

Korelacje wzrokowo-słuchowe

Bartosz Kunka

Katedra Systemów Multimedialnych
Wydział Elektroniki, Telekomunikacji i Informatyki
Politechnika Gdańska



Synestezja a percepcja wielomodalna

- **Synestezja**
 - pojęcie funkcjonujące w psychologii, w literaturze
 - stan, w którym stymulacja jednego zmysłu wywołuje doznania charakterystyczne dla innych zmysłów, np. barwa niebieska odczuwana jest jako chłodna
 - ujmowanie doznań właściwych dla jednego zmysłu w kategoriach doznań innego zmysłu, np. „ciepły głos”
- **Percepcja wielomodalna**
 - odbieranie bodźców pochodzących z co najmniej dwóch jednocześnie stymulowanych zmysłów, np. w systemach wirtualnej rzeczywistości: jednoczesna stymulacja wzroku, słuchu, dotyku

Percepcja wielomodalna

- Oszacowanie przepływności informacji „transmitowanej” przez poszczególne zmysły
 - podano przybliżoną liczbę komórek nerwowych związanych z każdym zmysłem
 - na podstawie przepływności pojedynczej komórki, oszacowano „szerokość pasma” poszczególnych zmysłów

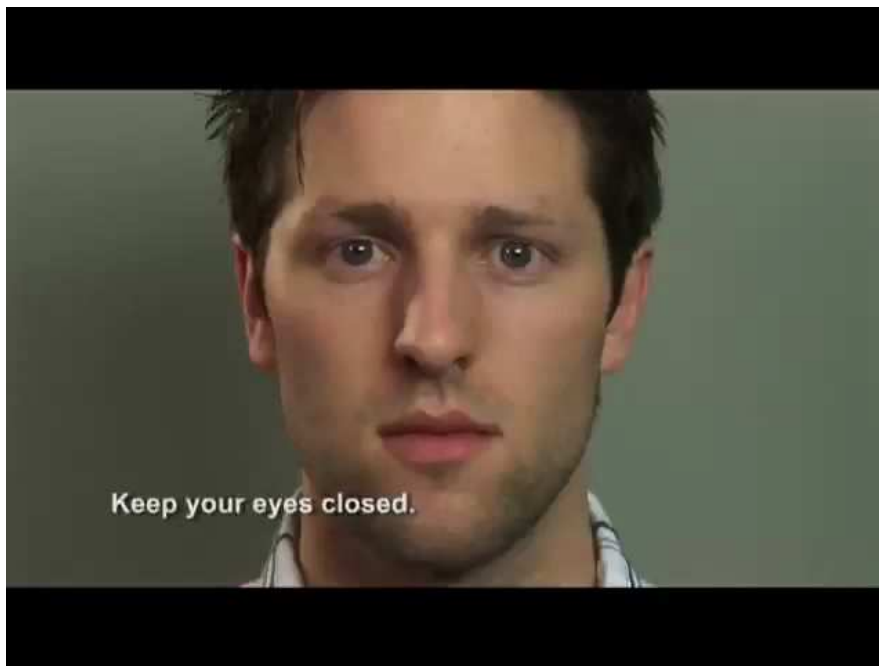
zmysł	przepływność [b/s]
wzrok	10 000 000
słuch	100 000
dotyk	1 000 000
smak	1 000
zapach	100 000
w sumie:	11 201 000

Percepcja wielomodalna

- ang. *multimodal (cross-modal) perception*
- w przypadku jednoczesnej stymulacji np. wzroku i słuchu
 - percepcja (odbiór) wrażenia wizyjno-fonicznego, różniącego się od złożenia wrażeń będących następstwem niezależnej stymulacji zmysłu wzroku i słuchu
- Korelacje wzrokowo-słuchowe w przypadku niedopasowania treści wizyjno-fonicznej mogą powodować:
 - poczucie dyskomfortu (wrażenie niespójności prezentowanej treści)
 - przekłamanie postrzeganego wrażenia (iluzja słuchowa)
- Typowy przykład przekłamania:
 - efekt McGurka

