

**prof. Bożena Kostek**

**Technologia nagrań – testy subiektywne**

**Politechnika Gdańska**

**Wydział Elektroniki, Telekomunikacji i Informatyki**

**Katedra Systemów Multimedialnych**

# Metody oceny subiektywnej

## Testy subiektywne

```
graph TD; A[Testy subiektywne] --> B[Metoda preferencji dwójkowych (test porównań parami)]; A --> C[Metoda parametryczna (test parametryczny)];
```

Metoda preferencji  
dwójkowych  
(test porównań parami)

Metoda parametryczna  
(test parametryczny)

# Metody oceny subiektywnej

## Testy subiektywne

```
graph TD; A[Testy subiektywne] --> B[Metoda preferencji dwójkowych (test porównań parami)]; A --> C[Metoda parametryczna (test parametryczny)];
```

Metoda preferencji  
dwójkowych  
(test porównań parami)

Metoda parametryczna  
(test parametryczny)

# Metoda preferencji dwójkowych

- Zasada: wybór jednego spośród dwóch fragmentów, lepszego w subiektywnym odczuciu eksperta (jeden z fragmentów może być wzorcem)
- Ocena w skali 2- lub 3-stopniowej:
  - lepszy - gorszy
  - lepszy - taki sam - gorszy

# Struktura testu porównań parami

Układ testowanych par:

A-B W-A-B A-B-A A-B-A-B

W-A-W-B

Ocena:

- większy – mniejszy (lepszy – gorszy)
- większy – taki sam- mniejszy (lepszy – taki sam – gorszy)
- Dla zespołu  $n$  obiektów konieczne jest wykonanie  $q$  porównań:

$$q = \frac{n \cdot (n - 1)}{2}$$

# Procedura testu

- Prezentacja kolejnych par obiektów zgodnie z zasadą “każdy z każdym”
- Powtórna prezentacja całej serii w celu zwiększenia obiektywizmu oceny
- Losowa kolejność par sygnałów w każdej serii, umożliwiająca weryfikację stabilności sądów ekspertów

# Struktura testu porównań parami

- Jedna para: 8-sekundowe (8-15 s) fragmenty, przedzielone 2-sekundową przerwą
- 5-sekundowy odstęp między kolejnymi parami
- 15 par w jednej serii
- Druga seria po 3-4 minutach przerwy (pary w innej kolejności)
- kilkusobowe grupy ekspertów

# Struktura testu porównań parami

- Czas trwania całej sesji odsłuchowej (15-20 min)

$$T_c = (q - 1) \cdot T_3 + q \cdot k \cdot T_1 + q \cdot (k - 1) \cdot T_2$$

- $T_c$  - całkowity czas trwania bloku zadań
- $T_1$  - czas trwania pojedynczego bloku
- $T_2$  - czas trwania przerwy między bodźcami
- $T_3$  - czas trwania przerwy między kolejnymi zadaniami dźwiękowymi
- $q$  - liczba zadań
- $k$  - liczba bodźców porównywanych w jednym zadaniu



# Metody oceny subiektywnej

## Testy subiektywne

```
graph TD; A[Testy subiektywne] --> B[Metoda preferencji dwójkowych (test porównań parami)]; A --> C[Metoda parametryczna (test parametryczny)];
```

Metoda preferencji  
dwójkowych  
(test porównań parami)

Metoda parametryczna  
(test parametryczny)

# Metoda parametryczna

- Zasada: ocena wybranych parametrów każdego z obiektów na podstawie jego pojedynczej prezentacji
- Ocena przykładowych parametrów w skali bezwzględnej:
  - szerość i ciągłość bazy,
  - przejrzystość, przestrzenność,
  - dynamika, jasność
  - brzmienia, ocena ogólna

# Struktura testu parametrycznego

- Fragmenty 25-sekundowe
- 15 sekund przerwy pomiędzy fragmentami
- Podział 7 parametrów na dwie grupy (po 3 i 4 parametry w dwóch seriach)
- Serie oddzielone 3-4 minutową przerwą

# Prezentacja wyników i sposobów ich analizy

## Testy subiektywne

```
graph TD; A[Testy subiektywne] --> B[Metoda preferencji dwójkowych (test porównań parami)]; A --> C[Metoda parametryczna (test parametryczny)];
```

Metoda preferencji  
dwójkowych  
(test porównań parami)

Metoda parametryczna  
(test parametryczny)

# Test porównań parami

- Analiza wyników - obliczenia:
  - parametry statystyczne związane z liczbą pomyłek ekspertów
  - krzywe preferencyjne:
    - łącznie dla wszystkich ocen
    - dla poszczególnych serii
    - dla kolejnych ekspertów

# Test porównań parami

- Dane podlegające analizie statystycznej, podczas której wykonuje się następujące obliczenia:
  - 1. Zsumowanie liczby głosów oddanych przez poszczególnych ekspertów na każdy z obiektów,
  - 2. Określenie stabilności wskazań każdego eksperta (parametr  $z_1$ ). Polega ona na określeniu liczby odmiennych wskazań w obrębie jednej pary (dla dwóch części testu),
  - 3. Wyznaczenie sumy głosów oddanych na każdy obiekt przez wszystkich ekspertów,
  - 4. Wyznaczenie liczby głosów oddanych na każdy obiekt przez wszystkich ekspertów osobno dla obu części testu,
  - 5. Wyznaczenie statystyki  $\chi^2$  porównującej wyniki obu części testu. Polega ona na stwierdzeniu czy istnieje istotna różnica między głosami oddanymi w obu częściach testu,
  - 6. Wyznaczenie liczby ekspertów, którzy daną parę interpretują odmiennie w zależności od części testu (parametr  $z_2$ ),
  - 7. Badanie istotności różnic pomiędzy obiektami tworzącymi daną parę, przy założonym poziomie istotności (parametr  $z_3$ ).

$$\chi^2 = n \cdot \sum_{i=1}^r \sum_{j=1}^s \frac{\left( n_{ij} - \frac{n_{i\cdot} \cdot n_{\cdot j}}{n} \right)^2}{n_{i\cdot} \cdot n_{\cdot j}}$$

gdzie:

- $r$  – liczba części testu,
- $s$  – liczba obiektów badanych,
- $n_{ij}$  – liczba obserwacji, które należą do  $i$ -tej części testu oraz  $j$ -tego obiektu. Określa ona, ile razy  $j$ -ty obiekt został wybrany w  $i$ -tej części testu,

gdzie:

- $n_{i\bullet} = \sum_{j=1}^s n_{ij}$  – liczba obserwacji należących do  $i$ -tej części testu,
- $n_{\bullet j} = \sum_{i=1}^r n_{ij}$  – liczba obserwacji należących do  $j$ -tego obiektu,
- $n = \sum_{j=1}^s \sum_{i=1}^r n_{ij}$  – ogólna liczba obserwacji.



