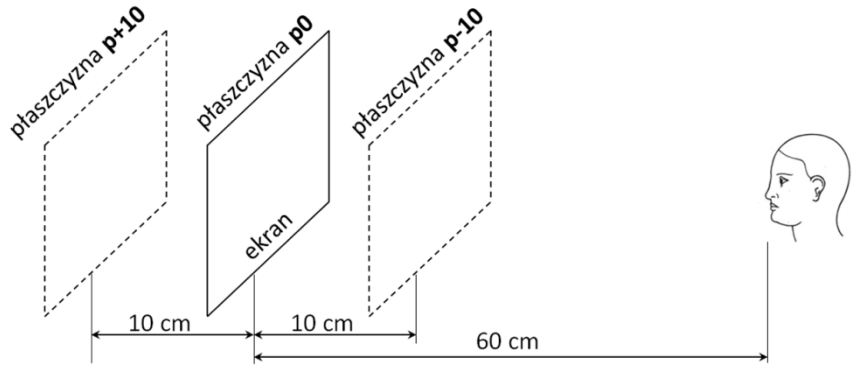
Rys. 4.22 Uśrednione punkty fiksacji wzroku wyznaczone dla kolejnych odległości zrzutowane na płaszczyznę ekranu Cyber-Oka

Położenie punktów na powyższym wykresie w ogólności odpowiada położeniu punktów testowych widzianych z perspektywy badanego. Jednak dokładna analiza wyznaczonych wartości współrzędnych punktów między kolejnymi odległościami, wskazuje na to, że zmiany wartości pomiędzy kolejnymi odległościami przyjmują charakter losowy. Podjęto decyzję o opracowaniu algorytmu wyznaczania współrzędnej *Z* na

podstawie współrzędnych punktów fiksacji oka lewego i prawego. Głównym celem opracowania algorytmu było przeprowadzenie eksperymentu w celu ostatecznej weryfikacji czy za pomocą systemu CO jest możliwe precyzyjne określenie odległości, w jakiej znajduje się obiekt, na którym osoba badana skupia swój wzrok. Na potrzeby eksperymentu powstała testowa wersja aplikacji, w której algorytm ten zaimplementowano. W pierwszej kolejności wprowadzone zostaną pojęcia istotne z punktu widzenia funkcjonalności algorytmu:

- płaszczyzna ‘p+10’^{s} – płaszczyzna równoległa do płaszczyzny ekranu, znajdująca się 10 cm za monitorem,
- płaszczyzna ‘p0’^{s} – płaszczyzna ekranu monitora,
- płaszczyzna ‘p-10’^{s} – płaszczyzna równoległa do płaszczyzny ekranu, znajdująca się 10 cm przed monitorem,
- offset^{s} – różnica między deklarowanymi a obliczonymi współrzędnymi punktu fiksacji,
- offset dla ‘p+10’^{s} – średni offset dla współrzędnej *Z* w płaszczyźnie ‘p+10’,
- offset dla ‘p-10’^{s} – średni offset dla współrzędnej *Z* w płaszczyźnie ‘p-10’.

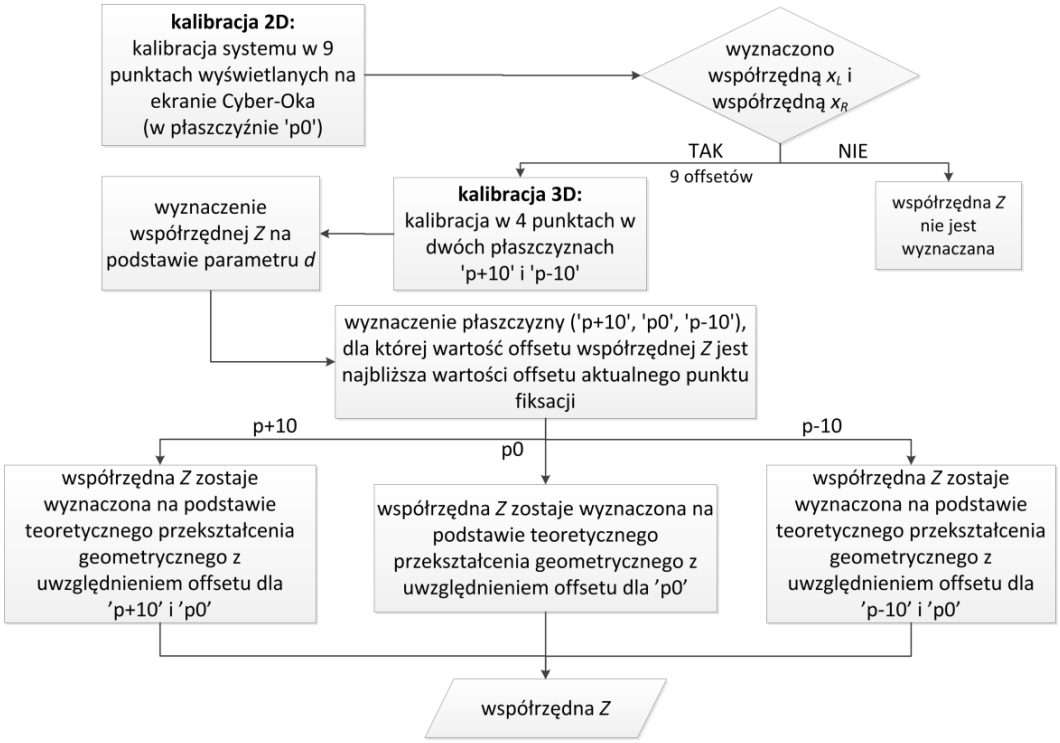
Na rys. 4.23 przedstawiono lokalizację poszczególnych płaszczyzn w przestrzeni.



Rys. 4.23 Lokalizacja płaszczyzn w badaniu punktu fiksacji wzroku w przestrzeni

W pierwszym kroku pracy z aplikacją wykonywana była standardowa kalibracja systemu Cyber-Oko (kalibracja 2D), w której badany skupiał wzrok na dziewięciu kolejno wyświetlanych punktach na ekranie. W wyniku tej kalibracji uzyskano dziewięć tak zwanych *offsetów*, czyli różnic pomiędzy deklarowanymi a wyznaczonymi współrzędnymi punktów fiksacji. Następnie wykonano kalibrację 3D, podczas której osoba

badana fiksowała wzrok w czterech punktach, znajdujących się w odległości 10 cm za płaszczyzną ekranu (płaszczyzna 'p+10') oraz w czterech punktach przed płaszczyzną ekranu, w tak zwanej płaszczyźnie 'p-10', przy czym rozkład punktów na płaszczyźnie był taki sam, jak na rys. 4.16. Po przeanalizowaniu danych uzyskanych w wyniku kalibracji 2D i 3D aplikacja jest w stanie wyznaczać współrzędną *Z*, zgodnie z algorytmem, którego schemat przedstawiono na rys. 4.24. W pierwszym kroku algorytmu wyznaczana jest przybliżona wartość współrzędnej *Z* na podstawie parametru *d* z uwzględnieniem poprawki kalibracyjnej obliczonej na początku pracy z aplikacją (kalibracja 2D). Następnie wyznacza się płaszczyznę, znajdującą się najbliżej oszacowanej wartości współrzędnej *Z*. Wybierana jest jedna z płaszczyzn: 'p+10', 'p0', lub 'p-10' – ta, dla której wartość offsetu współrzędnej *Z* jest najbliższa wartości offsetu aktualnego punktu fiksacji. W ostatnim kroku algorytmu współrzędna *Z* jest wyznaczana na podstawie teoretycznego przekształcenia geometrycznego z uwzględnieniem odpowiedniego dla wybranej płaszczyzny offsetu.



Rys. 4.24 Schemat procedury wyznaczania współrzędnej *Z* na podstawie punktów fiksacji oka lewego i prawego

W ostatnim eksperymencie badani byli proszeni o skupienie wzroku na obiekcie, który znajdował się dokładnie na środku wysokości i szerokości ramki imitującej ekran CO. Położenie punktu zmieniało się w płaszczyźnie przód-tył a odległości, w których

umieszczano obiekt ograniczono do następujących wartości: 80 cm (20 cm za ekranem), 70 cm (10 cm za ekranem), 40 cm (20 cm przed ekranem), 30 cm (30 cm przed ekranem) oraz 20 cm (40 cm przed ekranem). W eksperymencie udział wzięły trzy osoby. Stwierdzono jednak, iż wyznaczanie współrzędnych PoR w przestrzeni z wykorzystaniem systemu CO nie jest wystarczająco dokładne. Zaznaczyć należy, iż aplikacja umożliwiała wyświetlanie i zapis wartości chwilowych współrzędnej *Z* oraz wartości uśrednionych po określonej liczbie wyników. Wartości chwilowe współrzędnej *Z* wyznaczane przez aplikację dla kolejnych odległości punktu testowego od osoby badanej były bardzo niedokładne, często wręcz przypadkowe, pomimo stabilnego trzymania głowy osoby badanej w jednej pozycji. Na potrzeby eksperymentu wybrano zatem zapis wartości współrzędnej *Z* uśrednionej z 32 wyników (w przybliżeniu 6 sekund). W przypadku każdej badanej osoby do analizy wzięto pięć losowo wybranych wartości współrzędnej *Z* wyznaczonej dla każdej odległości. W tab. 4.7 zestawiono wyniki w postaci wartości średnich, obliczonych na podstawie pięciu pomiarów wartości współrzędnej *Z* oraz wartości odchylenia standardowego σ wyznaczonego dla tychże pomiarów. Ponadto, w tabeli zamieszczono wartości modułu błędu bezwzględnego obliczonego zgodnie z formułą 4.4.

$$|\Delta z_i| = |m_i - l_i| \tag{4.4}$$

gdzie: Δz_i – błąd bezwzględny; m_i – wartość średnia współrzędnej *Z*;
l_i – deklarowana odległość obiektu od ramki imitującej ekran CO

Analizując wyniki zestawione w tab. 4.7, można zauważyć, że w większości przypadków wartości współrzędnej *Z* wyznaczonej przez system CO znacznie odbiegają od wartości oczekiwanych. Co więcej, wartości błędu bezwzględnego mają losowy charakter w przypadku każdej badanej osoby. Nie jest zatem możliwe wyznaczenie stałego przesunięcia, którego wartość mogłaby posłużyć do kompensacji ewentualnego błędu powstałego przy wyznaczaniu współrzędnej *Z*. Warto zaznaczyć, że wartości w tab. 4.7 są obarczone stosunkowo małym błędem bezwzględnym. Wynika to przede wszystkim z faktu wyznaczania współrzędnej *Z* jako średniej 32 wartości chwilowych.

Tab. 4.7 Zestawienie uśrednionych wartości współrzędnej *Z*

	badany	1			2			3		
i	odległość	*m*	*σ*	$\lvert\measuredangle z\rvert$	*m*	*σ*	$\lvert\measuredangle z\rvert$	*m*	*σ*	$\lvert\measuredangle z\rvert$
1	20	25,50	11,30	5,50	-	30,97	-	24,04	1,08	4,04
2	10	18,41	12,81	8,41	15,27	7,56	5,27	9,36	5,45	0,64
3	0	6,57	5,49	6,57	-12,98	1,71	12,98	-2,03	3,42	2,03
4	-20	-15,98	0,46	4,02	-20,78	2,89	0,78	-14,74	4,82	5,26
5	-30	-25,55	1,55	4,45	-24,96	0,36	5,04	-25,76	0,76	4,24
6	-40	-36,44	0,37	3,56	-35,60	0,26	4,40	-38,02	0,45	1,98

Na podstawie wyników uzyskanych w drodze przeprowadzonych eksperymentów można stwierdzić, że wykorzystanie systemu CO do efektywnego wyznaczania punktu fiksacji wzroku w przestrzeni nie jest w praktyce możliwe. Wniosek ten odnosi się do obu zbadanych metod: metody środków źrenic PCD i tak zwanej metody paralaksy stereoskopowej. Zakłada się jednak, że obie metody pozwalają na śledzenie położenia wzroku w przestrzeni, ale wymagają systemu śledzenia wzroku charakteryzującego się zdecydowanie większą rozdzielczością przestrzenną niż system CO.

Wspomnieć należy również o innej metodzie wyznaczania PoR w przestrzeni. Pirri, Pizzoli i Rudi [107] zaproponowali obliczanie współrzędnych trójwymiarowego punktu fiksacji w oparciu o model geometryczny gałki ocznej. Uwzględniając jednak dokładność systemu CO w warunkach rzeczywistych (średnia wartość rozdzielczości przestrzennej: $3{,}35^{o}$), zrezygnowano z badania efektywności tej metody w kontekście zastosowania jej w śledzeniu uwagi wzrokowej widza na obrazie trójwymiarowym.

W następnym podrozdziale opisano podejście, które umożliwia śledzenie położenia wzroku na treści obrazu dwu- i trójwymiarowego, opierające się na wykorzystaniu systemu CO, którego dokładność w tym przypadku jest wystarczająca.

4.4.3 Indeksacja treści obrazu wizyjnego

W związku z tym, że przedstawione powyżej metody wyznaczania współrzędnych PoR w przestrzeni okazały się zbyt wymagające w aspekcie dokładności Cyber-Oka, zaproponowano inne rozwiązanie badania uwagi wzrokowej na obrazie trójwymiarowym. Głównym założeniem tej metody jest indeksacja treści oglądanego obrazu wizyj-

nego. Indeksacja treści polega na zdefiniowaniu obszarów obrazu, istotnych w kontekście przeprowadzanego badania oraz zdefiniowaniu przedziałów czasowych (interwałów), w których te obszary zawierają pożądaną treść. Takie podejście umożliwia śledzenie aktywności wzrokowej widza z wykorzystaniem dowolnego materiału wizyjnego – dwu- i trójwymiarowego. W przypadku obrazu 3D opis wybranego obszaru zainteresowania[s] (ang. *Region of Interest* – ROI) zostaje wzbogacony o dodatkowy atrybut wskazujący na położenie obiektu związanego z tym obszarem w trójwymiarowej scenie wyświetlanego obrazu. Ze względu na stabilne (w większości badanych próbek wizyjnych) położenie obiektów w scenie trójwymiarowej, zastosowano trzy wartości atrybutu określającego położenie obiektu w płaszczyźnie przód-tył (*depth3D*):

- „+” – za płaszczyzną ekranu,
- „0” – na płaszczyźnie ekranu,
- „-” – przed płaszczyzną ekranu.

Przypisanie do wybranego ROI atrybutu przyjmującego jedną z trzech powyżej wymienionych wartości pozwala w sposób efektywny śledzić położenie wzroku widza na trójwymiarowych elementach obrazu. Możliwe jest definiowanie wielu obszarów zainteresowania w wielu przedziałach czasowych w ramach jednej próbki wizyjno-fonicznej. Dane pozyskane w wyniku procesu indeksacji przechowywane są w strukturze XML i można je podzielić na trzy zasadnicze grupy:

- **metadane**[s] (reprezentowane przez znacznik *movieDescription*) – ogólne informacje o próbce wizyjno-fonicznej;
- **obszar**[s] (reprezentowany przez znacznik *area*) – określa wymiary ROI i przechowuje etykietę obszaru, wyrażany w pikselach;
- **przedział czasowy**[s] (reprezentowany przez znacznik *interval*) – określa przedział czasu, w którym wybrany obszar zainteresowania występuje, wyrażany w milisekundach.

Znacznik *movieDescription* przyjmuje następujące atrybuty: *width* – szerokość klatki obrazu wizyjnego w pikselach, *height* – wysokość obrazu wizyjnego w pikselach, *length* – długość próbki wizyjno-fonicznej w milisekundach, *filename* – nazwa pliku. Do atrybutów znacznika *interval* zalicza się dwa: *tstart* – określający początek przedziału czasowego, w

którym interesujący element obrazu wystąpił oraz *tend* – określający koniec tego przedziału. Wśród atrybutów znacznika *area* wymienić należy następujące: *x* – współrzędna odcięta punktu początkowego ROI, *y* – współrzędna rzędna punktu początkowego, przy czym punkt (0,0) znajduje się w lewym górnym rogu analizowanego obrazu, *width* – szerokość ROI, *height* – wysokość ROI, *label* – etykieta (krótki opis) ROI oraz *depth3D* – atrybut określający położenie obiektu w scenie trójwymiarowej. Przykładowe opisy próbek testowych (próbki nr 8, 14 i 20, przy czym próbka nr 8 jest pierwszą próbką wizyjno-foniczną, próbki 1-7 stanowią ścieżki dźwiękowe) zawarte w strukturze XML zamieszczono w Załączniku A, dołączonym do niniejszej rozprawy.

Graficzne przedstawienie procesu indeksacji obrazu wizyjnego, korespondujące ze strukturą XML, reprezentującego próbkę nr 14 przedstawiono na rys. 4.25. Celem indeksacji obrazu jest porównanie wartości współrzędnych punktów fiksacji wyznaczonych przez system CO ze współrzędnymi zawartymi w opisie poszczególnych obszarów zainteresowania.

Rys. 4.25 Indeksacja treści obrazu wizyjnego

W wyniku takiego porównania otrzymuje się informację o skupieniu wzroku osoby badanej na danym fragmencie obrazu w określonym przedziale czasu. Przykładowy wynikowy obraz wizyjny z naniesioną dynamiczną mapą przejść wskazującą na skupienie uwagi wzrokowej widza na twarzy bohatera przedstawiono w rozdziale 5. na rys. 5.13.

W rozdziale czwartym przedstawiono ogólne informacje dotyczące technik śledzenia wzroku, ale przede wszystkich scharakteryzowano wykorzystany w przeprowadzonych w ramach rozprawy badaniach system śledzenia wzroku CO. Przedstawiono algorytm przetwarzania obrazu wizyjnego, na podstawie którego system wyznacza kierunek patrzenia użytkownika. Scharakteryzowano część sprzętową systemu, odnosząc się do rekomendacji, zawierających informacje o dopuszczalnych przez normy dawkach emisji promieniowania podczerwonego, które jest wykorzystywane w systemie. W kontekście analizy aktywności wzrokowej widza oglądającego film trójwymiarowy przedstawiono metody wyznaczania punktu fiksacji wzroku w przestrzeni. Opisano badania, które w ramach tego wątku rozprawy zostały zaplanowane i przeprowadzone.

W następnym rozdziale scharakteryzowano wykorzystany w eksperymentach materiał badawczy oraz opisano wykonane badania – ich metodykę i sposób przeprowadzenia.

5 Badanie wpływu kierunku patrzenia na lokalizację pozornego źródła dźwięku

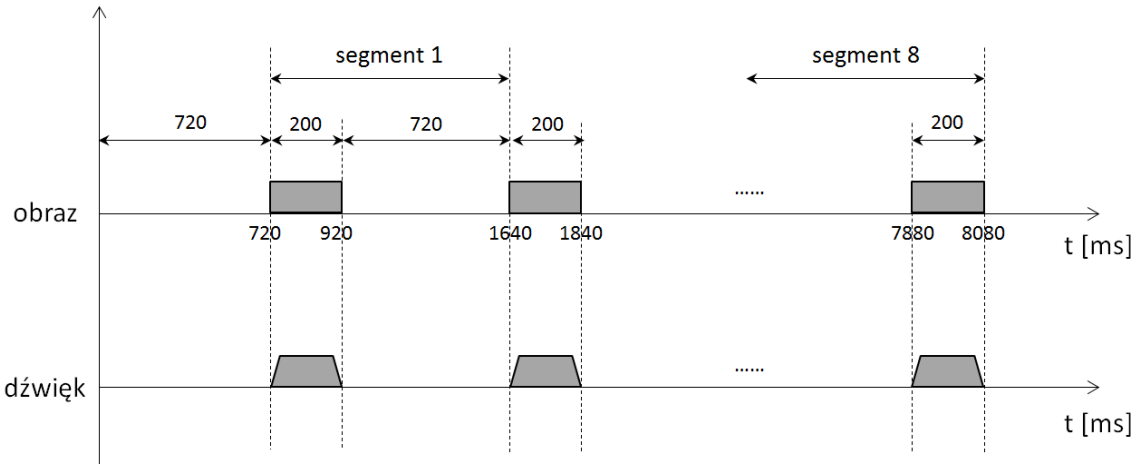
W niniejszym rozdziale zostały przedstawione badania, które przeprowadzono w ramach rozprawy doktorskiej. Celem wykonanych eksperymentów było zbadanie wpływu ściągającego obrazu na percepcję dźwięku, czyli zbadanie w sposób obiektywny wpływu kierunku patrzenia na lokalizację pozornego źródła dźwięku w panoramie stereofonicznej. Badania zostały przeprowadzone z wykorzystaniem dwóch systemów śledzenia wzroku. Pierwszą serię badań wykonano przy użyciu komercyjnego systemu Tobii T60. W drugiej serii wykorzystano natomiast system Cyber-Oko, opracowany w Katedrze Systemów Multimedialnych Politechniki Gdańskiej. Oba te systemy zostały dokładnie scharakteryzowane w rozdziale czwartym niniejszej rozprawy. Przeprowadzenie badań składało się z trzech etapów: opracowania materiału badawczego, skonfigurowania stanowiska badawczego oraz właściwych eksperymentów. Każdy z wyżej wymienionych etapów wymagał odpowiedniego przygotowania z uwzględnieniem wcześniej przyjętych założeń z nim związanych. W kolejnych podrozdziałach scharakteryzowano poszczególne etapy przeprowadzonych badań.

5.1 Opracowanie materiału badawczego

W pierwszym etapie związanym z przygotowaniem materiału badawczego sprecyzowano rodzaj materiału wizyjno-fonicznego, który został poddany badaniu. Założono, że badanie wpływu kierunku patrzenia na percepcję dźwięku powinno być przeprowadzone z wykorzystaniem trzech rodzajów próbek. Pierwszy rodzaj próbek umożliwił przeprowadzenie tak zwanego „testu podstawowego” (ang. *basic test*), na potrzeby niniejszej pracy określanego skrótem BT. BT może być przeprowadzany w różnych konfiguracjach [4] [8] [20] [124]. Wyodrębnić można jego charakterystyczne cechy: bodźce wzrokowe i słuchowe trwają bardzo krótko (120-200 ms), odstęp pomiędzy kolejnymi bodźcami jest równy wielokrotności czasu trwania pojedynczego bodźca. Ponadto, dźwięk i obraz są najczęściej prezentowane synchronicznie. Bodziec słuchowy może przyjmować różną postać. W badaniach opisanych w literaturze najczęściej

wykorzystywanym bodźcem wzrokowym były pojedyncze jednobarwne kwadraty lub koła (dyski), zaś bodźcem słuchowym – szum biały, dźwięk metronomu lub pojedyncze tony.

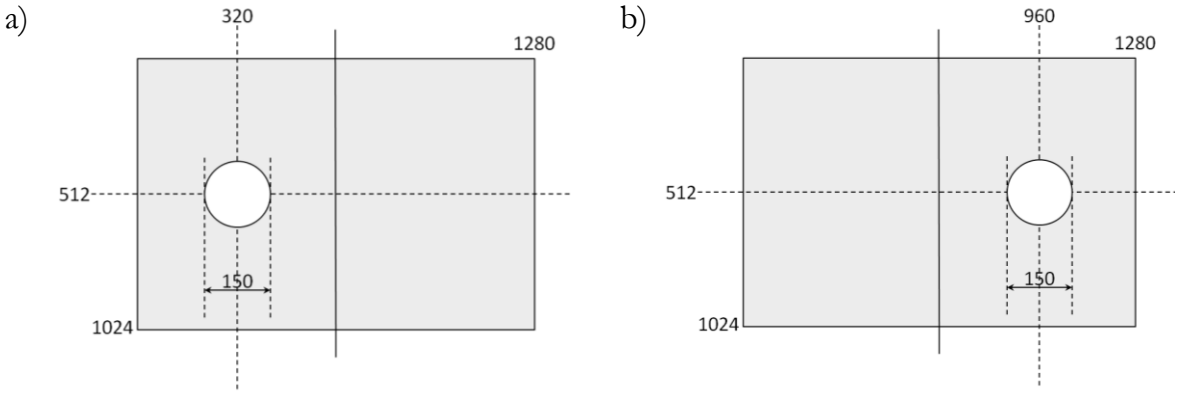
W przygotowanym na potrzeby badań rozprawy doktorskiej teście BT stymulatorem uwagi wzrokowo-słuchowej był biały dysk o średnicy 150 pikseli na czarnym tle, prezentowany synchronicznie z tonem prostym o częstotliwości 1 kHz. Próbka BT, podobnie jak pozostałe próbki wykorzystane w badaniu, była wyświetlana na ekranie monitora komputerowego o rozdzielczości 1280x1024 pikseli. Zastosowano łagodne zwiększanie i zmniejszanie poziomu dźwięku (ang. *fade in*, *fade out*) w celu wyeliminowania charakterystycznych trzasków. Czas trwania próbki wizyjno-fonicznej testu podstawowego wynosił 8080 ms. Próbka zawierała serię ośmiu segmentów „bodziec + przerwa”, przy czym przed pierwszym segmentem występowała przerwa o czasie trwania 720 ms. Rys. 5.1 przedstawia strukturę czasową testu podstawowego BT w sposób graficzny.



Rys. 5.1 Struktura czasowa testu podstawowego

Czas trwania bodźca wizyjno-fonicznego wynosił 200 ms, a wartość ta nie była dobrana w sposób przypadkowy. W związku z faktem, że próbka BT została przygotowana jako zwyczajny plik wizyjno-foniczny odtwarzany z częstotliwością 25 klatek na sekundę, czas trwania bodźca powinien być wielokrotnością liczby 25 w celu zachowania synchronizacji pomiędzy pobudzeniem wzrokowym i słuchowym. Próbka BT była odtwarzana w trzech różnych konfiguracjach w czasie badania. We wszystkich konfiguracjach ścieżka dźwiękowa próbki była taka sama, a bodziec słuchowy był monofo-

niczny, co oznacza, że w czasie odtwarzania próbki w dwukanałowym systemie stereofonicznym z lewego i prawego kanału emitowany był dźwięk o takim samym poziomie. W pierwszej konfiguracji prezentowana była tylko ścieżka dźwiękowa próbki BT. W drugiej konfiguracji ze ścieżką dźwiękową zsynchronizowany był bodziec wzrokowy w postaci białego dysku. Środek dysku pokrywał się z punktem o współrzędnych (320, 512), co oznacza, że współrzędna odcięta pokrywała się z ¼ szerokości ekranu, zaś współrzędna rzędna – z połową wysokości ekranu. W trzeciej konfiguracji próbki BT środek dysku pokrywał się z punktem (960, 512), czyli współrzędna odcięta leżała w ¾ szerokości ekranu. Na rys. 5.2 w sposób graficzny przedstawiono położenie białego dysku stymulującego uwagę wzrokową badanego zarówno w drugiej, jak i w trzeciej konfiguracji próbki BT. W punkcie 1 w Załączniku B niniejszej rozprawy scharakteryzowano poszczególne konfiguracje próbki BT.



Rys. 5.2 Położenie bodźca wzrokowego w badaniu podstawowym: a) druga konfiguracja próbki BT, b) trzecia konfiguracja próbki

Drugi rodzaj próbek wykorzystanych w przeprowadzonych badaniach stanowiły fragmenty rzeczywistych filmów. Przyjęto, że badaniu poddany zostanie zarówno konwencjonalny obraz wizyjny (film 2D), jak i wizyjny obraz stereoskopowy (3D) [41]. Badanie wpływu kierunku patrzenia na lokalizację pozornego źródła dźwięku (wpływ ściągający obrazu na percepcję dźwięku) z wykorzystaniem trójwymiarowego materiału wizyjnego stanowi innowację w stosunku do prowadzonych dotychczas badań korelacji wzrokowo-słuchowych w kontekście lokalizacji pozornego źródła dźwięku w panoramie stereofonicznej. Przygotowanie próbek wizyjno-fonicznych z obrazem 3D wymagało pozyskania filmów 3D w formacie umożliwiającym swobodny wybór pożądanych fragmentów nagrania. Zdecydowano się na wykorzystanie fragmentów filmów zapisa-

nych w formacie Blu-ray 3D[s] [157]. Rozdzielczość klatki każdej składowej (lewej i prawej) obrazu stereoskopowego zapisanego na nośniku Blu-ray w tym formacie wynosi 1920x1080 pikseli (powszechnie nazywana rozdzielczością „Full HD” – ang. *Full High Definition*). Ponadto, możliwe było przekonwertowanie filmu 3D z formatu Blu-ray 3D do tak zwanego formatu „side-by-side”. W formacie „side-by-side” lewa i prawa składowa obrazu są połączone ze sobą jedną krawędzią w płaszczyźnie poziomej z zachowaniem pełnego rozmiaru klatki każdej składowej. Dlatego rozdzielczość klatki obrazu stereoskopowego, którego każda ze składowych jest zapisana w rozdzielczości Full HD, wynosi 3840x1080 pikseli. Taki format zapisu obrazu stereoskopowego nie posiada w języku polskim odpowiedniego określenia, natomiast w literaturze anglojęzycznej nazywa się go formatem „side-by-side 100%”. Przykładowa klatka obrazu stereoskopowego w tym formacie została przedstawiona na rys. 5.3.

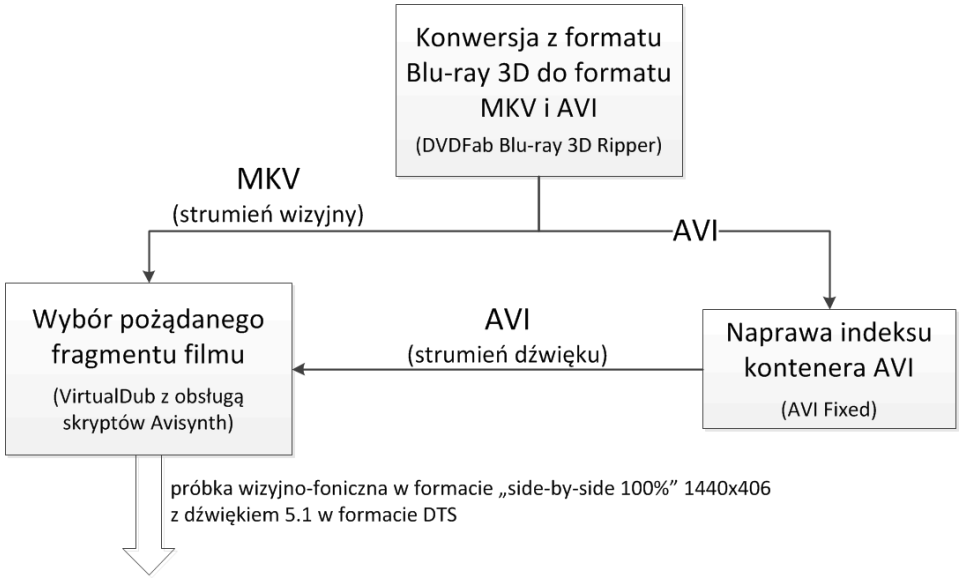
Rys. 5.3 Ramka obrazu 3D w formacie side-by-side 100%, rozdzielczość: 3840x1080 pikseli

W badaniach wykorzystano próbki wizyjno-foniczne stanowiące fragmenty filmów dostępnych na rynku (w czasie przygotowywania materiału badawczego) w formacie Blu-ray 3D. Pozyskano następujące filmy 3D: „Alicja w Krainie Czarów”, „Avatar”, „Piranha 3D” oraz „Resident Evil: Afterlife”. W celu przekonwertowania filmów z formatu Blu-ray 3D do łatwo edytowalnego formatu MKV czy AVI wykorzystano licencję na aplikację DVDFab Blu-ray 3D Ripper dostępną w zestawie narzędzi DVDFab 8. Program ten umożliwia konwersję materiału wizyjno-fonicznego zapisanego w formacie Blu-ray 3D do następujących formatów: MKV, AVI, MP4 oraz WMV. Pomimo, że program pozwala na zapisanie obrazu stereoskopowego w formacie „side-by-side 100%” o rozdzielczości 3840x1080 pikseli, zdecydowano się na zapisanie obrazu o rozmiarze klatki 1440x406 pikseli. Wybór klatki o takim rozmiarze wynikał bezpośred-

nio z ograniczonej mocy obliczeniowej systemu śledzenia wzroku, który jednocześnie wyświetlał próbkę wizyjno-foniczną oraz przetwarzał obraz z wbudowanej w nim kamery w celu wyznaczenia PoR. Problem ten dotyczył zarówno komercyjnego systemu Tobii T60, jak i opracowanego w Katedrze Systemów Multimedialnych systemu CO. Mniejszy rozmiar klatki umożliwił płynne odtwarzanie materiału badawczego przy jednoczesnym zapisie punktów fiksacji.

Poniżej przedstawiono kolejne kroki przygotowania materiału wizyjno-fonicznego. Ze względu na różne ograniczenia systemowo-technologiczne zastosowane podejście wymagało wielu działań. Po przekonwertowaniu filmów 3D do formatu AVI okazało się, że w przypadku każdego przekonwertowanego w ten sposób filmu, indeks kontenera AVI[s] jest niepoprawny. Choć możliwe było odtwarzanie pliku AVI z uszkodzonym indeksem, okazało się, że nie jest możliwe przewijanie filmu w trakcie odtwarzania oraz wybranie pożądanego fragmentu filmu i poprawne zapisanie go do nowego pliku. W związku z powyższym przekonwertowano wszystkie filmy 3D do formatu MKV, który pozwolił na swobodny wybór fragmentu filmu. Niestety, ze względu na ograniczenia wynikające z wymagań systemów Tobii T60 i CO, które współpracują jedynie z filmami w formacie AVI, konieczne okazało się zapisanie wybranego fragmentu filmu do formatu AVI. Wspomnieć należy, że wszystkie wymienione wyżej filmy posiadają ścieżkę dźwiękową przygotowaną w systemie dźwięku dookólnego 5.1. Założono, że w materiale badawczym pozyskanym z filmów 3D towarzyszący obrazowi dźwięk będzie przestrzenny (dookólny) – ang. *surround sound*. Program DVDFab Blu-ray 3D Ripper umożliwia zapisanie ścieżki dźwiękowej filmu w formacie DTS bez ingerencji w rozkład pozornych źródeł dźwięku w wielokanałowej panoramie stereofonicznej. Ścieżki dźwiękowe wybranych fragmentów filmów zapisano właśnie w formacie DTS. Program VirtualDub obsługujący skrypty AVS (Avisynth) pozwala na wczytanie filmów w formacie MKV i eksport do formatu AVI, ale niestety jednocześnie dokonuje automatycznej konwersji formatu ścieżki dźwiękowej „w dół" (ang. *downmix*) – z dźwięku sześciokanałowego (5.1) do dźwięku dwukanałowego (2.0). W związku z powyższym do programu VirtualDub wczytano dwa strumienie wizyjno-foniczne: jeden w formacie MKV, drugi w formacie AVI. Z pierwszego strumienia wyekstrahowano ścieżkę obrazu wizyjnego, z drugiego zaś – ścieżkę dźwiękową. Takie

podejście pozwoliło na pozyskanie materiału badawczego w postaci pożądanych fragmentów filmów w formacie „side-by-side 100%” o rozdzielczości klatki obrazu stereoskopowego 1440x406 pikseli i ścieżce dźwiękowej z dźwiękiem sześciokanałowym (5.1) zakodowanym w formacie DTS. Przedstawioną powyżej metodykę pozyskania próbek wizyjno-fonicznych w założonym wcześniej formacie zobrazowano w postaci schematu blokowego na rys. 5.4.



Rys. 5.4 Schemat blokowy pozyskania próbek wizyjno-fonicznych w formacie AVI w wyniku konwersji z formatu Blu-ray 3D

Niestety, pomimo przygotowania próbek wizyjno-fonicznych z dźwiękiem dookólnym, podczas badań odtwarzany był dźwięk dwukanałowy (2.0). Wynikało to bezpośrednio z faktu, że komercyjny system Tobii T60 nie był wyposażony w wielokanałową kartę dźwiękową, która umożliwiłaby odtworzenie dźwięku dookólnego zakodowanego w formacie DTS. Dlatego dla zachowania systematycznego charakteru badań w drugiej serii eksperymentów – z wykorzystaniem systemu CO, odtwarzano dźwięk dwukanałowy, pomimo że system śledzenia wzroku opracowany w Katedrze Systemów Multimedialnych umożliwia prowadzenie badań z wykorzystaniem materiału z dźwiękiem dookólnym. Jednak należy zaznaczyć, że praktycznie w całym przygotowanym materiale badawczym dźwięk dookólny nie odgrywał najważniejszej roli, ponieważ w kanałach LS (ang. *left surround*) oraz RS (ang. *right surround*), odpowiedzialnych za efekt przestrzenności dźwięku zawarta była jedynie informacja o tak zwanym tle akustycznym[s] prezentowanej sceny.

Ponieważ nie jest możliwe wyświetlanie obrazu trójwymiarowego na monitorze systemu Tobii T60 w technice innej niż anaglifowa, wybrane próbki wizyjno-foniczne przeznaczone do wyświetlania obrazu stereoskopowego, przygotowano właśnie w technice anaglifowej. Ostateczna rozdzielczość klatki obrazu stereoskopowego wynosiła zatem 720x406 pikseli. Poza tym, w przypadku wybranych próbek wizyjno-fonicznych stanowiących fragmenty filmów fabularnych poza próbkami z obrazem 3D przygotowano również próbki z obrazem 2D oraz próbki w wersji z napisami w języku polskim. Celem przygotowania różnych konfiguracji tej samej próbki było zbadanie wpływu trzeciego wymiaru obrazu oraz wpływu czytania polskich napisów na lokalizowanie pozornego źródła dźwięku w panoramie stereofonicznej. W Załączniku B w punktach 2–7 scharakteryzowano poszczególne próbki testowe, opracowane na podstawie fragmentów filmów 3D.

Poza próbkami testu podstawowego oraz fragmentami rzeczywistych filmów, badania przeprowadzono również z wykorzystaniem próbek wizyjno-fonicznych przygotowanych przez autora rozprawy. Próbki te stanowią fragmenty nagrania koncertu skrzypcowo-fortepianowego zarejestrowanego w technice stereoskopowej. Autor zarejestrował obraz trójwymiarowy z wykorzystaniem dwóch identycznych kamer (firmy Panasonic, model AG-HMC151E), umieszczonych na specjalnej podstawce, zapewniającej ich stabilne, równoległe położenie. W ramach dygresji warto zaznaczyć, iż wspomniana podstawka umożliwiająca równoległe zamontowanie dwóch kamer nie posiada jednoznacznego polskiego określenia czy terminu. Środowisko naukowe Politechniki Poznańskiej, opracowujące metody kodowania obrazów stereoskopowych konsekwentnie nazywa tę podstawkę „dubletem kamerowym" [37]. Poza tym określeniem można również spotkać się z terminem „rig kamerowy", który stanowi bezpośrednie przełożenie z języka angielskiego (ang. *camera rig*). To drugie określenie traktuje się jednak jako żargon techniczny środowiska filmowego. W związku z tym, że pojęcie „dublet" może kojarzyć się niejednoznacznie, podstawka umożliwiająca zamontowanie dwóch kamer na statywie będzie nazywana w ramach niniejszej rozprawy po prostu „podstawką kamerową"[s]. Dwie kamery zamontowane w sposób równoległy na „podstawce kamerowej" przedstawiono na rys. 5.5. Nagranie wizyjne koncertu zrealizowano zgodnie z podstawowymi zasadami rejestracji obrazu stereoskopowego. Kamery

umieszczone na „podstawce kamerowej” w minimalnej odległości od siebie. Odległość pomiędzy środkami obiektywów kamer, czyli tak zwana stereobaza[s] wynosiła 145 mm.