Percepcja wielomodalna

- Efekt McGurka



Percepcja wielomodalna

- w procesie percepcji obrazu i dźwięku biorą udział informacje przechowywane w pamięci
- do ich wydobycia konieczny jest odpowiedni bodziec, który zainicjalizuje proces przetwarzania – w filmie takim bodźcem jest dźwięk
- przykład:
 - scena przedstawiająca słabo oświetlone pomieszczenie z towarzyszącym dźwiękiem, np. krzykiem; u większości widzów taka scena wywoła poczucie niepokoju, a nawet strachu

Percepcja wielomodalna

- Poczucie niedopasowania treści wizyjnej i fonicznej wynikające z informacji przechowywanej w pamięci



Percepcja wielomodalna

- Hipoteza 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, np.
 - zmysł wzroku zazwyczaj dominuje nad zmysłem słuchu w przypadku określania odległości, czyli w analizie położenia źródeł dźwięku w przestrzeni
 - słuch dominuje nad wzrokiem w sytuacjach, gdy wymagana jest duża rozdzielczość czasowa

Geneza

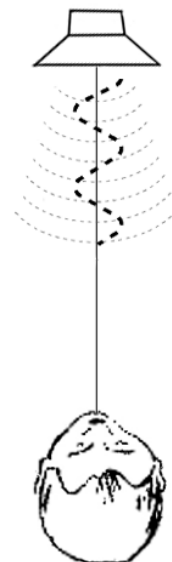
- Pierwsze badania w XIX w.
 - **Stratton** – prekursor badań w tej dziedzinie
 - bodźce wzrokowe mają istotny wpływ na lokalizację źródła dźwięku w przestrzeni
- Różne konteksty badania korelacji wzrokowo-słuchowych:
 - synchronizacja dźwięku i obrazu
 - kompresja obrazu wizyjnego na podstawie informacji zawartych w ścieżce dźwiękowej
 - odbiór treści wizyjno-fonicznej przez widza
 - zmiana percepcji kierunku źródła dźwięku na skutek stymulacji wzrokowej (przesunięcie pozornego źródła dźwięku)

Odbiór treści wizyjno-fonicznej przez widza

- badania Holliera i Voelckera
 - pogorszenie jakości obrazu nie wpływa na jej postrzeganie w przypadku prezentacji dźwięki o niezmienionej jakości
- badania Davisa
 - dźwięk dodany do obrazu znacząco zwiększa poczucie uczestnictwa w wirtualnej rzeczywistości (wirtualny świat wydaje się bardziej realistyczny, gdy obrazowi towarzyszą dźwięki)
- badania Stormsa i Zydy
 - dźwięk wysokiej jakości towarzyszący obrazowi wysokiej jakości:
 - obraz jest postrzegany jako lepszej jakości w porównaniu z tym samym obrazem bez towarzyszenia dźwięku

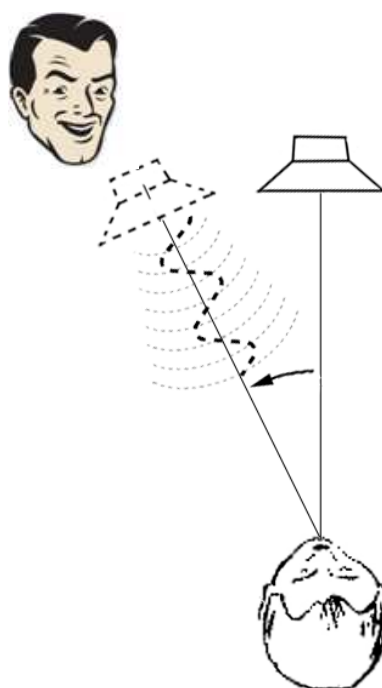
Przesunięcie pozornego źródła dźwięku...

- stymulacja słuchowa



Przesunięcie pozornego źródła dźwięku...

- stymulacja wzrokowo-słuchowa
 - zmiana percepcji kierunku źródła dźwięku na skutek stymulacji wzrokowej (przesunięcie pozornego źródła dźwięku w kierunku bodźca wzrokowego)



Przesunięcie pozornego źródła dźwięku...