# Statystyka

$\chi^2$

- Liczba stopni swobody / obliczana jest w oparciu o zależność:

$$l = (r - 1) \cdot (s - 1)$$

- Postawioną i następnie testowaną hipotezą jest statystyczna zgodność porównywanych wyników obu części testu. Jeżeli hipoteza ta jest prawdziwa, statystyka  $\chi^2$  nie powinna przyjmować zbyt dużej wartości. Obszar krytyczny jest określany na podstawie prawdopodobieństwa, że obliczona statystyka przekroczy wartość krytyczną:

$$P(\chi^2 > \chi^2_{\alpha}) = \alpha$$

# Statystyka



- W zależności powyższej  $\alpha$  jest poziomem istotności. Zatem zawsze, gdy wartość statystyki  $\chi^2$  przekroczy wartość krytyczną (odczytaną z tablic rozkładu  $\chi^2$  dla prawdopodobieństwa  $\alpha$  oraz dla  $l$  stopni swobody), hipotezę o zgodności wyników należy odrzucić na korzyść hipotezy alternatywnej, z prawdopodobieństwem błędnej decyzji równym  $\alpha$ .
- Interpretacja wyników testu odbyła się w oparciu o przyjęty poziom istotności  $\alpha = 0.05$ . Wartość krytyczna statystyki  $\chi^2$  jest dla 5 stopni swobody i obranego poziomu istotności równa  $\chi^2_{\alpha} = 11.070$
- W celu określenia istotności różnic (parametr  $z_3$ ) wykorzystuje się następującą zależność:

# Wzór obliczeniowy - istotność różnic

$$z_{ij} = \frac{|p_i - p_j|}{\sqrt{\frac{(p_i + p_j) \cdot (2 - p_i - p_j)}{2 \cdot N}}}$$

- $z_{ij}$  - określone prawdopodobieństwo
- $p_i, p_j$  - względne liczebności głosów opowiadających się za  $i$ -tymi i  $j$ -tymi obiektami  $p_i = \frac{r_i}{N}$   $p_j = \frac{r_j}{N}$
- $N$  - maks. liczba głosów możliwa do uzyskania przez jeden obiekt  $N = m \cdot (n - 1) \cdot 2$
- $m$  - liczebność grupy ekspertów,  $n$  - liczba porównywanych obiektów

# Wzór obliczeniowy - istotność różnic

- Statystykę  $\chi^2$  porównuje się z wartością graniczną z tablic rozkładu normalnego, która dla przyjętego poziomu istotności  $\alpha = 0.05$  wynosi  $z(\alpha) = 1.96$ . Jeżeli  $z_{ij} < z(\alpha)$ , to nie ma podstaw do odrzucenia hipotezy o braku istotnej różnicy między porównywanymi w danej parze obiektami (parametr  $z_3$  przyjmuje znak „-”). W przeciwnym wypadku należy przyjąć, że różnica pomiędzy obiektami tworzącymi daną parę jest statystycznie istotna (parametr  $z_3$  przyjmuje znak „+”).

# Wzór obliczeniowy – stabilność odpowiedzi ekspertów

- W celu zbadania stabilności odpowiedzi ekspertów stosuje się test polegający na porównaniu dwóch ciągów odpowiedzi każdego z nich (dla obu części, serii testu) i wyznaczeniu liczby odmiennych typowań w obrębie danej pary (parametr  $z_1$ ). Test taki nazywany jest testem znaków. Uzyskane wartości  $z_1$  porównuje się następnie z wartością krytyczną testu znaków, uzyskaną z tablic matematycznych, która dla poziomu istotności  $\alpha = 0.05$  oraz liczby obiektów np.  $n = 15$  wynosi 3.

# Test porównań parami - analiza

OBIEKT:	A	B	C	D	E	F	
EXP. 1	5	1	8	7	7	2	z1= 6
EXP. 2	2	0	9	8	5	6	z1= 3
EXP. 3	2	0	9	6	6	7	z1= 2
EXP. 4	3	1	8	4	4	10	z1= 3
EXP. 5	4	0	9	6	6	5	z1= 6
EXP. 6	8	1	7	2	6	6	z1= 2
EXP. 7	4	9	4	4	4	5	z1= 5
EXP. 8	5	1	6	4	8	6	z1= 6
EXP. 9	3	0	7	7	9	4	z1= 3
EXP.10	5	8	6	7	4	0	z1= 3
EXP.11	9	7	0	6	4	4	z1= 2
EXP.12	8	0	7	5	8	2	z1= 3
EXP.13	4	5	8	2	8	3	z1= 4
EXP.14	4	0	6	7	7	6	z1= 4
EXP.15	4	0	8	8	6	4	z1= 5
EXP.16	6	0	3	9	6	6	z1= 4
EXP.17	6	0	4	7	6	7	z1= 6
EXP.18	4	0	5	9	6	6	z1= 2
EXP.19	3	1	8	6	6	6	z1= 6
EXP.20	3	1	7	7	6	6	z1= 4

A - XY  
B - Stereosonic  
C - MS  
D - ORTF  
E - Card  
F - Omni

# Test porównań parami - analiza

```
suma = 92 35 129 121 122 101 600
cz.1 = 49 18 60 65 63 45 300
cz.2 = 43 17 69 56 59 56 300
```