Rys. 5.5 Rejestracja obrazu stereoskopowego z wykorzystaniem „podstawki kamerowej”

Szerokość stereobazy w przypadku wykorzystanego zestawu rejestrującego obraz stereoskopowy wynikała z wymiarów kamer. Odległość równa 145 mm stanowi minimalną odległość, w jakiej mogą znajdować się środki obiektywów wykorzystanych kamer. Istnieje reguła, pozwalająca na określenie szerokości stereobazy, dla której percypowany przez widza efekt 3D nie wywołuje poczucia dyskomfortu. Jest to tak zwana „reguła 3%”, którą wykazano w sposób empiryczny [96] [146]. Zgodnie z „regułą 3%” szerokość stereobazy powinna wynosić 1/30 odległości pomiędzy zestawem rejestrującym a pierwszym planem znajdującym się na scenie. Oznacza to, że na każdy 1 m tej odległości powinno przypadać 25 mm szerokości stereobazy. Regułę tę można odwrócić w celu ustalenia odległości zestawu rejestrującego od obiektu znajdującego się w pierwszym planie. Zatem, zgodnie z „regułą 3%” optymalna odległość zestawu rejestrującego o szerokości stereobazy równej 145 mm wynosi 5,8 m. Ze względu na wymiary pomieszczenia i panujące w nim warunki, odległość ta nie została zachowana. W rzeczywistości zarejestrowano obraz trójwymiarowy w mniejszej odległości zestawu rejestrującego od pierwszego planu, ale w procesie postprodukcji (przygotowania obrazu stereoskopowego z wykorzystaniem zarejestrowanej lewej i prawej składowej) zmieniono rozsunięcie składowych w płaszczyźnie poziomej, co pozwoliło na uzyskanie prawidłowego efektu 3D (poprawnej fuzji). Wspomniana „reguła 3%” ma zastosowanie tylko w przypadku tworzenia wizyjnego obrazu stereoskopowego, który w założe-

niu ma być wyświetlany na niewielkich wyświetlaczach, takich jak monitory komputerowe czy ekrany telewizyjne [96] [146].

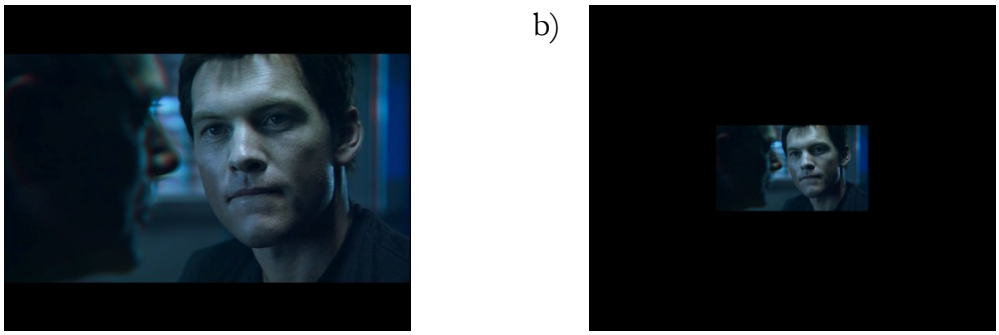
Należy pamiętać, że w przypadku rejestracji obrazu trójwymiarowego z wykorzystaniem dwóch kamer bardzo ważna jest kwestia synchronizacji zapisu obrazu i dźwięku przez kamery. W przypadku nagrania koncertu skrzypcowo-fortepianowego zastosowano jednocześnie dwie proste metody synchronizacji. Po pierwsze, zapis w kamerach uruchomiono za pomocą pilota. W nagraniu wykorzystano dwie identyczne kamery, zatem ich włączenie nastąpiło w tym samym momencie. Ponadto, po każdym włączeniu zapisu w kamerach klaśnięto dłońmi. Klaśnięcie jest widoczne w przebiegu czasowym sygnału dźwiękowego (ang. *waveform*) jako charakterystyczny „pik" o wysokiej amplitudzie i krótkim czasie trwania (charakter impulsowy). Dwa strumienie wizyjne lewej i prawej składowej obrazu stereoskopowego można w stosunkowo łatwy sposób z sobą zsynchronizować właśnie na podstawie położenia tego „piku" w ścieżkach dźwiękowych obu strumieni. Podczas rejestracji obrazu trójwymiarowego zadbano o zachowanie identycznych ustawień (na obu kamerach) następujących wielkości/parametrów: balans bieli (ang. *white balance* – WB), wartość przysłony (ang. *iris*), długość ogniskowej (ang. *focal length*). Zestaw rejestrujący ustawiono względem sceny w taki sposób, aby każdy z instrumentów zajmował odpowiednią część kadru – skrzypce znajdowały się w lewej części, zaś pianista – w prawej części kadru. Dźwięk zarejestrowano za pomocą przenośnego rejestratora dźwięku Zoom H4 (2 kanały, częstotliwość próbkowania: 48000 Hz, rozdzielczość bitowa: 16 bitów).

Jak wspomniano we Wprowadzeniu, jednym z wątków przeprowadzonych w ramach niniejszej rozprawy eksperymentów było zbadanie zjawiska skalowalności wpływu ściągającego obrazu na percepcję dźwięku, zgodnie z którym wielkość wyświetlacza, na którym prezentowany jest materiał wizyjny, nie wpływa istotnie na wpływ ściągający obrazu na percepcję dźwięku. W celu przeprowadzania badań skalowalności wybrane próbki testowe były prezentowane badanym w 3 różnych konfiguracjach. Wyjściowy rozmiar klatki obrazu miał wymiar 1280x1024 pikseli. Próbki wizyjne o takim rozmiarze klatki powstały w wyniku przeskalowania klatki obrazu o rozmiarze 720x406 pikseli do tego rozmiaru z zachowaniem oryginalnych proporcji obrazu. Materiał badawczy o takich wymiarach klatki wyświetlano w dwóch różnych ustawieniach stanowiska ba-

dawczego. W pierwszym ustawieniu stanowisko badawcze było skonfigurowane typowo, to znaczy obraz był wyświetlany na ekranie monitora systemu śledzenia wzroku, a badany był oddalony od ekranu o 60 cm. W drugim ustawieniu badany oglądał obraz wyświetlany na ekranie projektora i był oddalony od niego o 285 cm. Dokładniejszą charakterystykę tych dwóch konfiguracji stanowiska badawczego zamieszczono w podrozdziale 5.2. W trzeciej konfiguracji prezentacji materiału badawczego szerokość klatki obrazu prezentowanego badanemu była ponad 2,5 razy mniejsza w porównaniu z szerokością klatki w konfiguracji typowej i wynosiła 480 pikseli. Podobnie, jak w przypadku próbek wizyjno-fonicznych o wyjściowym rozmiarze klatki obrazu równym 1280x1024 pikseli, zastosowano tak zwaną kaszetę[s], czyli czarny margines w górnej i dolnej części ekranu (ang. *letterbox*) oraz odpowiednio szeroki czarny margines po lewej i prawej stronie obszaru wyświetlania właściwego obrazu. Takie rozwiązanie zapewniło poprawną projekcję materiału badawczego zarówno na monitorze systemu Tobii, jak i systemu CO. Warto zaznaczyć, że obraz w próbkach wizyjno-fonicznych, których szerokość klatki została pomniejszona 2,5-krotnie w stosunku do klatki wyjściowej przygotowano zgodnie z zasadą zmniejszania rozmiaru klatki obrazu z zachowaniem jego zawartości.

Współcześnie, dopasowanie zawartości obrazu do wymiarów wyświetlacza, niezależnie czy jest to ekran telewizora, monitora komputerowego czy przenośnego urządzenia multimedialnego, nie stanowi problemu. W przeszłości natomiast był to znaczący problem. Telewizyjna emisja filmu, który był realizowany z myślą o projekcji w kinie, wymagała specjalnego przygotowania materiału wizyjnego. Opracowana została wówczas technika nazywana w literaturze anglojęzycznej „pan & scan” (pol. – w wolnym tłumaczeniu – *panoramowanie i skanowanie obrazu*). Technika „pan & scan” polegała na dopasowaniu obrazu panoramicznego do wyświetlania na ekranie o innych proporcjach (najczęściej 4:3) poprzez usuwanie bocznych fragmentów obrazu. Często pozostawiany był środkowy fragment obrazu, jednak stosowało się również technikę kadrowania dynamicznego, polegającą na tym, że fragment obrazu poruszał się, podążając za akcją lub po prostu za istotnym elementem danej sceny [169]. Przykładowe klatki próbek testowych przedstawiono na rys. 5.6.

a) b)



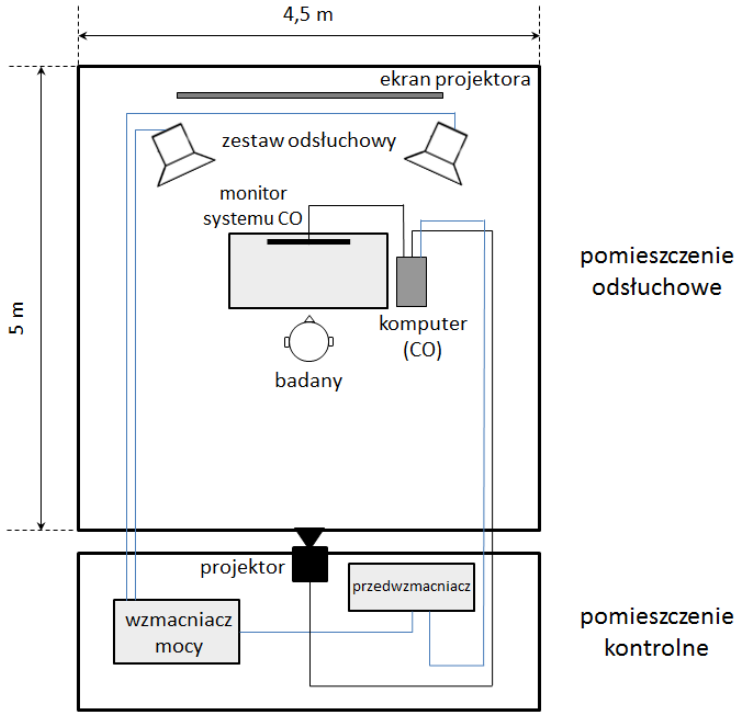
Rys. 5.6 Różne konfiguracje próbki testowej
a) szerokość klatki i szerokość obszaru wyświetlania: 1280 pikseli; b) szerokość klatki: 1280 pikseli, szerokość obszaru wyświetlania: 480 pikseli

Należy wspomnieć również o zastosowanym kodeku wizyjnym, którym skomprymowano próbki wizyjno-foniczne, stanowiące materiał badawczy przeprowadzonych eksperymentów. Wszystkie próbki zostały skomprymowane kodekiem XviD MPEG-4 w trybie kodowania jednoprzebiegowego (ang. *single pass*) z ustawieniem parametru 'jakość' na wartość maksymalną.

5.2 Stanowisko badawcze

Wszystkie badania w ramach niniejszej rozprawy zostały przeprowadzone w jednym pomieszczeniu z zachowaniem stabilnych warunków odsłuchowych. Eksperymenty przeprowadzono w sali laboratoryjnej nr 728 Katedry Systemów Multimedialnych. Sala ta składa się z dwóch pomieszczeń: pomieszczenia odsłuchowego (ang. *auditory room*) i kontrolnego (ang. *control room*). Część odsłuchowa, w której prowadzono badania miała wymiary: 4,5 x 5 m. W pomieszczeniu tym odbywają się zajęcia laboratoryjne, w ramach których studenci m.in. nagrywają lektorów. Ta część sali jest dobrze wytłumiona i można prowadzić w niej testy odsłuchowe z wykorzystaniem dwukanałowego (2.0), sześciokanałowego (5.1) czy ośmiokanałowego (7.1) systemu odsłuchowego. W związku z ograniczeniem systemu Tobii, który nie był wyposażony w wielokanałową kartę dźwiękową, właściwe badania przeprowadzono w powszechnie stosowanej stereofonii dwukanałowej. Wspomnieć należy, że pomieszczenie odsłuchowe jest również uniezależnione od światła dziennego, dlatego badania mogły odbywać się w przyciemnionym pomieszczeniu. Dzięki temu w trakcie eksperymentów badani mogli koncentrować się na treści wizyjnej materiału badawczego. W części odsłuchowej

znajdował się zestaw głośników stereofonii wielokanałowej NEXO, ekran projektora i stół, na którym stał monitor systemu Tobii (w pierwszej serii testów) lub systemu CO (w drugiej i trzeciej serii testów), specjalna podstawka, będąca elementem lampy szczelinowej (wykorzystywanej w gabinetach okulistycznych) oraz komputer przenośny, na którym badani uzupełniali formularz wskazując percypowane położenie pozornego źródła dźwięku. Na rys. 5.7 przedstawiono rozkład i wyposażenie pomieszczeń sali laboratoryjnej, w której przeprowadzono wszystkie badania związane z niniejszą rozprawą.

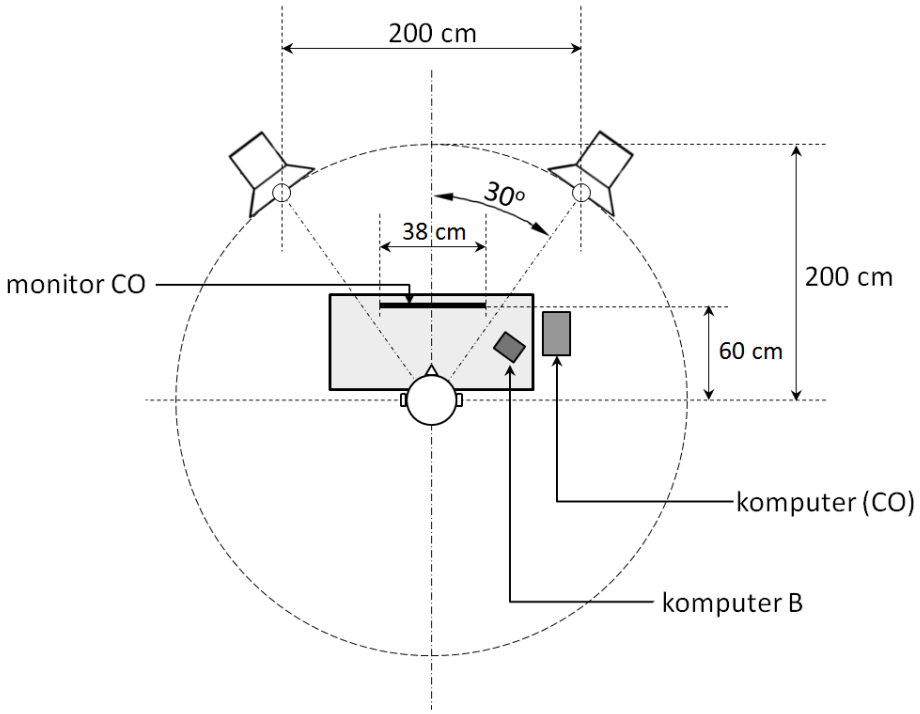


Rys. 5.7 Pomieszczenia sali 728 KSM, w której przeprowadzono badania korelacji wzrokowo-słuchowych (rzut z góry)

Jak wspomniano w podrozdziale 5.1, w związku z badaniem zjawiska skalowalności wpływu ściągającego obrazu na percepcję dźwięku, materiał badawczy był prezentowany osobom badanym w dwóch różnych konfiguracjach stanowiska badawczego. W pierwszej konfiguracji (wyjściowej) badani patrzyli na ekran monitora systemu CO, w drugiej zaś – na ekran projektora. Wraz ze zmianą położenia głowy badanego względem płaszczyzny ekranu zmieniały się warunki odsłuchowe. W pierwszej konfiguracji stanowiska badawczego punkt najlepszego odsłuchu (ang. *sweet spot*) znajdował się w odległości 200 cm od głośników zestawu odsłuchowego. W związku z tym, że druga konfiguracja wymagała odsunięcia stanowiska badawczego od ekranu, czyli zwiększenia

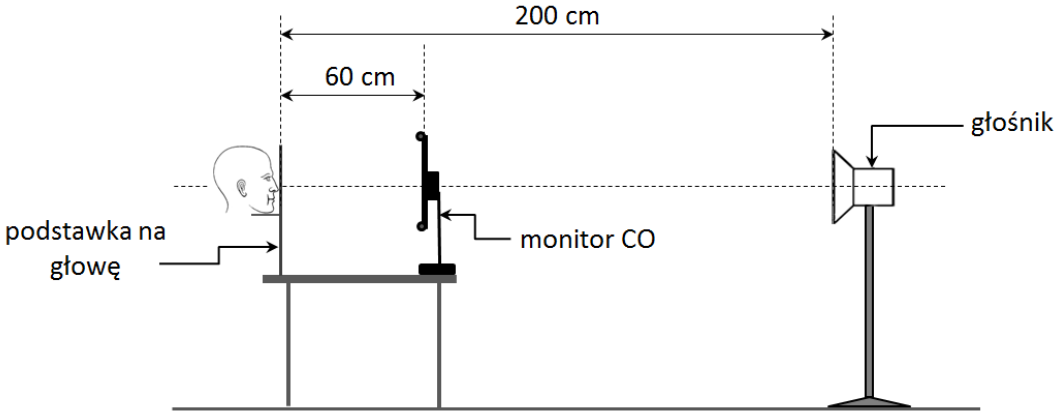
odległości pomiędzy ekranem a badanym, zmieniło się również rozstawienie zestawu odsłuchowego. Promień koła, na obrysie którego umieszczono głośniki miał długość 255 cm. Zarówno pierwsza, jak i druga konfiguracja dwukanałowego systemu odsłuchowego była zgodna ze specyfikacją normy ITU-R BS.1116-1 [60]. Według zaleceń tej rekomendacji promień koła, na którym rozstawione są głośniki, czyli tak zwana szerokość bazy (ang. *base width*) zawiera się w przedziale 200-300 cm. W przeprowadzonych w ramach rozprawy eksperymentach badano wpływ bodźca wzrokowego na percepcję bodźca słuchowego. Zatem z założenia w części odsłuchowej pomieszczenia, w którym prowadzone były eksperymenty, powinien znajdować się nie tylko zestaw odsłuchowy, ale również ekran, na którym mogą być wyświetlane próbki testowe. W celu poprawnego skonfigurowania stanowiska badawczego przeznaczonego do badań z wykorzystaniem bodźców wzrokowych i słuchowych zapoznano się ze specyfikacją normy ITU-R BS.1286 [61]. Zgodnie z tą rekomendacją odległość pomiędzy widzem a ekranem powinna być wielokrotnością wysokości ekranu zawierającą się w zakresie 3-6. Niestety, ani w pierwszej, ani w drugiej konfiguracji stanowiska badawczego warunek ten nie został do końca spełniony. Głównym powodem było ograniczenie wynikające z położenia głowy badanego względem monitora systemu CO. Jak wspomniano w rozdziale trzecim, optymalna odległość głowy badanego od płaszczyzny ekranu, zapewniająca największą dokładność wyznaczania PoR, wynosi 60 cm. Zatem spełnienie podstawowego warunku pracy systemu CO uniemożliwiło spełnienie normy, wskazującej na optymalną odległość między widzem i ekranem podczas prowadzenia badań z zakresu korelacji wzrokowo-słuchowych. Wysokość ekranu monitora systemu CO wynosiła 30 cm, zatem stosunek odległości widza od monitora do wysokości ekranu wynosił 2. W drugiej konfiguracji stanowiska badawczego odległość widza od ekranu projektora wynosiła 285 cm, a wysokość obszaru, na którym wyświetlany był obraz – 135 cm. Zatem, w tej konfiguracji stosunek odległości widza od monitora do wysokości ekranu był równy 2,11. Warto zaznaczyć, iż dołożono wszelkich możliwych starań, aby badania odbywały się w pełni optymalnych warunkach. Na rys. 5.8 przedstawiono rzut z góry i z boku pomieszczenia odsłuchowego w pierwszej konfiguracji stanowiska badawczego (z wykorzystaniem standardowego wyświetlacza – monitora systemu CO).

a)



komputer B – komputer przeznaczony do udzielania odpowiedzi na temat położenia pozornego źródła dźwięku (wypełnianie formularza)

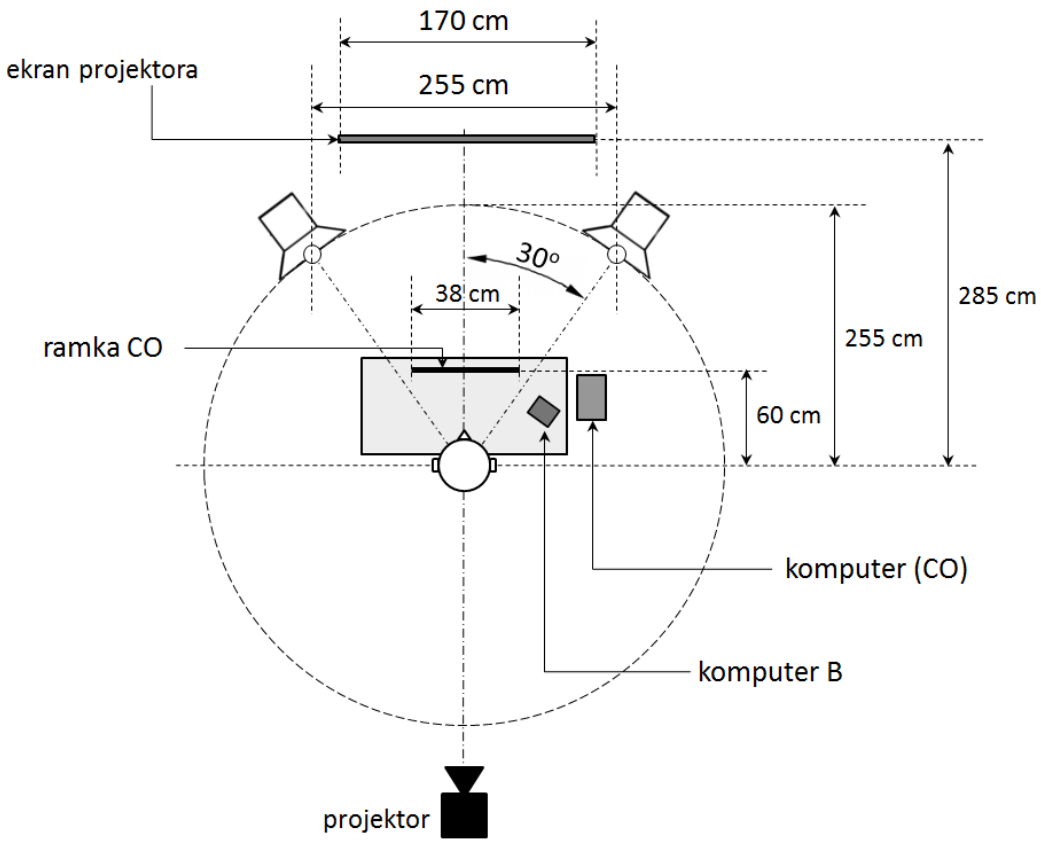
b)



Rys. 5.8 Pomieszczenie odsłuchowe w pierwszej konfiguracji stanowiska badawczego: a) rzut z góry; b) rzut z boku

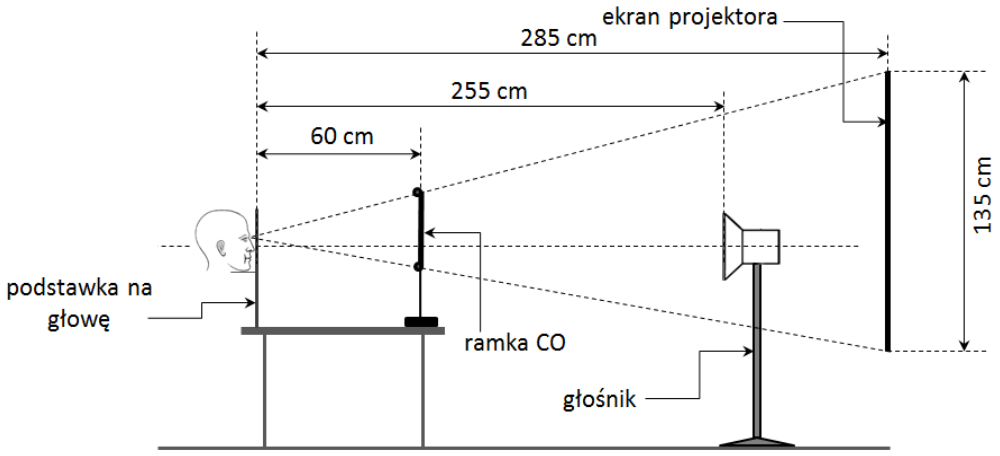
Jak wspomniano wcześniej, w drugiej konfiguracji stanowiska badawczego próbki wizyjno-foniczne były prezentowane badanym na ekranie projektora, oddalonym od punktu najlepszego odsłuchu o 285 cm. Pomieszczenie odsłuchowe w drugiej konfiguracji stanowiska badawczego w rzucie z góry i z boku przedstawiono na rys. 5.9.

a)



komputer B – komputer przeznaczony do udzielania odpowiedzi na temat położenia pozornego źródła dźwięku (wypełnianie formularza)

b)



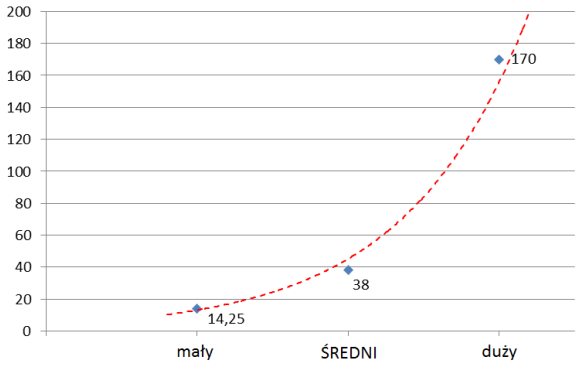
ramka CO – ramka imitująca ekran monitora systemu Cyber-Oko; w rogach ramki zamontowano diody podczerwieni

Rys. 5.9 Pomieszczenie odsłuchowe w drugiej konfiguracji stanowiska badawczego: a) rzut z góry; b) rzut z boku

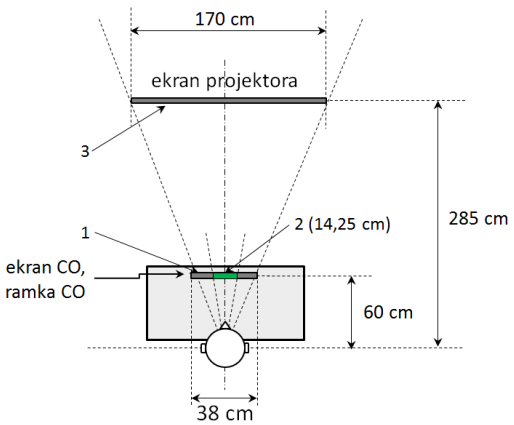
Warto dodać, że różne konfiguracje stanowiska badawczego przedstawiono również na fotografiach zamieszczonych w Załączniku E niniejszej rozprawy.

Jednym z wątków badawczych niniejszej rozprawy było zbadanie zjawiska skalowalności wpływu ściągającego obrazu na percepcję dźwięku. Zbadano wpływ bodźca słuchowego na percepcję kierunku dźwięku z wykorzystaniem trzech wielkości obszaru wyświetlania. Wyjściowy rozmiar obszaru wyświetlania (określany jako „średni”) odpowiadał wielkości ekranu monitora systemu CO, co oznacza, że jego szerokość była równa 38 cm. Ze względu na fakt, że wyjściowym obszarem wyświetlania był obszar ekranu monitora komputerowego, przyjęto, że będzie stanowił on odniesienie dla „małego” i „dużego” obszaru wyświetlania. W drugiej konfiguracji szerokość obszaru wyświetlania odpowiadała wielkości standardowych ekranów, montowanych między innymi w samolotach i wynosiła 14,25 cm („mały” obszar wyświetlania). W ostatniej konfiguracji materiał badawczy był prezentowany na ekranie projektora, którego szerokość wynosiła 170 cm („duży” obszar wyświetlania). Relację pomiędzy szerokościami obszarów wyświetlania wykorzystanych w teście skalowalności przedstawiono na rys. 5.10a. Kształt linii trendu poprowadzonej pomiędzy zaznaczonymi na wykresie punktami świadczy o wykładniczym wzroście szerokości obszarów wyświetlania wykorzystanych w eksperymencie.

a)



b)



Rys. 5.10 Zastosowane w teście skalowalności obszary wyświetlania:
a) relacja pomiędzy szerokościami ekranów;
b) relacje geometryczne pomiędzy obszarami wyświetlania materiału badawczego

Należy w tym miejscu również zaznaczyć, że w przypadku pierwszej i drugiej konfiguracji (średni i mały obszar wyświetlania), nie zmieniła się odległość widza od monitora. Wykorzystanie w badaniu systemu CO zdeterminowało bowiem zachowanie

stałej odległości widz-monitor równej 60 cm. Stosunek szerokości średniego obszaru wyświetlania (zaznaczonego na rys. 5.10b cyfrą „1") do odległości widza od monitora wyniósł 0,63, zaś w przypadku tak zwanego małego obszaru wyświetlania (oznaczonego cyfrą „2") stosunek ten wyniósł 0,24. W przypadku trzeciej konfiguracji odległość widza od monitora została dostosowana do zmienionej szerokości obszaru wyświetlania. Stosunek szerokości dużego obszaru wyświetlania („3") do odległości widz-monitor był równy 0,60. Można zatem przyjąć, że w rzeczywistych warunkach stosunek ten w przypadku pierwszej i trzeciej konfiguracji był zachowany. Natomiast w przypadku drugiej konfiguracji wartość tego stosunku była o 2,5 razy mniejsza od wartości stosunku dla konfiguracji wyjściowej, gdy szerokość obszaru wyświetlania wynosiła 38 cm. Na rys. 5.10b przedstawiono wzajemne relacje pomiędzy poszczególnymi obszarami wyświetlania. Warto zaznaczyć, że z wykorzystaniem systemu śledzenia punktu fiksacji wzroku Tobii T60 przeprowadzono badania w pierwszej i drugiej konfiguracji (średni i mały obszar wyświetlania). Natomiast z wykorzystaniem systemu CO przeprowadzono badania w pierwszej i trzeciej konfiguracji (średni i duży obszar wyświetlania). Zastosowanie systemu Tobii T60 w drugiej konfiguracji (mały ekran) w celu zbadania aktywności wzrokowej było uzasadnione większą rozdzielczością przestrzenną systemu w porównaniu z CO. Z drugiej zaś strony, zbadanie aktywności wzrokowej w trzeciej konfiguracji (projekcja materiału badawczego na ekranie projektora) było możliwe jedynie z wykorzystaniem systemu CO. Umożliwiał on bowiem śledzenie punktów fiksacji wzroku na ekranie projektora dzięki zastosowaniu specjalnej ramki imitującej ekran CO.

5.3 Testy subiektywne

Założono, że w ramach badań zostaną przeprowadzone testy z wykorzystaniem dwóch różnych systemów śledzenia punktu fiksacji wzroku: komercyjnego systemu Tobii T60 oraz opracowanego w Katedrze Systemów Multimedialnych systemu CyberOko. Uzyskane w ten sposób wyniki będą wolne od osobniczych cech systemu. Takie podejście ma pozytywny wpływ na weryfikację tez, postawionych w niniejszej rozprawie. W związku z tym, że w badaniach wykorzystano dwa systemy śledzenia punktu fiksacji wzroku, przeprowadzono dwie serie eksperymentów. Pierwsza seria składała

się z jednej sesji, podczas której zbadano 15 osób z wykorzystaniem systemu Tobii T60. Wiek osób biorących udział w badaniu zawierał się w przedziale 23-30 lat (wartość średnia: 24,27, odchylenie standardowe: 1,8). W grupie badanych znalazło się 5 kobiet i 10 mężczyzn. Żaden z badanych nie miał ubytku słuchu, a poziom dźwięku zapewniający komfortowe warunki odsłuchu dla wszystkich uczestników eksperymentu był taki sam. Należy zaznaczyć, że uczestnikami badań nie byli ani doktoranci, ani pracownicy Katedry Systemów Multimedialnych. Autorowi rozprawy zależało na tym, aby po pierwsze – badani nie byli zaznajomieni z tematyką i celem prowadzonych przez niego badań, po drugie – aby nie byli w pełni świadomi natury zjawisk percepcyjnych, które zachodzą w mózgu w procesie stymulacji polisensorycznej (tzw. eksperci). Okazuje się bowiem, że obserwowany wpływ ściągający w przypadku tak zwanych ekspertów jest zdecydowanie mniejszy w porównaniu z wielkością badanego wpływu ściągającego u osób nie będących ekspertami. Powyższą zależność zauważyli Dybski i Napierski w swojej pracy dyplomowej, której autor rozprawy był konsultantem [39]. Czas trwania badania z wykorzystaniem systemu Tobii T60 wynosił dokładnie 30 minut na każdą osobę. W związku z tym, że pierwsza seria składała się tylko z jednej sesji, wszyscy zostali przebadani w ciągu jednego dnia. Wiązało się to z zachowaniem odpowiedniej dyscypliny prowadzenia badań, ze względu na fakt, iż system Tobii T60 został wypożyczony tylko na jeden dzień, nie było fizycznej możliwości przeprowadzenia eksperymentów z udziałem większej liczby osób.

Druga seria badań odbyła się w odstępie kilku dni od pierwszej. Liczba badanych w tej serii również wyniosła 15 osób (te same osoby), przy czym składała się ona z dwóch sesji. W pierwszej sesji zbadano 10 osób, w drugiej 5. Przeprowadzenie badań w dwóch sesjach wynikało z faktu, że czas pojedynczego badania z wykorzystaniem systemu CO wynosił 45 minut. Autor rozprawy z góry zaplanował czas trwania badania w drugiej serii, biorąc pod uwagę konieczność zmiany konfiguracji stanowiska badawczego podczas badania, a tym samym – potrzebę podwójnej kalibracji systemu. W Załączniku E zamieszczono fotografie przedstawiające stanowisko badawcze w różnych konfiguracjach.

W prowadzonych eksperymentach bardzo ważne było zwrócenie uwagi badanych na źródło dźwięku, którego kierunek badani określali po projekcji każdej próbki. Autor

rozprawy, który był obecny podczas wszystkich sesji badawczych, informował uczestników eksperymentu nie tylko o poddawanym ocenie źródle dźwięku, ale także o tym, na co konkretnie badani powinni zwrócić uwagę. W większości próbek zadaniem badanych było wskazanie kierunku pozornego źródła dźwięku w dwukanałowej panoramie stereofonicznej (płaszczyzna lewo-prawo). W jednej z próbek wizyjno-fonicznych (fragment filmu „Alicja w Krainie Czarów") dodatkowo badano położenie dwóch źródeł dźwięku również w płaszczyźnie przód-tył (w głębi). Informacja o tym, co jest przedmiotem badania w danej próbce była istotna, ponieważ badani rzeczywiście koncentrowali się na lokalizowaniu pozornego źródła dźwięku. W innym przypadku można by przypuszczać, że wskazywany kierunek dźwięku odpowiadałby raczej położeniu bodźca wzrokowego w obrazie aniżeli percypowanemu kierunkowi pozornego źródła dźwięku związanego z tym bodźcem.

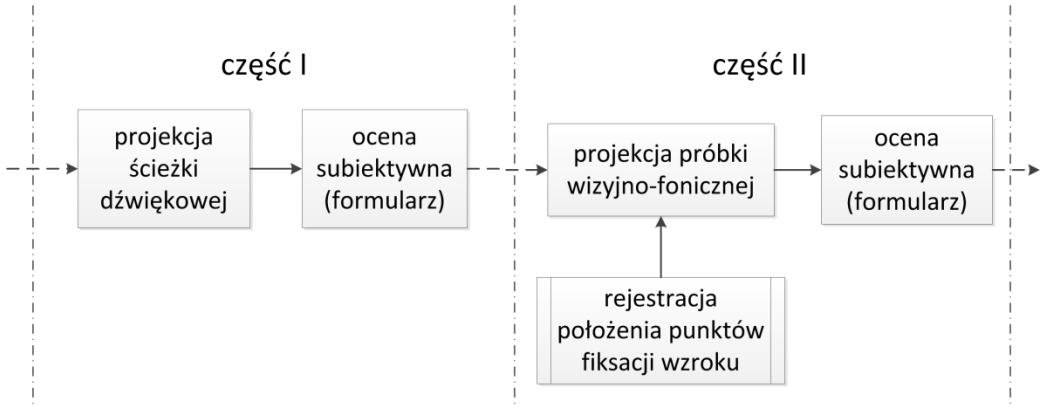
Przebieg eksperymentu

Każde badanie składało się z dwóch części (etapów). W pierwszej części ocenie badanych poddane zostały ścieżki dźwiękowe próbek wizyjno-fonicznych. W tej części prezentowanych było siedem próbek (1-7). W drugim etapie badanym prezentowano materiał badawczy składający się ze ścieżek dźwiękowych, poddanych ocenie w części pierwszej, przy czym ścieżce dźwiękowej towarzyszył obraz w różnych konfiguracjach. W zależności od próbki wizyjno-fonicznej obraz był prezentowany z polskimi napisami, jako konwencjonalny obraz 2D lub jako obraz trójwymiarowy, wyświetlany w technice anaglifowej. W tej części z wykorzystaniem systemu Tobii T60 ocenie badanych poddano 18 próbek, zaś z wykorzystaniem systemu CO – 20 próbek. Warto zwrócić uwagę na fakt, że – w przypadku systemu Tobii – cztery ostatnie próbki były prezentowane w tak zwanej drugiej konfiguracji, czyli w małym obszarze wyświetlania. Natomiast w przypadku systemu CO sześć ostatnich próbek wyświetlono w tak zwanej trzeciej konfiguracji, czyli na ekranie projektora.

Jak wspomniano wcześniej, badani wskazywali położenie pozornego źródła dźwięku bezpośrednio po projekcji każdej próbki, wypełniając elektroniczną wersję formularza ocen subiektywnych. Wskazywanie kierunku dźwięku odbywało się poprzez ustawienie (za pomocą myszy komputerowej) suwaka w miejscu odpowiadającym kie-

runkowi dochodzenia dźwięku. Przedział wartości, w którym badani wskazywali kierunek dźwięku zawierał się w zakresie -30 do +30. Przedział ten odpowiadał zakresowi dwukanałowej panoramy stereofonicznej: -30^{o} – $+30^{o}$. Możliwe było wskazanie kierunku dźwięku z dokładnością do 1^{o}. W przypadku określania położenia pozornego źródła dźwięku w płaszczyźnie przód-tył posłużono się skalą z wartościami w zakresie 0-10, gdzie „0” oznaczało lokalizację pozornego źródła dźwięku blisko widza, zaś „10” – położenie daleko od widza za płaszczyzną ekranu. Koncepcję określania kierunku dźwięku oraz wskazywania położenia pozornego źródła dźwięku w płaszczyźnie przód-tył przedstawiono w Załączniku C niniejszej rozprawy.

Przebieg eksperymentu dla większości próbek wizyjno-fonicznych można przedstawić na schemacie z rys. 5.11. Materiał badawczy był prezentowany osobom badanym w dwóch konfiguracjach: tylko próbka dźwiękowa oraz próbka dźwiękowa z towarzyszeniem obrazu wizyjnego. Wyjątek stanowiły dwie próbki wizyjno-foniczne, których ścieżka dźwiękowa zawierała głównie dźwięk efektowy. Były one prezentowane w jednej konfiguracji – ścieżka dźwiękowa z towarzyszeniem obrazu.