- **Witkin**

- kierunek głosu spikera tv
- 2 etapy eksperymentu: w pierwszym badani widzieli twarz spikera i słyszeli jego głos, zaś w drugim słyszeli tylko jego głos
- w pierwszym przypadku badani lokalizowali głos spikera jako „dochodzący ze środka”
- w drugim – badani słyszeli jego głos jako dochodzący z lewej lub prawej strony

- **Komiyama i Nakabayashi**

- płeć badanego determinuje jego ocenę lokalizacji źródła dźwięku w panoramie stereofonicznej
- zdecydowanie większy wpływ ściągający występował, gdy mężczyźni prezentowano obraz spikerki, a kobiecie obraz spikera, niż w przypadku, gdy badany i spiker byli tej samej płci

Zastosowanie śledzenia wzroku

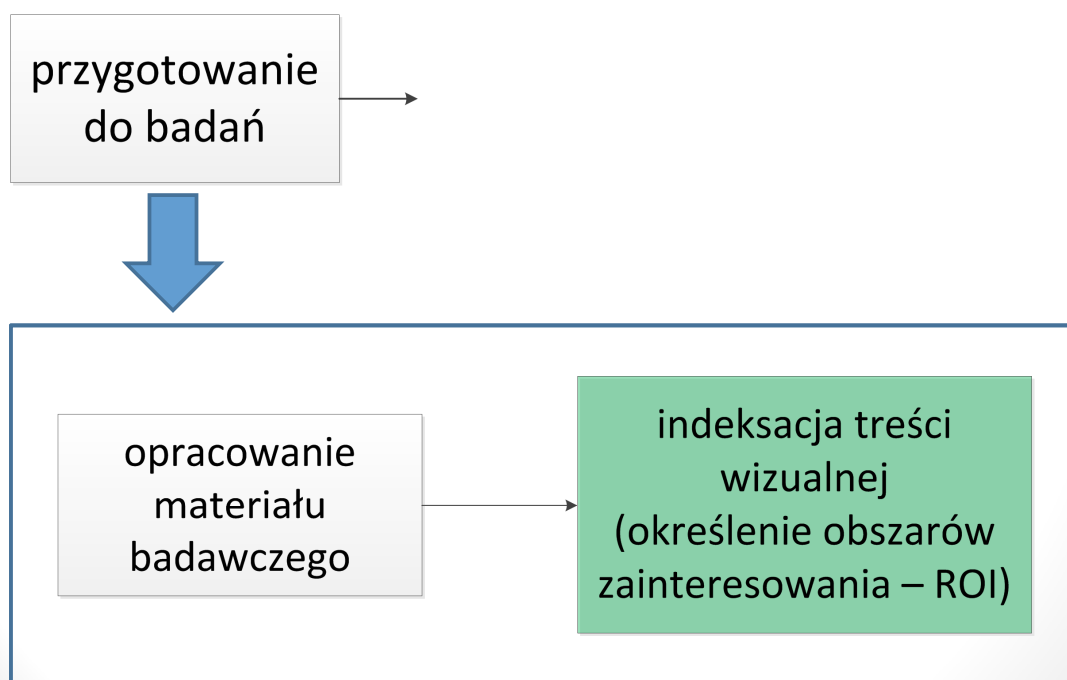
- **Badania subiektywne**

- wnioski na podstawie wyników ankiet

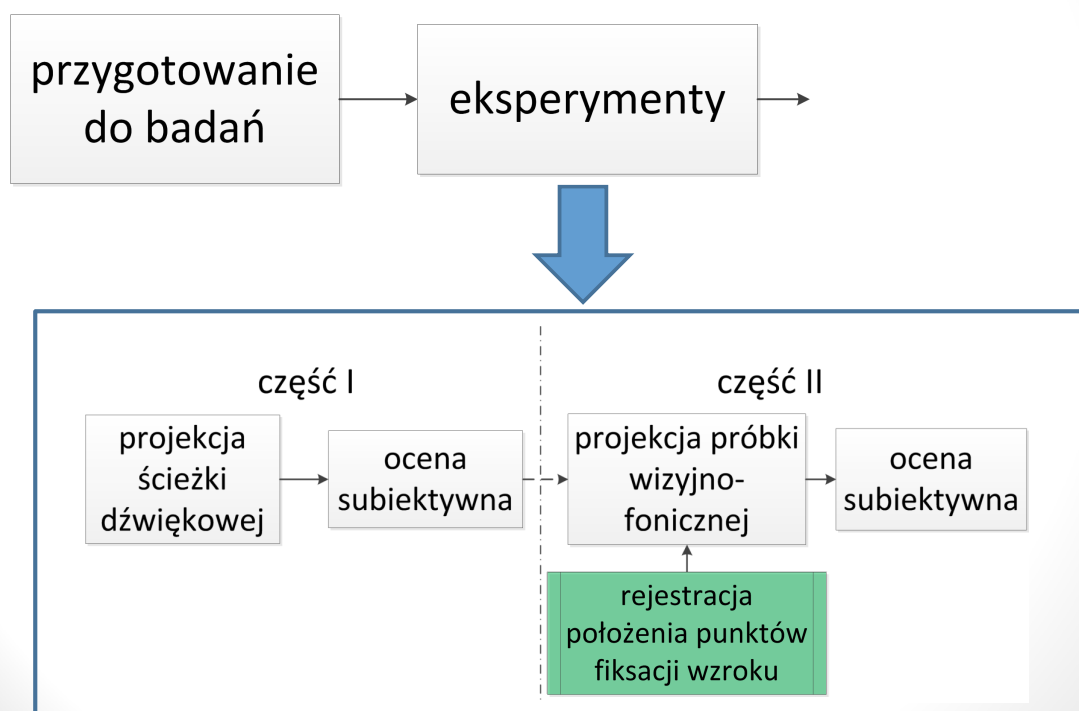


zastosowanie systemu śledzenia punktu fiksacji wzroku
– obiektywizacja badań subiektywnych

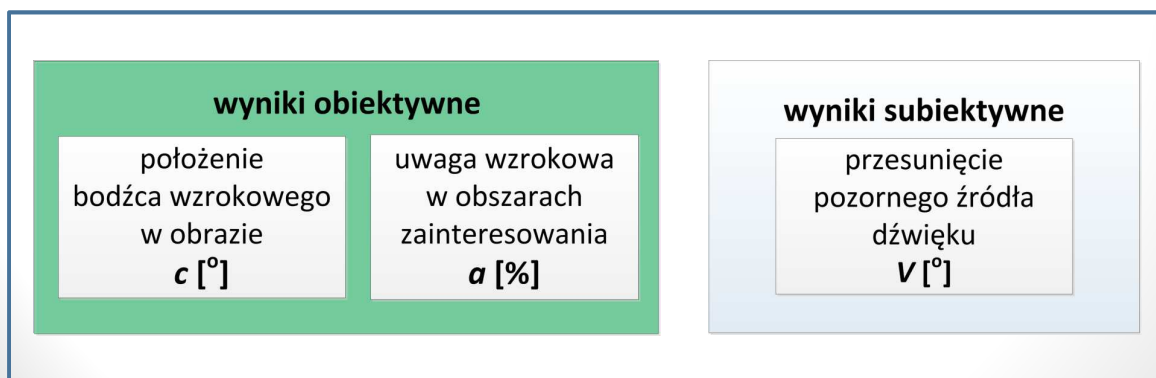
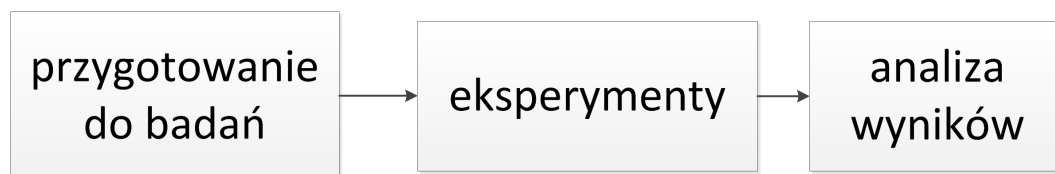
Metodologia