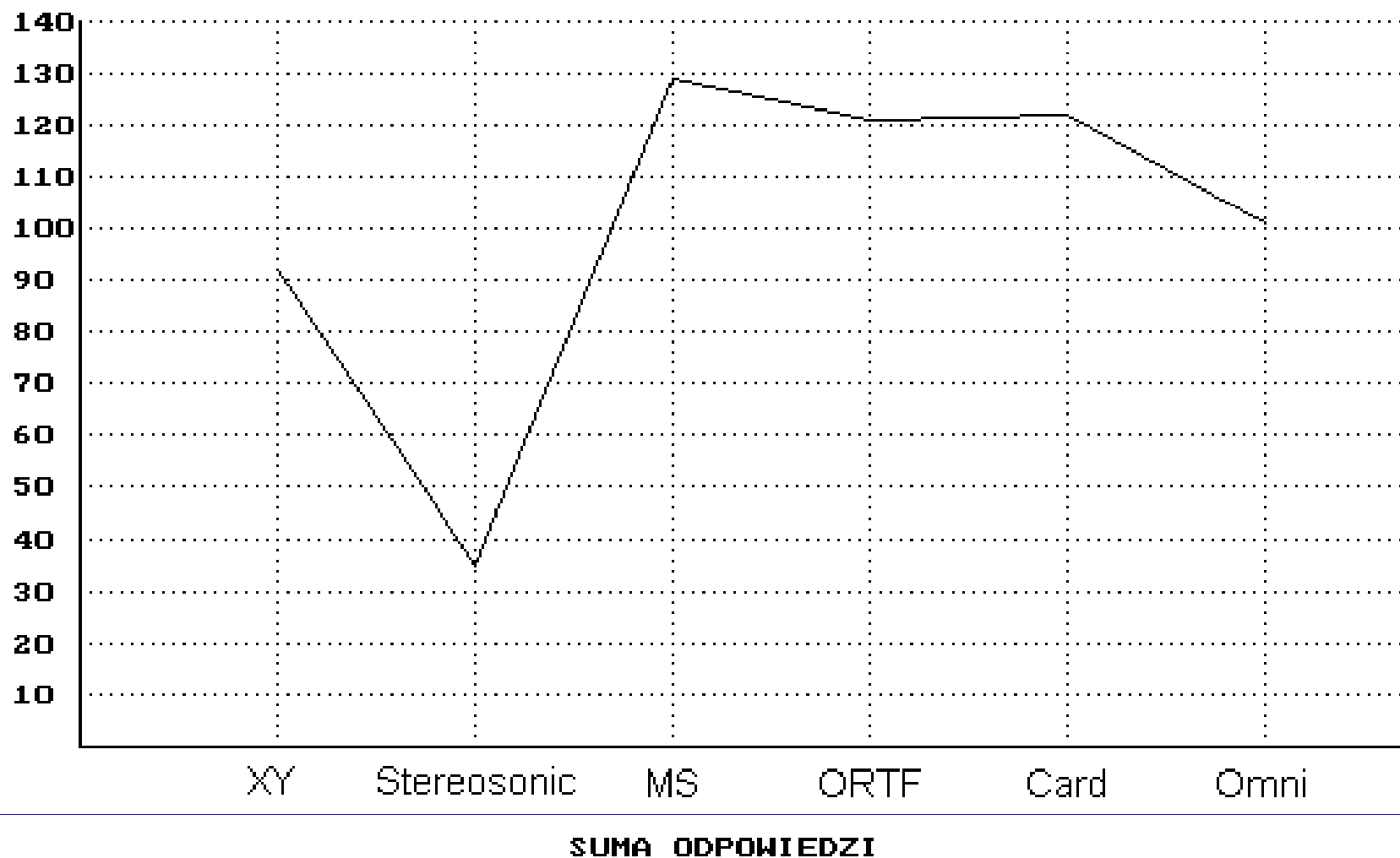
Porównanie pomiędzy częściami - PAR\_1.TOE :  
chi-kwadrat = 3.046

PARY:	A	A	A	A	A	B	B	B	B	C	C	C	D	D	E
	B	C	D	E	F	C	D	E	F	D	E	F	E	F	F
z2 =	3	4	5	7	9	1	2	1	4	9	10	3	10	7	4
z3 =	+	+	+	+	-	+	+	+	+	-	-	+	-	+	+

[5%]

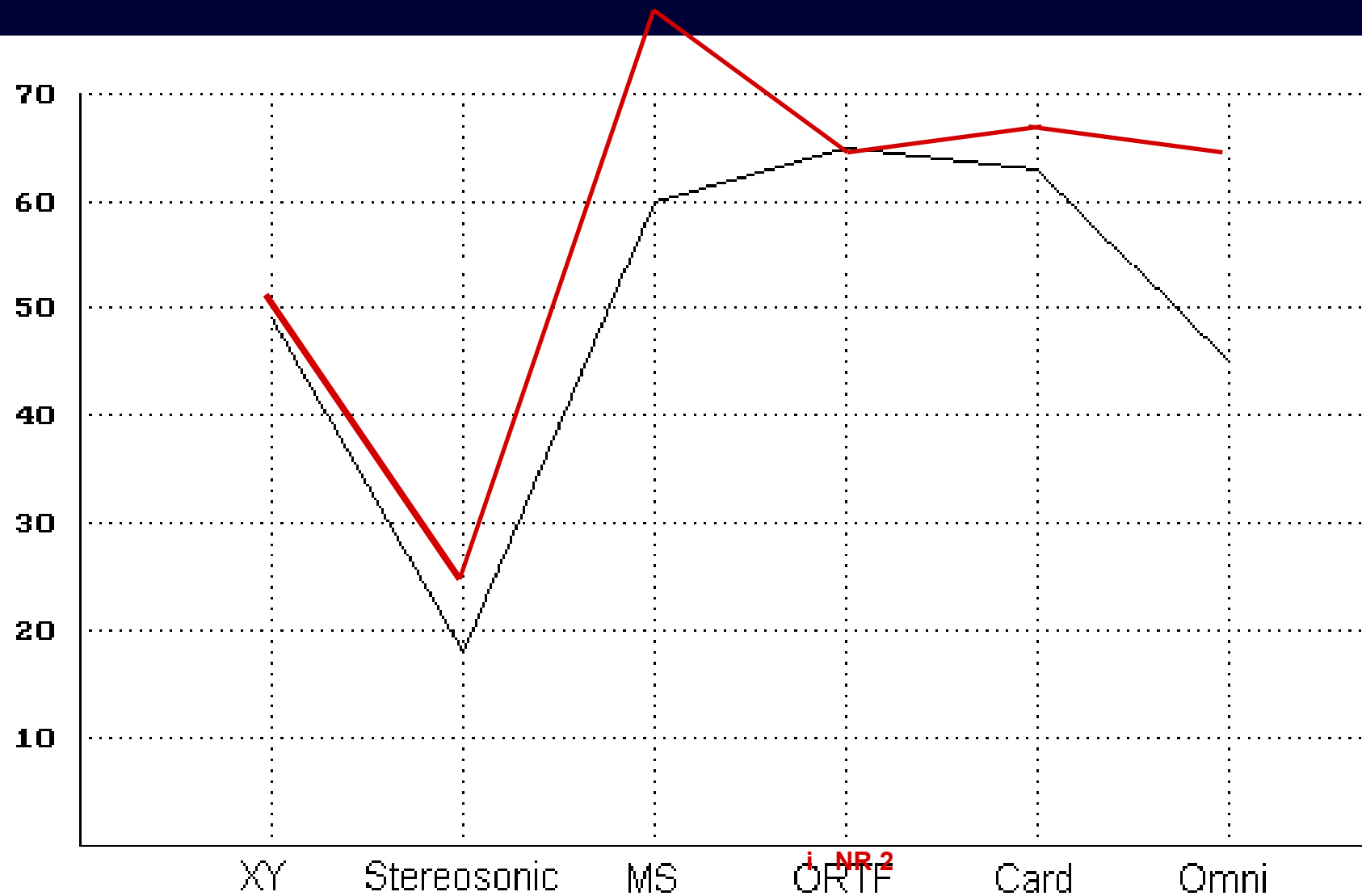
A - XY                    D - ORTF  
B - Stereosonic        E - Card  
C - MS                    F - Omni