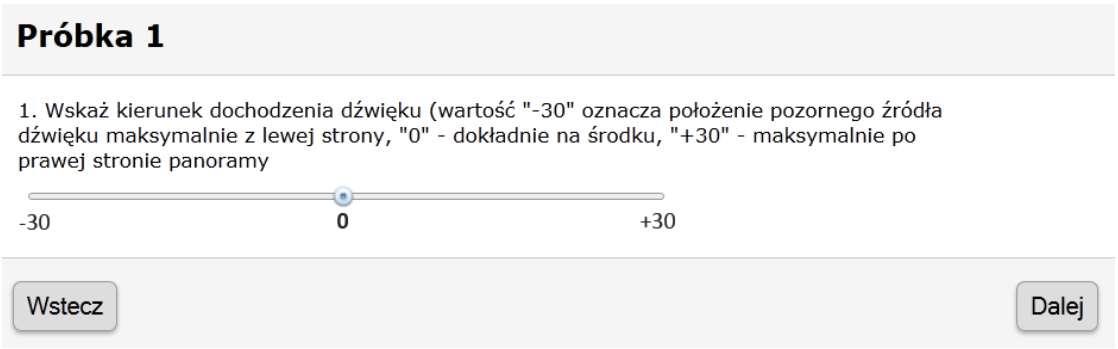


Rys. 5.11 Przebieg eksperymentu dla pojedynczej próbki

5.4 Wyniki testów

Zaprezentowany na rys. 5.11 model przebiegu eksperymentu wyraźnie wskazuje, że z każdą prezentowaną w trakcie badania próbką wizyjno-foniczną związane były dwa rodzaje danych. W ogólności można je sklasyfikować jako dane subiektywne i

obiektywne. Dane subiektywne reprezentują również próbkę foniczną, prezentowaną w pierwszej części eksperymentu. Na wyniki subiektywne, uzyskane w drodze wypełniania formularza ankiety związanej z badaniem, składają się oceny położenia pozornego źródła dźwięku w panoramie stereofonicznej (płaszczyzna lewo-prawo, zakres wartości: od -30 do +30) oraz w płaszczyźnie przód-tył (zakres wartości: 0-10). Na rys. 5.12 przedstawiono fragment formularza w wersji elektronicznej, który w trakcie eksperymentu wypełniali badani.



Rys. 5.12 Fragment formularza ocen subiektywnych

Dane obiektywne uzyskane w wyniku badania są natomiast współrzędnymi (*x, y*) PoR na ekranie komputera w jednostce czasu. Wśród nich wyróżnić można dane zapisane przez system Tobii (w formacie TSV) oraz dane zapisane przez system CO (format XML). Warto w tym miejscu zwrócić uwagę na różnice występujące w formacie danych zapisywanych przez system Tobii i CO. Zarówno system Tobii, jak i system CO traktuje płaszczyznę ekranu, na który patrzy osoba badana, jako układ współrzędny, którego punkt (0,0) znajduje się w lewym górnym rogu ekranu. Pierwsza różnica polega na przyjętych wartościach zakresów osi OX i osi OY. System Tobii zapisuje współrzędne punktów fiksacji wzroku w zakresie 0-1280 (na osi OX) oraz 0-1024 (na osi OY). Natomiast system CO wyznacza współrzędne punktów fiksacji w zakresie 0-100 zarówno na osi OX, jak i osi OY. Druga różnica w zapisie danych między dwoma systemami jest związana z ich rozdzielczością czasową. W przypadku systemu Tobii na jedną sekundę przypada 60 wartości współrzędnych punktu fiksacji, zaś w przypadku systemu CO – 5 wartości. Różnice, występujące pomiędzy tymi systemami, należy brać pod uwagę w analizie uzyskanych przez nie wyników.

Na potrzeby analizy danych pozyskanych z obu systemów śledzenia punktu fiksacji wzroku przygotowana została aplikacja zwracająca względne wartości czasu skupienia wzroku wyrażone w [%]. Aplikacja ta „porównuje” współrzędne punktów fiksacji zebrane podczas oglądania materiału testowego (dane w formacie TSV oraz XML) ze zdefiniowanymi obszarami zainteresowania zamieszczonymi w pliku opisu filmu (*movieDescription.xml*). Na tej podstawie otrzymano wartości uwagi wzrokowej widza, skupionej na poszczególnych bodźcach wzrokowych. Wartości uwagi wzrokowej (UW), związanej z odpowiednimi bodźcami zamieszczono na tzw. wykresach pudełkowych w Załączniku D. Sposób wyznaczania uwagi wzrokowej widza związanej z konkretnym bodźcem wzrokowym w materiale wizyjnym jest podobny do zaproponowanego przez Kocejkę i Wtorka (naukowców z Katedry Inżynierii Biomedycznej Politechniki Gdańskiej) sposobu wyznaczania uwagi wzrokowej na obrazie statycznym [69].

Warto również wspomnieć o formie prezentacji danych obiektywnych – próbkach wizyjno-fonicznych z naniesionymi punktami fiksacji wzroku. Pomimo, że ta forma wyników badania nie była bezpośrednio wykorzystana w analizie danych, stanowi ona przejrzystą wizualizację aktywności wzrokowej badanego na prezentowanym obrazie wizyjnym. Tę formę prezentacji uwagi wzrokowej na fragmentach prezentowanego obrazu nazwano w ramach niniejszej rozprawy „dynamiczną mapą przejść”. Na rys. 5.13 przedstawiono przykładowe klatki próbek wizyjnych z dynamiczną mapą przejść wygenerowaną przez system Tobii oraz CO.

a)

b)

Rys. 5.13 Przykładowe dynamiczne mapy przejść: a) wygenerowana przez system Tobii T60, b) wygenerowana przez system CO

W niniejszym rozdziale przedstawiono metodykę przeprowadzonych badań z uwzględnieniem opracowanego materiału badawczego wraz ze stanowiskiem oraz opisem właściwych eksperymentów. W następnym rozdziale przedstawiono wyniki analizy statystycznej uzyskanych wyników badań dla poszczególnych próbek wizyjno-fonicznych.

6 Analiza wyników

W niniejszym rozdziale przedstawiona została analiza statystyczna uzyskanych wyników. Opracowane wyniki podzielono na dwie części. W pierwszej części skoncentrowano się na analizie istotności statystycznej wpływu ściągającego obrazu na percepcję dźwięku w poszczególnych próbkach. W tym celu posłużono się analizą wariancji, czyli tak zwanym testem ANOVA, badającym różnice pomiędzy średnimi z kilku prób (zmiennych). W drugiej części opracowania skoncentrowano się na zbadaniu związku pomiędzy obserwowanym wpływem ściągającym a wybranymi parametrami: położeniem bodźca wzrokowego na ekranie, czasem skupienia wzroku na danym bodźcu, czyli tak zwanej uwadze wzrokowej, wyznaczonej przez system śledzenia punktu fiksacji wzroku oraz wielkością wyświetlanego obiektu (związaną bezpośrednio z szerokością bodźca wzrokowego). W analizie statystycznej drugiej części opracowania wykorzystanym narzędziem statystycznym był przede wszystkim współczynnik korelacji rang Spearmana. Wszystkie metody statystyczne, istotne w kontekście wykonanej analizy, scharakteryzowano w następnym podrozdziale.

Należy w tym miejscu wyraźnie zaznaczyć, że w opracowaniu wyników przyjęto założenie, zgodnie z którym szerokość zakresu panoramy stereofonicznej odniesiono do szerokości obszaru wyświetlania. Przyjęto, że założenie to zostało spełnione w tak zwanej pierwszej (średni rozmiar ekranu) i trzeciej (ekran projektora) konfiguracji obszaru wyświetlania. Jak wynika z rysunków 5.8a oraz 5.9a relacja pomiędzy szerokością panoramy stereofonicznej a szerokością obszaru wyświetlania nie jest w proporcji 1:1, niemniej jednak na potrzeby interpretacji uzyskanych w niniejszej pracy wyników, takie podejście jest uzasadnione. W myśl tego założenia na wykresach pudełkowych, przedstawiających rozkład odpowiedzi dotyczących lokalizacji pozornego źródła dźwięku w panoramie, zaznaczono dodatkowo kolorem zielonym położenie i wielkość wyświetlanego obiektu, przykuwającego uwagę wzrokową widza. Wykresy pudełkowe dla poszczególnych próbek zamieszczono w Załączniku D niniejszej rozprawy.

Analizę statystyczną uzyskanych wyników wykonano przy użyciu dwóch programów. Największą część obliczeń przeprowadzono z wykorzystaniem specjalistycznego pakietu oprogramowania STATISTICA [172]. Obliczenia weryfikujące poprawność

uzyskanych wyników oraz tzw. wykresy pudełkowe (inaczej: wykresy „ramka-wąsy" – ang. *box and whisker plots*) wykonano w środowisku MATLAB z zastosowaniem odpowiedniego pakietu do analizy statystycznej (ang. *statistics toolbox*).

6.1 Wybrane metody statystyczne

W niniejszym podrozdziale scharakteryzowano statystyki, które wykorzystano w analizie wyników. Posłużono się statystykami, które są najczęściej stosowane w interpretacji wyników badań związanych z wpływem bodźca wzrokowego na lokalizację pozornego źródła dźwięku. Ponadto, na wybór odpowiedniej statystyki w analizie uzyskanych wyników miał wpływ również przedmiot badania.

Poziom istotności statystycznej

Poziom istotności stanowi maksymalne dopuszczalne prawdopodobieństwo popełnienia błędu I rodzaju (zazwyczaj oznaczane symbolem α). Poziom istotności wskazuje tym samym na maksymalne „ryzyko błędu", jakie osoba przeprowadzająca eksperyment jest skłonna zaakceptować. Najczęściej przyjmowana wartość α wynosi 0,05. Należy w tym miejscu zaznaczyć, że wszystkie dane uzyskane w wyniku przeprowadzonych w ramach niniejszej rozprawy eksperymentów analizowano z przyjętym poziomem istotności równym 0,05. Wartość założonego poziomu istotności jest porównywana z wyliczonym z testu statystycznego poziomem p (ang. *p-value*). Jeżeli wartość p jest większa od przyjętego poziomu istotności α, to przyjmuje się, że nie ma powodu do odrzucenia tzw. hipotezy zerowej $\mathbf{H_0}$ (ang. *null hypothesis*), zgodnie z którą badany efekt jest dziełem przypadku [127].

Analiza wariancji (ang. *Analysis of Variance* – ANOVA)

Zasadniczym celem analizy wariancji jest badanie różnic pomiędzy średnimi analizowanych grup lub zmiennych. Analiza wariancji, popularnie nazywana testem ANOVA, stanowi rozszerzenie testu t-Studenta w przypadku porównywania większej liczby grup lub zmiennych. Test ANOVA pozwala na weryfikację tego czy porówny-

wane średnie różnią się od siebie w sposób istotny z przyjętym poziomem istotności [25] [120] [133]. Wyróżnia się dwa rodzaje testu ANOVA: test jednoczynnikowy oraz test dwuczynnikowy. W związku z tym, że w niniejszej analizie wyników porównywano średnie wartości kilku zmiennych wyróżnionych ze względu na jedną cechę (tzw. zmienną grupującą) w dalszej części niniejszego podrozdziału scharakteryzowano tylko jednoczynnikową analizę wariancji.

Należy zaznaczyć, że wykonanie testu ANOVA zakłada spełnienie dwóch podstawowych warunków:

1. rozkład każdej analizowanej zmiennej jest zgodny z rozkładem normalnym,
2. wariancje analizowanych zmiennych są homogeniczne (równe).

Niespełnienie jednego z powyższych założeń uniemożliwia wykonanie testu ANOVA. Jednakże pozostaje możliwość wykonania nieparametrycznej alternatywy analizy wariancji – testu Kruskala-Wallisa, opisanego w dalszej części niniejszego podrozdziału. W celu weryfikacji pierwszego założenia testu ANOVA stosuje się test Shapiro-Wilka. Natomiast do sprawdzenia warunku o równości wariancji stosowany jest test Levene'a. Oba testy zostały scharakteryzowane poniżej.

Wykonanie testu ANOVA wiąże się z postawieniem i weryfikacją następującej hipotezy zerowej $\mathbf{H_0}$: średnie wartości analizowanych zmiennych są równe. W wyniku analizy wariancji w programie STATISTICA oraz w środowisku MATLAB uzyskane wyniki testu są zestawione w tabeli. W kontekście weryfikacji hipotezy zerowej najważniejsze są następujące wartości: wartość testu F oraz poziom p. Jak wspomniano wcześniej, dla wszystkich analizowanych wyników przyjęto ten sam poziom istotności równy 0,05. Interpretacja wartości uzyskanych w wyniku przeprowadzenia testu ANOVA, sprowadza się do porównania obliczonej wartości poziomu p z przyjętym poziomem istotności α. Gdy $p>\alpha$, to nie ma podstaw do odrzucenia hipotezy zerowej. Jeżeli zaś $p<\alpha$, to przyjmuje się, że na poziomie istotności 0,05 średnie wartości analizowanych zmiennych są różne.

Test Shapiro-Wilka

Test Shapiro-Wilka jest wykonywany w celu weryfikacji rozkładu normalnego analizowanej zmiennej. W tym celu sprawdzane są następujące hipotezy:

- hipoteza zerowa $\mathbf{H_0}$: rozkład badanej zmiennej jest zgodny z rozkładem normalnym,
- hipoteza alternatywna $\mathbf{H_A}$: rozkład badanej zmiennej nie jest rozkładem normalnym.

Wartość przyjętego poziomu istotności $\alpha=0,05$. Jeżeli wyznaczony poziom p jest większy od 0,05, to nie ma podstaw do odrzucenia hipotezy o zgodności rozkładu analizowanej zmiennej z rozkładem normalnym. W przeciwnym przypadku ($p<\alpha$) hipoteza zerowa zostaje odrzucona, co jest równoznaczne z przyjęciem hipotezy alternatywnej, zgodnie z którą rozkład analizowanej zmiennej nie jest rozkładem normalnym.

Test Levene'a

Test Levene'a pozwala zbadać homogeniczność, czyli jednorodność (równość) wariancji porównywanych zmiennych. W teście tym stawiane i weryfikowane są następujące hipotezy:

- $\mathbf{H_0}$: wariancje analizowanych zmiennych są statystycznie równe,
- $\mathbf{H_A}$: wariancje różnią się istotnie.

Jeżeli wyznaczona w wyniku testu wartość p jest duża, to można wnioskować, że na poziomie istotności 0,05 brak jest podstaw do odrzucenia hipotezy o homogeniczności wariancji. W przypadku, gdy $p<0,05$, hipoteza zerowa zostaje odrzucona, co oznacza, że wariancje analizowanych zmiennych różnią się istotnie.

Jeżeli w wyniku przeprowadzenia testu Shapiro-Wilka i testu Levene'a okazuje się, że jedno z założeń (rozkład normalny zmiennych albo równość wariancji) nie zostało spełnione, pozostaje wykonanie alternatywnego testu Kruskala-Wallisa.

Test Kruskala-Wallisa

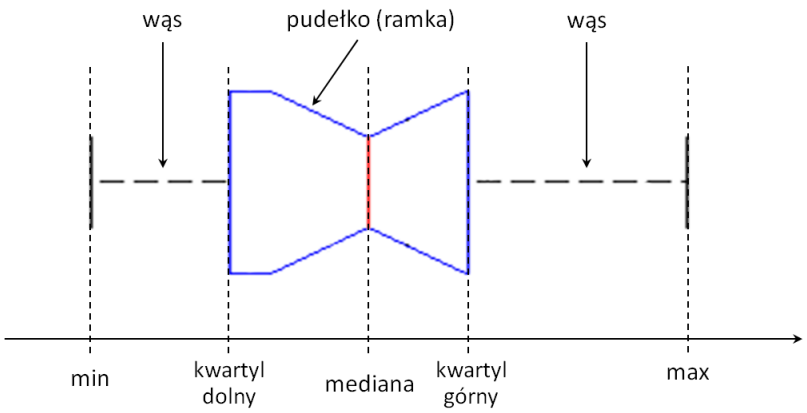
Przeznaczeniem nieparametrycznego testu Kruskala-Wallisa (K-W) jest testowanie różnic pomiędzy średnimi z kilku prób niezależnych. Innymi słowy test ten pozwala sprawdzić czy *n* niezależnych próbek pochodzi z tej samej populacji. Stawiana jest następująca hipoteza zerowa $\mathbf{H_0}$: dystrybuanty rozkładów porównywanych zmiennych są równe. W przypadku, gdy wyznaczony poziom $p>\alpha$ nie ma podstaw do odrzucenia hipotezy o równości dystrybuant rozkładów. Natomiast, gdy poziom *p* jest mniejszy od przyjętego poziomu istotności statystycznej, należy odrzucić hipotezę zerową.

Współczynnik korelacji rang Spearmana

W przypadku wyznaczania siły związku (współczynnika korelacji) pomiędzy zmiennymi, których zależność nie jest prostoliniowa, stosuje się współczynnik rang Spearmana, zamiast współczynnika korelacji Pearsona. Korelacja rangowa przyjmuje zawsze wartości z przedziału [-1, +1], a ich interpretacja jest podobna do klasycznego współczynnika korelacji Pearsona. Ponadto, należy zaznaczyć, że wartość wyznaczonego współczynnika korelacji jest istotna statystycznie, gdy poziom $p<0{,}05$.

Wykres pudełkowy

Wykres pudełkowy, nazywany również wykresem „ramka-wąsy” (ang. *box and whisker plot*) w niniejszej rozprawie stanowi graficzną ilustrację rozkładu odpowiedzi udzielonych przez badanych podczas eksperymentów. Wykresy pudełkowe są opracowywane w oparciu o wartości statystyk opisowych i charakteryzują się dużą przejrzystością i zwięzłością. Wykresy pudełkowe stosuje się przy porównaniu rozkładów dwóch lub więcej zmiennych. Wykres „ramka-wąsy” najczęściej jest stosowany w celu zdefiniowania rozproszenia danej cechy oraz wsparcia analizy i interpretacji danych statystycznych. Wyróżnia się aż cztery typy wykresów pudełkowych, jednakże w niniejszej rozprawie zastosowano typ wykresu przedstawiającego medianę, kwartyle oraz wartość minimalną i maksymalną [166], jak pokazano na rys. 6.1.



Rys. 6.1 Przykładowy wykres pudełkowy

Długość pudełka (ramki) reprezentuje rozstęp ćwiartkowy (ang. *interquartile range* – IQR). Z definicji rozstęp ćwiartkowy obejmuje 50% wszystkich obserwacji. Pudełko jest rozdzielone pionową linią koloru czerwonego, wyznaczającą wartość mediany. Linia ta dzieli przedział IQR na dwa obszary, w których znajduje się 25% obserwacji. Tak zwane „wąsy" łączą pudełko z najmniejszą (min) i największą (max) wartością badanej zmiennej. W pierwszym przedziale znajduje się 25% obserwacji o wartościach niższych od dolnego kwartyla, w drugim zaś przedziale – 25% obserwacji o wartościach wyższych od górnego kwartyla [166].

6.2 Analiza statystyczna wyników badania wpływu obrazu na percepcję dźwięku

W niniejszym rozdziale przeprowadzono analizę statystyczną wyników uzyskanych podczas wykonanych badań. Analizę wyników podzielono na dwie części. W pierwszej części, w podrozdziale 6.2.1, opisano wnioski z analizy tzw. danych subiektywnych, pozyskanych w wyniku wypełniania przez badanych ankiety w wersji elektronicznej. Fragment interfejsu ankiety przedstawiono na rys. 5.12. W drugiej części – w podrozdziale 6.2.2 – zbadano związek pomiędzy przeanalizowanymi danymi subiektywnymi a danymi obiektywnymi. Wśród danych obiektywnych wyróżnić należy: położenie bodźca wzrokowego (związane ze współrzędną odciętą jego środka ciężkości), uwagę wzrokową widza na danym bodźcu wzrokowym, wyznaczoną przez system śledzenia wzroku oraz wielkość wyświetlanego obiektu (rozumianą jako szerokość wy-

świetlanego bodźca wzrokowego). Warto zaznaczyć, że położenie oraz szerokość bodźca wzrokowego wyrażono w stopniach kątowych. Założono, że środek ekranu (punkt o współrzędnych (640; 512)) leży na prostej 0^o, gdzie znajduje się również głowa osoby badanej. Przyjęcie powyższego założenia było możliwe, ponieważ w płaszczyźnie poziomej głowa osoby badanej znajdowała się dokładnie na osi symetrii ekranu dzięki zastosowaniu specjalnej podstawki przedstawionej na rys. E.2 w Załączniku E. Zakresy kątowe bodźców wzrokowych znormalizowano do przyjętego wcześniej zakresu kątowego bazy stereofonicznej [-30^o, $+30^o$].

6.2.1 Analiza istotności statystycznej wpływu ściągającego

W niniejszym podrozdziale zostały przedstawione wnioski szczegółowej analizy statystycznej przeprowadzonej dla zebranych danych subiektywnych w kontekście badania wpływu ściągającego obrazu na percepcję dźwięku oraz zjawiska skalowalności tego wpływu. Ponadto, w ostatniej części niniejszego podrozdziału przeprowadzono analizę wpływu ściągającego dla materiału wizyjnego wyświetlanego w technice stereoskopowej.

W pierwszej kolejności zamieszczono ostateczne wnioski wysnute na podstawie analizy statystycznej przeprowadzonej dla materiału badawczego. Szczegółową analizę statystyczną wykonaną dla poszczególnych próbek wizyjno-fonicznych zamieszczono w Załączniku D niniejszej rozprawy. W tab. 6.1 zestawiono informacje o istotności statystycznej wyników uzyskanych dla kolejnych próbek w dwóch seriach eksperymentu: z wykorzystaniem systemu Tobii T60 oraz systemu CO. Numeracja próbek w tab. 6.1 koresponduje z numeracją próbek zastosowaną w Załączniku B.

W związku z tym, że w obu seriach eksperymentu wzięły udział te same osoby, założono, że wyniki uzyskane podczas badania z wykorzystaniem systemu Tobii, jak i systemu CO, powinny mieć ten sam charakter – w jednym i drugim przypadku powinny być istotne, bądź nieistotne statystycznie. Założenie to nie zostało jednak spełnione w przypadku czterech z osiemnastu analizowanych próbek testowych. Rozbieżność w zachowaniu istotności statystycznej wyników dotyczy próbek: „3 głos (pan, średni)” – próbka nr 3, w której badano położenie głosu bohaterki w panoramie stereofonicznej

w średnim obszarze wyświetlania, „5 królowa (pan, średni)" – próbka nr 5, w której badano położenie głosu królowej w panoramie stereofonicznej w średnim obszarze wyświetlania, „5 królowa (głębia, średni)" – próbka nr 5, w której badano położenie głosu królowej w głębi (w płaszczyźnie przód-tył) w średnim obszarze wyświetlania oraz „9 skrzypce (pan, średni)" – próbka nr 9, w której badano półożenie skrzypiec w panoramie stereofonicznej w średnim obszarze wyświetlania. Dokładną charakterystykę próbek wizyjno-fonicznych zestawionych w tab. 6.1 zawarto w Załączniku B. Oznaczenie „-" w poniższej tabeli oznacza brak możliwości przygotowania danej konfiguracji próbki. Zgodnie z przyjętym założeniem, w serii eksperymentu z wykorzystaniem systemu Tobii T60 nie prezentowano próbek w dużym obszarze wyświetlania, natomiast w serii eksperymentu z wykorzystaniem systemu CO nie prezentowano treści wizyjnej w małym obszarze wyświetlania.

Tab. 6.1 Zestawienie informacji o istotności statystycznej wyników dla poszczególnych konfiguracji próbek

	Czy wyniki uzyskane dla próbki są istotne statystycznie?	
próbka	**Tobii T60**	**Cyber-Oko**
1 L (pan, średni)	TAK	TAK
1 R (pan, średni)	TAK	TAK
2 głos (pan, średni)	TAK	TAK
2 głos (pan, mały)	TAK	-
2 głos (pan, duży)	-	TAK
3 głos (pan, średni)	TAK	NIE
5 królowa (pan, średni)	TAK	NIE
5 królowa (głębia, średni)	NIE	TAK
5 alicja (pan, średni)	TAK	TAK
5 alicja (głębia, średni)	NIE	NIE
6 głos (pan, średni)	TAK	TAK
8 fortepian (pan, średni)	TAK	TAK
9 skrzypce (pan, średni)	NIE	TAK
9 skrzypce (pan, mały)	TAK	-
9 skrzypce (pan, duży)	-	TAK
9 fortepian (pan, średni)	NIE	NIE
9 fortepian (pan, mały)	NIE	-
9 fortepian (pan, duży)	-	NIE

wyjaśnienie wybranych oznaczeń:

1 L (pan, średni) – próbka nr 1, w której badano położenie źródła dźwięku w panoramie stereofonicznej, przy czym bodziec wzrokowy znajdował się w lewej części kadru; badanie w średnim obszarze wyświetlania;

5 królowa (pan, średni) – próbka nr 5, w której badano położenie głosu królowej w panoramie stereofonicznej w średnim obszarze wyświetlania;

9 skrzypce (pan, mały) – próbka nr 9, w której badano lokalizację skrzypiec w panoramie stereofonicznej w małym obszarze wyświetlania.

Przyjęto, że na zaobserwowaną rozbieżność w zachowaniu istotności statystycznej wyników wpłynął charakter próbki (np. w próbce „3 głos (pan, średni)” postać bohaterki stanowiąca element przykuwający uwagę widza dynamicznie zmieniała swoje położenie w czasie trwania próbki) oraz stosunkowo nisko oceniony efekt głębi stereoskopowej (w przypadku pozostałych próbek). Na uwagę zasługuje fakt, że wyniki istotne statystycznie, uzyskane zarówno w serii z wykorzystaniem systemu Tobii, jak i systemu CO, otrzymano dla dziesięciu próbek testowych. Można zatem przyjąć, że próbki te stanowią odpowiedni materiał badawczy przeznaczony do prowadzenia badań nad wpływem obrazu wizyjnego na percepcję dźwięku.

Przeanalizowano wykresy pudełkowe próbek testowych (zamieszczone w Załączniku D), dla których uzyskano wyniki istotne statystycznie w każdej z dwóch przeprowadzonych serii eksperymentu. Przedmiotem niniejszej analizy było ustalenie, w jakim stopniu bodziec wzrokowy, na którym osoba badana fiksuje wzrok, „ściąga” położenie pozornego źródła dźwięku w swoim kierunku. W przypadku serii eksperymentu z wykorzystaniem systemu Tobii T60 wyniki wskazują na to, że położenie bodźca słuchowego odpowiada położeniu bodźca wzrokowego dla pięciu z ośmiu analizowanych próbek. Dla pozostałych trzech próbek zaobserwowano wpływ ściągający w kierunku bodźca wzrokowego, jednak położenie pozornego źródła dźwięku nie pokrywa się z położeniem bodźca wzrokowego. W przypadku badania z wykorzystaniem systemu CO zaobserwowano podobną tendencję, jednakże ścisły związek między lokalizacją pozornego źródła dźwięku w bazie stereofonicznej a położeniem bodźca wzrokowego w obrazie odnotowano dla siedmiu z dziewięciu analizowanych próbek wizyjno-fonicznych.

Badanie wpływu ściągającego z wykorzystaniem materiału stereoskopowego

Jak wynika z opisu wykorzystanego w badaniach materiału testowego, zawartego w Załączniku B, próbki stereoskopowe były prezentowane badanym aż czternastokrotnie. Uwzględniając jednak fakt, że oglądanie obrazu trójwymiarowego w małym i dużym obszarze wyświetlania nie zawsze było możliwe (ze względu na znaczący spadek dokładności systemu CO po nałożeniu przez użytkownika okularów anaglifowych), poniżej skoncentrowano się jedynie na wynikach uzyskanych dla średniego obszaru wyświetlania. Na ekranie monitora systemu śledzącego wzrok trójwymiarowe próbki wizyjne wyświetlono dziesięciokrotnie. Jak wynika z analizy ocen percypowanej głębi obrazu stereoskopowego średnia wartość efektu 3D (głębi stereoskopowej), podczas badania z wykorzystaniem systemu Tobii wyniosła 4,98 (w zakresie ocen 0-10), przy odchyleniu standardowym 1,62, natomiast w przypadku badania z wykorzystaniem systemu CO: 5,21 przy wartości odchylenia standardowego równej 1,53. Na podstawie powyższych wyników można uznać, że percypowany efekt przestrzenności prezentowanego obrazu stereoskopowego został oceniony stosunkowo nisko.

Biorąc pod uwagę fakt, że stereoskopowy obraz wizyjny podczas projekcji materiału badawczego był wyświetlany w technice anaglifowej, dla której efekt przestrzenności obrazu oceniono stosunkowo nisko, zdecydowano się przeprowadzić dodatkowy eksperyment, w którym obraz wizyjny był wyświetlany w technice polaryzacyjnej. W badaniu udział wzięło siedem losowo wybranych osób, a jego celem było zbadanie wpływu ściągającego obrazu 3D prezentowanego w technice polaryzacyjnej na lokalizację pozornego źródła dźwięku. Obraz stereoskopowy wyświetlono na monitorze polaryzacyjnym Zalman Trimon ZM-M220W, będącym na wyposażeniu Katedry Systemów Multimedialnych. W eksperymencie wykorzystano najbardziej charakterystyczną pod względem badania efektu 3D próbkę wizyjno-foniczną nr 5, według numeracji Załącznika B. Próbka nr 5 umożliwiła zbadanie wpływu obrazu przestrzennego na lokalizację pozornego źródła dźwięku zarówno w dwukanałowej bazie stereofonicznej, jak i w płaszczyźnie przód-tył. Źródłem dźwięku poddanym ocenie badanych w tej próbce był głos bohaterki nr 1 (królowej) oraz głos bohaterki nr 2 (Alicji). Bohaterki znajdowały się w różnych miejscach w płaszczyźnie przód-tył prezentowanej sceny, stąd możliwe było badanie wpływu bodźca wzrokowego na lokalizację pozornego źró-

dla dźwięku w tej płaszczyźnie. Próbkę nr 5 wyświetlono zarówno w technice polaryzacyjnej, jak i anaglifowej w celu oceny percypowanej w obrazie głębi stereoskopowej. Otrzymane wyniki przeanalizowano statystycznie. Przy sprawdzaniu pierwszego warunku testu ANOVA okazało się, że rozkład zmiennej reprezentującej wyniki ocen dla techniki anaglifowej nie odpowiada rozkładowi normalnemu. W związku z powyższym obliczono wartość testu *H* Kruskala-Wallisa, badającego istotność różnic pomiędzy średnimi analizowanych zmiennych. Otrzymana wartość testu *H* wyniosła 5,700, przy wartości poziomu p=0,02. Zatem przeprowadzony eksperyment wykazał istotność różnic pomiędzy ocenami efektu 3D dla obrazu wizyjnego prezentowanego w technice polaryzacyjnej i technice anaglifowej. Warto dodać, że średnia wartość percypowanego efektu 3D (w skali 0-10) dla próbki nr 5 w technice polaryzacyjnej wyniosła 7,57 przy wartości odchylenia standardowego 1,92, natomiast w technice anaglifowej 4,57 przy odchyleniu standardowym równym 1,40. W przypadku badanego wpływu bodźca wzrokowego (prezentowanego w technice polaryzacyjnej) na lokalizację pozornego źródła dźwięku, uzyskane wyniki wskazują, że istnieje wpływ ściągający, choć w sensie statystycznym nie jest on istotny. Można zatem wnioskować, iż pomimo istotnemu statystycznie wzrostowi percypowanego w obrazie efektu 3D, wyświetlanie obrazu stereoskopowego w technice polaryzacyjnej nie prowadzi do istotnych statystycznie różnic w obserwowanym wpływie ściągającym obrazu na percepcję dźwięku. Jednocześnie należy stwierdzić, że uzyskane wyniki mogą nie odzwierciedlać w pełni zjawiska wpływu ściągającego w przypadku prezentacji materiału wizyjnego w technice polaryzacyjnej, ponieważ percypowany efekt przestrzenności obrazu oceniono średnio na 7,57 w skali 0-10.

Biorąc pod uwagę wyniki uzyskane dla prezentowanego materiału badawczego należy stwierdzić, iż **w przypadku prezentacji próbki wizyjno-fonicznej w technice anaglifowej oraz polaryzacyjnej obserwuje się wpływ ściągający obrazu przestrzennego na percepcję położenia źródła dźwięku w bazie stereofonicznej, choć obserwowany wpływ ściągający nie różni się istotnie (w sensie statystycznym) w porównaniu z prezentacją próbki testowej z konwenconalnym obrazem 2D**. Powyższy wniosek jest istotny w kontekście wykorzystanego w badaniach obrazu trójwymiarowego. Biorąc pod uwagę fakt, że opisane powyżej eksperymenty przepro-

wadzono z wykorzystaniem dwóch systemów śledzenia punktu fiksacji wzroku (Tobii T60 i Cyber-Oko), dzięki którym możliwa była ocena aktywności wzrokowej osoby badanej na bodźcu wzrokowym, **dowiedzono drugiej postawionej w rozprawie tezy.**

Badanie skalowalności wpływu ściągającego

Dla przypomnienia kąt widzenia obrazu obiektu (rozumiany jako szerokość obszaru zajmowanego przez obiekt, który jest widziany przez obserwatora) decyduje o wpływie ściągającym obrazu na percepcję dźwięku niezależnie od wielkości wyświetlanego obrazu. W tym kontekście rozumiane jest w niniejszej rozprawie pojęcie „skalowalności".

W badaniu zjawiska skalowalności wykorzystano dwie grupy próbek. Przedmiotem badania pierwszej grupy próbek był wpływ dźwięku efektowego (słyszanego przed pojawieniem się źródła dźwięku w obrazie) na kierunek patrzenia. W grupie tej znalazły się dwie nieanalizowane dotychczas próbki – próbka nr 4 i próbka nr 7 (zgodnie z numeracją zastosowaną w Załączniku B). Przedmiotem badania drugiej grupy próbek było natomiast zbadanie zjawiska skalowalności wpływu ściągającego obrazu na percepcję dźwięku w kontekście lokalizowania pozornego źródła dźwięku w panoramie stereofonicznej.

Grupa I: próbki z dźwiękiem efektowym

W pierwszej grupie próbek przeanalizowano średnie wartości tak zwanej uwagi wzrokowej widzów, wyrażonej w [%] (wyznaczonej przez system śledzenia wzroku) oraz wyznaczono współczynniki korelacji dla zmiennych reprezentowanych przez: subiektywną ocenę wpływu bodźca słuchowego na uwagę wzrokową oraz uwagę wzrokową widza, będącą odzwierciedleniem wyników uzyskanych z systemu śledzenia wzroku. Zestawienie średnich wartości względnych czasów skupienia wzroku na bodźcu wzrokowym (uwagi wzrokowej) dla systemu Tobii zamieszczono w tab. 6.2, zaś dla systemu CO – w tab. 6.3.

Tab. 6.2 Zestawienie średnich wartości uwagi wzrokowej dla systemu Tobii T60

	obszar wyświetlania	L	m_{UW} [%]
próbka nr 4	średni	5	13
	mały	13	82
próbka nr 7	średni	14	29
	mały	14	72

gdzie:

L – liczba osób, które zwróciły swój wzrok na obszar zainteresowania (ROI) w przedziale czasu przed pojawieniem się źródła dźwięku w obrazie (przy czym całkowita liczba osób w badanej grupie wynosiła 15),

m_{UW} – wartość średnia względnego czasu skupienia wzroku (uwagi wzrokowej) w ROI dla wszystkich osób, które skupiły swój wzrok w obszarze zainteresowania.

Tab. 6.3 Zestawienie średnich wartości uwagi wzrokowej dla systemu Cyber-Oko

	obszar wyświetlania	L	m_{UW}
próbka nr 4	średni	10	67%
	mały	12	81%
próbka nr 7	średni	5	34%
	mały	13	73%

Należy przyjąć, że w przypadku tego badania miarodajne wyniki uzyskano tylko w przypadku pierwszej serii eksperymentu. Charakter próbek testowych poddanych badaniu wymagał braku ich wcześniejszej znajomości. Dlatego, choć wyniki zawarte w tab. 6.3 są dość optymistyczne, muszą zostać pominięte w formułowaniu wniosków związanych z niniejszym badaniem. Jak wynika z wartości zamieszczonych w tab. 6.2, uwaga wzrokowa badanych w miejscu potencjalnego pojawienia się źródła dźwięku (sugerowanego przez kierunek dochodzenia bodźca słuchowego) była stosunkowo niska. Dodatkowo, w przypadku próbki nr 4 prezentowanej w średnim obszarze wyświetlania, jedynie 5 spośród 15 badanych osób zwróciło swój wzrok w kierunku ROI. Obserwacje te pozwalają wysnuć wniosek, iż dźwięk efektowy słyszany przed pojawieniem się w obrazie źródła tego dźwięku nie wpływa zasadniczo na kierunek patrzenia widza.

Drugim etapem analizy pierwszej grupy próbek było wyznaczenie odpowiednich współczynników korelacji. Współczynniki korelacji zostały wyznaczone dla zmiennych reprezentujących subiektywną ocenę wpływu bodźca słuchowego na uwagę wzrokową oraz uwagę wzrokową widza. Wartości współczynników zamieszczono w tab. 6.4 i 6.5.

Tab. 6.4 Zestawienie wartości współczynników korelacji dla systemu Tobii T60

	obszar wyświetlania	
	średni	mały
próbka nr 4	0,26	0,28
próbka nr 7	-0,04	0,02

Tab. 6.5 Zestawienie wartości współczynników korelacji dla systemu Cyber-Oko

	obszar wyświetlania	
	średni	ekran projektora
próbka nr 4	0,32	0,22
próbka nr 7	-0,13	-0,05

Wartości współczynników korelacji zawarte w powyższych tabelach wskazują na brak związku pomiędzy subiektywnymi ocenami wpływu bodźca wzrokowego na uwagę wzrokową a czasem skupienia wzroku w obszarze potencjalnego pojawienia się źródła dźwięku. W związku z powyższym, w przypadku próbek z dźwiękiem efektowym nie można mówić o wystąpieniu wpływu ściągającego dźwięku na kierunek patrzenia, zatem tym bardziej nie można stwierdzić, że dla tych próbek wpływ ściągający jest skalowalny.

Grupa II: lokalizacja źródła dźwięku w panoramie stereofonicznej

Poniżej przedstawiono wyniki badania skalowalności dla próbek nr 2 i nr 9, zgodnie z numeracją próbek przyjętą w Załączniku B niniejszej rozprawy. W badaniu drugiej grupy próbek nie będą brane pod uwagę wyniki ocen subiektywnych otrzymanych dla położenia fortepianu w panoramie stereofonicznej. Wyniki te okazały się bo-

wiem nieistotne statystycznie. Warto zaznaczyć, że każdą próbkę scharakteryzowano w dwóch etapach: dla małego i średniego oraz średniego i dużego obszaru wyświetlania. W ramach analizy statystycznej wykonano test ANOVA. „Skalowalność” zostanie potwierdzona wówczas, gdy nie zostanie odrzucona hipoteza zerowa o równości średnich analizowanych zmiennych.

Próbka nr 2

Mały i średni obszar wyświetlania

W pierwszej kolejności przeanalizowano wyniki ocen subiektywnych uzyskanych dla małego i średniego obszaru wyświetlania. W niniejszej próbce badani wskazywali położenie głosu bohatera w bazie stereofonicznej. Otrzymane wyniki przeanalizowano za pomocą testu ANOVA. Zgodnie z założeniami testu ANOVA scharakteryzowanymi w podrozdziale 6.1, rozkład analizowanych zmiennych musi być rozkładem normalnym, a wariancje tych zmiennych muszą być równe. W celu sprawdzenia pierwszego warunku przeprowadzono test Shapiro-Wilka, natomiast równość (homogeniczność) wariancji zbadano za pomocą testu Levene’a.