Metodologia



Metodologia

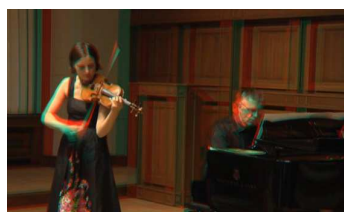


Badania – założenia

- wykorzystanie 2 systemów śledzenia wzroku (Tobii, Cyber-Oko)
- opracowanie materiału badawczego
 - fragmenty rzeczywistych filmów 3D (pozyskane z płyt w formacie Blu-ray 3D)



- próbki autorskie (nagranie koncertu duetu: skrzypce-fortepian)



Badania – założenia

- 2 konfiguracje stanowiska badawczego (wielkość obszaru wyświetlania)

- konfiguracja nr 1: ekran monitora
(mały i średni obszar wyświetlania)
- konfiguracja nr 2: ekran projektora
(duży obszar wyświetlania)

badanie „skalowalności” wpływu ściągającego obrazu na percepcję dźwięku

„Skalowalność” – kąt widzenia obrazu obiektu decyduje o wpływie ściągającym niezależnie od wielkości wyświetlanego obrazu

Badania – założenia

- druga konfiguracja stanowiska badawczego (duży obszar wyświetlania)

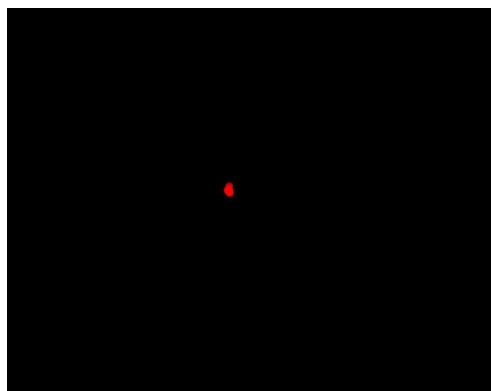


Badania

Wizualizacja uwagi wzrokowej widza

- dynamiczna mapa przejść

Tobii T60



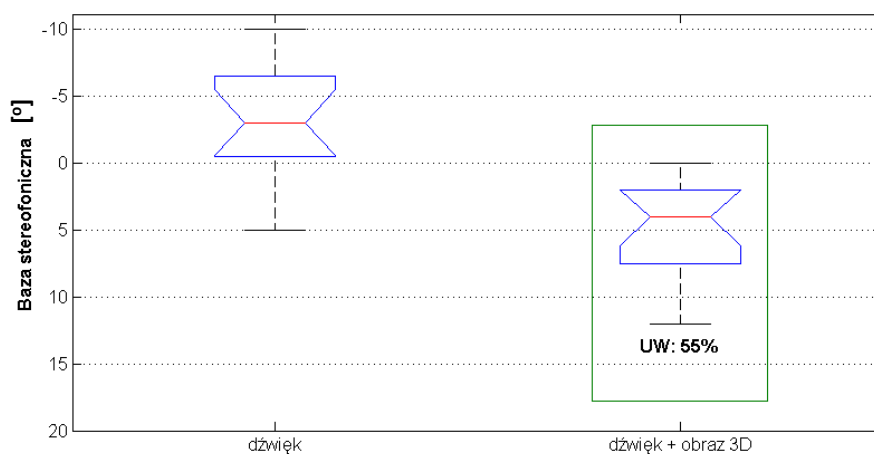
Cyber-Oko



Badania – wyniki

- wykres pudełkowy (wpływ ściągający istotny statystycznie)
próbka 2 (pan, średni) – głos bohatera, fragment filmu „Avatar”:

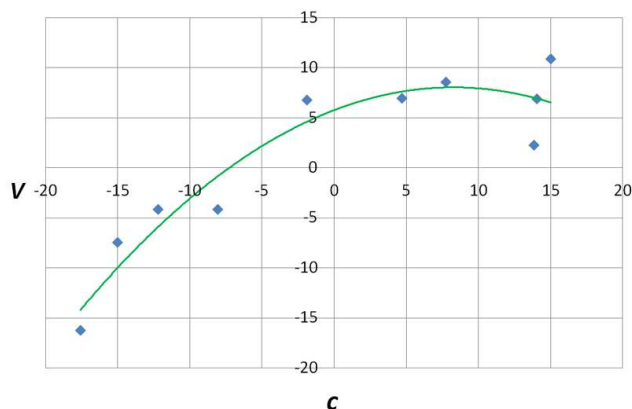
$$F(1, 28) = 33,09; p = 0,000004$$



Badania – wyniki

Wpływ ściągający a lokalizacja bodźca wzrokowego w obrazie

- $V = f(c)$

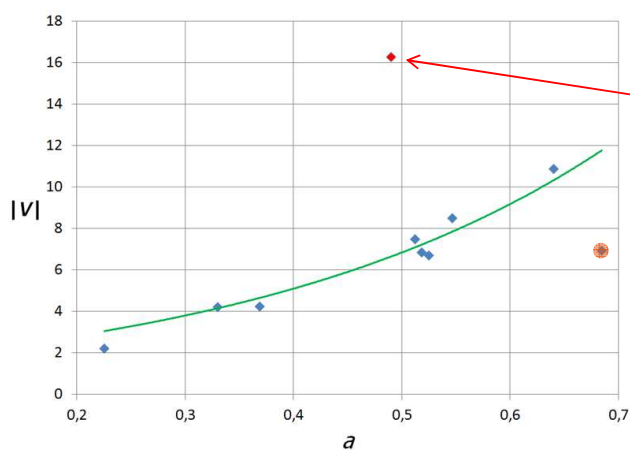


$R = 0,855; p = 0,002$

Badania – wyniki

Wpływ ściągający a uwaga wzrokowa widza

- $|V| = f(a)$



wynik dla **próbki nr 6** stanowiącej fragment filmu „Piranha 3D” – bodziec wzrokowy w lewej części ekranu, znacznie oddalony od środka: $|c| \gg 0$

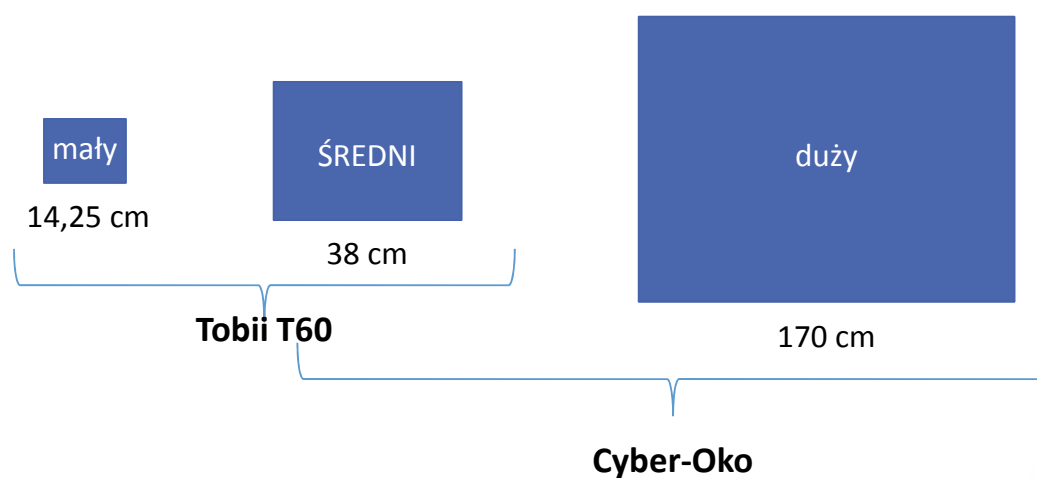


położenie bodźca wzrokowego jest czynnikiem dominującym

$R = 0,8; p = 0,01$

Badania – wyniki

„Skalowalność” wpływu ściąającego



- kąt widzenia obrazu obiektu decyduje o wpływie ściąającym **niezależnie od wielkości wyświetlanego obrazu**

Podsumowanie

- istnieje związek pomiędzy położeniem bodźca wzrokowego w obrazie a wpływem ściąającym
- istnieje zależność pomiędzy czasem skupienia wzroku na bodźcu wzrokowym a obserwowanym wpływem ściąającym
- wpływ ściąający obrazu na percepcję dźwięku jest zjawiskiem skalowalnym