# Test porównań parami - analiza





# Test porównań parami - analiza



CZESC NR 1

# Prezentacja wyników i sposobów ich analizy

## Testy subiektywne

```
graph TD; A[Testy subiektywne] --> B[Metoda preferencji dwójkowych (test porównań parami)]; A --> C[Metoda parametryczna (test parametryczny)];
```

Metoda preferencji  
dwójkowych  
(test porównań parami)

Metoda parametryczna  
(test parametryczny)

# Analiza wyników testu parametrycznego

- Uśrednienie wartości poszczególnych parametrów dla wszystkich systemów
- Analiza statystyczna obliczenia następujących zależności pomiędzy poszczególnymi parametrami:
  - kowariancję
  - korelację
  - statystykę  $t$
- Analiza w oparciu o logikę rozmytą

# Test parametryczny

- Zasada: ocena każdego fragmentu ze względu na wybrane parametry
- Możliwe parametry:  
przestrzenność, przejrzystość, spójność, dynamika, wyważenie dynamiczne, szerokość i ciągłość bazy, jasność i ciepło brzmienia, potęga brzmienia, ocena ogólna



Mean opinion score (MOS)

# MOS – Mean Opinion Score

- Ocena parametrów w skali 1-5

Mean opinion score (MOS)

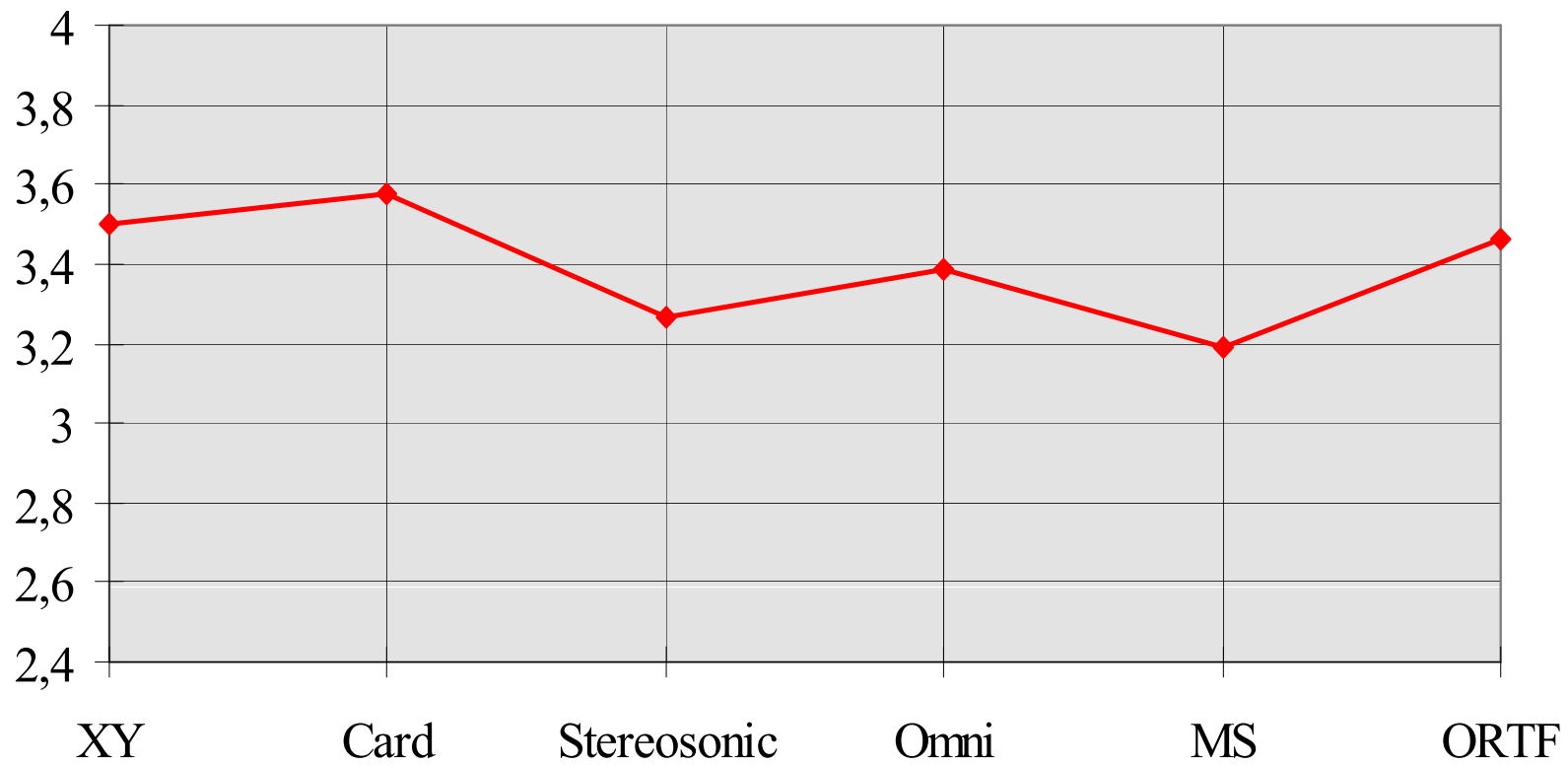
MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

# Analiza wyników

- Analiza w oparciu o teorię zbiorów rozmytych
- Cel: wyliczenie oceny ogólnej dla każdego z badanych systemów stereofonicznych, umożliwiające ich klasyfikację
- Wnioskowanie na podstawie oceny ogólnej oraz wyników pomiarów poszczególnych parametrów