Należy w tym miejscu zaznaczyć, że wyznaczone wartości parametrów poszczególnych testów statystycznych zostały zapisane w Załączniku D z dokładnością do szóstego miejsca po przecinku. W niniejszym rozdziale natomiast wartości odpowiednich parametrów zapisano z dokładnością do trzeciego miejsca po przecinku. W interpretacji wyników taka dokładność jest bowiem wystarczająca.

Tab. 6.6 Zestawienie wartości testu Shapiro-Wilka dla zmiennych reprezentujących położenie pozornego źródła dźwięku (mały i średni obszar wyświetlania)

obszar wyświetlania	***N***	***W***	***p***
mały	15	0,892	0,071
średni	15	0,929	0,268

gdzie:

N – liczba obserwacji,

W – wartość statystyki Shapiro-Wilka,

p – poziom *p*, poziom krytyczny testu.

Wyniki testu Shapiro-Wilka wskazują na spełnienie warunku o rozkładzie normalnym analizowanych zmiennych.

Tab. 6.7 Zestawienie wartości testu Levene'a dla zmiennych reprezentujących położenie pozornego źródła dźwięku (mały i średni obszar wyświetlania)

	SS Efekt	**df Efekt**	**MS Efekt**	**SS Błąd**	**df Błąd**	**MS Błąd**	***F***	***p***
lokalizacja poz.źr.dźw.	2,352	1	2,352	88,277	28	3,153	0,746	0,395

gdzie:

SS Efekt – suma kwadratów odchyleń pomiędzy grupami; inaczej: wariancja wyjaśniona przez eksperyment (wariancja kontrolowana),

df Efekt – liczba stopni swobody pomiędzy grupami,

MS Efekt – efekt średniokwadratowy, zmienność pomiędzy grupami,

SS Błąd – suma kwadratów odchyleń wewnątrz grup, prawdziwy błąd losowy,

df Błąd – liczba stopni swobody wewnątrz grup,

MS Błąd – błąd średniokwadratowy, zmienność wewnątrz grup,

F – wartość testu F, związana z rozkładem F (Fischera-Snedeckora).

Warunek jednorodności wariancji został spełniony, ponieważ poziom $p>0,05$.

Tab. 6.8 Zestawienie wartości testu ANOVA dla zmiennych reprezentujących położenie pozornego źródła dźwięku (mały i średni obszar wyświetlania)

	SS Efekt	**df Efekt**	**MS Efekt**	**SS Błąd**	**df Błąd**	**MS Błąd**	***F***	***p***
lokalizacja poz.źr.dźw.	14,7	1	14,7	354,267	28	12,652	1,162	0,29

Przy założonym poziomie $p>0,05$ nie ma podstaw do odrzucenia hipotezy zerowej. Oznacza to, że średnie wartości ocen subiektywnych udzielonych po projekcji próbki w małym i średnim obszarze wyświetlania są równe. W związku z powyższym w przypadku próbki nr 2 wykazano istnienie zjawiska skalowalności dla małego i średniego obszaru wyświetlania.

Średni i duży obszar wyświetlania

W drugim kroku badania zjawiska skalowalności dla próbki nr 2 wykonano test ANOVA dla zmiennych reprezentujących wyniki ocen subiektywnych, udzielonych po projekcji próbki na średnim i dużym obszarze wyświetlania. Warto przypomnieć, iż duży obszar wyświetlania oznacza wyświetlanie materiału wizyjnego na ekranie projektora. Analizowane poniżej wyniki uzyskano podczas eksperymentu z wykorzystaniem systemu CO.

Tab. 6.9 Zestawienie wartości testu Shapiro-Wilka dla zmiennych reprezentujących położenie pozornego źródła dźwięku (średni i duży obszar wyświetlania)

obszar wyświetlania	*N*	*W*	*p*
średni	15	0,983	0,985
duży (projektor)	15	0,958	0,658

Wyniki testu Shapiro-Wilka wskazują na spełnienie warunku o rozkładzie normalnym analizowanych zmiennych.

Tab. 6.10 Zestawienie wartości testu Levene’a dla zmiennych reprezentujących położenie pozornego źródła dźwięku (średni i duży obszar wyświetlania)

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	5,125	1	5,125	242,287	28	8,653	0,592	0,448

Zgodnie z wartościami zawartymi w powyższej tabeli wariancje analizowanych zmiennych są równe. Można zatem wykonać test ANOVA.

Tab. 6.11 Zestawienie wartości testu ANOVA dla zmiennych reprezentujących położenie pozornego źródła dźwięku (średni i duży obszar wyświetlania)

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	14,7	1	14,7	654,667	28	23,381	0,629	0,435

Wartość testu $F(1, 28) = 0{,}629$ przy $p<0{,}05$ oznacza brak podstaw do odrzucenia hipotezy zerowej. Można zatem uznać, że średnie wartości ocen subiektywnych udzielonych po projekcji próbki w średnim i dużym obszarze wyświetlania są równe.

Wniosek: w próbce nr 2, w której lokalizowano położenie głosu bohatera w panoramie stereofonicznej, stwierdzono zjawisko skalowalności wpływu ściągającego obrazu na percepcję dźwięku. Wykazana skalowalność wystąpiła między małym, średnim i dużym obszarem wyświetlania.

Próbka nr 9

W przypadku próbki nr 9 badano wpływ ściągający obrazu na percepcję dźwięku dla dwóch źródeł dźwięku: skrzypiec i fortepianu. W związku z tym, że wartość zbadanego wpływu ściągającego dla fortepianu okazała się nieistotna statystycznie (jak wynika z tab. 6.1), pominięto badanie zjawiska skalowalności dla tego instrumentu. Zamieszczona poniżej analiza statystyczna odnosi się jedynie do wpływu ściągającego obrazu skrzypiec na percepcję dźwięku.

Mały i średni obszar wyświetlania

Podobnie jak w przypadku próbki nr 2, w pierwszej kolejności przeanalizowano wyniki ocen subiektywnych uzyskanych dla małego i średniego obszaru wyświetlania.

Tab. 6.12 Zestawienie wartości testu Shapiro-Wilka dla zmiennych reprezentujących położenie pozornego źródła dźwięku (mały i średni obszar wyświetlania)

obszar wyświetlania	*N*	*W*	*p*
mały	15	0,963	0,747
średni	15	0,946	0,456

Wyniki testu Shapiro-Wilka wskazują na spełnienie warunku o rozkładzie normalnym analizowanych zmiennych.

Tab. 6.13 Zestawienie wartości testu Levene'a dla zmiennych reprezentujących położenie pozornego źródła dźwięku (mały i średni obszar wyświetlania)

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	10,482	1	10,482	276,57	28	9,877	1,061	0,312

Wynik testu Levene'a wskazuje, że warunek równości wariancji został spełniony.

Tab. 6.14 Zestawienie wartości testu ANOVA dla zmiennych reprezentujących położenie pozornego źródła dźwięku (mały i średni obszar wyświetlania)

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	1,633	1	1,633	984,667	28	35,167	0,046	0,831

Wartości testu ANOVA zestawione w powyższej tabeli wskazują na brak podstaw do odrzucenia hipotezy zerowej. Oznacza to równość średnich wartości ocen subiektywnych udzielonych po projekcji próbki w małym i średnim obszarze wyświetlania. W ten sposób wykazano istnienie zjawiska skalowalności dla małego i średniego obszaru wyświetlania próbki nr 9.

Średni i duży obszar wyświetlania

W drugim kroku analizy zbadano wystąpienie zjawiska skalowalności wpływu ściągającego obrazu na percepcję dźwięku dla średniego i dużego obszaru wyświetlania. Zgodnie z założeniami testu ANOVA w pierwszej kolejności wykonano test Shapiro-Wilka badający rozkład normalny analizowanych zmiennych.

Tab. 6.15 Zestawienie wartości testu Shapiro-Wilka dla zmiennych reprezentujących położenie pozornego źródła dźwięku (średni i duży obszar wyświetlania)

obszar wyświetlania	*N*	*W*	*p*
średni	15	0,87	0,033
duży (projektor)	15	0,913	0,152

Rozkład zmiennej odpowiadającej wynikom ocen subiektywnych uzyskanych dla obszaru średniego nie pokrywa się z rozkładem normalnym. Niespełnienie podstawowego warunku, związanego z rozkładem normalnym zmiennej, uniemożliwia wykonanie testu ANOVA. Oznacza to konieczność przeprowadzenia alternatywnego testu Kruskala-Wallisa, który nie wymaga rozkładu normalnego analizowanych zmiennych.

Test Kruskala-Wallisa:

$$H = 0{,}34, \text{ przy } p = 0{,}56$$

W związku z tym, że poziom $p>0{,}05$ nie ma podstaw do odrzucenia hipotezy o równości rozkładów. Wynik testu Kruskala-Wallisa wykazał, że średnie wartości ocen subiektywnych udzielonych po projekcji próbki w średnim i dużym obszarze wyświetlania są równe.

Wniosek: w próbce nr 9 stwierdzono istnienie zjawiska skalowalności wpływu ściągającego obrazu na percepcję dźwięku. Podobnie, jak w przypadku próbki nr 2, wykazana skalowalność wystąpiła między małym, średnim i dużym obszarem wyświetlania.

6.2.2 Związek pomiędzy wpływem ściągającym a wybranymi parametrami bodźca wzrokowego

W niniejszym podrozdziale zbadano relację pomiędzy przeanalizowanymi danymi subiektywnymi a danymi obiektywnymi. Dane subiektywne rozumie się w tym kontekście jako obserwowany wpływ ściągający, wyznaczone na podstawie subiektywnych ocen badanych, wskazujących lokalizację pozornego źródła dźwięku. W dalszej części niniejszej pracy wpływ ściągający będzie oznaczany literą *V*. Wśród analizowanych danych obiektywnych wymienić należy:

- **położenie bodźca wzrokowego**, związane ze współrzędną odciętą jego środka ciężkości – oznaczane literą c [°],
- **uwagę wzrokową widza**, czyli względną wartość czasu skupienia wzroku na danym bodźcu wzrokowym – w dalszej części pracy oznaczaną literą a [%],

– **wielkość wyświetlanego obiektu**, rozumianą jako szerokość wyświetlanego bodźca wzrokowego – oznaczaną literą w [°].

Poszukiwanie związku pomiędzy subiektywnie wyznaczonym wpływem ściągającym a scharakteryzowanymi powyżej parametrami obiektywnymi można zapisać zgodnie z formułą 6.1.

$$V = f(c, a, w) \tag{6.1}$$

Uwzględniając jednak charakter analizowanych zmiennych (w, $a > 0$), formułę 6.1 przedstawiono za pomocą formuły 6.2 i 6.3.

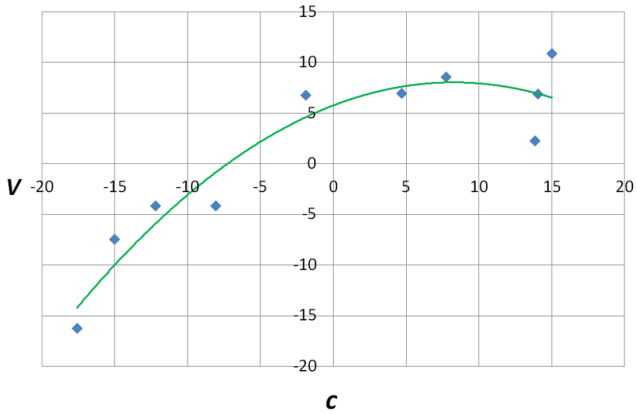
$$V = f(c) \tag{6.2}$$

$$|V| = f(a, w) \tag{6.3}$$

Analizę powyżej postawionego problemu przeprowadzono niezależnie dla wyników uzyskanych z wykorzystaniem systemów: Tobii oraz Cyber-Oko.

Analiza wyników: Tobii

W pierwszej kolejności zbadano związek pomiędzy obserwowanym wpływem ściągającym a położeniem bodźca wzrokowego w obrazie. W celu zastosowania prawidłowego współczynnika korelacji sporządzono tzw. wykres rozrzutu zmiennych V i c.



Rys. 6.2 Wykres rozrzutu dla zmiennych V i c

Linia trendu poprowadzona między zaznaczonymi na wykresie punktami wskazuje na wykładniczy charakter funkcji wiążącej zmienne V i c. Jak wspomniano w opisie metod statystycznych w podrozdziale 6.1, w przypadku, gdy zależność pomiędzy zmiennymi

nie jest prostoliniowa, lepszym „narzędziem" do określania siły związku między nimi jest współczynnik rang Spearmana, nie zaś współczynnik korelacji Pearsona.

Tab. 6.16 Współczynnik korelacji rang Spearmana: *V* i *c*

Para zmiennych	***N* ważnych**	***R* Spearman**	**poziom *p***
V i *c*	10	0,855	0,002

gdzie:

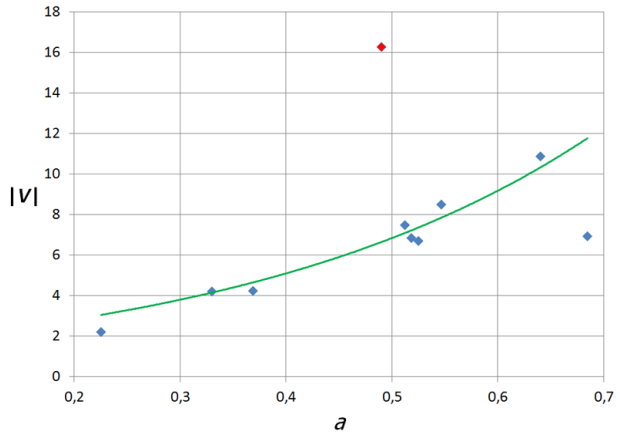
***N* ważnych** – liczba ważnych obserwacji,

***R* Spearman** – wartość współczynnika korelacji rang Spearmana, zakres wartości współczynnika [-1, +1].

Siła związku pomiędzy wpływem ściągającym a lokalizacją bodźca wzrokowego na płaszczyźnie ekranu wynosi 0,855 i jest istotna statystycznie, ponieważ $p<0,05$. Na tej podstawie można wnioskować, iż zmienne *V* i *c* są ze sobą silnie związane. Ponadto warto zauważyć, że wartość współczynnika korelacji *R* jest dodatnia, co oznacza, że ze wzrostem jednej zmiennej, np. z oddaleniem się bodźca wzrokowego od środka ekranu, wzrasta druga zmienna, czyli wpływ ściągający. Sytuacja taka niewątpliwie ma miejsce w przypadku, gdy bodziec wzrokowy przykuwający uwagę widza znajduje się w lewej części kadru, jak wynika z wykresu na rys. 6.2. Warto zauważyć, że zmiana obserwowanego wpływu ściągającego w przypadku, gdy bodziec wzrokowy jest zlokalizowany w prawej części kadru, ma inny charakter w porównaniu z sytuacją, gdy bodziec znajduje się po lewej stronie. W ogólności, wraz z oddalaniem się bodźca wzrokowego od środka ekranu w kierunku prawym wpływ ściągający nieznacznie wzrasta. Niemniej jednak, w przypadku dwóch próbek wizyjno-fonicznych: próbki nr 5 alicja (pan) oraz próbki nr 9 fortepian (pan) (zgodnie z numeracją zastosowaną w Załączniku B) zaobserwowano spadek badanego wpływu ściągającego wraz z oddaleniem się bodźca wzrokowego od środka ekranu.

Po wykazaniu wysokiego współczynnika korelacji pomiędzy wpływem ściągającym obrazu a położeniem bodźca wzrokowego w obrazie, można przejść do analizy związku zależności pomiędzy obserwowanym wpływem ściągającym ($|V|$) a czasem skupienia wzroku na danym bodźcu wzrokowym (*a*), oznaczanym również jako UW.

W celu wybrania odpowiedniego współczynnika korelacji sporządzony został wykres rozrzutu dla zmiennych $|V|$ i a.



Rys. 6.3 Wykres rozrzutu dla zmiennych $|V|$ i a

Zależność między zmienną $|V|$ i a jest nieliniowa, zatem w następnym kroku wyznaczony zostanie współczynnik korelacji rang Spearmana.

Tab. 6.17 Współczynnik korelacji rang Spearmana: $|V|$ i a

Para zmiennych	*N* ważnych	*R* Spearman	poziom *p*
$\|V\|$ i a	10	0,6	0,067

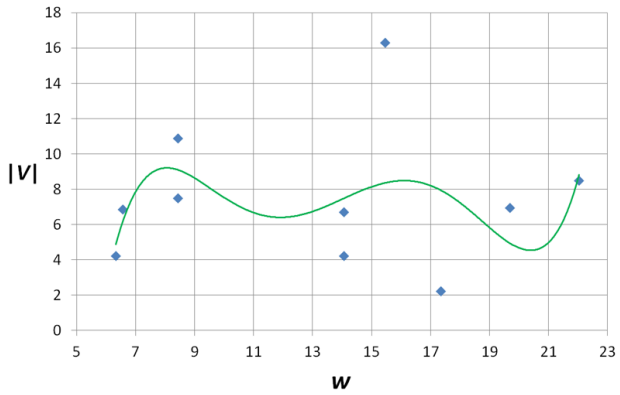
Siła związku pomiędzy obserwowanym wpływem ściągającym a uwagą wzrokową widza (rozumianą jako względny czas skupienia wzroku) na danym bodźcu wzrokowym wynosi 0,6 i w związku z tym, że $p>\alpha$, uzyskany wynik nie jest istotny statystycznie. Warto jednak zwrócić uwagę na położenie punktu, oznaczonego na rys. 6.3 kolorem czerwonym, który znacząco odbiega od pozostałych punktów pomiaru. Punkt ten reprezentuje wynik pomiaru dla próbki nr 6 głos (pan) (wg numeracji Załącznika B). Próbka nr 6 charakteryzuje się lokalizacją bodźca wzrokowego (bohatera) w skrajnie lewej części kadru. Można zatem wywnioskować, iż w przypadku badania zależności $|V|=f(a)$ próbki testowe, w których $|c|>>0$ nie stanowią odpowiedniego materiału badawczego, ponieważ wyznaczony dla nich wpływ ściągający jest zdeterminowany przez położenie bodźca wzrokowego w kadrze. Uwzględniając powyższy wniosek, zdecydowano się na ponowne wyznaczenie współczynnika korelacji rang Spearmana z pominięciem pomiaru dla próbki nr 6. Otrzymane wyniki zestawiono w tab. 6.18.

Tab. 6.18 Współczynnik korelacji rang Spearmana (z pominięciem pomiaru dla próbki nr 6): | *V*| i *a*

Para zmiennych	*N* ważnych	*R* Spearman	poziom *p*
$\lvert V \rvert$ i *a*	9	0,8	0,01

Wyznaczona wartość współczynnika korelacji rang Spearmana zmiennych $\lvert V \rvert$ i *a* w przypadku pominięcia wyniku uzyskanego dla próbki nr 6 wynosi: 0,8 i jest to wartość istotna statystycznie, ponieważ $p<0,05$. Można zatem stwierdzić, że w przypadku pierwszej serii eksperymentu (z wykorzystaniem systemu Tobii T60) zaobserwowano silną zależność pomiędzy czasem skupienia wzroku na danym bodźcu wzrokowym a obserwowanym wpływem ściągającym.

W ostatnim etapie badania zależności wyrażonej formułą 6.3 przeanalizowano związek pomiędzy zmiennymi $\lvert V \rvert$ i *w*. W pierwszej kolejności sporządzony zostaje wykres rozrzutu tych zmiennych, przedstawiony na rys. 6.4.



Rys. 6.4 Wykres rozrzutu dla zmiennych | *V*| i *w*

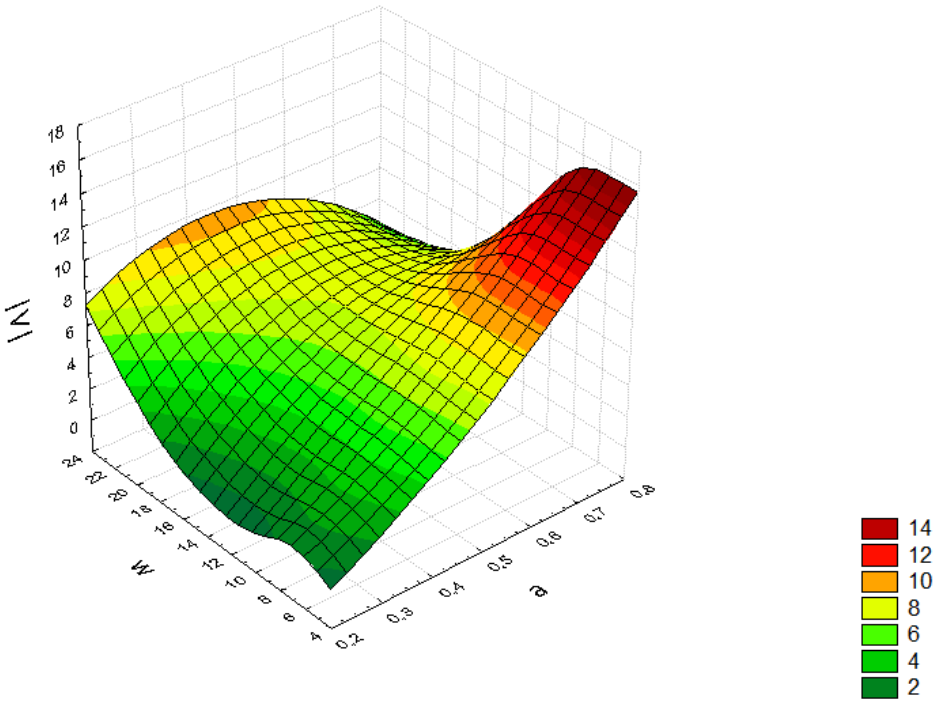
Jak wynika z rysunku 6.4 zależność pomiędzy analizowanymi zmiennymi nie jest prostoliniowa. W związku z powyższym w następnym kroku wyznaczony zostanie współczynnik korelacji rang Spearmana.

Tab. 6.19 Współczynnik korelacji rang Spearmana: | *V*| i *w*

Para zmiennych	*N* ważnych	*R* Spearman	poziom *p*
$\lvert V \rvert$ i *w*	10	0,140	0,699

Siła związku pomiędzy obserwowanym wpływem ściągającym (rozumianego jako wartość wychylenia kątowego) a wielkością wyświetlanego obiektu wynosi 0,140 i nie jest istotna statystycznie, ponieważ $p>0,05$.

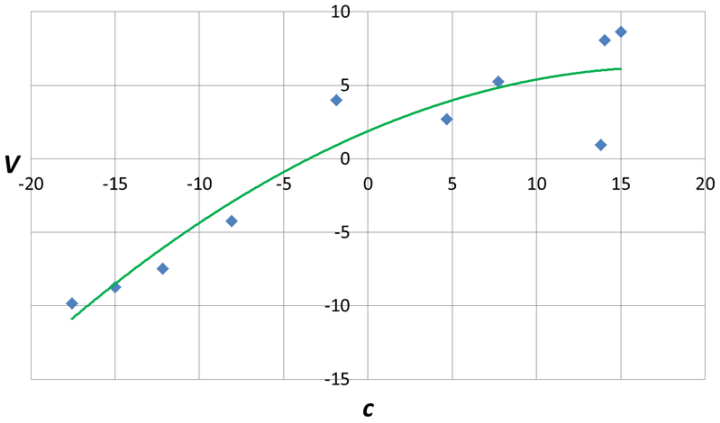
Na rys. 6.5 przedstawiono trójwymiarowy wykres prezentujący w sposób graficzny zależności pomiędzy zmiennymi $|V|$, a i w dla materiału badawczego z wyjątkiem próbki nr 6, w której wpływ ściągający jest zdeterminowany zmienną c. Warto zwrócić uwagę na zaobserwowany powyżej związek zmiennej $|V|$ ze zmienną a. Jak wykazano wzrost czasu skupienia wzroku na bodźcu wzrokowym powoduje wzrost wpływu ściągającego obrazu na percepcję dźwięku. Zależność tę odzwierciedla wartość współczynnika korelacji Spearmana równa 0,8.



Rys. 6.5 Wykres funkcji $|V| = f(w, a)$ (Tobii)

Analiza wyników: Cyber-Oko

W drugim etapie przeprowadzona została analiza wyników otrzymanych w serii eksperymentu z wykorzystaniem systemu CO. W pierwszej kolejności, podobnie jak w przypadku analizy dla systemu Tobii, zbadano relację pomiędzy obserwowanym wpływem ściągającym a położeniem bodźca wzrokowego w obrazie. W celu zastosowania prawidłowego współczynnika korelacji sporządzono tzw. wykres rozrzutu zmiennych.



Rys. 6.6 Wykres rozrzutu dla zmiennych *V* i *c*

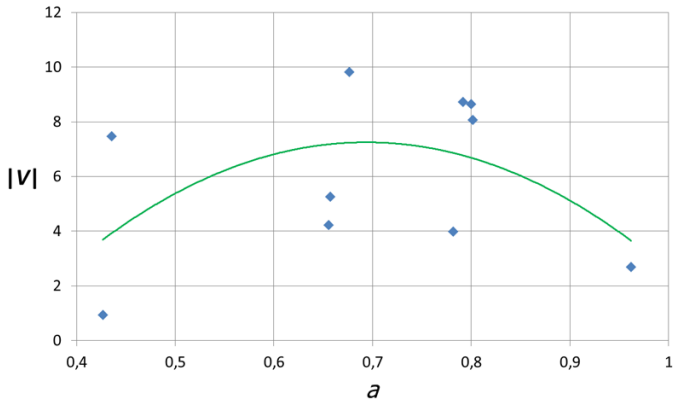
Jak wynika z wykresu na rys. 6.6 relacja między zmienną V i c jest nieliniowa. W związku z powyższym w dalszej analizie obliczony zostanie współczynnik korelacji rang Spearmana.

Tab. 6.20 Współczynnik korelacji rang Spearmana: *V* i *c*

Para zmiennych	*N* ważnych	*R* Spearman	poziom *p*
V i c	10	0,915	0,0002

Jak wynika z zestawienia wartości w tab. 6.20 współczynnik korelacji rang Spearmana wynosi 0,915 i jest istotny statystycznie. Tak wysoka wartość współczynnika (bliska jedności) oznacza istnienie silnej zależności pomiędzy obserwowanym wpływem ściągającym a lokalizacją bodźca wzrokowego w obrazie. Tym samym potwierdzony zostaje wniosek, sprecyzowany na podstawie wyników uzyskanych w eksperymencie z wykorzystaniem systemu Tobii.

W następnym etapie przeprowadzono analizę związku zmiennych $|V|$ i a. Na rys. 6.7 przedstawiono wykres rozrzutu dla zmiennej reprezentującej obserwowany wpływ ściągający oraz zmiennej reprezentującej uwagę wzrokową widza, wyznaczoną przez system śledzenia wzroku.



Rys. 6.7 Wykres rozrzutu dla zmiennych | *V*| i *a*

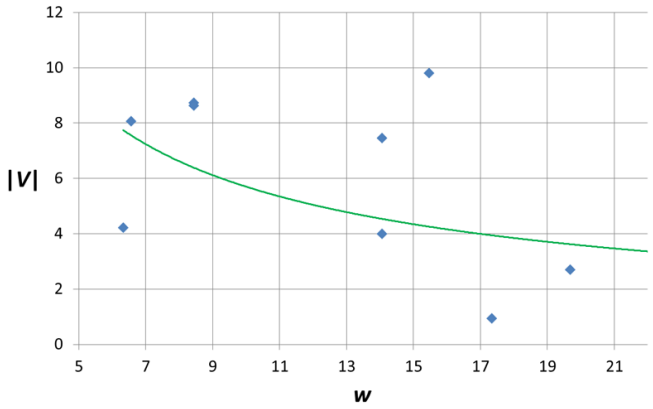
W związku z nieliniową zależnością między zmiennymi $|V|$ i a wyznaczono współczynnik korelacji rang Spearmana, wskazujący na siłę związku tych zmiennych.

Tab. 6.21 Współczynnik korelacji rang Spearmana: | *V*| i *a*

Para zmiennych	*N* ważnych	*R* Spearman	poziom *p*
$\|V\|$ i a	10	0,249	0,489

Wartość współczynnika korelacji R wynosi 0,249 i nie jest istotna statystycznie. Zatem wyniki uzyskane w serii eksperymentu z wykorzystaniem systemu CO nie wskazują na istotną statystycznie zależność między skupieniem wzroku na bodźcu wzrokowym a wpływem ściągającym. Warto jednak zauważyć, że wykreślona linia trendu (wykres rozrzutu na rys. 6.7) miałaby rosnący charakter, gdyby nie punkt pomiarowy, w którym wartość a=0,96 przy $|V|$=2,7. Punkt ten reprezentuje wyniki uzyskane dla próbki nr 3 głos (pan), wg numeracji w Załączniku B. Próbka nr 3 charakteryzuje się bodźcem wzrokowym, który dynamicznie zmienia swoje położenie w kadrze. Można zatem wywnioskować, że próbka wizyjno-foniczna, w której bodziec wzrokowy zmienia swoje położenie w kadrze, nie stanowi odpowiedniego materiału badawczego do prowadzenia eksperymentów z zakresu wpływu obrazu na percepcję dźwięku.

Ostatnim etapem badania zależności 6.3 była analiza związku pomiędzy zmiennymi $|V|$ i w. W pierwszej kolejności sporządzony został wykres rozrzutu tych zmiennych, przedstawiony na rys. 6.8.



Rys. 6.8 Wykres rozrzutu dla zmiennych | *V*| i *w*

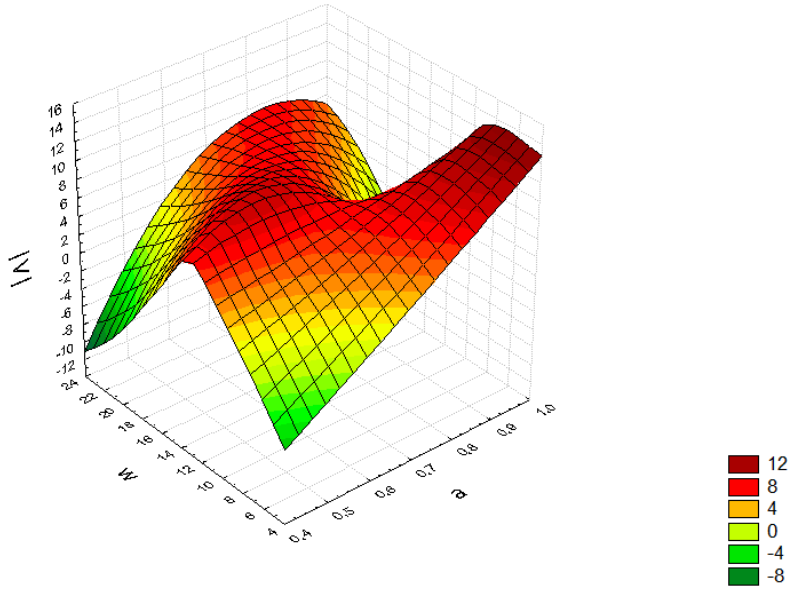
Charakter wykresu rozrzutu dla zmiennych | *V*| i *w* wskazuje na nieliniową zależność między tymi zmiennymi. W tab. 6.22 zamieszczono wartość współczynnika korelacji rang Spearmana otrzymanego w wyniku przeprowadzonych obliczeń.

Tab. 6.22 Współczynnik korelacji rang Spearmana: | *V*| i *w*

Para zmiennych	*N* ważnych	*R* Spearman	poziom *p*
	V| i *w*	10	-0,36	0,307

Współczynnik korelacji dla zmiennych | *V*| i *w* przyjmuje wartość -0,36 i nie jest to wartość istotna statystycznie. Warto zwrócić uwagę na fakt, że obliczony współczynnik korelacji jest ujemny. Oznacza to, że w tym przypadku wzrost wielkości wyświetlanego obiektu w ogólności zmniejszał wpływ ściągający obrazu na percepcję dźwięku.

Na poniższym rysunku przedstawiono w sposób graficzny relacje pomiędzy zmiennymi | *V*|, *a* oraz *w*. Analizując jedynie wartości zmiennej | *V*| w zależności od zmiennej *a*, można zauważyć wzrost wpływu ściągającego (reprezentowanego przez zmienną | *V*|) związany ze wzrostem czasu skupienia wzroku na bodźcu wzrokowym (reprezentowanym przez zmienną *a*).



Rys. 6.9 Wykres funkcji $|V| = f(w, a)$ (Cyber-Oko)

Wnioski wspólne:

Badania przeprowadzone z wykorzystaniem zarówno systemu Tobii, jak i systemu CO wykazały istnienie silnego związku między wpływem ściągającym a położeniem bodźca wzrokowego w obrazie. W przypadku badań przeprowadzonych z wykorzystaniem systemu Tobii, współczynnik korelacji wyniósł 0,855, zaś w przypadku systemu CO: 0,915. Zarówno w pierwszym, jak i drugim przypadku wyznaczone wartości są istotne statystycznie. W związku z powyższym można potwierdzić istnienie relacji między zmiennymi V i c, wyrażonej formułą 6.2. Jednocześnie należy zaznaczyć, iż nie można w sposób precyzyjny wyznaczyć wzoru funkcji opisującej tę zależność. Na podstawie wykresów rozrzutu przedstawionych na rys. 6.2 i 6.6 wyznaczono wzory funkcji aproksymujących trend rozrzutu zmiennych V i c w przypadku badania z wykorzystaniem systemu Tobii oraz systemu CO. Formuła 6.4 przedstawia wzór funkcji aproksymującej rozrzut dla zmiennych V i c w przypadku badania z wykorzystaniem systemu Tobii.

$$V = -0{,}0307c^2 + 0{,}5622c + 4{,}9969 \tag{6.4}$$

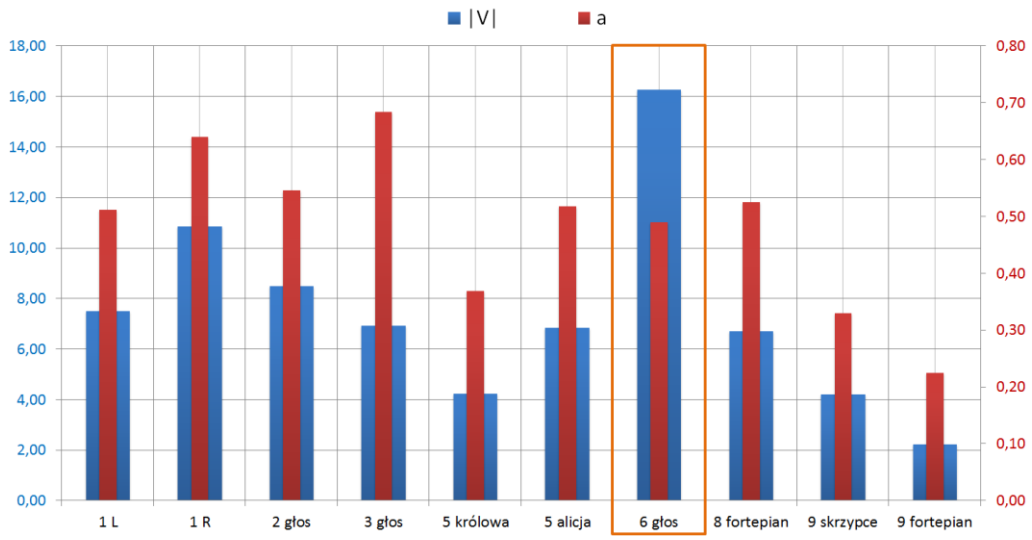
Funkcję aproksymującą trend rozrzutu zmiennych V i c w przypadku badania z wykorzystaniem systemu CO opisuje formuła 6.5.

$$V = -0{,}011c^2 + 0{,}524c + 1{,}607 \tag{6.5}$$

Pomimo, że nie można wyznaczyć formuły precyzyjnie opisującej zależność zmiennych V i c, można zauważyć, że funkcja linii trendu wyznaczonej na wykresie rozrzutu tych zmiennych ma charakter wykładniczy. Choć parametr a funkcji kwadratowej przyjmuje wartość ujemną w obu przypadkach, w analizowanym zakresie wartości znajduje się zbocze rosnące funkcji wykładniczej. Potwierdza to wcześniejszą obserwację, zgodnie z którą **ze wzrostem odległości bodźca od środka ekranu wzrasta obserwowany wpływ ściągający obrazu na percepcję dźwięku**, przy czym – wzrost wpływu ściągającego w przypadku oddalania się bodźca wzrokowego w kierunku lewej krawędzi ekranu jest większy w porównaniu z oddalaniem się bodźca wzrokowego w kierunku prawej krawędzi ekranu. W ten sposób wykazano istnienie zależności, którą badano w ramach celu cząstkowego rozprawy.

Analiza relacji wpływu ściągającego i uwagi wzrokowej widza pozwala wysnuć wniosek, iż **w ogólności wpływ ściągający zależy od czasu skupienia wzroku na bodźcu przykuwającym uwagę wzrokową**. Należy jednak zaznaczyć, iż zależność ta nie została potwierdzona w przypadku próbki testowej, dla której wpływ ściągającegy był zdeterminowany położeniem bodźca w kadrze. Ponadto, warto zauważyć, że zarówno w pierwszej, jak i drugiej serii eksperymentu wzięły udział te same osoby, a prezentowany im materiał badawczy, zgodnie z przyjętym założeniem, nie uległ zmianie. Takie podejście prawdopodobnie miało wpływ na czas skupienia wzroku osób badanych na fragmentach obrazu przykuwających wzrok w drugiej serii eksperymentu. Osoby, którym materiał badawczy był prezentowany po raz drugi, w pewnym sensie znały treść wizyjną próbek. W związku z powyższym, należy przyjąć, że wartości związane z uwagą wzrokową widzów na zdefiniowanych obszarach zainteresowania, wyznaczone przez system CO, nie odzwierciedlają badanego zjawiska w pełni. Dlatego przyjęto, że w precyzowaniu wniosku dotyczącego zależności pomiędzy czasem skupienia wzroku osoby badanej na bodźcu wzrokowym (uwagą wzrokową) a obserwowanym wpływem ściągającym, powinny być brane pod uwagę tylko wyniki pierwszej serii eksperymentu – z wykorzystaniem systemu Tobii. Współczynnik korelacji zmiennych reprezentujących uwagę wzrokową widza (a) i obserwowany wpływ ściągający ($|V|$) wyniósł 0,8 i jest to wartość istotna statystycznie. Ponadto, jak wynika z wykresów przedstawionych na rys. 6.5 i 6.9 wartości zmiennej $|V|$, reprezentującej badany

wpływ ściągający wzrastają wraz ze wzrostem zmiennej *a*, reprezentującej uwagę wzrokową widza. Na rys. 6.10 zestawiono wyznaczony wpływ ściągający (kolor niebieski) wraz z względnymi wartościami czasu skupienia wzroku na bodźcach wzrokowych poszczególnych próbek wizyjno-fonicznych (kolor czerwony). Nazwy poszczególnych próbek korespondują z nazwami próbek w Załączniku B, a zestawione wyniki zebrano podczas projekcji obrazu w średnim obszarze wyświetlania (na ekranie monitora komputerowego). Analiza poniższego wykresu pozwala potwierdzić wniosek, że w przypadku próbek, dla których obserwowany wpływ ściągający nie jest zdeterminowany położeniem bodźca wzrokowego w obrazie (na wykresie na rys. 6.10 próbka „6 głos”, której wartości otoczono pomarańczową ramką), wpływ ściągający wzrasta wraz ze wzrostem czasu skupienia wzroku na bodźcu wzrokowym. **Dowodzi to pierwszej postawionej w rozprawie tezy**.