# Analiza wyników

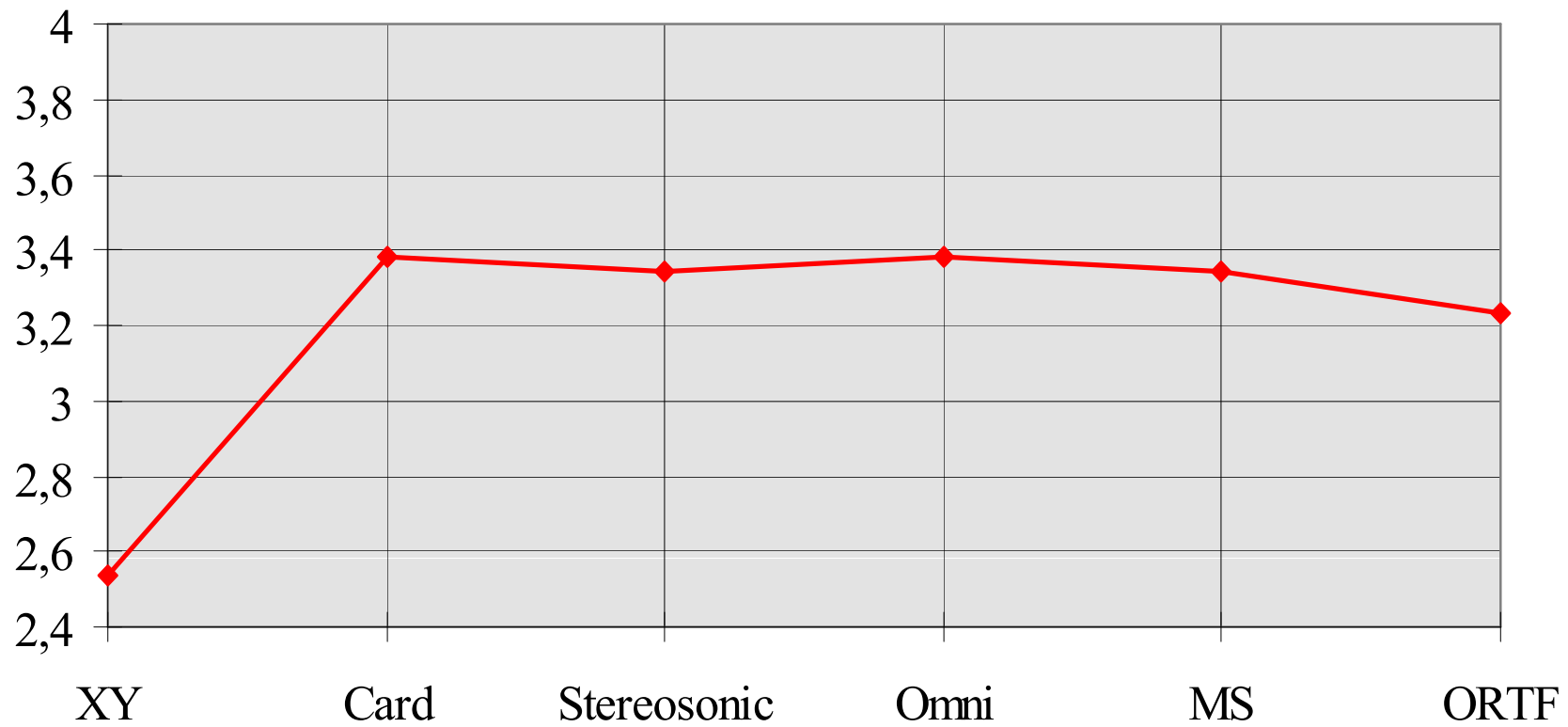
Szerokość bazy





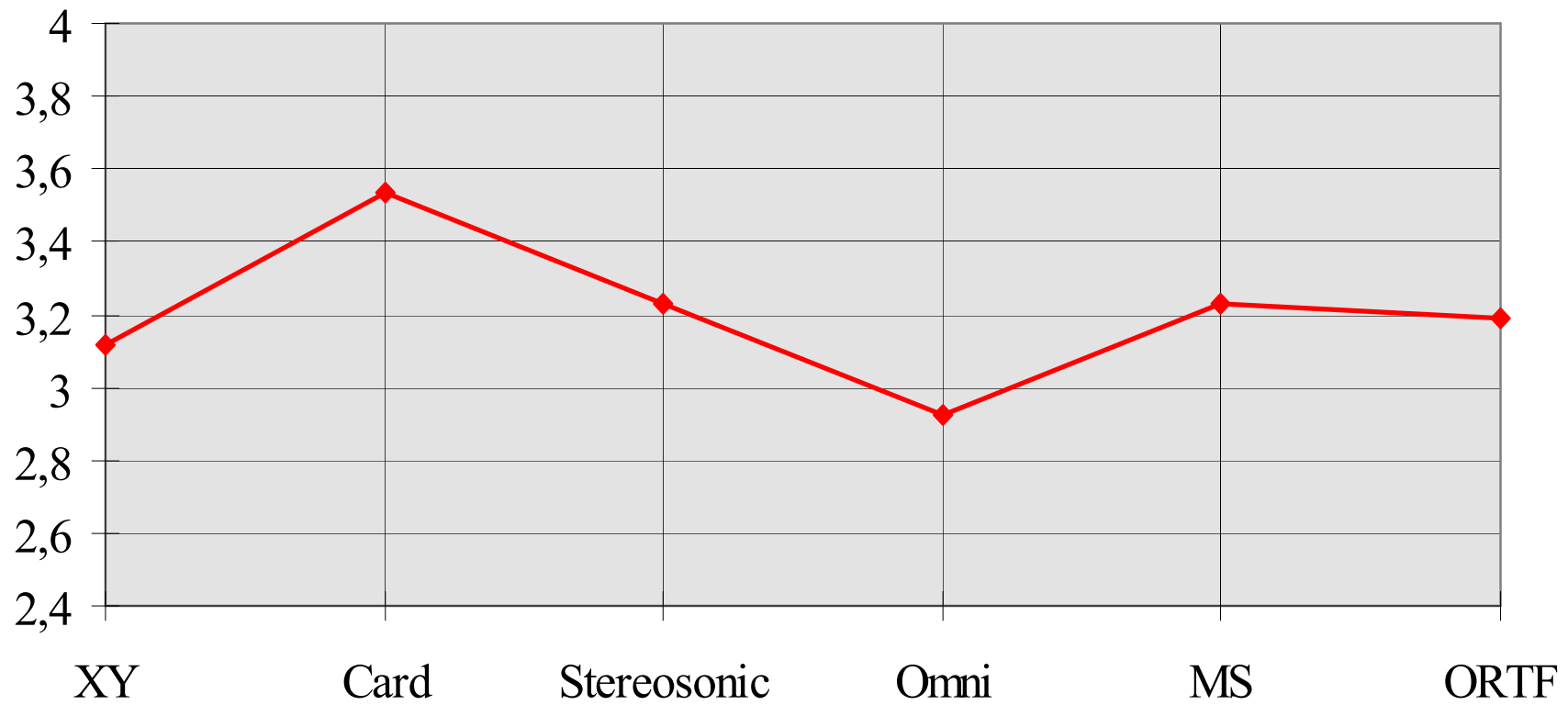