Rys. 6.10 Porównanie obserwowanego wpływu ściągającego z czasem skupienia wzroku osób badanych na bodźcach wzrokowych poszczególnych próbek testowych

Analizując zależność pomiędzy obserwowanym wpływem ściągającym a wielkością wyświetlanego obiektu, wyznaczono współczynniki korelacji Spearmana, których wartości w obu przypadkach (eksperyment z wykorzystaniem Tobiego i CO) są stosunkowo niskie. Jednocześnie w obu seriach eksperymentu uzyskane wartości współczynnika korelacji nie są istotne statystycznie. Oznacza to, że na podstawie przeprowadzonych badań **brak jest podstaw zarówno do potwierdzenia, jak i odrzucenia związku pomiędzy wielkością wyświetlanego na ekranie obiektu i obserwowanym wpływem ściągającym**.

7 Podsumowanie

W rozprawie zaproponowano oraz zweryfikowano autorską metodologię prowadzenia badań z zakresu korelacji wzrokowo-słuchowych w kontekście lokalizacji pozornego źródła dźwięku w panoramie stereofonicznej. Nowatorskim aspektem zaproponowanej metodologii jest analiza danych subiektywnych wraz z danymi obiektywnymi, głównie współrzędnymi PoR, odzwierciedlającymi uwagę wzrokową osoby badanej, pozyskiwanymi z systemu śledzenia wzroku. Dotychczas badania prowadzone w tym obszarze opierały się głównie na wynikach ocen subiektywnych pozyskiwanych z ankiety.

W przeprowadzonych w ramach rozprawy badaniach wykorzystano dwa różne systemy śledzenia wzroku: komercyjny system Tobii T60 oraz system Cyber-Oko, opracowany w Katedrze Systemów Multimedialnych. Przeprowadzenie dwóch serii eksperymentów umożliwiło weryfikację wniosków dotyczących wpływu ściągającego, badanego w poszczególnych próbkach testowych i przyczyniło się do ich obiektywizacji. Dla poszczególnych parametrów obiektywnych bodźca wzrokowego otrzymano podobne wykresy rozrzutu zmiennych zarówno w badaniu z wykorzystaniem systemu Tobii T60, jak i Cyber-Oko, co również wpłynęło na wzrost obiektywności uzyskanych wyników. Ponadto, możliwe stało się porównanie funkcjonalności obu systemów w kontekście badania wpływu obrazu na lokalizację pozornego źródła dźwięku w panoramie stereofonicznej. Po zakończeniu każdej serii eksperymentu badani udzielali odpowiedzi na temat rozpraszania uwagi oraz komfortu pracy z systemem. Okazało się, że w przypadku systemu Tobii 14 z 15 badanych uznało, że system nie rozpraszał ich podczas badania, zaś średnia ocen badanych dotycząca komfortu pracy z systemem (w skali 0-10) wyniosła 6,73, przy wartości odchylenia standardowego równej 1,91. W przypadku systemu CO natomiast, wszyscy badani uznali, że system nie wpływał na ich koncentrację w czasie przeprowadzania eksperymentu, a średnia wartość ocen wskazujących na komfort pracy z systemem wyniosła 7,53, przy odchyleniu standardowym równym 1,41. Jak wynika z analizy zestawionych powyżej wartości, subiektywna ocena pracy z systemem Tobii jest porównywalna z subiektywną oceną pracy z wykorzystaniem systemu CO. Jednocześnie warto zwrócić uwagę na fakt, że w kontekście prze-

prowadzanych badań nie zauważono różnicy w dokładności działania systemu CO w porównaniu z systemem Tobii. Ponadto, wymienić można kilka zalet systemu CO, świadczących o jego większej funkcjonalności w prowadzeniu badań nad wpływem obrazu na percepcję dźwięku w stosunku do systemu Tobii T60. Po pierwsze, należy wspomnieć o możliwości wyświetlania obrazu stereoskopowego w technice polaryzacyjnej, która charakteryzuje się większą efektywnością w stosunku do techniki anaglifowej. Jak wykazano w ramach eksperymentu dodatkowego opisanego w podrozdziale 6.2.1 system Tobii, umożliwia wyświetlanie trójwymiarowego obrazu wizyjnego jedynie w technice anaglifowej. Ponadto, system CO umożliwia prowadzenie badań korelacji wzrokowo-słuchowych przy wyświetlaniu obrazu na ekranie projektora (duży obszar wyświetlania) dzięki zastosowaniu specjalnej ramki imitującej ekran monitora CO. Wreszcie należy wspomnieć o wykorzystaniu materiału badawczego z dźwiękiem dookólnym. Na potrzeby prowadzonych w ramach rozprawy badań przygotowano próbki testowe z dźwiękiem przestrzennym (5.1), jednakże nie wykorzystano ich w eksperymencie, ponieważ pierwsza seria badań odbyła się z wykorzystaniem systemu Tobii, który nie miał możliwości odtworzenia dźwięku w tym systemie. Warto w tym miejscu zaznaczyć, iż niezależnie od badań prowadzonych w ramach niniejszej rozprawy, autor przeprowadził eksperyment (z wykorzystaniem systemu CO), w którym badał wpływ bodźca wzrokowego zmieniającego swoje położenie w płaszczyźnie przód-tył (obraz 3D) na lokalizację pozornego źródła dźwięku w wielokanałowej panoramie stereofonicznej [83].

Przytoczone powyżej argumenty uzasadniają wniosek, iż w badaniu korelacji wzrokowo-słuchowych w kontekście lokalizacji pozornego źródła dźwięku w panoramie stereofonicznej zdecydowanie większą funkcjonalnością charakteryzuje się system CO przy zachowaniu wystarczającej dokładności wyznaczania PoR.

Zgodnie z wynikami analizy danych subiektywnych i obiektywnych przeprowadzonej w rozdziale 6., wykazano istnienie zależności między obserwowanym wpływem ściągającym a czasem skupienia wzroku na badanym bodźcu wzrokowym – wartość wyznaczonego współczynnika korelacji wyniosła 0,8. W tym kontekście stosowanie systemu śledzenia wzroku w tego typu badaniach jest uzasadnione. Co więcej – prowadzi do obiektywizacji wyników badania, uzyskanych na podstawie ocen subiektywnych.

Przykładowo posiadanie informacji o tym, że widz nie koncentrował wzroku na danym bodźcu wzrokowym lub czas skupienia wzroku na bodźcu był zbyt krótki, uzasadnia pominięcie otrzymanego wyniku oceny subiektywnej w analizie wpływu ściągającego obrazu na percepcję dźwięku. Dowodzi to pierwszej postawionej tezy:

1. **Zastosowanie systemu śledzenia punktu fiksacji wzroku do prowadzenia badań nad wpływem obrazu na percepcję dźwięku prowadzi do obiektywizacji ich wyników.**

Jednocześnie, warto zaznaczyć, iż w przypadku badania zależności pomiędzy obserwowanym wpływem ściągającym a czasem skupienia wzroku osoby badanej na bodźcu wzrokowym z wykorzystaniem próbek wizyjno-fonicznych, dla których $|c|>>0$, wartość wyznaczonego wpływu ściągającego jest zdeterminowana przez położenie bodźca wzrokowego w kadrze. Zatem w przypadku badań z zakresu percepcji bodźców wzrokowych i słuchowych należy brać pod uwagę również wpływ czynników „niemierzalnych", związanych z psychofizjologiczną naturą badanego zjawiska. Ponadto, czas skupienia wzroku na bodźcu wzrokowym potrzebny do wywołania określonego „przesunięcia" pozornego źródła dźwięku związanego z tym bodźcem, może zależeć od konkretnej badanej osoby.

Jak wykazano w rozdziale 6., możliwe jest prowadzenie eksperymentów korelacji wzrokowo-słuchowych, w których osobie badanej prezentowany jest obraz przestrzenny z wykorzystaniem systemu śledzenia punktu fiksacji wzroku. Odnosząc się do wyników uzyskanych dla wykorzystanego w eksperymentach materiału badawczego, należy stwierdzić, iż w przypadku prezentacji próbki wizyjno-fonicznej w technice anaglifowej oraz polaryzacyjnej obserwuje się przesunięcie pozornego źródła dźwięku w kierunku bodźca wzrokowego, prezentowanego w technice stereoskopowej. W ten sposób udowodniono drugą postawioną w rozprawie tezę:

2. **Śledzenie punktu fiksacji wzroku umożliwia prowadzenie eksperymentów nad wpływem ściągającym obrazu przestrzennego na percepcję dźwięku.**

W ramach rozprawy stworzone zostało stanowisko badawcze, oparte na systemie śledzenia punktu fiksacji wzroku, przeznaczone do prowadzenia zobiektywizowanych badań korelacji wzrokowo-słuchowych. Opracowano również odpowiedni materiał badawczy, zawierający próbki wizyjno-foniczne z konwencjonalnym obrazem 2D, jak również obrazem przestrzennym (3D), szczegółowo scharakteryzowany w Załączniku B. W oparciu o opracowany materiał badawczy prezentowany z wykorzystaniem stworzonego stanowiska badawczego przeprowadzono badania, które pozwoliły na udowodnienie postawionych we Wprowadzeniu tez. W związku z powyższym **osiągnięty został cel niniejszej rozprawy**.

Analiza wyników badań przeprowadzonych z wykorzystaniem systemu Tobii oraz systemu CO wykazała istnienie silnego związku między obserwowanym wpływem ściągającym a położeniem bodźca wzrokowego w obrazie. Współczynniki korelacji wyznaczone dla zmiennych reprezentujących wpływ ściągający oraz lokalizację bodźca wzrokowego w obrazie były bardzo wysokie w przypadku dwóch serii eksperymentu, a obliczona wartość poziomu *p* wskazywała na ich istotność statystyczną. Zatem wykazując zależność **pomiędzy położeniem bodźca wzrokowego, na którym skupiony jest wzrok osoby badanej a obserwowanym wpływem ściągającym obrazu na percepcję dźwięku** osiągnięto pierwszy cel cząstkowy rozprawy.

Interesującym aspektem przeprowadzonych w ramach rozprawy eksperymentów było zbadanie wpływu ściągającego obrazu na percepcję dźwięku w przypadku prezentacji materiału badawczego w małym, średnim i dużym obszarze wyświetlania. Okazało się, że **kąt widzenia obrazu obiektu (rozumiany jako szerokość obszaru zajmowanego przez obiekt, który jest widziany przez obserwatora) decyduje o wpływie ściągającym niezależnie od wielkości wyświetlanego obiektu**. Zaobserwowane zjawisko skalowalności wpływu ściągającego wykazano dla dwóch rodzajów bodźca słuchowego: zarówno dla głosu bohatera, jak i instrumentu muzycznego – skrzypiec. Wykazując powyższą właściwość, osiągnięto drugi cel cząstkowy rozprawy.

W tym miejscu należy również wspomnieć o podjętym w przeprowadzonych badaniach wątku, związanym z wpływem czytania napisów wyświetlanych w obrazie na zmianę percepcji kierunku pozornego źródła dźwięku, wskazywanego przez osoby ba-

dane. Zaobserwowano, iż występowanie polskich napisów w anglojęzycznym filmie nie wpływa istotnie na lokalizację pozornego źródła dźwięku związanego z bodźcem wyświetlanym w obrazie. Ponadto, analiza skupienia wzroku widzów na napisach wskazuje na fakt, iż badani bardzo rzadko je czytali, pomimo że próbki wizyjno-foniczne z polskimi napisami były prezentowane w pierwszej kolejności.

Jednym z wątków badawczych podjętych przez autora rozprawy było także opracowanie metody wyznaczania PoR w przestrzeni. Tak zwaną metodę paralaksy stereoskopowej scharakteryzowano w podrozdziale 4.4.2. Wyniki przeprowadzonych eksperymentów wskazały, że efektywne wyznaczanie punktu fiksacji wzroku w trzech wymiarach w oparciu o tę metodę z wykorzystaniem systemu CO nie jest możliwe w praktyce. Niemniej jednak, warto zwrócić uwagę na fakt, że zaproponowana przez autora metoda pozwala na wyznaczanie punktu fiksacji wzroku w przestrzeni, niezależnie od tego, w której płaszczyźnie osoba badana zmienia kierunek patrzenia.

Podsumowując wnioski wynikające z przeprowadzonych w ramach rozprawy eksperymentów, należy podkreślić, że zaproponowana metodologia prowadzenia badań nad wpływem obrazu na percepcję dźwięku jest słuszna. W związku z powyższym, może być ona stosowana przez badaczy, zajmujących się odkrywaniem zjawisk jednoczesnej percepcji bodźców wzrokowych i słuchowych, którzy dążą do obiektywizacji wyników uzyskanych w drodze przeprowadzanych przez nich eksperymentów. Ponadto, należy podkreślić, że badania nad wzajemnym wpływem percepcji bodźca wzrokowego i słuchowego mogą przyczynić się do zmiany podejścia, związanego z przygotowywaniem treści wizyjno-fonicznych. Możliwe jest bowiem poszerzenie zakresu wrażeń odbieranych przez osobę oglądającą film bądź użytkownika systemu multimedialnego przy uwzględnieniu zjawiska percepcji wielomodalnej.

Poniżej przedstawiono kolejne etapy metodologii, zgodnie z którą zostały przeprowadzone badania w ramach rozprawy doktorskiej. Nowatorskie elementy niniejszej metodologii zaznaczono czcionką wytłuszczoną.

Metodologia prowadzenia badań nad wpływem obrazu na percepcję dźwięku z wykorzystaniem systemu śledzenia punktu fiksacji wzroku:

I etap: przygotowanie

1. wybór/przygotowanie materiału badawczego (próbki jednoujęciowe ze statycznym bodźcem wzrokowym),
2. **indeksacja treści wizualnej** materiału badawczego – określenie obszarów zainteresowania (ROI), czyli obszarów związanych z bodźcem wzrokowym przykuwającym uwagę wzrokową widza (zarówno na konwencjonalnym obrazie 2D, jak również na **obrazie stereoskopowym**);

II etap: badanie

3. projekcja ścieżek dźwiękowych próbek testowych i poddanie ich subiektywnej ocenie badanych, związanej z lokalizacją pozornego źródła dźwięku w bazie stereofonicznej,
4. projekcja próbek wizyjno-fonicznych wraz z **zapisem współrzędnych punktów fiksacji wzroku (PoR)**, odzwierciedlających uwagę wzrokową osoby badanej, przy czym:
    a. przed projekcją próbki: informacja o bodźcu słuchowym (lub bodźcach słuchowych), na którym osoba badana powinna się skoncentrować,
    b. po projekcji próbki – ocena subiektywna dotycząca lokalizacji pozornego źródła dźwięku,

III etap: analiza danych

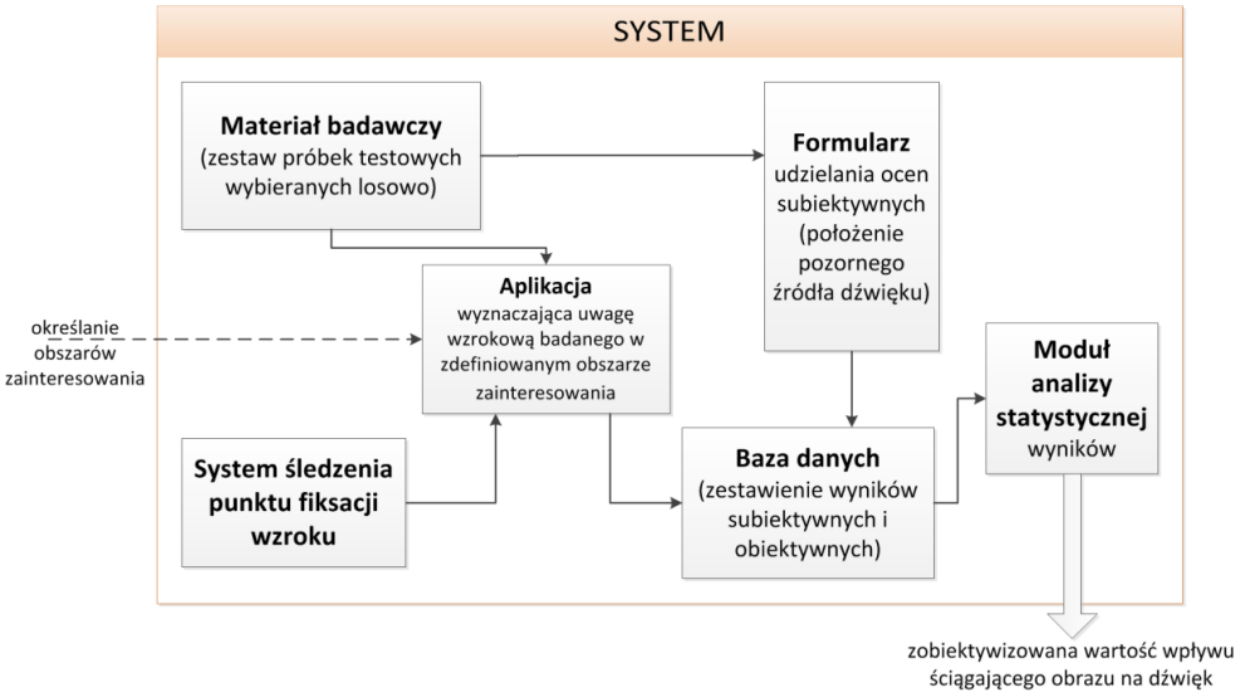
5. **wyznaczenie względnego czasu skupienia wzroku** osoby badanej w zdefiniowanym ROI,
6. zestawienie wyników subiektywnych i obiektywnych:
    - subiektywne:
    - położenie pozornego źródła dźwięku (ścieżka dźwiękowa) [°],
    - położenie pozornego źródła dźwięku (próbka wizyjno-foniczna) [°],

- **obiektywne:**
  - **uwaga wzrokowa widza w danym ROI [%]**,
  - **położenie bodźca wzrokowego w obrazie [°]**.

Perspektywy rozwoju

W przyszłości powinny zostać przeprowadzone badania korelacji wzrokowo-słuchowych z wykorzystaniem dźwięku sześciokanałowego z towarzyszeniem wizyjnego obrazu stereoskopowego wyświetlanego na monitorze polaryzacyjnym. Jak wskazują wyniki eksperymentu dodatkowego przeprowadzonego w ramach rozprawy, wyświetlanie obrazu wizyjnego w technice polaryzacyjnej zapewnia wzrost ocen percypowanej głębi stereoskopowej w stosunku do techniki anaglifowej. Porównanie funkcjonalności i ocen subiektywnych wskazujących na komfort pracy z systemem Tobii i CO uzasadnia założenie następnych badań korelacji wzrokowo-słuchowych, zgodnie z którym system CO spełnia wymagania dotyczące funkcjonalności i dokładności wyznaczania PoR.

Warto zaznaczyć, iż większość etapów przeprowadzonych badań była wykonywana niezależnie. Dla przykładu, w pierwszej kolejności wybrano zestaw próbek wizyjno-fonicznych stanowiących materiał badawczy. Następnie wykonywano badanie, po czym wyznaczano niezależnie czas skupienia wzroku w obszarze zainteresowania oraz zestawiano wyniki ocen subiektywnych dla każdej badanej próbki. Przebieg większości etapów eksperymentu można zautomatyzować. Na rys. 7.1 przedstawiono schemat ideowy proponowanego systemu, w którym zintegrowano wszystkie moduły związane z poszczególnymi etapami przebiegu eksperymentu. Autor rozprawy wyraża nadzieję, że w przyszłości będzie możliwe przeprowadzanie badania wpływu obrazu na percepcję dźwięku z wykorzystaniem w pełni modyfikowalnego systemu przeznaczonego do prowadzenia zobiektywizowanych badań korelacji wzrokowo-słuchowych w kontekście lokalizacji pozornego źródła dźwięku w panoramie stereofonicznej. I dodatkowo z zastosowaniem metod śledzenia punktu fiksacji wzroku w przestrzeni w badaniu wpływu trójwymiarowego obrazu na percepcję dźwięku.



Rys. 7.1 Schemat ideowy proponowanego systemu do obiektywizacji badań korelacji wzrokowo-słuchowych

Bibliografia

[1] A. Abel et al., "Maximising Audiovisual Correlation with Automatic Lip Tracking and Vowel Based Segmentation", Biometric ID Management and Multimodal Communication, 16-18, Madrid, Spain, 2009:
Link: http://www.isir.upmc.fr/UserFiles/File/Fabien_Ringeval/Abel_paper_42.pdf

[2] J. S. Agustin, H. Skovsgaard, J. P. Hansen, D. W. Hansen, "Low-Cost Gaze Interaction: Ready to Deliver the Promises", Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, 2009.
Dostępny pod: http://portal.acm.org/citation.cfm?doid=1520340.1520682

[3] J. S. Agustin, E. Mollenbach, M. Barret et al., "Evaluation of a Low-Cost Open-Source Gaze Tracker", Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, 2010.
Dostępny pod: http://portal.acm.org/citation.cfm?id=1743685

[4] D. Alais, D. Burr, "The Ventriloquist Effect Results from Near-Optimal Bimodal Integration", Current Biology, vol. 14, 257-262, 2004.

[5] A. Andrzejewski, "Postprodukcja nagrania wideofonicznego w wysokiej rozdzielczości z towarzyszeniem dźwięku w systemie 5.1", praca dyplomowa, promotor – prof. B. Kostek, konsultant – dr P. Odya, Katedra Systemów Multimedialnych, Politechnika Gdańska, 2009.

[6] P. Augustyniak, R. Tadeusiewicz, „Investigation of Human Interpretation Process Based on Eyetrack Features of Biosignal Visual Inspection", Proc. 2005 IEEE, Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, 2005.

[7] P. Augustyniak, R. Tadeusiewicz, „Ubiquitous Cardiology: Emerging Wireless Telemedical Applications", wyd. Medical Information Science Reference, rozdz. "Investigation of focus attention distribution during visual ECG inspection", 180-201, New York 2009.

[8] P. W. Battaglia, R. A. Jacobs, R. N. Aslin, "Bayesian integration of visual and auditory signals for spatial localization", vol. 20, No 7, Journal Optical Society of America, 1391-1397, 2003.

[9] S. Bech, V. Hansen, W. Woszczyk, "Interaction Between Audio-Visual Factors in a Home Theater System: Experimental Results", 99th Audio Eng. Soc. Conv., New York, Preprint No. 4096, 1995

[10] J. G. Beerends, F. E. de Caluwe, "The Influence of Video Quality on Perceived Audio Quality and Vice Versa", J. Audio Eng. Soc., vol. 47, No. 5, 355-362, 1999.

[11] J. C. Bennett, K. Barker, F. O. Edeko, "A New Approach to the Assessment of Stereophonic Sound System Performance", J. Audio Eng. Soc., vol. 33, No. 5, 314, 1985.

[12] R. I. Bermant, R. B. Welch, "The effect of degree of visual–auditory stimulus separation and eye position upon the spatial interaction of vision and audition", Perceptual & Motor Skill, **43**, 487-493, 1976.

[13] P. Bertelson, "The cognitive architecture behind auditory – visual interaction in scene analysis and speech identification" Current Psychology of Cognition, **13**, 69-75, 1994.

[14] P. Bertelson, "Starting from the ventriloquist: The perception of multimodal event", w: M. Sabourin, F. I. M. Craik, M. Robert (red.), Advances in psychological science: vol. 1. Biological and cognitive aspects, Hove, U.K.: Psychology Press., 419-439, 1998.

[15] P. Bertelson, G. Aschersleben, "Automatic visual bias of perceived auditory location", Psychonomic Bulletin & Review, **5**, 482-489, 1998.

[16] P. Bertelson, M. Radeau, "Ventriloquism, sensory interaction, and response bias: Remarks on the paper by Choe, Welch, Gilford, and Juola", Perception & Psychophysics, **19**, 531-535, 1976.
[17] P. Bertelson, M. Radeau, "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance", Perceptions and Psychophysics, vol. 29 (6), 578-584, 1981.
[18] P. Bertelson, J. Vroomen, B. de Gelder, J. Driver, "The ventriloquist effect does not depend on the direction of deliberate visual attention", Percept Psychophys., vol. 62 (2), 321-332, 2000.
[19] R. Bhola, "Binocular vision", M.Sc. Thesis, Department of Ophtamology and Visual Sciences, The University of Iowa, 2006.
[20] M. Bischoff, B. Walter, C. R. Blecker, K. Morgen, D. Vaitl, G. Sammer, "Utilizing the ventriloquism-effect to investigate audio-visual binding", Neuropsychologia, Nr 45, 578-586, 2007.
[21] J. Blauert, "Spatial hearing - revised edition: the psychophysics of human sound localization", The MIT Press, Rev Sub edition, 1996.
[22] J. Blauert, "Spatial Hearing: The Psychophysics of Human Sound", rozdz. 2 – Spatial hearing with one sound source, The MIT Press, Cambridge, MA, 1997.
[23] J. Blauert, "Spatial Hearing: The Psychophysics of Human Sound Localization", The MIT Press, Cambridge, 2001.
[24] A. Bochenek, M. Reicher, "Anatomia człowieka", tom V, Państwowy Zakład Wydawnictw Lekarskich, wyd. III, 507-512, Warszawa 1989.
[25] M. Bogdan, „Analiza wriancji (ANOVA)", materiały do wykładu, Instytut Matematyki i Informatyki, Politechnika Wrocławska, Wrocław 2010.
Link: http://www.im.pwr.wroc.pl/~mbogdan/dydak/In/wyklady/wyklad9.ppt
[26] M. Bogdanowicz, „Integracja Percepcyjno-Motoryczna, teoria – diagnoza – terapia", Centrum Metodyczne Pomocy Psychologiczno-Pedagogicznej, wyd. III, Warszawa 2000.
[27] R. A. Bolt, "Gaze-orchestrated dynamic windows", Proceedings of the 8th Annual Conference on Computer Graphics and interactive Techniques, SIGGRAPH '81, ACM Press, str. 109-119, 1981.
[28] A. S. Bregman, "Auditory scene analysis: the perceptual organization of sound", The MIT Press, Cambridge, MA, 1990.
[29] M. Brook, L. Danilenko, W. Strasser, "Wie bewertet der Zuschauer das stereofone Fernsehes", 13 Tonemeistertagung; Internationaler Kongres, 367-377, 1984.
[30] G. T. Buswell, "Fundamental reading habits: A study of their development", IL: University of Chicago Press, Chicago, 1922.
[31] L. M. Chalupa, J. S. Werner, "The Visual Neurosciences", vol. 1, Massachusetts Institute of Technology, 2004.
[32] T. Chen, R. R. Rao, "Audio-visual integration in multimodal communication", Proceedings of the IEEE, vol. 86, No. 5, 1998
[33] B. Cyganek, J. P. Siebert, "An introduction to 3D computer vision techniques and algorithms", wyd. John Wiley & Sons, 2010.
[34] A. Czyżewski, B. Kunka, M. Kurkowski, R. Branchat, "Comparison of developed gaze point estimation methods", NTAV/SPA 2008, 133-136, Poznań, 2008.
[35] E. T. Davis, K. Scott, J. Pair, L. F. Hodges, J. Oliverio, "Can audio enhance visual perception and performance in a virtual environment?", 43rd Human Factors and Ergonomics Society Annual Meeting, Houston, 1999.
[36] A. Dobrucki, P. Plaskota, "Computational modelling of head-related transfer function", Archives of Acoustics, vol. 32, 2007.
[37] M. Domański, "Telewizja trójwymiarowa – stan badań i perspektywy rozwoju", Przegląd Telekomunikacyjny, nr 6, 223-228, 2010.

[38] H. Drewes, "Eye Gaze Tracking for Human Computer Interaction", LFE Medien-Informatik der Ludwig-Maximilians-Universität, Monachium 2010.

[39] P. Dybski, K. Napierski, "Stworzenie bazy nagrań multimedialnych na potrzeby system do badania korelacji wzrokowo-słuchowych", praca dyplomowa, promotor – prof. B. Kostek, konsultant – B. Kunka, Katedra Systemów Multimedialnych, Politechnika Gdańska, 2010.

[40] EN 60825-1:2005, norma: „Bezpieczeństwo urządzeń laserowych Część 1: Klasyfikacja sprzętu, wymagania i przewodnik użytkownika", data publikacji: 27 kwietnia 2005

[41] L. Fauster, "Stereoscopic Techniques in Computer Graphics", mat. Nr 0425241, TU Wien, 2007.
Link: http://www.cg.tuwien.ac.at/research/publications/2006/Fauster-06-st/Fauster-06-st-.pdf

[42] P. M. Fitts, R. E. Jones, J. L. Milton, "Eye movements of aircraft pilots during instrument-landing approaches" wydany w *Aeronautical Engineering Review*, 9(2), str. 24–29, 1950 (zgodnie z odniesieniem do literatury w [63])

[43] J. D. Fix, "Neuroanatomia", wyd. I polskie, red. J. Moryś, wyd. Urban & Partner, 188-191, Wrocław 1997.

[44] A. Florek, P. Szczuko, „Badanie korelacji słuchowo-wzrokowych dla obrazu cyfrowego i dźwięku dookólnego", praca dyplomowa, promotor – prof. B. Kostek, konsultanci – P. Odya, A. Kornacki, Katedra Systemów Multimedialnych, Politechnika Gdańska, 2002.

[45] I. Frissen, J. Vroomen, B. de Gelder, P. Bertelson, "The aftereffects of ventriloquism: Generalization across sound-frequencies", Acta Psychologica, vol. 118 (1-2), 93-100, 2004.

[46] A. Gaggioli, M. Bassi, A. D. Fave, "Quality of Experience in virtual environments: Chapter 8", fragment publikacji: G. Riva, F. Davide, W.A IJsselsteijn (Eds.), "Concepts, effects and measurement of user presence in synthetic environments", wyd. IOS Press, Amsterdam, 2003.

[47] M. B. Gardner, "Proximity image effect in sound localization", J. Acoust. Soc. Amer. vol. 43, 163, 1968.

[48] L. Gooding, "The effect of viewing distance and disparity on perceived depth", Stereoscopic Displays and Applications – II, Spie Proceedings, vol. 1457, 259-266, 1991.

[49] D. W. Grantham, B. W. Y. Hornsby, E. A. Erpenbeck, "Auditory spatial resolution in horizontal, vertical, and diagonal planes", The Journal of the Acoustical Society of America, 114(2), 1009-1022, 2003.

[50] H. Haas, "The influence of a single echo on the audibility of speech", Journal of Audio Eng. Soc., vol. 20 (2), 146-159, 1972.

[51] A. Hajdukiewicz, „Badanie wpływu telewizyjnego obrazu spikera i pianisty na postrzeganie kierunku głosu i dzwięków fortepianu", materiały konf. 32. Otwartego Seminarium z Akustyki OSA'85, 371-374, Krakow.

[52] A. Hajdukiewicz, „Współzależności realizacji dźwięku i obrazu w studyjnej technice telewizyjnej", rozprawa doktorska, promotor: doc. dr inż. Gustaw Budzyński, Katedra Systemów Multimedialnych, Politechnika Gdańska, 1987.

[53] W. M. Hartmann, "On the minimum audible angle - a descision theory approach", The Journal of the Acoustical Society of America, 85(5), 2031-2041, 1989.

[54] H. Hautzel, J. G. Taylor, B. J. Krause, N. Schmitz, L. Tellmann, "The motion aftereffect: more than area V5/MT? Evidence from 15O-butanol PET studies", Brain Research., vol. 892, issue 2, 281–292, 2001.

[55] P. M. Hofman, A. J. Van Opstal, "Spectro-temporal factors in two-dimensional human sound localization" Journl of Acoustical Society of America, vol. 103, No. 5, 1998.

[56] M. P. Hollier, A. N. Rimell, D. S. Hands, R. M. Voelcker, "Multi-modal perception", BT Technology Journal, vol. 17, No. 1, 35-46, 1999.

[57] M. P. Hollier, R. Voelcker, "Objective performance assessment: video quality as an influence on audio perception", 103rd Eng. Soc. Conv., New York, Preprint No. 4590, 1997.
[58] E. B. Huey, "The psychology and pedagogy of reading", MA: MIT Press, Cambridge 1968. (oryginalne wydanie z roku 1908)
[59] ATSC Implementation Subcommittee IS-191: „Relative timing of sound and vision for broadcast operations", norma określająca dopuszczalne wartości przesunięcia dźwięku w stosunku do obrazu na potrzeby transmisji w telewizji cyfrowej (DTV), 2003. Link:http://www.atsc.org/cms/index.php/standards/document-download/doc_download/47-is-191-relative-timing-of-sound-and-vision-for-broadcast-operations
[60] ITU-R BS.1116-1: "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems", 16-19.
[61] ITU-R BS.1286: "Methods for the subjective assessment of audio systems with accompanying picture".
[62] ITU-R BT.1359, norma określająca progi wykrywalności przesunięć czasowych pomiędzy dźwiękiem a obrazem, 1998.
[63] R. J. K. Jacob, K. S. Karn, "Eye tracking in Human-Computer Interaction and usability research: ready to deliver the promises (section commentary)" zawarty w "*The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements"*, (J. Hyona, R. Radach, H. Deubel), wydawnictwo Elsevier Science, Oxford 2003.
[64] J. Ki, Y.-M. Kwon, K. Sohn, "3D gaze tracking and analysis for attentive Human Computer Interaction", Frontiers in the Convergence of Bioscience and Information Technologies, 617-621, 2007.
[65] J. Ki, Y.-M. Kwon, "3D gaze estimation and interaction", 3DTV Conference: The True Vision – Capture, Transmission and Display of 3D Video, 373 – 376, Istanbul, 2008.
[66] O. Klemm, "Untersuchungen ueber die Lokalisation von Schalrezien III: Ueber den Anteil des beideohrigen Horens", Arch. Ges. Psychol. 38, 1918.
[67] T. Kocejko, A. Bujnowski, J. Wtorek, „Dual camera based eye gaze tracking system", 4th European Conference of the International Federation for Medical and Biological Engineering, IFMBE Proceedings, vol. 22, 10, 1459-1462, Antwerp, 2009.
[68] T. Kocejko, A. Bujnowski, J. Wtorek, „Eye Mouse for disabled", Advances in Soft Computing, Human-Computer Systems Interaction, 109-122, Springer, Berlin 2009.
[69] T. Kocejko, J. Wtorek, A. Bujnowski, J. Rumiński, A. Poliński, "Authentication for elders and disabled using eye tracking", 3rd Conference on Human System Interactions (HSI), 629-633, Rzeszów, 2010.
[70] S. Komiyama, "Subjective evaluation of angular displacement between picture and sound directions for HDTV sound systems", J. Audio Eng. Soc., **37** (4), 210-214, 1989
[71] S. Komiyama, K. Nakabayashi, S. Nikaido, "Experiments on the interaction between a sound image and a video image", Society for Acoustic Research, 1981 (odwołanie do artykułu w: US Patent No. 6343132).
[72] S. Konturek, "Fizjologia człowieka", t. IV: „Neurofizjologia", Wydawnictwo Uniwersytetu Jagiellońskiego, wyd. VI, 221-223, Kraków 1998.
[73] B. Kostek, „Rough-neuro approach to testing the influence of visual cues on surround sound perception", rozdział w książce pt.: "Rough-Neuro Computing: A Way To Computing With Words" (S. K. Pal, L. Polkowski, A. Skowron eds.), Springer Verlag, Series on Artificial Intelligence, 555 - 572, 2004.
[74] B. Kostek, "Perception-based data processing in acoustics. Applications to music information retrieval and psychophysiology of hearing", wyd. Springer Verlag, Series on Cognitive Technologies, 389-400, Berlin, Heidelberg, New York 2005.

[75] B. Kostek, B. Kunka, "Application of gaze tracking technology to quality of experience domain", IEEE International Conference on Multimedia Communications, Services and Security, 134-139, Kraków, 2010.

[76] J. Kotus, B. Kunka, A. Czyzewski, P. Szczuko, P. Dalka, R. Rybacki, "Gaze-tracking and acoustic vector sensors technologies for PTZ camera steering and acoustic event detection", Database and Expert Systems Applications (DEXA), str. 276 – 280, Bilbao, 2010.

[77] Ł. Kulasek, B. Kunka, A. Czyżewski, „Badanie rozpoznawania twarzy przez człowieka z wykorzystaniem systemu śledzenia fiksacji wzroku Cyber-Oko", Elektronika, nr 1/2011, wyd. Sigma-Not, 2011

[78] M. Kulesza, „Specyfikacja produkcyjna warstwy sprzętowej systemu detekcji punktu fiksacji wzroku na monitorze komputerowym", raport wewnętrzny Katedry Systemów Multimedialnych, nr raportu MM004/04/09, str. 39-40, Gdańsk, 2009.

[79] B. Kunka, B. Kostek, "A new method of audio-visual correlation analysis", materiały konferencyjne: 2nd International Symposium on Multimedia – Applications and Processing MMAP'09, vol. 4, 497-502, Mrągowo, 2009.

[80] B. Kunka, B. Kostek, "Exploiting audio-visual correlation by means of gaze tracking", International Journal of Computer Science and Applications, vol. 7, nr 3, 104-123, 2010.

[81] B. Kunka, B. Kostek, "Objectivization of audio-video correlation assessment experiments", 128. Konwencja Audio Engineering of Society, Convention Paper No. 8148, Londyn, 2010.

[82] B. Kunka, B. Kostek, M. Kulesza, P. Szczuko, A. Czyżewski, "Gaze-tracking based audio-visual correlation analysis employing quality of experience methodology", Intelligent Decision Technologies Journal, vol. 4, nr 3, 217 – 227, 2010.

[83] B. Kunka, "Badanie korelacji słuchowo-wzrokowych w dziedzinie 3d", 14th International Symposium on Sound Engineering and Tonmeistering, Wrocław, 2011.

[84] Y.-M. Kwon, K.-W. Jeon, J. Ki, Q. M. Shahab, S. Jo, S.-K. Kim, „3D gaze estimation and interaction to stereo display", The International Journal of Virtual Reality, vol. 5 (3), 41-45, 2006.

[85] Y.-M. Kwon, J. K. Shul, "Experimental researches on gaze-based 3d interaction to stereo image display", Computer Science, 3942, Springer-Verlag, 1112-1120, 2006.

[86] J.-S. Lee, F. De Simone, T. Ebrahimi, "Efficient video coding based on audio-visual focus of attention", Journal of Visual Communication and Image Representation, Elsevier, doi:10.1016/j.jvcir.2010.11.002, 2010

[87] D. J. Levitin, K. MacLean, M. Mathews, L. Chu, "The perception of cross-modal simultaneity", Proceedings of International Journal of Computing Anticipatory Systems, Belgia, 1999.

[88] J. Lewald, „Eye-position effects in directional hearing", Behavioural Brain Research, vol. 87, 35-48, 1997.

[89] J. Lewald, "Rapid adaptation to auditory-visual spatial disparity" Learn. Mem., vol. 9, 268–278, 2002.

[90] J. Lewald, W. H. Ehrenstein, "Auditory-visual spatial integration: a new psychophysical approach using laser pointing to acoustic targets", J. Acoust. Soc. Am., vol. 104, Issue 3, 1586-1597, 1998.

[91] Y. Liu, Y. Sato, „Recovering audio-to-video synchronization by audiovisual correlation analysis", 19th International Conference on Pattern Recognition (ICPR 2008), Tampa, Florida, USA, 2008.
Link: http://figment.csee.usf.edu/~sfefilat/data/papers/MoAT4.4.pdf

[92] P. Majdak, M. J. Goupell, B. Laback, "3-D localization of virtual sound sources: effects of visual environment, pointing method, and training", Atten. Percept. Psychophys., vol. 72 (2), 454-469, 2010.