# Analiza wyników

Ciągłość bazy



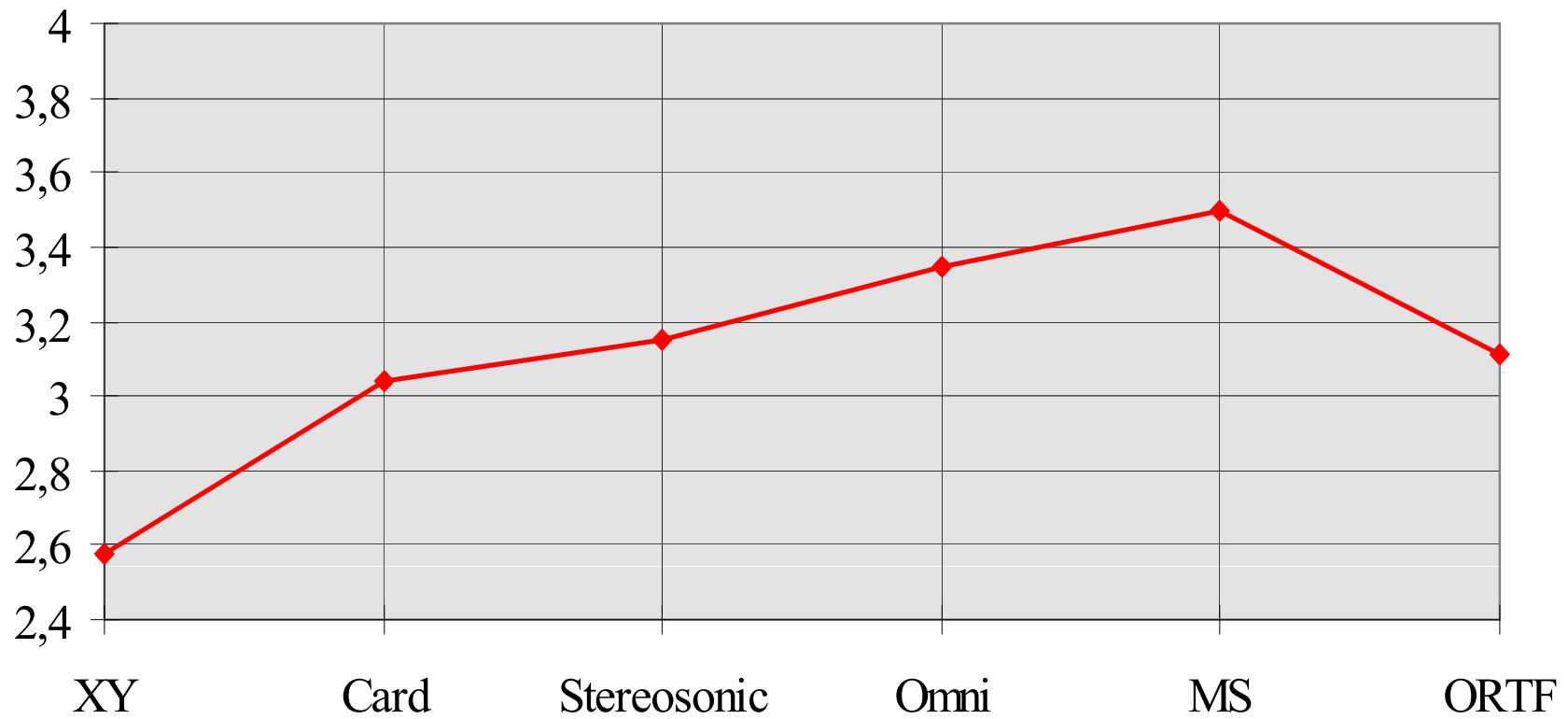
# Analiza wyników

## Przejrzystość



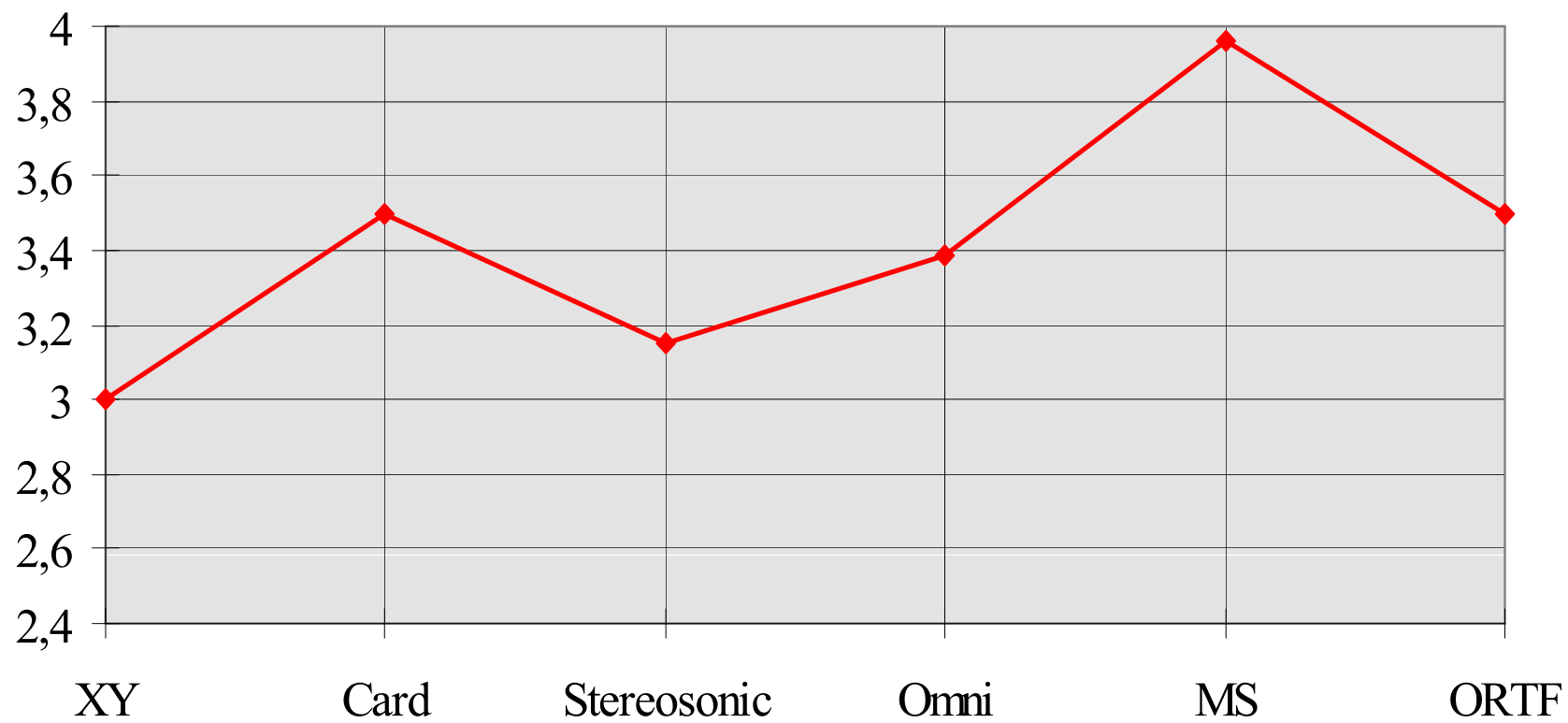