[93] J. MacDonald, McGurk, "Visual influences on speech perception processes," Perception and Psychophysics, vol. 24, nr 3, 253–257, 1978.
[94] H. McGurk, J. MacDonald, "Hearing lips and seeing voices", Nature, vol. 264, 746–748, 1976.
[95] D. J. Meares, "Perceptual attributes of multichannel sound", The Proceedings of 1st Audio Eng. Soc. International Conference, 171-179, Copenhagen, Denmark, 28-30 June 1993.
[96] B. Mendiburu, "3D movie making: stereoscopic digital cinema from script to screen", Focal Press, 2009.
[97] A. W. Mills, "On the minimum audible angle", The Journal of the Acoustical Society of America, 30, 237-246, 1958.
[98] B. Moore, „Wprowadzenie do psychologii słyszenia", Warszawa, 1997.
[99] J. D. Moore, "The development of a design tool for 5-speaker surround sound decoders", rozprawa doktorska, 98-101, University of Huddersfield, Wielka Brytania, 2009.
[100] M. Mujal, R. L. Kirlin, „Compression enhancement of video motion of mouth region using joint audio and video coding", 5th IEEE Southwest Symposium on Image Analysis and Interpretation, 2002.
[101] Y. Nakayama, K. Watanabe, S. Komiyama, F. Okano, Y. Izumi, "A method of 3-D sound image localization using loudspeaker arrays", 114 Audio Eng. Soc. Convention, Paper No. 5793, 2003.
[102] M. H. Niżankowska, „Podstawy okulistyki", wydawnictwo VOLUMED Wrocław, wyd. II, 1-5, Wrocław 2000.
[103] P. Odya, A. Czyzewski, B. Kostek, "Determination of influence of visual cues on perception of spatial sound", 110th Audio Eng. Soc. Convention, Preprint No. 5311, Amsterdam, 2001.
[104] P. Odya, B. Kostek, A. Czyzewski, "Discovering the influence of visual stimulation the perception of surround sound using genetic algorithms", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, 2001.
[105] H. Ono, "Binocular visual directions of an object when seen as single or double", Vision and Visual Dysfunction, vol. 9, red. D. Regan, Macmillian, London, 1-18, 1991.
[106] H. Pashler, "The psychology of attention", Cambridge, MA: MIT Press., 1998.
[107] F. Pirri, M. Pizzoli, A. Rudi, "A general method for the Point of regard estimation in 3D space", materiały konferencyjne: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, 2011.
[108] P. Plaskota, A. Dobrucki, "Head-related transfer function calculation using boundary element method", 122nd Convention AES, Convention paper 7101, Wiedeń, 2007.
[109] P. Plaskota, A. Dobrucki, "Reduction of numerical head model for calculation of head-related transfer function", Signal Processing Algorithms, Architectures, Arrangements, and Applications Conf. (SPA), Poznań 2007.
[110] P. Plaskota, A. Dobrucki, "The influence of pinna flare angle on Head-Related Transfer Function", Signal Processing Algorithms, Architectures, Arrangements, and Applications Conf. (SPA), Poznań 2008.
[111] Z. W. Pylyshyn, "Is vision continuous with cognition? The case for cognitive impenetrability of visual perception", Behavioural & Brain Sciences, 22, 341-423, 1999.
[112] M. Radeau, P. Bertelson, "The after-effects of ventriloquism", Q. J. Exp. Psychol., vol. 26, 63–71, 1974.
[113] M. Radeau, P. Bertelson, „Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations", Perceptions and Psychophysics, vol. 22 (2), 137-146, 1977.
[114] R. R. Rao, T. Chen, "Cross-modal prediction in audio-visual communication", IEEE International Conference on Acoustics, Speech, and Signal Processing, materiały konferencyjne, vol. 4, 1996.

[115] R. R. Rao, T. Chen, "Exploiting audio-visual correlation in coding of talking head sequences", IEEE Trans. on Industrial Electronics, 1996.
[116] G. H. Recanzone, "Rapidly induced auditory plasticity: the ventriloquism aftereffect", Proc. Natl. Acad. Sci., vol. 95, 869–875, 1998.
[117] G. H. Recanzone, M. L. Sutter, "The Biological Basis of Audition", artykuł w "Annual Review of Psychology", vol. 59, 127-128, San Diego, 2008.
[118] D. Rocchesso, P. Polotti, "Sound to Sense, Sense to Sound – A State of the Art in Sound and Music Computing", wyd. Information Society Technologies, 318–322, 2007.
[119] C. Rorden, J. Driver, "Does auditory attention shift in the direction of an upcoming saccade?", Pergamon, Neuropsychologia 37, 357-377, 1999.
[120] L. Rutkowska, J. Socha, "Statystyczna analiza danych z wykorzystaniem programu STATISTICA", skrypt do wykładu, Katedra Dendrometrii, Wydział Leśny Akademii Rolniczej w Krakowie, Kraków 2005.
Link: http://www.ar.krakow.pl/les/dendro/SAN/SAN.pdf
[121] K. Saberi, Y. Takahashi, M. Konishi, Y. Albeck, B. J. Arthur, H. Farahbod, "Effects of Interaural Decorrelation on Neural and Behavioral Detection of Spatial Cues", wyd. Neuron, vol. 21, 789-798, 1998.
[122] N. Sakamoto, T. Gotoh, T. Kogure, M. Shimbo, "Controlling Sound-Image Localization in Stereophonic Reproduction", J. Audio Eng. Soc., **29**(11), 794-798, 1981.
[123] N. Sakamoto, T. Gotoh, T. Kogure, M. Shimbo, "Controlling Sound-Image Localization in Stereophonic Reproduction: Part II", J. Audio Eng. Soc., **30**(10), 719-721, 1982.
[124] J. Sato, K. Fukue, Y. Kinoshita, K. Ozawa, "Relationship between gaze direction and sound localization in ventriloquism effect", The Acoustical Society of Japan, vol. 32, 1, 40-42, 2011.
[125] K. M. Schreiber, J. M. Hillis, H. R. Filippini, C. M. Schor, M. S. Banks, "The surface of the empirical horopter", Journal of Vision, vol. 8, No. 3, 1-20, 2008.
[126] M. Schutz, S. D. Lipscomb, "Influence of visual information on auditory perception of marimba stroke type", In S.D. Lipscomb, R. Ashley, R.O. Gjerdingen, & P. Webster (Eds.), Proceedings of the 8th International Conference on Music Perception & Cognition, 76-80. Sydney, 2004.
[127] M. Sobczyk, „Statystyka. Podstawy teoretyczne, przykłady, zadania", wydawnictwo UMCS, wyd. I, Lublin 2000.
[128] B. E. Stein, M. A. Meredith, "The merging of the senses", MIT Press, 1993.
[129] R. M. Steinman, "Gaze control under natural conditions" w książce "The visual neurosciences" [31], rozdział 90, 1339–1356, 2004.
[130] S. B. Steinman, B. A. Steinman, R. P. Garzia, „Foundations of Binocular Vision: a Clinical Perspective", McGraw-Hill, New York, 2000.
[131] R. L. Storms, M. J. Zyda, "Interactions in perceived quality of auditory-visual displays", Presence: Teleoperators and Virtual Environment, vol. 9, No. 6, 557-580, 2000.
[132] K. Szymków, J. Wilkiewicz, A. Zieleń, „Postprodukcja nagrania wideofonicznego w systemie 5.1", praca dyplomowa, promotor – prof. B. Kostek, konsultant – dr P. Odya, Katedra Systemów Multimedialnych, Politechnika Gdańska, 2009.
[133] "The ANOVA Procedure" (chapter 17), rozdział książki w wersji elektronicznej: SAS OnlineDoc, wersja 8, 337-342.
Link: http://www.okstate.edu/sas/v8/saspdf/stat/chap17.pdf
[134] G. J. Thomas, "Experimental study of the influence of vision of sound localization", J. Exp. Psych., vol. 28, 163-177, 1941.
[135] R. B. Tootell, J. B. Reppas, A. M. Dale, R. B. Look, M. I. Sereno, "Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging", Nature, vol. 375, 139–141, 1995.

[136] D. Vieth, „Über die Rechtung der Augen”, Wiley Verlag, vol. 58, **3**, 233–253, Weinheim, 1818.
[137] J. Vroomen, B. de Gelder, “Perceptual Effects of Cross-modal Stimulation: Ventriloquism and the Freezing Phenomenon”, rozdział w książce “Handbook of multisensory processes”, MIT Press, 2004. Link: http://www.beatricedegelder.com/documents/Vroomen2004Perceptualeffects.pdf
[138] M. Waltl, C. Timmerer, H. Hellwagner, “Improving the Quality of Multimedia Experience through sensory effects”, Proceedings of the 2nd International Workshop on Quality of Multimedia Experience (QoMEX2010), Trondheim, Norway, 2010.
[139] J. D. Warren, B. A. Zielinski, G. G. R. Green, J. P. Rauschecker, T. D. Griffiths, „Perception of Sound-Source Motion by the Human Brain”, artykuł w “Neuron” vol. 34, 139-148, wyd. Cell Press, 2002.
[140] R. B. Welch, D. H. Warren, “Intersensory interactions”, rozdział w książce: K. R. Boff, L. Kaufman, J. P. Thomas, “Handbook of perception and human performance – Volume 1: Sensory processes and perception”, wyd. JohnWiley & Sons, 1–36, New York 1986
[141] H. A. Witkin, S. Wapner, T. Leventhal, “Sound localization with conflicting visual and auditory coues”, Journal of Experimental Psychology, vol. 43, 58-67, 1952.
[142] T. M. Woods, G. H. Recanzone, “Visually induced plasticity of auditory spatial perception in macaques”, Curr. Biol, vol. 14, 1559–1564, 2004.
[143] A. L. Yarbus, “Eye movements and vision”, Plenum, New York 1967 (oryginalnie wydana w jęz. rosyjskim w 1962 r.)
[144] S. Morein-Zamir, S. Soto-Faraco, A. Kingstone, “Auditory capture of vision: examining temporal ventriloquism”, Cognitive Brain Research, vol. 17, 154–163, 2003.
[145] M. Zhang, W. Zhang, R. A. Kennedy, T. D. Abhayapala, “HRTF Measurements of a KEMAR Dummy-Head Microphone”, Proceedings of ACOUSTICS 2009, Adelaide, Australia, 2009.
[146] F. Zilly, J. Kluger, P. Kauff, “Production rules for stereo acquisition”, Proceedings of the IEEE, vol. 99, No. 4, 590-606, 04.2011.
[147] U. Zimmer, E. Macaluso, “High binaural coherence determines successful sound localization and increased activity in posterior auditory areas”, wyd. Neuron, vol. 47, 893-905, 2005.
[148] D. Y. N. Zotkin, J. Hwang, R. Duraiswaini, L. S. Davis, “HRTF personalization using anthropometric measurements”, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 157-160, 2004.

STRONY INTERNETOWE:
Stan podanych poniżej adresów stron internetowych z dnia 08.09.2011 r.

[149] http://www.alea-technologies.de/pages/en/products/intelligaze.php
Specyfikacja techniczna i zdjęcie systemu IntelliGaze.
[150] www.alexnoyes.com/page1/files/Sound_Perception.pdf
Artykuł pt. „About the perception of sound”, K. Dykhoff. W artykule opisano percepcję dźwięku w kontekście percepcji obrazu. Dane dotyczące pojemności zmysłów.
[151] http://www.apcthai.com/apc/page.php?id=22
Podstrona Advance Progressive Addition Lenses Club. Rysunek przedstawiający przestrzeń fuzyjną Panuma.
[152] http://www.brainconnection.com/topics/?main=anat/vision-work
Strona portalu medycznego BrainConnection. Informacje nt. percepcji trzeciego wymiaru.

[153] http://ccrs.nrcan.gc.ca/resource/tutor/stereo/chap3/chapter3_4_e.php
Strona edukacyjna. Ogólne informacje nt. techniki polaryzacyjnej.
[154] http://www.cogain.org/
Strona projektu COGAIN.
[155] http://www.cogain.org/wiki/Eye_Tracker_Erica
Podstrona portalu COGAIN, zawierająca specyfikację techniczną "gaze trackera" ERICA.
[156] http://www.cogain.org/wiki/File:COGAIN2008-Windows-control-with_ERICA-by-Eye-Response-Technologies.jpg
Link do podstrony portalu projektu COGAIN przedstawiający zdjęcie interfejsu Erica.
[157] http://en.wikipedia.org/wiki/Blu-ray_3D#Blu-ray_3D
Ogólne informacje nt. formatu Blu-ray 3D.
[158] http://www.eyegaze.com/content/eyefollower%E2%84%A220
Specyfikacja techniczna oraz zdjęcie interfejsu Eyefollower 2.0.
[159] http://www.eyetechds.com/
Strona główna EyeTech Digital Systems – producenta interfejsu EyeTech TM3.
[160] http://www.gazegroup.org/downloads/23-gazetracker
Strona z informacjami nt. aplikacji służącej do śledzenia wzroku ITU Gaze Tracker, dostępnej jako oprogramowanie otwarte.
[161] http://glossar.hs-augsburg.de/Binokulares_Sehen
Informacje na temat horoptera geometrycznego wraz z graficzną reprezentacją.
[162] http://hal.archives-ouvertes.fr/hal-00215967/PDF/Rapport_interne_1.pdf
Odnośnik do artykułu: A. O. Mohamed, M. P. Da Silva, V. Courboulay, "A history of eye gaze tracking".
[163] http://www.inference.phy.cam.ac.uk/opengazer/
Strona z podstawowymi informacjami nt. oprogramowania Opengazer.
[164] https://instruct.uwo.ca/psychology/215b-002/115-8-Depth.pdf
Podstrona Uniwersytetu w Zachodnim Ontario. Prezentacja poświęcona percepcji głębi. Informacje na temat czynników wpływających na percepcję głębi.
[165] http://www.inventa.com.au/Audio-Video%20Out%20of%20Sync%20in%20PC%20Video%20Capture.htm
Podstrona firmy Inventa Technologies. Informacje na temat rozsynchronizowania dźwięku i obrazu.
[166] http://mfiles.pl/pl/index.php/Wykres_pude%C5%82kowy
Podstrona serwisu Encyklopedia Zarządzania. Charakterystyka wykresu pudełkowego.
[167] http://www.nacinc.com/products/Eye-Tracking-Products/Voxer/
Strona NAC Image Technology, producenta interfejsu EMR-AT Voxer.
[168] http://pl.wikipedia.org/wiki/Anaglif
Ogólne informacje na temat techniki anaglifowej.
[169] http://www.reelclassics.com/Techtalk/panscan-article.htm
Odnośnik do artykułu: J. Cunningham, "Letterboxing vs. Pan-and-Scan: Most Viewers Don't Know What They're Missing", 1997.
[170] http://www.see-tech.de/english/index.php
Strona główna firmy Humanelektronik, produkującej interfejs wzrokowy SeeTech PRO.
[171] http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/red-red250-red-500.html
Podstrona firmy SMI – producenta interfejsu wzrokowego SMI RED500. Na stronie znajduje się specyfikacja techniczna interfejsu.
[172] http://www.statsoft.pl/
Strona główna firmy StatSoft, producenta i dystrybutora oprogramowania do analizy danych statystycznych STATISTICA.

[173] http://www.swiatlo.tak.pl/1/index.php/funkcje-wzroku-akomodacja-adaptacja-zbieznosc/
Podstrona portalu „Światło i oświetlenie”, zawierającego informacje nt. funkcji wzroku.
[174] http://www.swiatlo.tak.pl/pts/pts-oko-proces-widzenia.php
Opis procesu widzenia.
[175] http://thediemproject.wordpress.com/
Informacje nt. projektu DIEM. Opis metodologii przeprowadzania badania na filmie z wykorzystaniem systemu śledzenia wzroku.
[176] http://thirtysixthspan.com/openEyes/
Strona z podstawowymi informacjami nt. oprogramowania openEyes.
[177] http://www.tobii.com/en/analysis-and-research/global/products/
Podstrona firmy Tobii. Opis techniczny różnych modeli systemów śledzenia punktu fiksacji wzroku firmy.
[178] http://www.tobii.com/en/group/news-and-events/press-releases/the-worlds-first-eye-controlled-laptop/
Podstrona firmy Tobii. Opis techniczny laptopa Lenovo z wbudowanym systemem śledzenia punktu fiksacji wzroku firmy Tobii.
[179] http://www.trojmiasto.pl/wiadomosci/Awatar-na-Politechnice-Gdanskiej-n36747.html?strona=4&sort=up
Informacja prasowa na temat Laboratorium Zanurzonej Wizualizacji Przestrzennej, powstającego w Politechnice Gdańskiej.
[180] http://www.tutorgig.com/ed/Alternate-frame_sequencing
Ogólne informacje nt. techniki migawkowej.
[181] http://www.unmc.edu/physiology/Mann/mann8.html
Podstrona Uniwersytetu w Nebrasce. Informacje na temat percepcji dźwięku i budowy traktu słuchowego.

Załącznik A

Przykładowy opis (indeksacja) próbek testowych w strukturze XML

Opis próbki nr 8:

```
<movieDescription width="1280" height="1024"
length="8000" filename = "08_basic_AV_Left">

  <interval tstart="0" tend="8000">
     <area x="230" y="430" width="180" height="165"
     label="kropkaLewa" depth3D="0"/>
  </interval>

</movieDescription>
```

Opis próbki nr 14:

```
<movieDescription width="1280" height="1024"
length="27000" filename = "koncert">

  <interval tstart="0" tend="27000">
     <area x="200" y="250" width="270" height="150"
     label="skrzypce" depth3D="0" />

     <area x="700" y="320" width="300" height="480"
     label="fortepian" depth3D="+" />
  </interval>

</movieDescription>
```

Opis próbki nr 20:

```
<movieDescription width="1280" height="1024"
length="26000" filename = "20_alicja01">

     <interval tstart="0" tend="1710">
          <area x="460" y="380" width="276"
          height="260" label="krolowa" depth3D="-
          "/>
     </interval>

     <interval tstart="1800" tend="6700">
          <area x="870" y="240" width="140"
          height="144" label="Alicja" depth3D="+"/>
     </interval>

     <interval tstart="9100" tend="14800">
          <area x="630" y="230" width="340"
          height="380" label="Alicja" depth3D="-"/>
     </interval>

     <interval tstart="14805" tend="18000">
          <area x="400" y="300" width="135"
          height="145" label="krolowa"
          depth3D="+"/>
     </interval>

     <interval tstart="19350" tend="21200">
          <area x="400" y="300" width="135"
          height="145" label="krolowa"
          depth3D="+"/>
     </interval>

</movieDescription>
```

Załącznik B
Charakterystyka materiału badawczego

I Test podstawowy

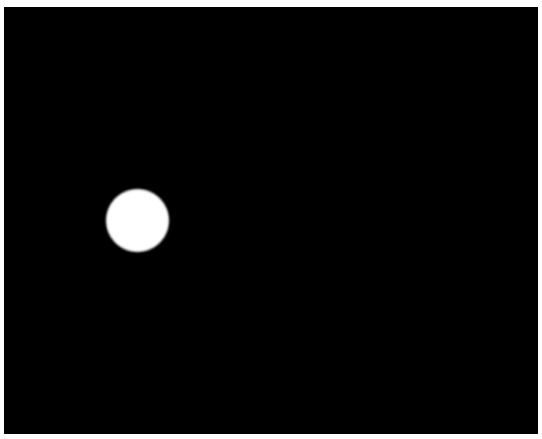
1. Test podstawowy – próbka BT (basic)

Numery próbek związane z materiałem badawczym: **1**, **8**, **11**,

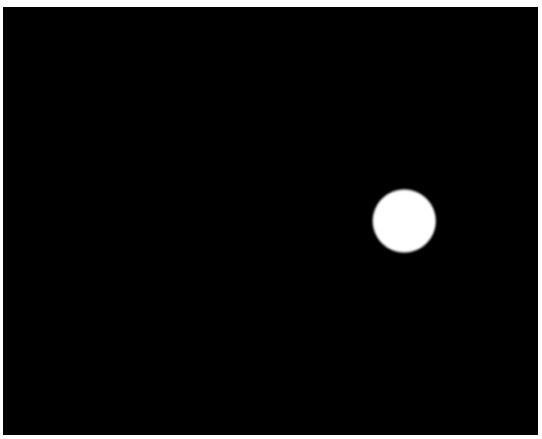
dodatkowo w badaniu z wykorzystaniem systemu Cyber-Oko: **22**, **27**

Czas trwania próbki:	8 s
Konfiguracje próbek:	1. dźwięk *(1 kHz)* 2. obraz wizyjny 2D + dźwięk (bodziec wzrokowy w lewej części kadru) 3. obraz wizyjny 2D + dźwięk (bodziec wzrokowy w prawej części kadru)
Tryby wyświetlania:	monitor (rozmiar średni), ekran projektora
Przedmiot badania:	– badanie wpływu ściągającego bodźca wzrokowego na lokalizację pozornego źródła dźwięku w panoramie stereofonicznej – badanie skalowalności wpływu ściągającego
Zadania badanego:	– wskazanie za pomocą suwaka (w zakresie -30 do +30) położenia źródła dźwięku

Bodziec wzrokowy w lewej części kadru



Bodziec wzrokowy w prawej części kadru



II Fragmenty rzeczywistych filmów 3D

2. Pierwsza próbka z filmu *Avatar* (avatar01)

Numery próbek związane z materiałem badawczym: **7**, **15**,

dodatkowo:

- w badaniu z wykorzystaniem systemu Tobii: **24**

- w badaniu z wykorzystaniem systemu Cyber-Oko: **25**

Czas trwania próbki	27 s
Konfiguracje próbek:	1. dźwięk *(głos bohatera)* 2. wideo 3D + dźwięk – mały rozmiar 3. wideo 3D + dźwięk – średni rozmiar
Rodzaj sceny	– 5 ujęć – ujęcia zrealizowane z wykorzystaniem jazdy kamerowej (nieznaczne ruchy)
Tryby wyświetlania:	monitor (mały, średni rozmiar) + ekran projektora
Przedmiot badania:	– badanie wpływu ściągającego twarzy bohatera na lokalizację jego głosu w panoramie stereofonicznej – badanie skalowalności wpływu ściągającego
Zadania badanego:	– ocena jakości efektu 3D prezentowanego obrazu – wskazanie za pomocą suwaka (w zakresie -30 do +30) położenia źródła dźwięku

3. Druga próbka z filmu *Avatar* (avatar02)

Numery próbek związane z materiałem badawczym: **3**, **10**, **19**

Czas trwania próbki:	15 s
Konfiguracje próbek:	1. dźwięk *(głos bohatera)* 2. wideo 3D + dźwięk + NAPISY 3. wideo 3D + dźwięk
Rodzaj sceny	– 5 ujęć – dynamiczne zmiany ujęć
Tryby wyświetlania:	monitor (średni rozmiar)
Przedmiot badania:	– badanie wpływu ściągającego twarzy bohaterki na lokalizację jej głosu w panoramie stereofonicznej – badanie wpływu czytania polskich napisów w filmie na percepcję kierunku dźwięku
Zadania badanego:	– ocena jakości efektu 3D prezentowanego obrazu – wskazanie za pomocą suwaka (w zakresie -30 do +30) położenia źródła dźwięku, – udzielenie odpowiedzi na pytanie czy w trakcie trwania próbki zmieniało się położenie pozornego źródła dźwięku w panoramie

4. Trzecia próbka z filmu *Avatar* (avatar03)

Numery próbek związane z materiałem badawczym: **16**,

dodatkowo:

- w badaniu z wykorzystaniem systemu Tobii: **23**

- w badaniu z wykorzystaniem systemu Cyber-Oko: **24**

Czas trwania próbki:	8 s
Konfiguracje próbek:	1. wideo 3D + dźwięk – mały rozmiar 2. wideo 3D + dźwięk – średni rozmiar
Rodzaj sceny	– 1 ujęcie – ujęcie zrealizowane z wykorzystaniem jazdy kamerowej
Tryby wyświetlania:	monitor (mały, średni rozmiar) + ekran projektora
Przedmiot badania:	– badanie wpływu dźwięku efektowego na kierunek patrzenia widza – badanie skalowalności wpływu ściągającego
Zadania badanego:	– ocena jakości efektu 3D prezentowanego obrazu – ocena wpływu bodźca słuchowego na uwagę wzrokową (kierunek patrzenia), przedział ocen: 0-10

5. Fragment filmu *Alicja w Krainie Czarów* (alicja01)

Numery próbek związane z materiałem badawczym: **2**, **9**, **13**, **20**

Czas trwania próbki:	26 s
Konfiguracje próbek:	1. dźwięk *(dialog bohaterek)* 2. wideo 3D + dźwięk + NnapisY 3. wideo 2D + dźwięk 4. wideo 3D + dźwięk
Rodzaj sceny	– 9 ujęć (z przebitkami) – ujęcia statyczne
Tryby wyświetlania:	monitor (średni rozmiar)
Przedmiot badania:	– badanie wpływu ściągającego twarzy bohaterek na lokalizację ich głosów w panoramie stereofonicznej (w scenie dialogowej) i głębi – badanie skalowalności wpływu ściągającego
Zadania badanego:	– ocena jakości efektu 3D prezentowanego obrazu – wskazanie za pomocą suwaka (w zakresie -30 do +30) położenia źródła dźwięku w panoramie stereofonicznej – wskazanie za pomocą suwaka (w zakresie 0 do 10) położenia źródła dźwięku w głębi (płaszczyzna przód-tył)

6. Fragment filmu *Piranha 3D* (piranha01)

Numery próbek związane z materiałem badawczym: **5**, **17**

Czas trwania próbki:	10 s
Konfiguracje próbek:	1. dźwięk *(głos bohatera)* 2. wideo 3D + dźwięk
Rodzaj sceny	– 1 ujęcie (statyczne)
Tryby wyświetlania:	monitor (średni rozmiar)
Przedmiot badania:	– badanie wpływu ściągającego postaci bohatera na lokalizację jego głosu w panoramie stereofonicznej
Zadania badanego:	– ocena jakości efektu 3D prezentowanego obrazu – wskazanie za pomocą suwaka (w zakresie -30 do +30) położenia źródła dźwięku (głosu bohatera) w panoramie stereofonicznej

7. Fragment filmu *Resident Evil: Afterlife* (resident01)

Numery próbek związane z materiałem badawczym: **21**,

dodatkowo:

- w badaniu z wykorzystaniem systemu Tobii: **25**

- w badaniu z wykorzystaniem systemu Cyber-Oko: **26**

Czas trwania próbki:	14 s
Konfiguracje próbek:	1. wideo 3D + dźwięk – mały rozmiar 2. wideo 3D + dźwięk – średni rozmiar
Rodzaj sceny	– 2 ujęcia
Tryby wyświetlania:	monitor (mały, średni rozmiar) + ekran projektora
Przedmiot badania:	– badanie wpływu dźwięku efektowego na kierunek patrzenia widza – badanie skalowalności wpływu ściągającego
Zadania badanego:	– ocena jakości efektu 3D prezentowanego obrazu – ocena wpływu bodźca słuchowego na uwagę wzrokową (kierunek patrzenia), przedział ocen: 0-10

III Autorski materiał badawczy z obrazem 3D

8. Pierwsza próbka z nagrania koncertu (koncert01)

Numery próbek związane z materiałem badawczym: **4, 18**

Czas trwania próbki:	22 s
Konfiguracje próbek:	1. dźwięk (*fortepian*) 2. wideo 3D + dźwięk
Rodzaj sceny	– 1 ujęcie (statyczne)
Tryby wyświetlania:	monitor (średni rozmiar)
Przedmiot badania:	– badanie wpływu ściągającego bodźca wzrokowego na lokalizację pozornego źródła dźwięku (fortepianu) w panoramie stereofonicznej
Zadania badanego:	– ocena jakości efektu 3D prezentowanego obrazu – wskazanie za pomocą suwaka (w zakresie -30 do +30) położenia źródła dźwięku (fortepianu) w panoramie stereofonicznej

9. Druga próbka z nagrania koncertu (koncert02)

Numery próbek związane z materiałem badawczym: **6**, **14,**

dodatkowo:

- w badaniu z wykorzystaniem systemu Tobii: **22**

- w badaniu z wykorzystaniem systemu Cyber-Oko: **23**

Czas trwania próbki:	27 s
Konfiguracje próbek:	1. dźwięk (*fortepian i skrzypce*) 2. wideo 3D + dźwięk – mały rozmiar 3. wideo 3D + dźwięk – średni rozmiar
Rodzaj sceny	– 1 ujęcie (statyczne)
Tryby wyświetlania:	monitor (mały, średni rozmiar) + ekran projektora
Przedmiot badania:	– badanie wpływu ściągającego bodźców wzrokowych na lokalizację pozornych źródeł dźwięku (fortepianu i skrzypiec) w panoramie stereofonicznej – badanie skalowalności wpływu ściągającego
Zadania badanego:	– ocena jakości efektu 3D prezentowanego obrazu – wskazanie za pomocą suwaka (w zakresie -30 do +30) położenia źródeł dźwięku (fortepianu i skrzypiec) w panoramie stereofonicznej

Załącznik C

Fragment projektu formularza do subiektywnej oceny położenia pozornego źródła dźwięku

(badani wypełniali formularz w wersji elektronicznej)

Badania korelacji wzrokowo-słuchowych

Formularz

– strona 1 –

Część I

ścieżki dźwiękowe próbek testowych

– strona 2 –

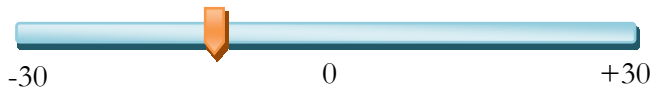
Próbka nr 1: (basic_audio)

1. Wskaż kierunek dochodzenia dźwięku
   (wartość „-30” oznacza położenie pozornego źródła dźwięku maksymalnie z lewej strony, „0” – dokładnie na środku, „30” – maksymalnie po prawej stronie panoramy)

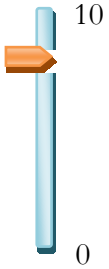
– strona 3 –

Próbka nr 2: (alicja01_audio)

2. Wskaż położenie głosu bohaterki nr 1 (królowej) w panoramie stereofonicznej (wartość „-30” oznacza położenie pozornego źródła dźwięku w maksymalnie z lewej strony, „0” – dokładnie na środku, „30” – maksymalnie po prawej stronie panoramy)

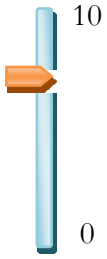


3. Wskaż położenie głosu bohaterki nr 1 (królowej) w płaszczyźnie przód-tył (w głębi)
   (wartość „0” oznacza lokalizację źródła blisko widza, wartość „10” – oznacza lokalizację daleko od widza)



4. Wskaż położenie głosu bohaterki nr 2 (Alicji) w panoramie stereofonicznej

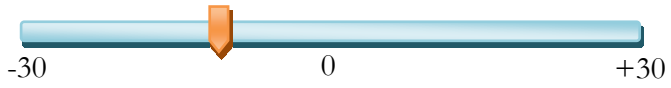
5. Wskaż położenie głosu bohaterki nr 2 (Alicji) w płaszczyźnie przód-tył (w głębi) (wartość „0" oznacza lokalizację źródła blisko widza, wartość „10" – oznacza lokalizację daleko od widza)



– strona 4 –

Próbka nr 3: (avatar02)

6. Wskaż położenie głosu bohaterki w panoramie stereofonicznej



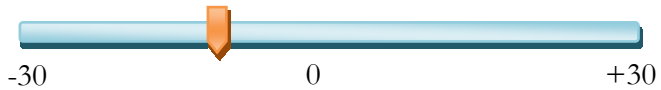
7. Czy położenie pozornego źr. dźwięku zmienia się w trakcie trwania próbki?

◯ NIE ◯ TAK

– strona 5 –

Próbka nr 4: (koncert01)

8. Wskaż lokalizację **fortepianu** w panoramie stereofonicznej



Dalsza część formularza zgodna z przedstawioną powyżej koncepcją…

Załącznik D

Analiza statystyczna wyników – test ANOVA

W niniejszym załączniku zamieszczono szczegółową analizę statystyczną wyników ocen subiektywnych dla każdej badanej próbki wizyjno-fonicznej. Wyniki zawarto w dwóch częściach. W pierwszej części zamieszczono wyniki uzyskane w eksperymencie przeprowadzonym z wykorzystaniem systemu Tobii T60. W drugiej zaś części zamieszczono wyniki otrzymane podczas badania z wykorzystaniem systemu Cyber-Oko.

Uwaga!
Numeracja próbek w niniejszym dokumencie odpowiada numeracji próbek zastosowanej w Załączniku B niniejszej rozprawy.

Część I
- wyniki dla systemu Tobii T60

Próbka 1:
próbka BT (bodziec wzrokowy w lewej części kadru)

Test Shapiro-Wilka (rozkład normalny):

rodzaj próbki	*N*	*W*	*p*
dźwięk	14	0,957171	0,676356
dźwięk + obraz 2D	14	0,873405	0,046856

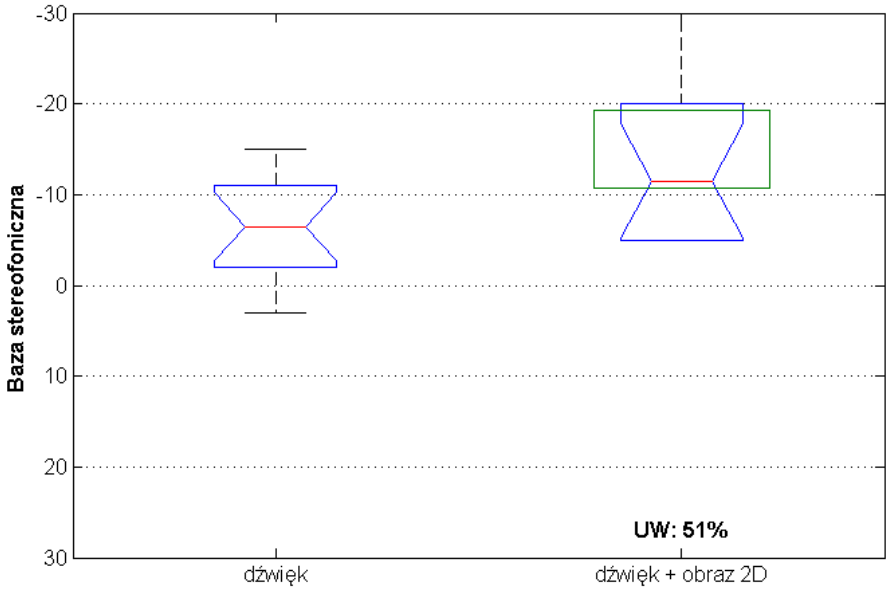
W przypadku zmiennej ‘dźwięk’ poziom $p>\alpha$, stąd na poziomie istotności 0,05 brak jest podstaw do odrzucenia hipotezy o zgodności jej rozkładu z rozkładem normalnym.
Niestety, zmienna ‘dźwięk + obraz 2D’ nie charakteryzuje się rozkładem normalnym. W tym przypadku zmienna nie spełnia pierwszego warunku koniecznego do przeprowadzenia testu ANOVA. W związku z powyższym przeprowadzony zostanie test alternatywny badający różnice pomiędzy średnimi zmiennych – test Kruskala-Wallisa.

Test Kruskala-Wallisa (K-W):
Hipoteza zerowa: rozkłady położenia pozornego źródła dźwięku w badanych próbkach są takie same.
Przyjęty poziom istotności: $\alpha = 0,05$.
W wyniku przeprowadzonej analizy uzyskuje się wartość testu:
$\boldsymbol{H = 3,921}$, przy $\boldsymbol{p = 0,0477}$

W związku z tym, że $p<0,05$, odrzuca się hipotezę o równości rozkładów.
Oznacza to, że zaobserwowany w niniejszej próbce wpływ ściągający obrazu na percepcję dźwięku jest istotny statystycznie.

Uwaga!

1. Na większości wykresów pudełkowych ramką koloru zielonego zaznaczono położenie bodźca wzrokowego na ekranie, odpowiadające lokalizacji pozornego źródła dźwięku w panoramie stereofonicznej.
2. Skrót „UW" oznacza „uwagę wzrokową" widza, zatem oznaczenie „UW: 51%" należy rozumieć następująco: „średnia wartość skupienia wzroku wszystkich badanych w zdefiniowanym ROI wyniosła 51% w odniesieniu do czasu projekcji próbki".



Wykres D.1 Wykres pudełkowy dla próbki BT, w której bodziec wzrokowy znajduje się w lewej części kadru

Próbka 1:
próbka BT (bodziec wzrokowy w prawej części kadru)

Test Shapiro-Wilka (rozkład normalny):

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,951405	0,546849
dźwięk + obraz 2D	15	0,947189	0,481311

Obie zmienne ('dźwięk' oraz 'dźwięk + obraz 2D') mają rozkład normalny, ponieważ $p>0,05$. W następnym kroku wykonany zostanie test Levene'a sprawdzający równość wariancji analizowanych zmiennych.

Test Levene'a (homogeniczność wariancji):

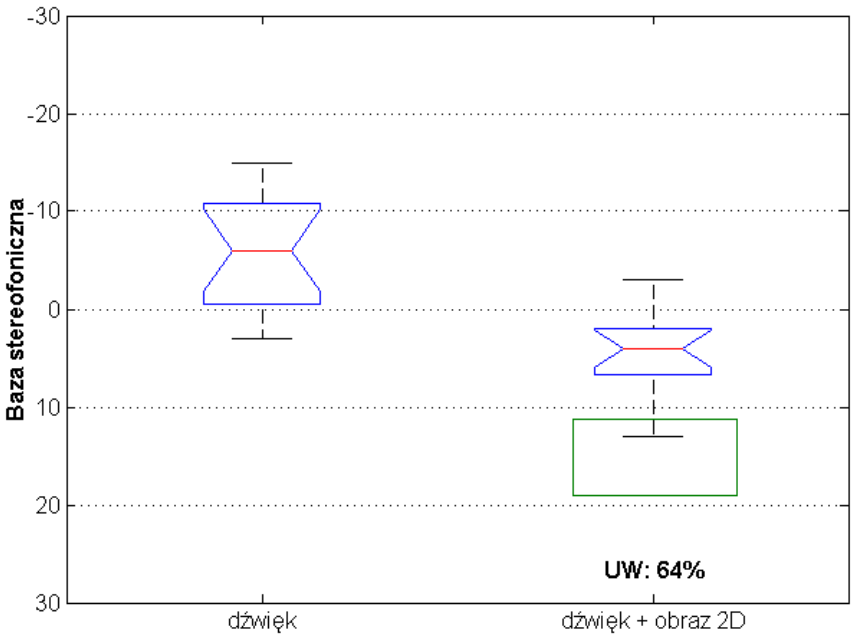
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	6,912000	1	6,912000	280,0160	28	10,00057	0,691161	0,412804

Wartość testu *F* wynosi: $F = 0{,}691$, $p>0{,}05$.
Warunek homogeniczności (jednorodności) wariancji został spełniony. Spełnienie dwóch podstawowych warunków pozwala na przeprowadzenie testu ANOVA.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	885,6333	1	885,6333	853,7333	28	30,49048	29,04623	0,000010

Wartość testu ANOVA $F(1, 28)=29{,}046$, przy $p<0{,}05$ oznacza odrzucenie hipotezy zerowej. Dlatego różnice wartości średnich między zmiennymi 'dźwięk' i 'dźwięk + obraz 2D' są istotne statystycznie na poziomie 0,05.



Wykres D.2 Wykres pudełkowy dla próbki BT, w której bodziec wzrokowy znajduje się w prawej części kadru

Próbka 2 (fragment filmu *Avatar*):
Położenie głosu bohatera w panoramie stereofonicznej (średni rozmiar obszaru wyświetlania)

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,973031	0,900123
dźwięk + obraz 3D	15	0,929427	0,267545

Analizowane powyżej zmienne mają rozkład normalny.

Test Levene'a:

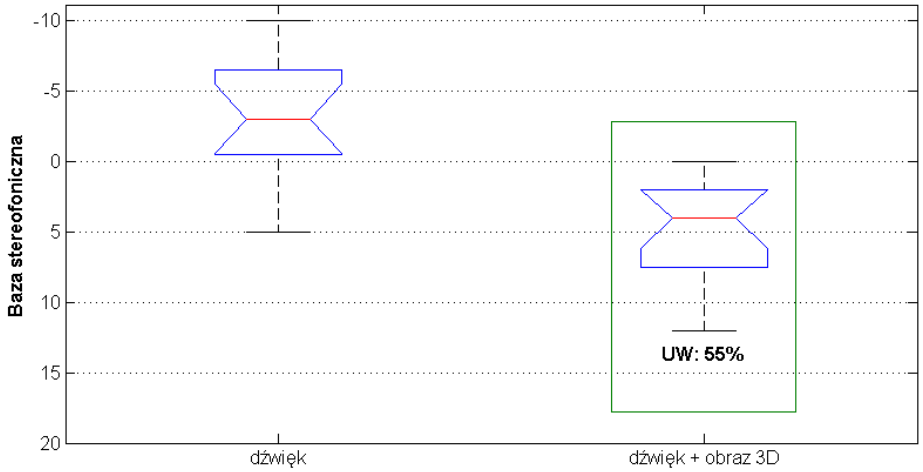
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	0,606815	1	0,606815	123,6130	28	4,414751	0,137452	0,713618

Wartość $p>0,05$, co oznacza, że warunek jednorodności wariancji został spełniony.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	488,0333	1	488,0333	412,9333	28	14,74762	33,09235	0,000004

$F(1, 28) = 33,092$, przy bardzo małym p oznacza odrzucenie hipotezy zerowej. Zatem zaobserwowany w niniejszej próbce wpływ ściągający obrazu na percepcję dźwięku istnieje i jest istotny statystycznie.



Wykres D.3 Wykres pudełkowy dla próbki 2 przy badaniu położenia głosu bohatera w panoramie stereofonicznej (średni rozmiar ekranu)

Próbka 2:
Położenie głosu bohatera w panoramie stereofonicznej (mały rozmiar obszaru wyświetlania)

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,973031	0,900123
dźwięk + obraz 3D	15	0,891794	0,071386

Rozkład każdej z wyżej analizowanych zmiennych jest rozkładem normalnym.

Test Levene'a:

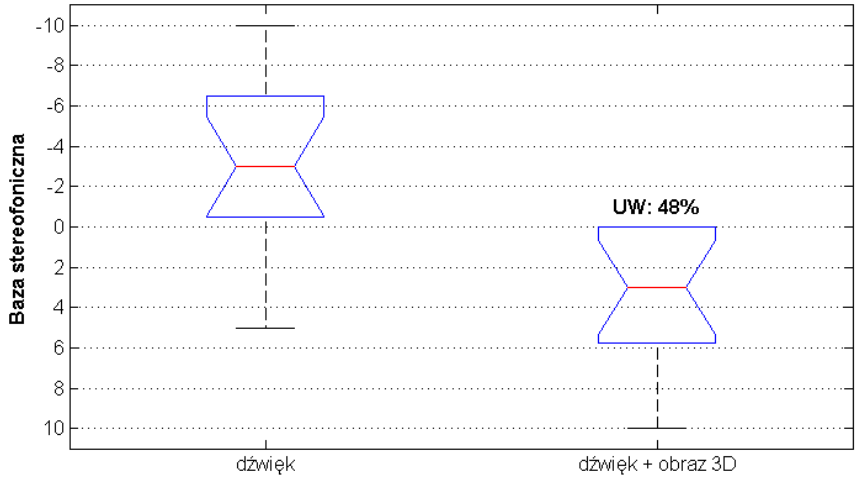
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	0,569481	1	0,569481	113,0157	28	4,036275	0,141091	0,710030

Warunek homogeniczności wariancji został spełniony.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	333,3333	1	333,3333	352,5333	28	12,59048	26,47504	0,000019

Różnica średnich lokalizacji pozornego źródła dźwięku (głosu bohatera) ocenianych po projekcji ścieżki dźwiękowej oraz po projekcji treści wizyjno-fonicznej jest stosunkowo duża i jest istotna statystycznie.



Wykres D.4 Wykres pudełkowy dla próbki 2 przy badaniu położenia głosu bohatera w panoramie stereofonicznej (mały rozmiar ekranu)

Próbka 3:
położenie głosu bohaterki w panoramie stereofonicznej

Uwaga!
Niniejsza próbka wizyjno-foniczna charakteryzuje się tym, że bodziec wzrokowy przykuwający uwagę widza (postać Neytri – bohaterki filmu „Avatar") porusza się stosunkowo dynamicznie. Pomimo to wyznaczono zakres kątowy, obejmujący obszar ekranu, w którym bohaterka się porusza.

Dodatkowo – w przypadku tej próbki – zadaniem badanych było udzielenie odpowiedzi („tak" lub „nie") na pytanie czy położenie pozornego źródła dźwięku związanego z bodźcem wzrokowym zmieniało się w trakcie trwania próbki.

Test Shapiro-Wilka:

rodzaj próbki	***N***	***W***	***p***
dźwięk	15	0,919864	0,191714
dźwięk + obraz z napisami	15	0,887607	0,061704
dźwięk + obraz 3D	15	0,906800	0,120945

Zgodnie z wartościami zawartymi w powyższej tabeli wszystkie analizowane zmienne mają rozkład normalny (wartość poziomu *p* jest większa od przyjętego poziomu istotności 0,05).

Test Levene'a:

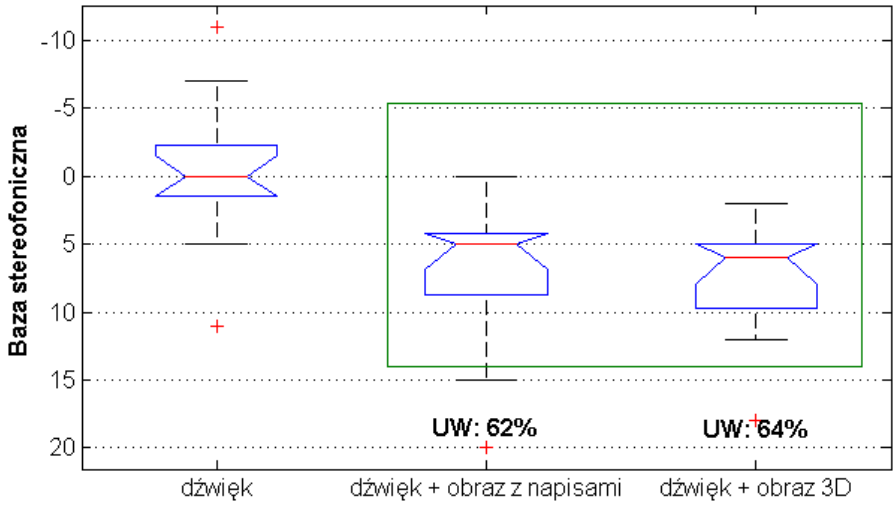
	SS Efekt	**df Efekt**	**MS Efekt**	**SS Błąd**	**df Błąd**	**MS Błąd**	***F***	***p***
lokalizacja poz.źr.dźw.	3,397531	2	1,698765	463,0074	42	11,02399	0,154097	0,857671

Wartość $p>\alpha$, zatem warunek homogeniczności analizowanych zmiennych został spełniony.

Test ANOVA:

	SS Efekt	**df Efekt**	**MS Efekt**	**SS Błąd**	**df Błąd**	**MS Błąd**	***F***	***p***
lokalizacja poz.źr.dźw.	541,1111	2	270,5556	998,0000	42	23,76190	11,38611	0,000112

Różnice średnich analizowanych zmiennych są istotne statystycznie.



Wykres D.5 Wykres pudełkowy dla próbki 3, w której badano położenie głosu bohaterki w panoramie stereofonicznej

W poniższej tabeli zestawiono odpowiedzi badanych na następujące pytanie:
Czy położenie pozornego źródła dźwięku (głosu bohaterki) zmieniało się w trakcie trwania próbki?

	dźwięk	dźwięk + obraz z napisami	dźwięk + obraz 3D
TAK	3	10	10
NIE	12	5	5

Analizując wyniki zestawione w powyższej tabeli, można zauważyć, że pod wpływem bodźca wzrokowego zmienił się sposób percepcji prezentowanego bodźca słuchowego.

Próbka 5:
położenie głosu bohaterki nr 1 (królowej) w panoramie stereofonicznej

Test Shapiro-Wilka:

rodzaj próbki	***N***	***W***	***p***
dźwięk	15	0,973308	0,903659
dźwięk + obraz z napisami	15	0,895790	0,082094
dźwięk + obraz 2D	15	0,956877	0,638312
dźwięk + obraz 3D	15	0,951484	0,548123

Wszystkie zmienne (odpowiadające poszczególnym rodzajom próbek) mają rozkład normalny.

Test Levene'a:

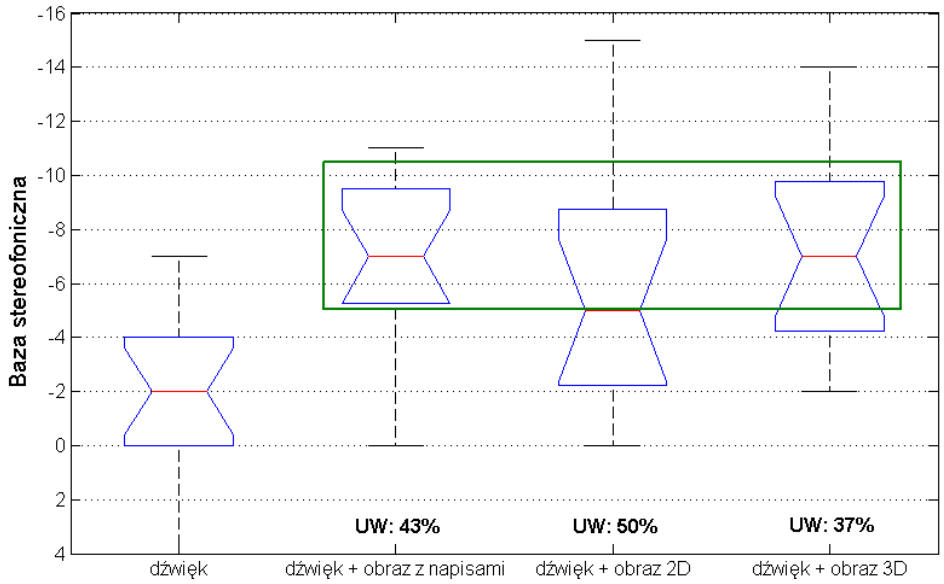
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	23,25215	3	7,750716	416,3052	56	7,434021	1,042601	0,380894

Wartość testu *F* wynosi: F=1,043 przy p>0,05, zatem nie ma podstaw do odrzucenia hipotezy zerowej o jednorodności wariancji. Można wykonać test ANOVA.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	282,7333	3	94,24444	989,2000	56	17,66429	5,335310	0,002645

$F(3, 56) = 5{,}335$, przy p=0,0026, zatem należy odrzucić hipotezę zerową. W związku z powyższym udowodniono, że średnie położenie pozornego źródła dźwięku w próbce fonicznej porównane ze średnim położeniem pozornego źródła dźwięku w próbkach wizyjno-fonicznych różni się w sposób istotny statystycznie.



Wykres D.6 Wykres pudełkowy dla próbki 5 przy badaniu położenia głosu bohaterki nr 1 (królowej) w panoramie stereofonicznej

Próbka 5:
położenie głosu bohaterki nr 1 (królowej) w płaszczyźnie przód-tył (w głębi)

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,953869	0,587311
dźwięk + obraz z napisami	15	0,924098	0,222362
dźwięk + obraz 2D	15	0,929813	0,271129
dźwięk + obraz 3D	15	0,935754	0,331966

Wszystkie analizowane zmienne mają rozkład normalny, co oznacza spełnienie pierwszego warunku koniecznego do wykonania testu ANOVA.

Test Levene'a:

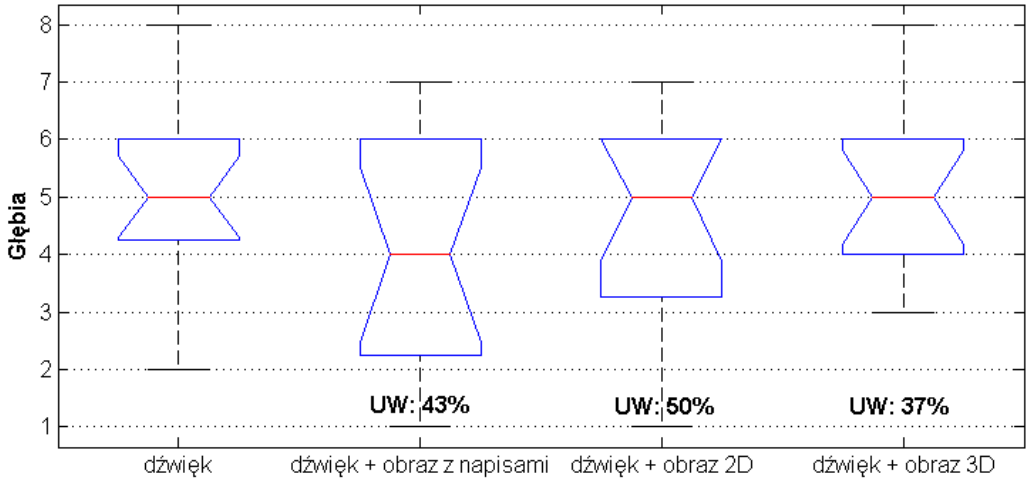
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	1,724148	3	0,574716	50,60919	56	0,903735	0,635934	0,594993

Zgodnie z wartościami zawartymi w powyższej tabeli poziom *p* jest większy od przyjętego poziomu istotności (0,05), nie ma zatem podstaw do odrzucenia hipotezy zerowej o jednorodności wariancji.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	6,733333	3	2,244444	175,6000	56	3,135714	0,715768	0,546716

Poziom $p>0,05$ nie można zatem odrzucić hipotezy zerowej. W związku z powyższym można przyjąć, że pomiędzy analizowanymi próbkami nie ma statystycznie istotnych różnic.
Potwierdzeniem wyniku testu ANOVA jest graficzne zestawienie wyników ocen subiektywnych na poniższym wykresie pudełkowym.



Wykres D.7 Wykres pudełkowy dla próbki 5 przy badaniu położenia głosu bohaterki nr 1 (królowej) w płaszczyźnie przód-tył

Próbka 5:
położenie głosu bohaterki nr 2 (Alicji) w panoramie stereofonicznej

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,979242	0,964065
dźwięk + obraz z napisami	15	0,856627	0,021583
dźwięk + obraz 2D	15	0,982024	0,981444
dźwięk + obraz 3D	15	0,895269	0,080610

Zmienne odpowiadające różnym konfiguracjom analizowanej próbki wizyjno-fonicznej, z wyjątkiem zmiennej 'dźwięk + obraz z napisami' mają rozkład normalny. Dlatego też w dalszej analizie statystycznej ta zmienna została pominięta.

Test Levene'a:

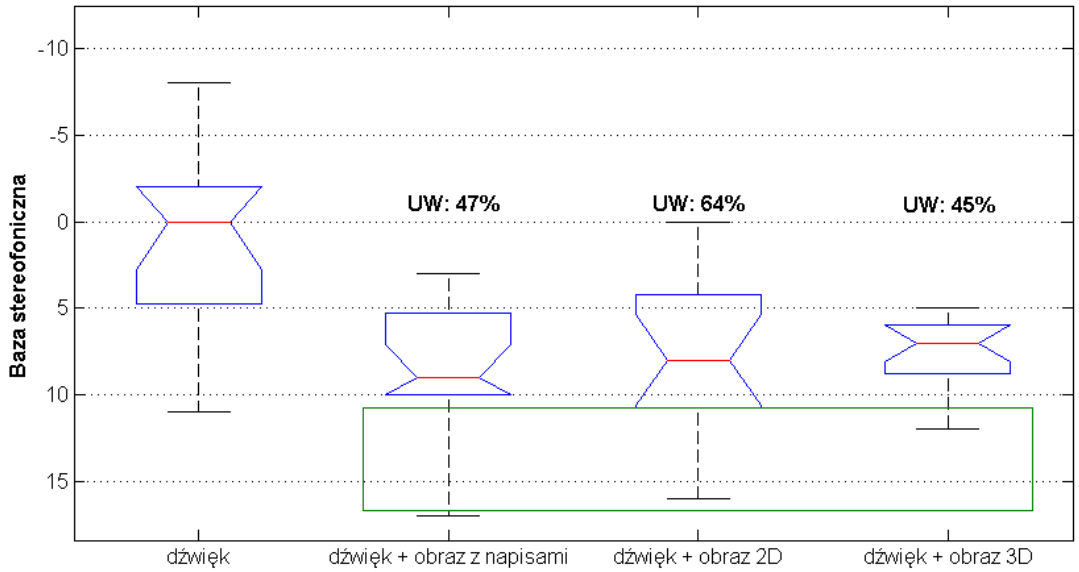
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	31,68316	2	15,84158	237,2130	42	5,647929	2,804847	0,071884

Wartość testu *F* wynosi: F=2,805 przy p>0,05. Można zatem przyjąć, że wariancje analizowanych zmiennych są jednorodne.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	499,5111	2	249,7556	693,7333	42	16,51746	15,12070	0,000011

Poziom $p<0,05$ zatem są podstawy do odrzucenia hipotezy zerowej. Oznacza to, że różnice wartości średnich analizowanych zmiennych są istotne statystycznie.



Wykres D.8 Wykres pudełkowy dla próbki 5 przy badaniu położenia głosu bohaterki nr 2 (Alicji) w panoramie stereofonicznej

---

Próbka 5:
położenie głosu bohaterki nr 2 (Alicji) w płaszczyźnie przód-tył (w głębi)

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,965613	0,788643
dźwięk + obraz z napisami	15	0,935983	0,334531
dźwięk + obraz 2D	15	0,888818	0,064357
dźwięk + obraz 3D	15	0,909642	0,133709

Wszystkie przeanalizowane powyżej zmienne mają rozkład normalny.

Test Levene'a:

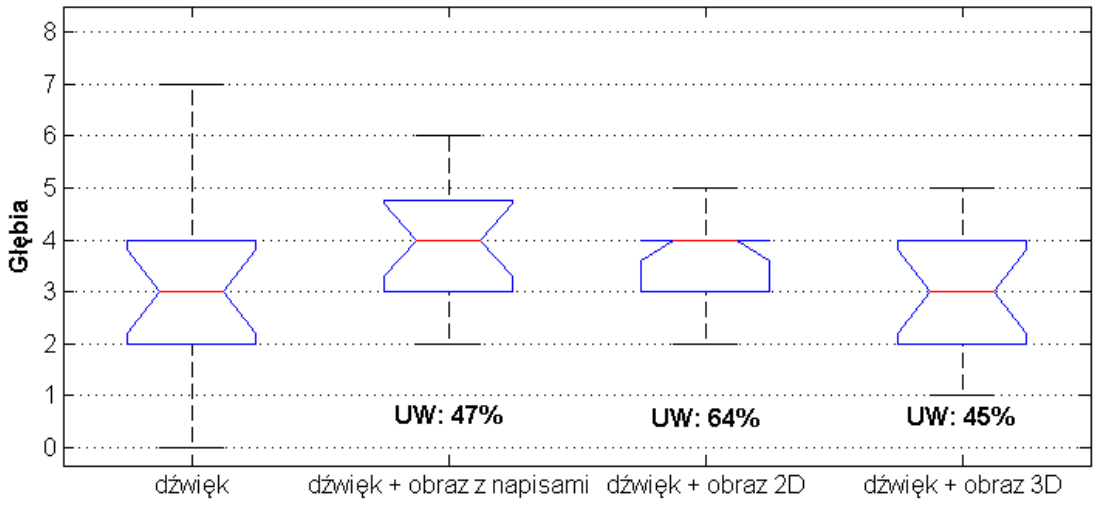
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	2,162963	3	0,720988	33,09452	56	0,590974	1,220000	0,310960

Wartość poziomu p=0,311 pozwala założyć, że warunek homogeniczności wariancji analizowanych zmiennych został spełniony.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	4,800000	3	1,600000	99,60000	56	1,778571	0,899598	0,447232

Wartość testu $F(3, 56) = 0{,}900$ przy $p>0{,}05$ pozwala przyjąć, że średnie wskazujące na położenie pozornego źródła dźwięku w płaszczyźnie przód-tył są równe. Zatem różnice ocen dla poszczególnych próbek, które można zaobserwować na poniższym wykresie nie są istotne statystycznie.



Wykres D.9 Wykres pudełkowy dla próbki 5 przy badaniu położenia głosu bohaterki nr 2 (Alicji) w płaszczyźnie przód-tył

Próbka 6 (fragment filmu *Piranha*):
Położenie głosu bohatera w panoramie stereofonicznej

Test Shapiro-Wilka:

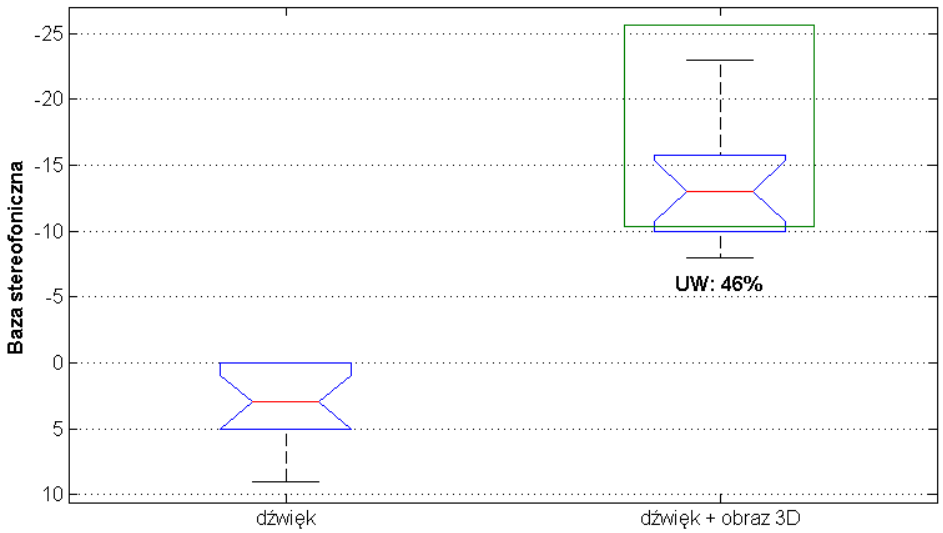
rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,875417	0,040556
dźwięk + obraz 3D	15	0,924560	0,225980

Rozkład zmiennej ‘dźwięk’ nie jest rozkładem normalnym. Nie został spełniony pierwszy warunek konieczny do przeprowadzenia testu ANOVA.
W związku z powyższym zostanie przeprowadzony alternatywny test Kruskala-Wallisa, testujący różnice pomiędzy średnimi i wymagający rozkładu normalnego analizowanych zmiennych.

Test Kruskala-Wallisa:

$H = 21{,}980$, przy $p = 0{,}0000$

W związku z tym, że $p<0{,}05$, odrzucamy hipotezę o równości rozkładów.
Wniosek ten, podobnie jak poniższy rysunek, potwierdza istnienie stosunkowo silnego wpływu ściągającego obrazu na percepcję dźwięku w niniejszej próbce. Zaobserwowana różnica jest istotna statystycznie.



Wykres D.10 Wykres pudełkowy dla próbki 6, w której badano położenie głosu bohatera panoramie stereofonicznej

Próbka 8:
Położenie pozornego źródła dźwięku (fortepianu) w panoramie stereofonicznej

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,967289	0,816083
dźwięk + obraz 2D	15	0,929540	0,268591
dźwięk + obraz 3D	15	0,918115	0,180283

Wszystkie powyżej analizowane zmienne charakteryzują się rozkładem normalnym.

Test Levene'a:

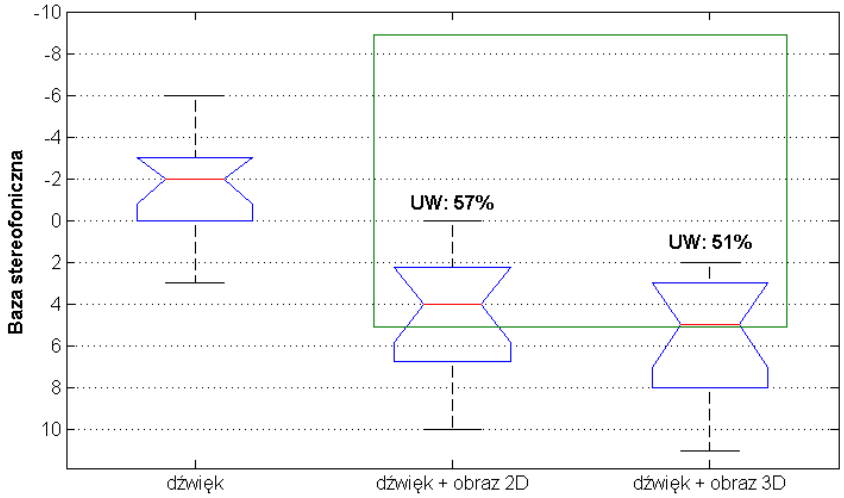
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	2,086716	2	1,043358	102,5843	42	2,442483	0,427171	0,655154

W związku z tym, że wartość $p>0,05$ brak podstaw do odrzucenia hipotezy zerowej, a zatem można przyjąć, że warunek o jednorodności wariancji został spełniony.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	420,4000	2	210,2000	378,8000	42	9,019048	23,30623	0,000000

Wartość *p* jest bardzo mała. Oznacza to, że przyjętą hipotezę zerową można odrzucić, zatem zaobserwowany w niniejszej próbce wpływ ściągający jest istotny statystycznie.



Wykres D.11 Wykres pudełkowy dla próbki 8, w której badano położenie fortepianu w panoramie stereofonicznej

Próbka 9:
Położenie skrzypiec w panoramie stereofonicznej (średni rozmiar obszaru wyświetlania)

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,889512	0,065931
dźwięk + obraz 3D	15	0,945468	0,456091

Zarówno zmienna odpowiadająca próbce fonicznej, jak i zmienna odpowiadająca próbce wizyjno-fonicznej charakteryzują się rozkładem normalnym.

Test Levene'a:

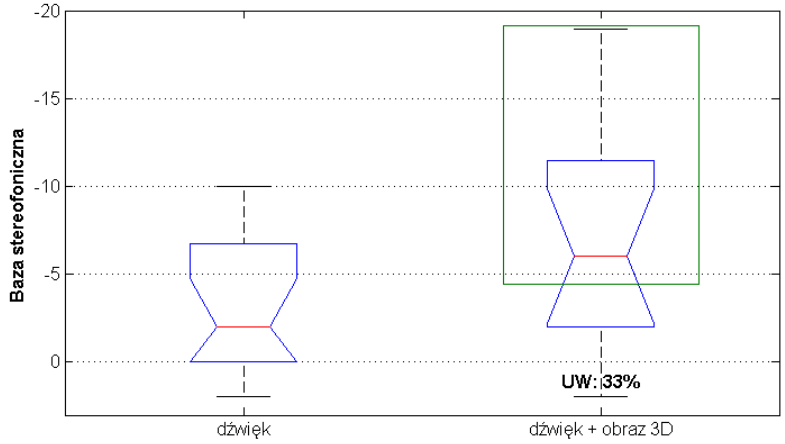
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	25,63793	1	25,63793	188,5250	28	6,733037	3,807780	0,061085

Wyznaczony poziom *p* w teście Levene'a jest większy od przyjętego poziomu istotności, zatem warunek jednorodności wariancji analizowanych zmiennych został spełniony.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	97,20000	1	97,20000	818,6667	28	29,23810	3,324430	0,078951

Wartość testu $F(1, 28)=3{,}324$ przy $p>\alpha$ sprzyja zachowaniu hipotezy zerowej, co wskazuje na równość średnich analizowanych zmiennych. Zatem zaobserwowane różnice w lokalizowaniu skrzypiec (w niniejszej próbce) w panoramie stereofonicznej nie są istotne statystycznie.



Wykres D.12 Wykres pudełkowy dla próbki 9 przy badaniu położenia skrzypiec w panoramie stereofonicznej (średni rozmiar ekranu)

Próbka 9:

Położenie skrzypiec w panoramie stereofonicznej (mały rozmiar obszaru wyświetlania)

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,889512	0,065931
dźwięk + obraz 3D	15	0,963148	0,746907

Zmienne odpowiadające próbce fonicznej i próbce wizyjno-fonicznej mają rozkład normalny.

Test Levene'a:

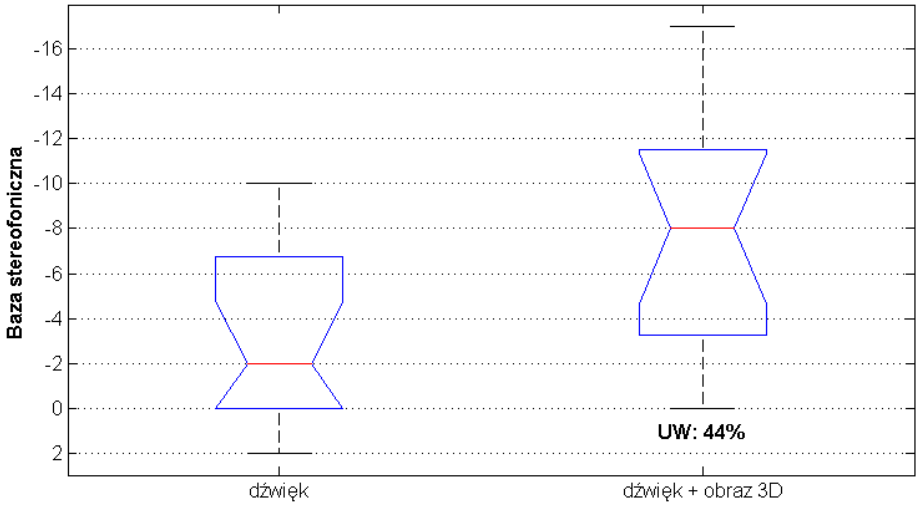
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	3,333333	1	3,333333	146,3532	28	5,226899	0,637727	0,431257

Poziom $p>0,05$, stąd przyjmuje się spełnienie warunku homogeniczności wariancji.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	124,0333	1	124,0333	605,4667	28	21,62381	5,735961	0,023554

Wartość $p<\alpha$, stąd można przyjąć, że zaobserwowany w niniejszym materiale badawczym wpływ ściągający jest istotny statystycznie.



Wykres D.13 Wykres pudełkowy dla próbki 9 przy badaniu położenia skrzypiec w panoramie stereofonicznej (mały rozmiar ekranu)

Próbka 9:
Położenie fortepianu w panoramie stereofonicznej (średni rozmiar obszaru wyświetlania)

Test Shapiro-Wilka:

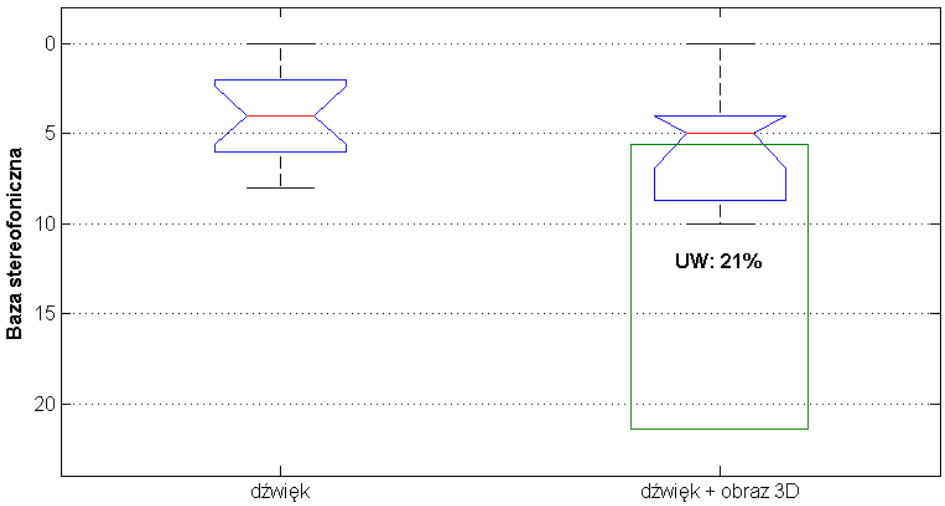
rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,864872	0,028392
dźwięk + obraz 3D	15	0,925747	0,235517

Dla zmiennej ‘dźwięk’ poziom $p<0,05$, co oznacza, że zmienna ta nie ma rozkładu normalnego. Dlatego w następnym kroku analizy statystycznej przeprowadzony zostanie test Kruskala-Wallisa.

Test Kruskala-Wallisa:

$H=2,324$, przy $p=0,1274$

W związku z tym, że $p>0,05$ nie ma podstaw do odrzucenia hipotezy o równości rozkładów. Zatem zaobserwowany w niniejszej próbce wpływ ściągający nie jest istotny statystycznie. Wniosek ten potwierdza również poniższy wykres.



Wykres D.14 Wykres pudełkowy dla próbki 9 przy badaniu położenia fortepianu w panoramie stereofonicznej (średni rozmiar ekranu)

Próbka 9:
Położenie fortepianu w panoramie stereofonicznej (mały rozmiar obszaru wyświetlania)

Test Shapiro-Wilka:

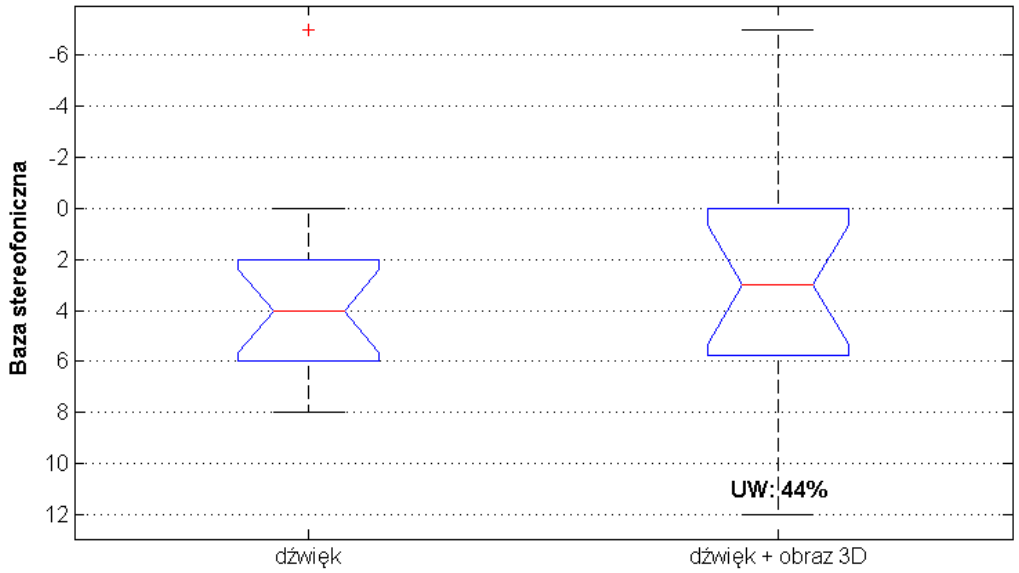
rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,864872	0,028392
dźwięk + obraz 3D	15	0,968192	0,830436

W związku z tym, że zmienna 'dźwięk' nie ma rozkładu normalnego dalsza analiza statystyczna będzie oparta na teście Kruskala-Wallisa.

Test Kruskala-Wallisa:

H=0,3178, przy p=0,5730

W związku z tym, że p>0,05 nie ma podstaw do odrzucenia hipotezy o równości rozkładów. Zatem zaobserwowany w niniejszej próbce wpływ ściągający nie jest istotny statystycznie.



Wykres D.15 Wykres pudełkowy dla próbki 9 przy badaniu położenia fortepianu w panoramie stereofonicznej (mały rozmiar ekranu)

Część II
- wyniki dla systemu Cyber-Oko

Próbka 1:
próbka BT (bodziec wzrokowy w lewej części kadru)

Test Shapiro-Wilka (rozkład normalny):

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,974309	0,915944
dźwięk + obraz 2D	15	0,936159	0,336520

Rozkład powyżej analizowanych zmiennych pokrywa się z rozkładem normalnym.

Test Levene'a (homogeniczność wariancji):

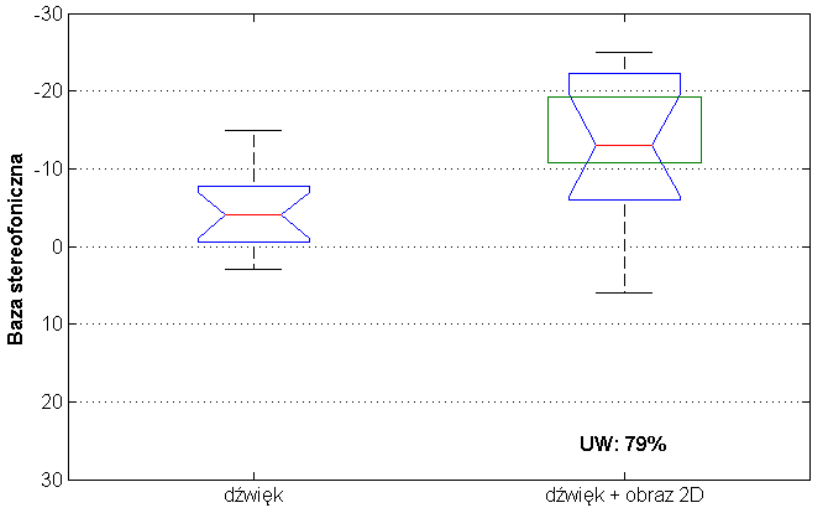
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	89,21126	1	89,21126	533,3197	28	19,04713	4,683711	0,039131

W związku z niespełnieniem warunku jednorodności wariancji nie można wykonać testu ANOVA. Przeprowadzony zostanie test alternatywny – test Kruskala-Wallisa.

Test Kruskala-Wallisa (testowanie różnic pomiędzy średnimi):

H=7,180, przy p=0,0074

W związku z tym, że p<0,05 należy odrzucić hipotezę o równości rozkładów. Zatem zaobserwowany w niniejszej próbce wpływ ściągający jest istotny statystycznie.