# Analiza wyników

## Przestrzenność



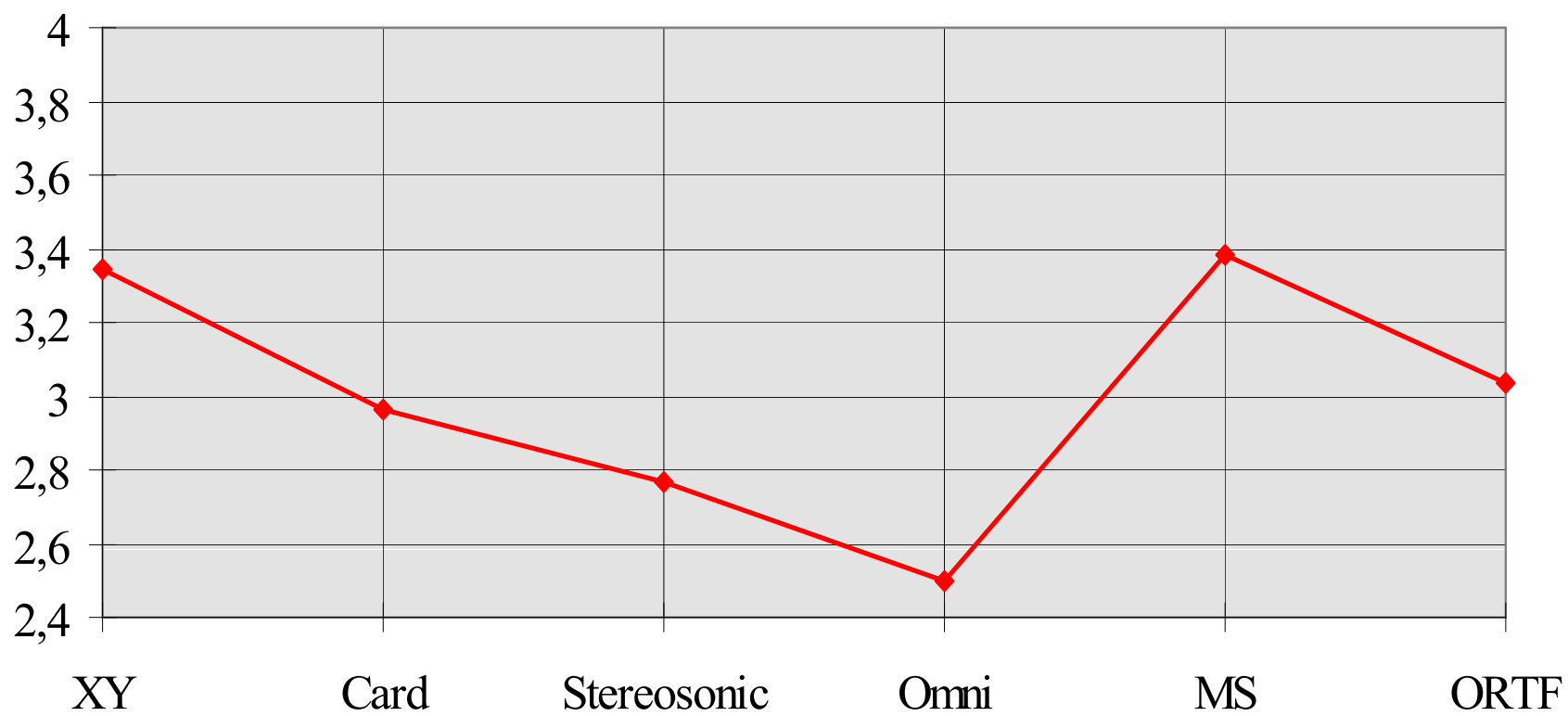
# Analiza wyników

Dynamika