Wykres D.16 Wykres pudełkowy dla próbki BT, w której bodziec wzrokowy znajduje się w lewej części kadru

Próbka 1:
próbka BT (bodziec wzrokowy w prawej części kadru)

Test Shapiro-Wilka:

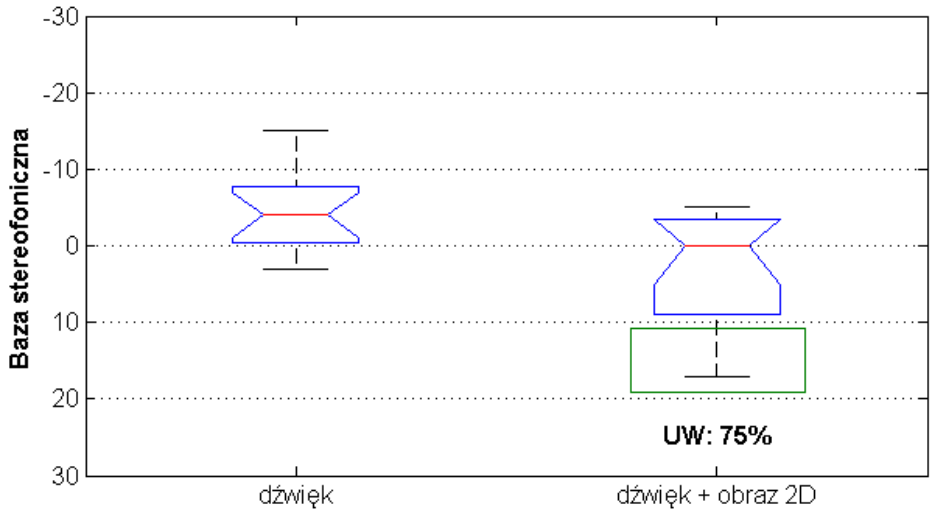
rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,974309	0,915944
dźwięk + obraz 2D	15	0,865773	0,029263

W związku z tym, że rozkład zmiennej 'dźwięk + obraz 2D' nie jest rozkładem normalnym, nie można wykonać testu ANOVA. Poniżej przedstawiono wyniki dla alternatywnego testu Kruskala-Wallisa.

Test Kruskala-Wallisa:

H=8,039, przy p=0,0046

W związku z tym, że p<0,05 należy odrzucić hipotezę o równości rozkładów. Zatem zaobserwowany w niniejszej próbce wpływ ściągający jest istotny statystycznie.



Wykres D.17 Wykres pudełkowy dla próbki BT, w której bodziec wzrokowy znajduje się w prawej części kadru

---

Próbka 2 (fragment filmu *Avatar*):
Położenie głosu bohatera w panoramie stereofonicznej (średni rozmiar obszaru wyświetlania)

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,945261	0,453133
dźwięk + obraz 3D	15	0,982803	0,985077

Warunek rozkładu normalnego zmiennych został spełniony.

Test Levene'a:

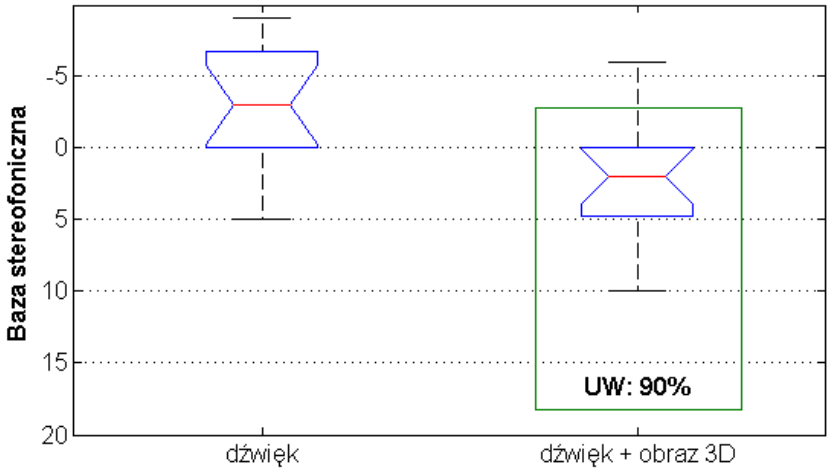
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	0,002370	1	0,002370	155,5123	28	5,554011	0,000427	0,983664

Warunek jednorodności wariancji analizowanych zmiennych został spełniony. Można zatem wykonać test ANOVA.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	208,0333	1	208,0333	478,2667	28	17,08095	12,17926	0,001618

Wartość testu ANOVA przy $p<\alpha$ wskazuje na to, że różnice położenia pozornego źródła dźwięku (głosu bohatera) w panoramie stereofonicznej dla analizowanych zmiennych są istotne statystycznie.



Wykres D.18 Wykres pudełkowy dla próbki 2 przy badaniu położenia głosu bohatera w panoramie stereofonicznej (średni rozmiar ekranu)

Próbka 2:
Położenie głosu bohatera w panoramie stereofonicznej (obraz wyświetlany na ekranie projektora)

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,945261	0,453133
dźwięk + obraz 3D	15	0,958023	0,658079

Wyniki testu Shapiro-Wilka wskazują na spełnienie warunku o rozkładzie normalnym analizowanych zmiennych.

Test Levene'a:

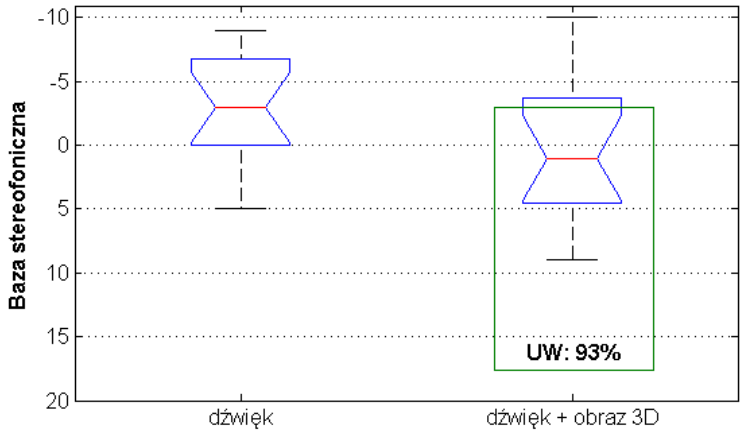
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	4,907259	1	4,907259	216,9381	28	7,747788	0,633375	0,432816

Warunek jednorodności wariancji został spełniony.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	112,1333	1	112,1333	631,0667	28	22,53810	4,975280	0,033903

Zaobserwowane różnice średnich wskazują na to, że przesunięcie percypowanego położenia dźwięku (głosu bohatera) w kierunku bodźca wzrokowego (twarzy bohatera) jest istotne statystycznie.



Wykres D.19 Wykres pudełkowy dla próbki 2 przy badaniu położenia głosu bohatera w panoramie stereofonicznej (ekran projektora)

Próbka 3:
położenie głosu bohaterki w panoramie stereofonicznej

Test Shapiro-Wilka:

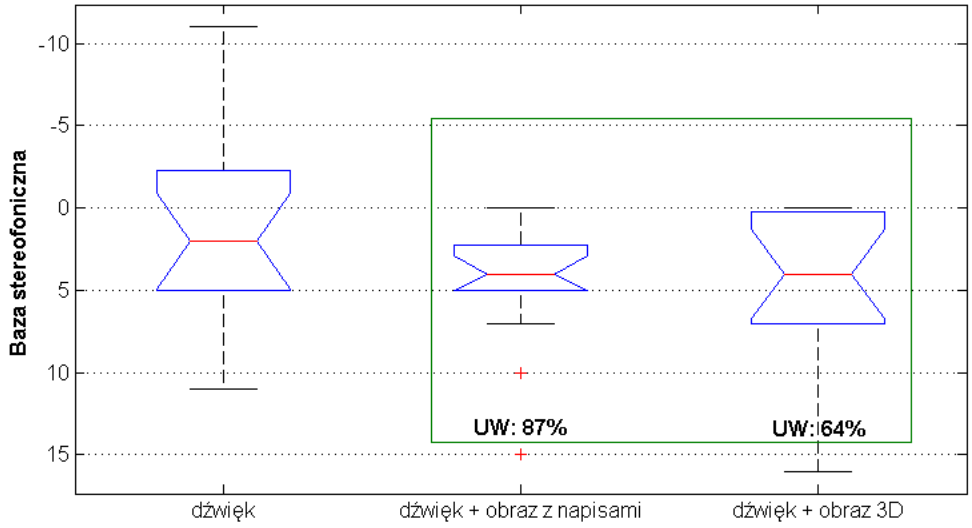
rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,966297	0,799952
dźwięk + obraz 2D	15	0,865098	0,028608
dźwięk + obraz 3D	15	0,863264	0,026905

W związku z tym, że rozkład zmiennej ‘dźwięk + obraz 2D’ oraz zmiennej ‘dźwięk + obraz 3D’ nie jest rozkładem normalnym, nie można wykonać testu ANOVA. Poniżej przedstawiono wyniki dla alternatywnego testu Kruskala-Wallisa.

Test Kruskala-Wallisa:

H=2,890, przy p=0,2358

W związku z tym, że p>0,05 nie ma podstaw do odrzucenia hipotezy zerowej (o równości rozkładów). Zatem zaobserwowany w niniejszej próbce efekt przesunięcia percepcji kierunku bodźca słuchowego w kierunku bodźca wzrokowego nie jest istotny statystycznie.



Wykres D.20 Wykres pudełkowy dla próbki 3, w której badano położenie głosu bohaterki w panoramie stereofonicznej

Zestawienie odpowiedzi na pytanie:
Czy położenie pozornego źródła dźwięku zmieniało się w czasie trwania próbki?

	dźwięk	dźwięk + obraz z napisami	dźwięk + obraz 3D
TAK	0	5	6
NIE	15	10	9

Próbka 5:
położenie głosu bohaterki nr 1 (królowej) w panoramie stereofonicznej

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,927347	0,248978
dźwięk + obraz z napisami	15	0,958488	0,666133
dźwięk + obraz 2D	15	0,953209	0,576325
dźwięk + obraz 3D	15	0,928778	0,261622

Rozkład wszystkich analizowanych powyżej zmiennych pokrywa się z rozkładem normalnym, zatem w następnym kroku analizy zostanie przeprowadzony test Levene'a w celu zbadania homogeniczności wariancji.

Test Levene'a:

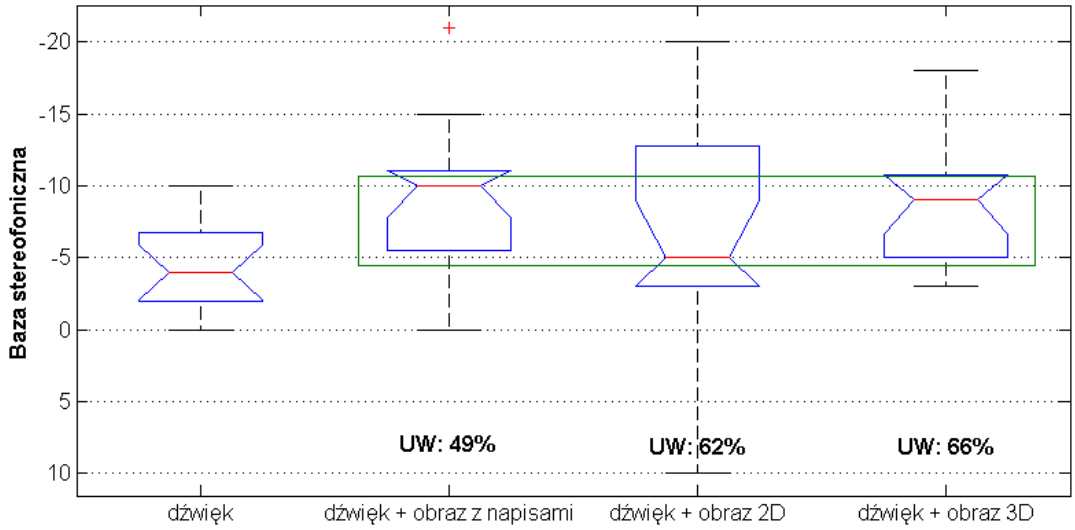
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	90,08474	3	30,02825	631,0572	56	11,26888	2,664706	0,056588

W związku ze spełnieniem warunku jednorodności wariancji ($p>\alpha$) w następnym kroku może być wykonany test ANOVA.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	204,0500	3	68,01667	1696,133	56	30,28810	2,245657	0,092978

Wartość testu $F(3, 56)=2{,}246$ przy $p>\alpha$ sprzyja zachowaniu hipotezy zerowej, co wskazuje na równość średnich analizowanych zmiennych. Zatem zaobserwowane różnice w lokalizowaniu głosu bohaterki w panoramie stereofonicznej (dla różnych konfiguracji próbek) nie są istotne statystycznie.



Wykres D.21 Wykres pudełkowy dla próbki 5 przy badaniu położenia głosu bohaterki nr 1 (królowej) w panoramie stereofonicznej

Próbka 5:
położenie głosu bohaterki nr 1 (królowej) w płaszczyźnie przód-tył

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,917672	0,177490
dźwięk + obraz z napisami	15	0,902187	0,102790
dźwięk + obraz 2D	15	0,924511	0,225596
dźwięk + obraz 3D	15	0,905410	0,115154

Wszystkie zmienne analizowane w ramach niniejszej próbki mają rozkład normalny.

Test Levene'a:

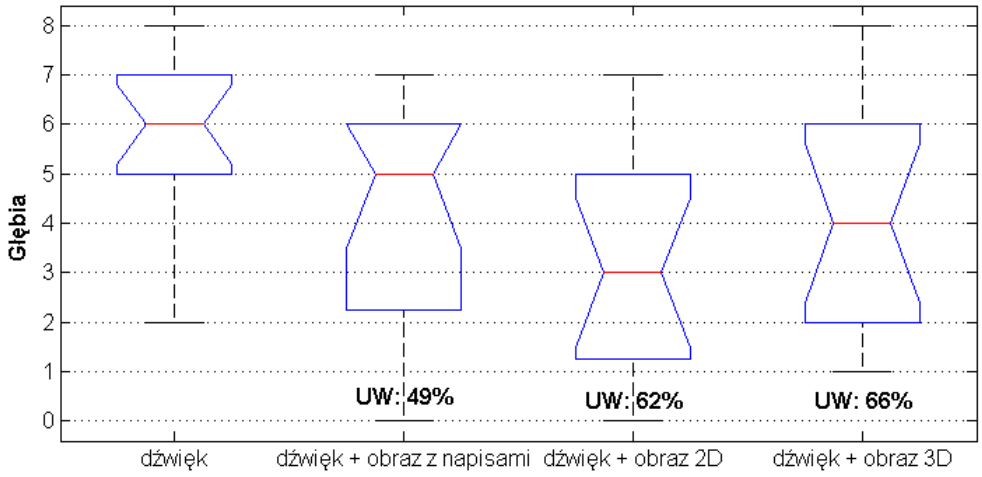
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	4,893630	3	1,631210	55,76296	56	0,995767	1,638144	0,190877

Wartość $p>0,05$, co oznacza, że warunek jednorodności wariancji został spełniony.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	41,93333	3	13,97778	240,0000	56	4,285714	3,261481	0,028033

Wartość $p<\alpha$, stąd można przyjąć, że zaobserwowane w niniejszym materiale badawczym różnice w lokalizowaniu źródła dźwięku w płaszczyźnie przód-tył są istotne statystycznie.



Wykres D.22 Wykres pudełkowy dla próbki 5 przy badaniu położenia głosu bohaterki nr 1 (królowej) w płaszczyźnie przód-tył

Próbka 5:
położenie głosu bohaterki nr 2 (Alicji) w panoramie stereofonicznej

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,892998	0,074452
dźwięk + obraz z napisami	15	0,909234	0,131797
dźwięk + obraz 2D	15	0,926849	0,244711
dźwięk + obraz 3D	15	0,972345	0,891138

Zgodnie z wartościami zawartymi w powyższej tabeli wszystkie analizowane zmienne mają rozkład normalny.

Test Levene'a:

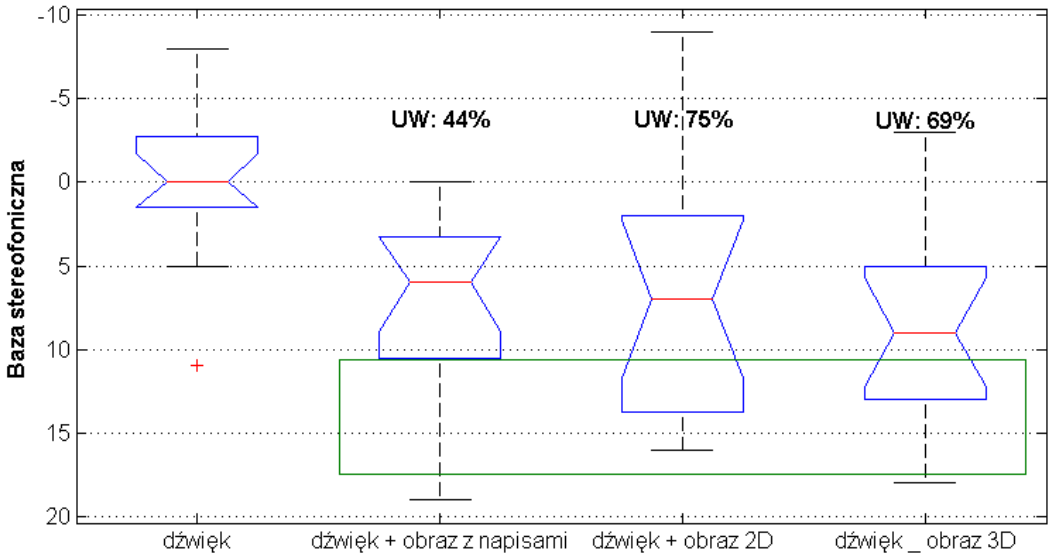
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	27,49244	3	9,164148	738,8450	56	13,19366	0,694587	0,559229

Wartość testu *F* wynosi: F=0,695, przy p>0,05, zatem nie ma podstaw do odrzucenia hipotezy zerowej o jednorodności wariancji. Można wykonać test ANOVA.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	572,1833	3	190,7278	2049,467	56	36,59762	5,211480	0,003033

Wartość p=0,003, co oznacza, że należy odrzucić hipotezę o równości średnich analizowanych zmiennych. W związku z powyższym zaobserwowane w niniejszej próbce różnice we wskazaniach położenia głosu bohaterki w panoramie stereofonicznej są istotne statystycznie.



Wykres D.23 Wykres pudełkowy dla próbki 5 przy badaniu położenia głosu bohaterki nr 2 (Alicji) w panoramie stereofonicznej

Próbka 5:
położenie głosu bohaterki nr 2 (Alicji) w płaszczyźnie przód-tył

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,890623	0,068530
dźwięk + obraz z napisami	15	0,908457	0,128228
dźwięk + obraz 2D	15	0,957889	0,655759
dźwięk + obraz 3D	15	0,914789	0,160346

Wszystkie zmienne (określające poszczególne rodzaje próbek) mają rozkład normalny.

Test Levene'a:

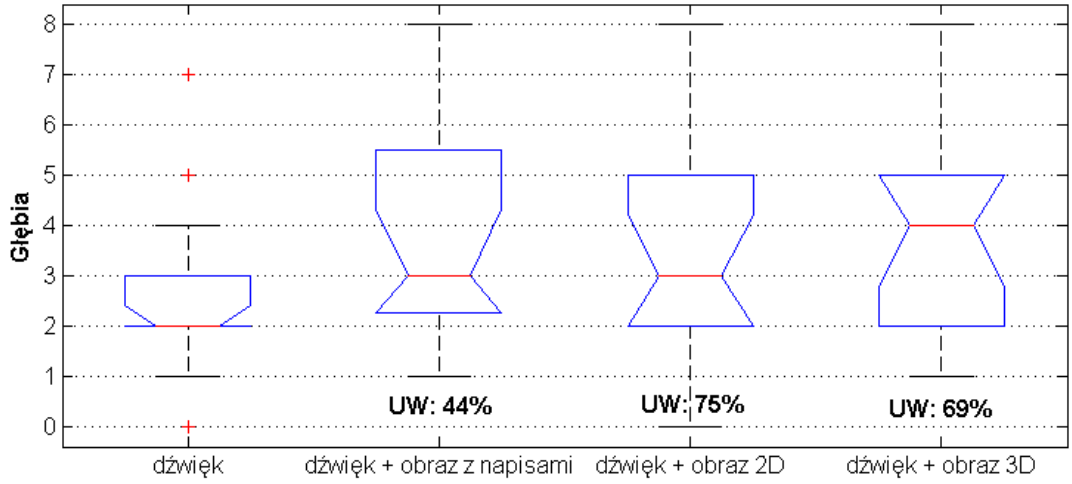
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	3,656000	3	1,218667	85,67467	56	1,529905	0,796564	0,500978

Warunek jednorodności wariancji analizowanych zmiennych został spełniony.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	13,73333	3	4,577778	254,6667	56	4,547619	1,006632	0,396704

Wartość *p* jest większa od założonego poziomu istotności, stąd przyjmuje się, że nie ma podstaw do odrzucenia hipotezy zerowej. Zatem zaobserwowane w niniejszej próbce wizyjno-fonicznej różnice pomiędzy średnimi wskazującymi na położenie pozornego źródła dźwięku w płaszczyźnie przód-tył nie są istotne statystycznie.



Wykres D.24 Wykres pudełkowy dla próbki 5 przy badaniu położenia głosu bohaterki nr 2 (Alicji) w płaszczyźnie przód-tył

Próbka 6 (fragment filmu *Piranha*):
Położenie głosu bohatera w panoramie stereofonicznej

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,957339	0,646265
dźwięk + obraz 3D	15	0,958241	0,661856

Zgodnie z wartościami zawartymi w powyższej tabeli rozkłady zmiennych ‘dźwięk’ oraz ‘dźwięk + obraz 3D’ pokrywają się z rozkładem normalnym.

Test Levene'a (homogeniczność wariancji):

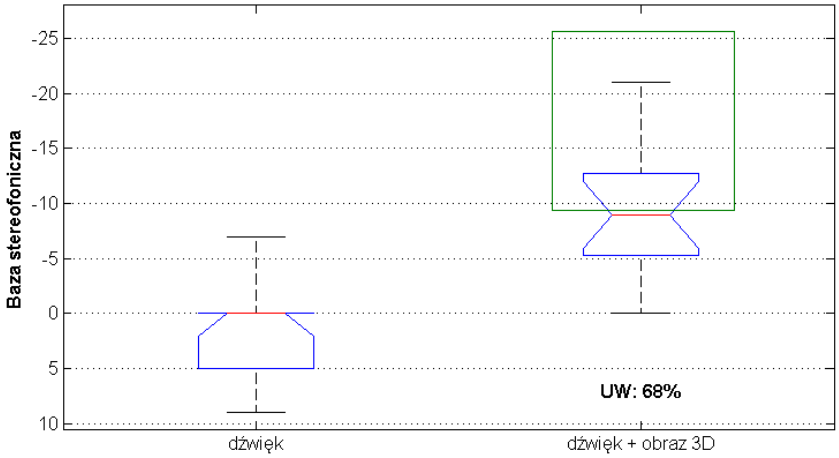
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	3,513481	1	3,513481	261,5052	28	9,339471	0,376197	0,544596

Wartość *p* jest większa od przyjętego poziomu istotności, stąd nie ma podstaw do odrzucenia hipotezy o równości wariancji analizowanych zmiennych.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	940,8000	1	940,8000	754,6667	28	26,95238	34,90601	0,000002

Wartość testu ANOVA $F(1, 28)=34,906$, przy $p<0,05$, stąd przyjmuje się, że zaobserwowane różnice średnich analizowanych zmiennych są istotne statystycznie.



Wykres D.25 Wykres pudełkowy dla próbki 6, w której badano położenie głosu bohatera panoramie stereofonicznej

Próbka 8:

Położenie pozornego źródła dźwięku (fortepianu) w panoramie stereofonicznej

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,902935	0,105532
dźwięk + obraz 2D	15	0,923020	0,214141
dźwięk + obraz 3D	15	0,978515	0,958352

Rozkłady wszystkich zmiennych analizowanych w ramach niniejszej próbki pokrywają się z rozkładem normalnym.

Test Levene'a:

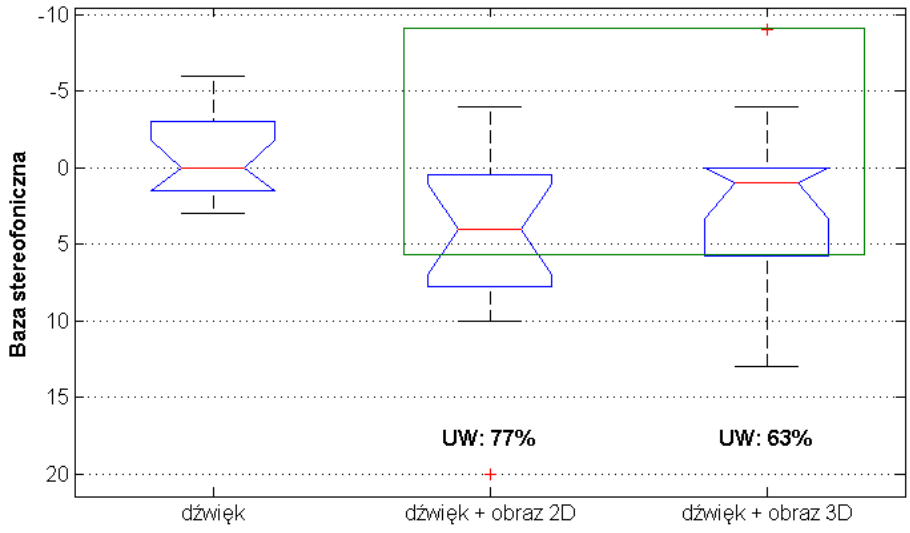
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	33,21442	2	16,60721	395,6966	42	9,421347	1,762721	0,184034

Zgodnie z wartościami zawartymi w powyższej tabeli, warunek homogeniczności wariancji został spełniony.

Test ANOVA:

	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	252,5778	2	126,2889	1034,400	42	24,62857	5,127739	0,010173

Zgodnie z wartościami w powyższej tabeli, wartość testu ANOVA wskazuje na to, że zaobserwowane różnice wartości średnich poszczególnych zmiennych są istotne statystycznie.



Wykres D.26 Wykres pudełkowy dla próbki 8, w której badano położenie fortepianu w panoramie stereofonicznej

Próbka 9:
Położenie skrzypiec w panoramie stereofonicznej (średni rozmiar obszaru wyświetlania)

Test Shapiro-Wilka:

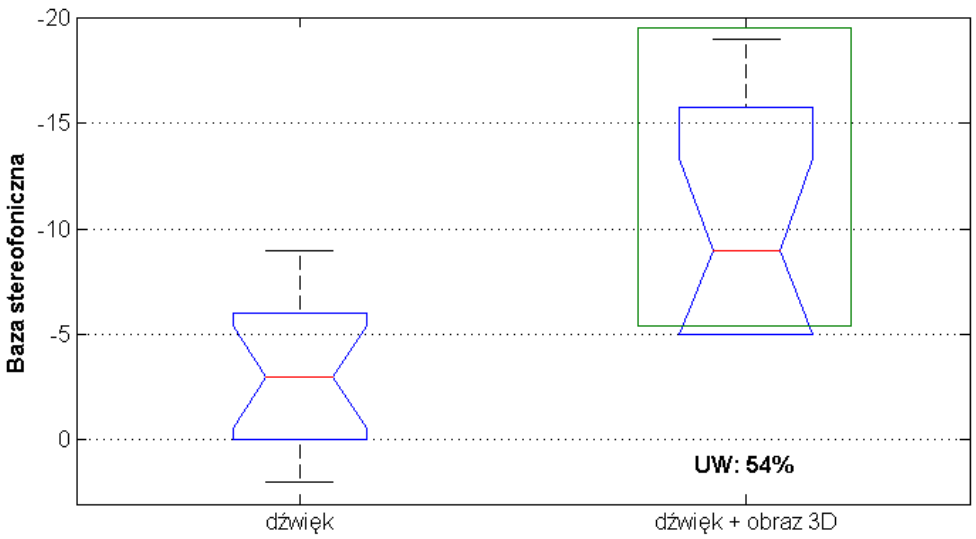
rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,921055	0,199902
dźwięk + obraz 3D	15	0,869472	0,033144

Rozkład zmiennej 'dźwięk + obraz 3D' nie pokrywa się z rozkładem normalnym, stąd nie można wykonać testu ANOVA. Poniżej przedstawiono wyniki alternatywnego testu Kruskala-Wallisa, badającego różnice pomiędzy średnimi.

Test Kruskala-Wallisa:

H=11,262, przy p=0,0008

W związku z tym, że p<0,05 należy odrzucić hipotezę zerową (o równości rozkładów). Zatem zaobserwowany w niniejszej próbce efekt przesunięcia percepcji kierunku bodźca słuchowego w kierunku bodźca wzrokowego jest istotny statystycznie.



Wykres D.27 Wykres pudełkowy dla próbki 9 przy badaniu położenia skrzypiec w panoramie stereofonicznej (średni rozmiar ekranu)

Próbka 9:
Położenie skrzypiec w panoramie stereofonicznej (obraz wyświetlany na ekranie projektora)

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,921055	0,199902
dźwięk + obraz 3D	15	0,913213	0,151675

Rozkłady powyżej analizowanych zmiennych pokrywają się z rozkładem normalnym, zatem w następnym kroku analizy wykonany zostanie test Levene'a.

Test Levene'a:

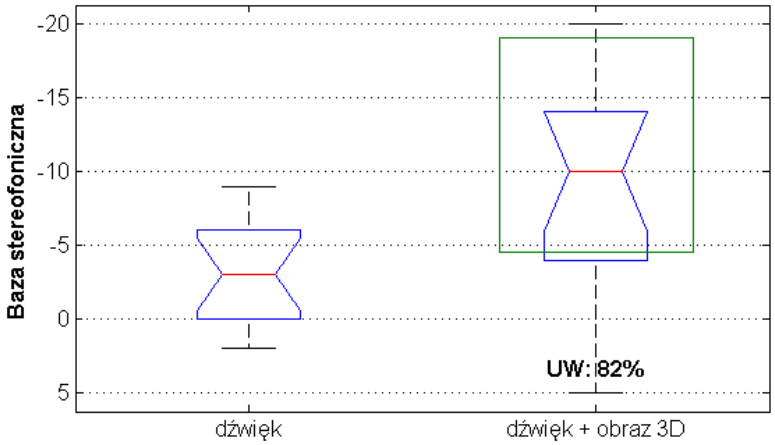
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	91,06015	1	91,06015	306,1641	28	10,93443	8,327834	0,007436

Warunek równości wariancji analizowanych zmiennych nie został spełniony, dlatego w następnym kroku wykonano test Kruskala-Wallisa.

Test Kruskala-Wallisa:

H=5,1432, przy p=0,0233

W związku z tym, że p<0,05 należy odrzucić hipotezę zerową (o równości rozkładów). Zatem zaobserwowany w niniejszej próbce efekt przesunięcia percepcji kierunku bodźca słuchowego w kierunku bodźca wzrokowego jest istotny statystycznie.



Wykres D.28 Wykres pudełkowy dla próbki 9 przy badaniu położenia skrzypiec w panoramie stereofonicznej (ekran projektora)

Próbka 9:
Położenie fortepianu w panoramie stereofonicznej (średni rozmiar obszaru wyświetlania)

Test Shapiro-Wilka (rozkład normalny):

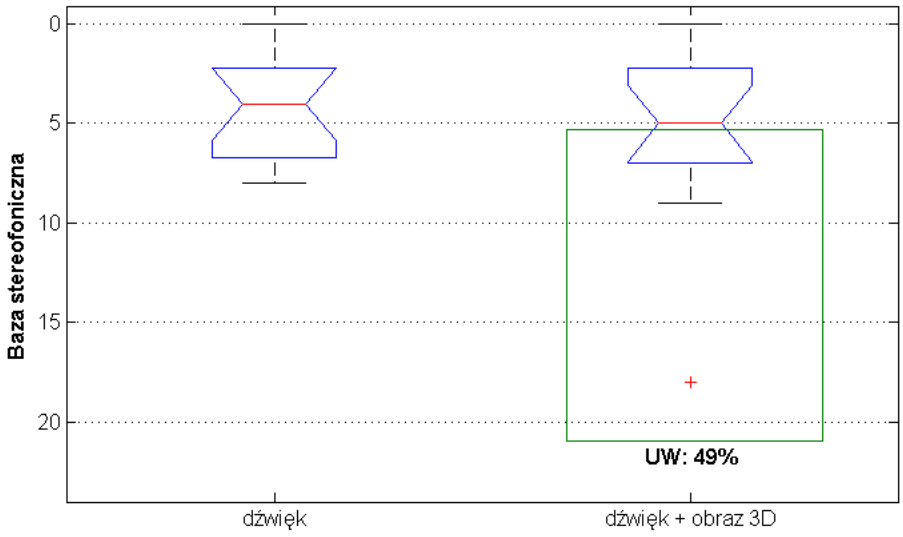
rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,933695	0,309641
dźwięk + obraz 3D	15	0,871197	0,035136

Zmienna 'dźwięk + obraz 3D' nie ma rozkładu normalnego, zatem w następnym kroku analizy wykonano test Kruskala-Wallisa.

Test Kruskala-Wallisa (testowanie różnic pomiędzy średnimi):

H=0,1418, przy p=0,7065

W związku z tym, że p>0,05 nie ma podstaw do odrzucenia hipotezy zerowej. Zaobserwowany w niniejszej próbce wpływ bodźca wzrokowego na lokalizację kierunku bodźca słuchowego nie jest istotny statystycznie.



Wykres D.29 Wykres pudełkowy dla próbki 9 przy badaniu położenia fortepianu w panoramie stereofonicznej (średni rozmiar ekranu)

Próbka 9:
Położenie fortepianu w panoramie stereofonicznej (obraz wyświetlany na ekranie projektora)

Test Shapiro-Wilka:

rodzaj próbki	*N*	*W*	*p*
dźwięk	15	0,933695	0,309641
dźwięk + obraz 3D	15	0,975861	0,933387

Rozkłady zmiennych analizowanych w ramach niniejszej próbki wizyjno-fonicznej pokrywają się z rozkładem normalnym.

Test Levene'a:

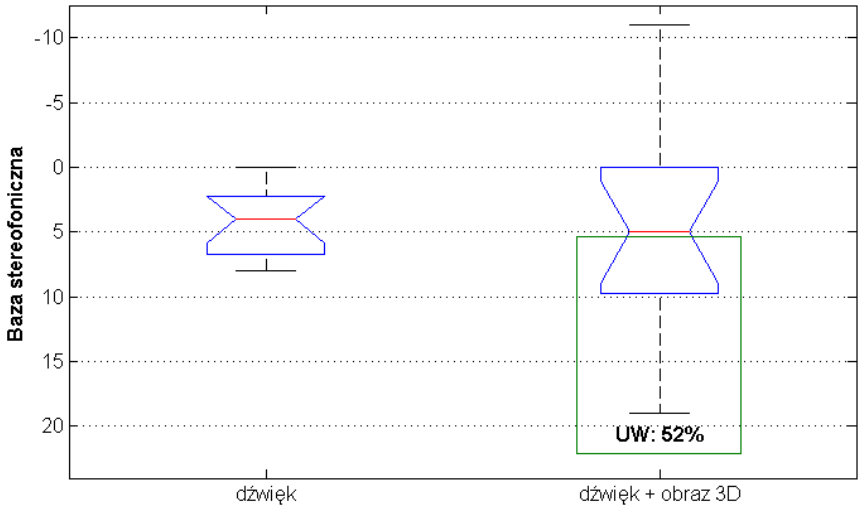
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	*F*	*p*
lokalizacja poz.źr.dźw.	97,68059	1	97,68059	329,4181	28	11,76493	8,302691	0,007516

Niespełnienie warunku jednorodności wariancji determinuje przeprowadzenie testu Kruskala-Wallisa w kolejnym kroku analizy statystycznej.

Test Kruskala-Wallisa:

H=0,3642, przy p=0,5462

W związku z tym, że p>0,05 nie ma podstaw do odrzucenia hipotezy o równości rozkładów. Zatem zaobserwowany w niniejszej próbce efekt przesunięcia percepcji kierunku bodźca słuchowego w kierunku bodźca wzrokowego nie jest istotny statystycznie.



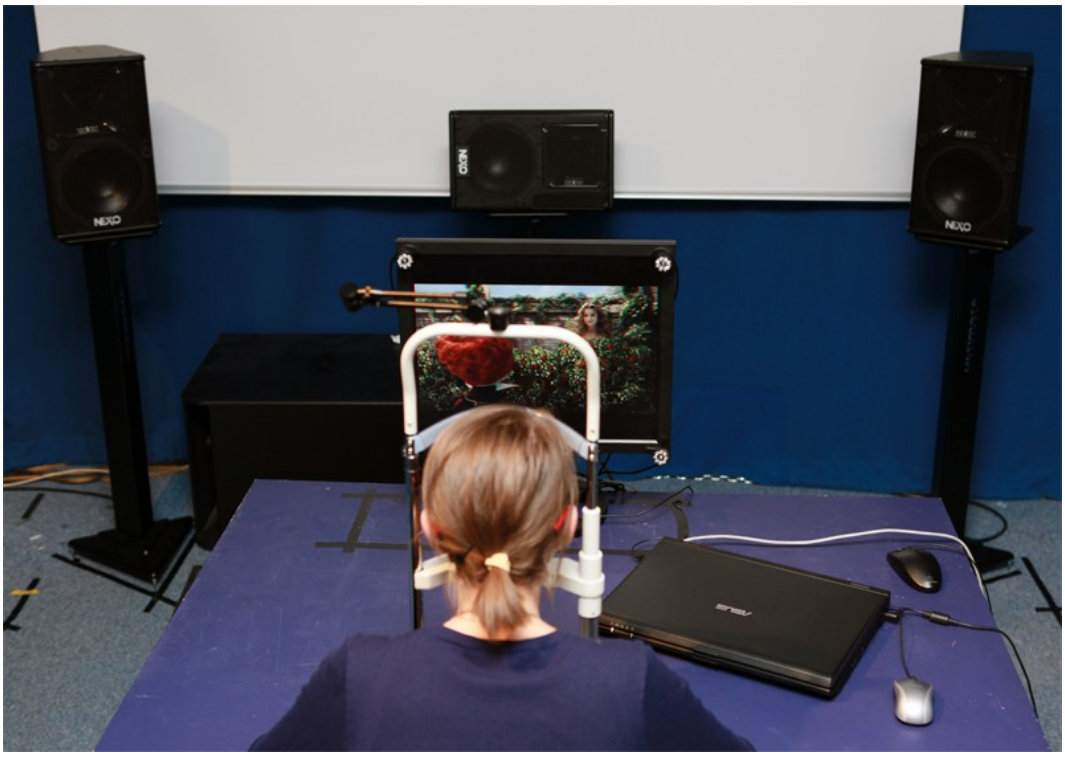
Wykres D.30 Wykres pudełkowy dla próbki 9 przy badaniu położenia fortepianu w panoramie stereofonicznej (ekran projektora)

Załącznik E

Stanowisko badawcze – fotografie

Rys. E.1 Stanowisko badawcze z wykorzystaniem systemu Tobii T60

Rys. E.2 Konfiguracja stanowiska badawczego z wykorzystaniem systemu Tobii T60



Rys. E.3 Stanowisko badawcze z wykorzystaniem systemu Cyber-Oko pracującego w trybie normalnym

Rys. E.4 Stanowisko badawcze z wykorzystaniem systemu Cyber-Oko pracującego w trybie wyświetlania obrazu na ekranie projektora

Rys. E.5 Widok obrazu wyświetlanego na ekranie projektora z perspektywy osoby badanej

Załącznik F
Zawartość płyty DVD-ROM

Zawartość płyty (z uwzględnieniem katalogów i podkatalogów):

1. Materiał badawczy
2. Opisy (indeksacja) próbek testowych
3. Wyniki
   - 3.1 Aktywność wzrokowa badanych
     - 3.1.1 Cyber-Oko
     - 3.1.2 Tobii T60
   - 3.2 Przykładowe próbki z naniesioną dynamiczną mapą przejść
     - 3.2.1 Cyber-Oko
     - 3.2.2 Tobii T60
   - 3.3 Zestawienia wyników ocen subiektywnych
     - 3.3.1 Cyber-Oko
     - 3.3.2 Tobii T60