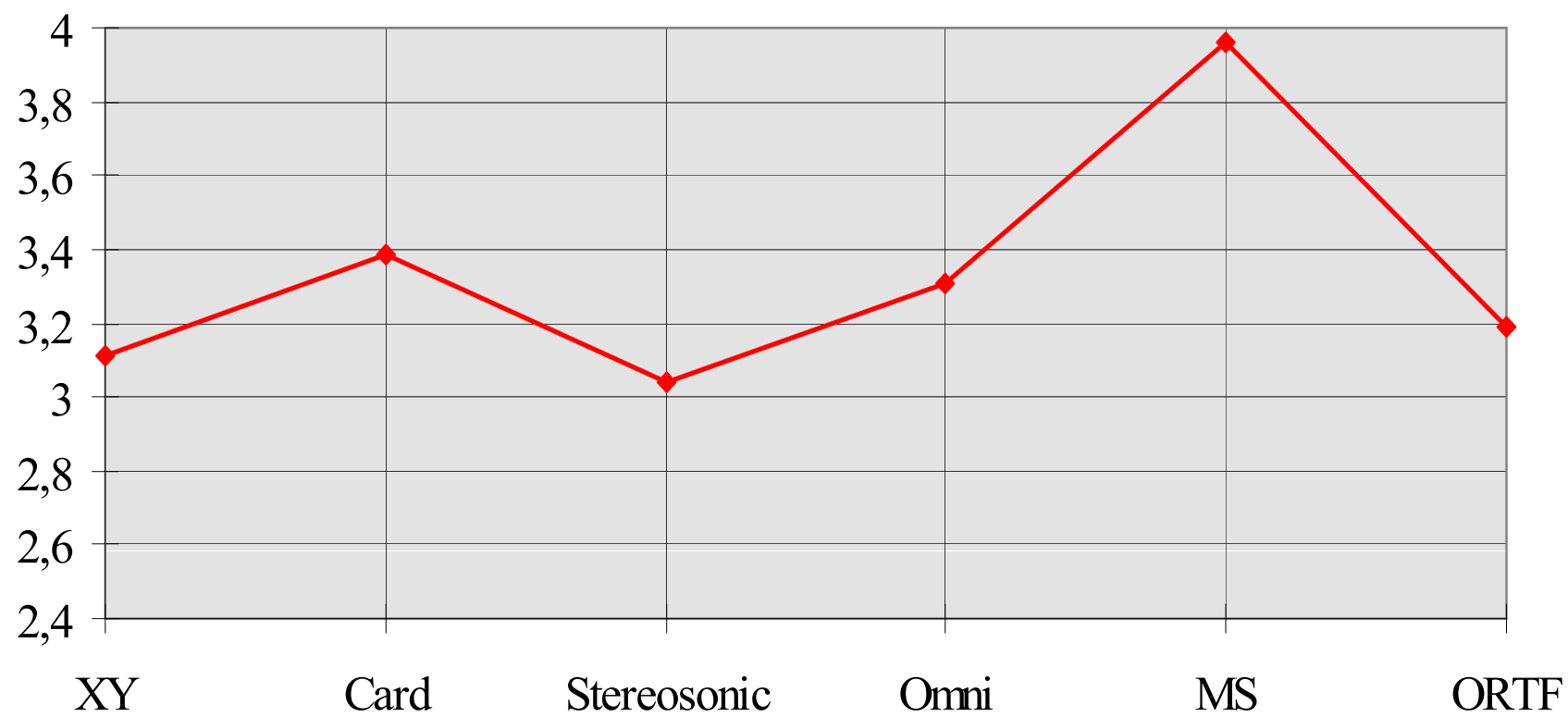
# Analiza wyników

Jasność brzmienia

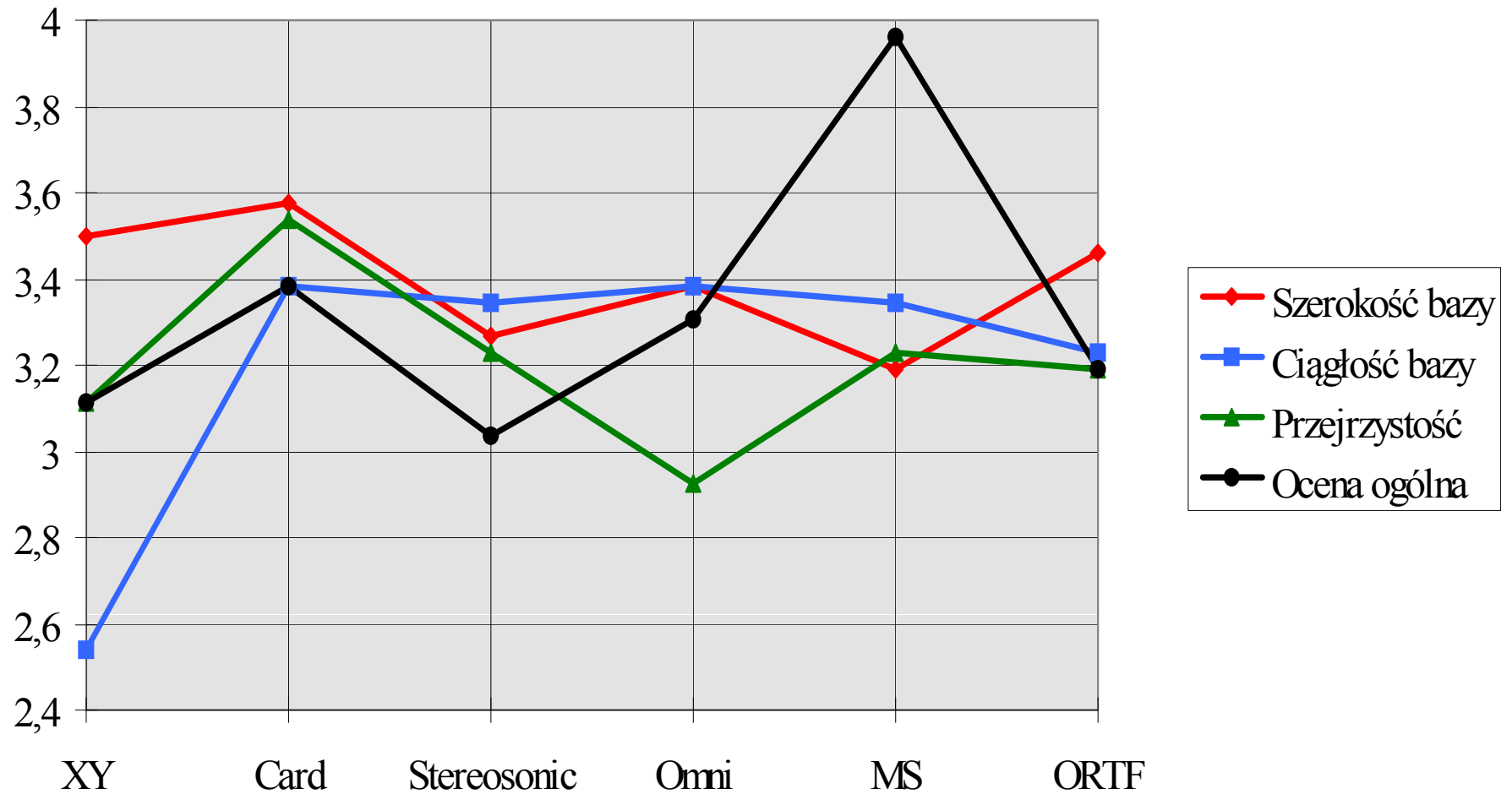


# Analiza wyników

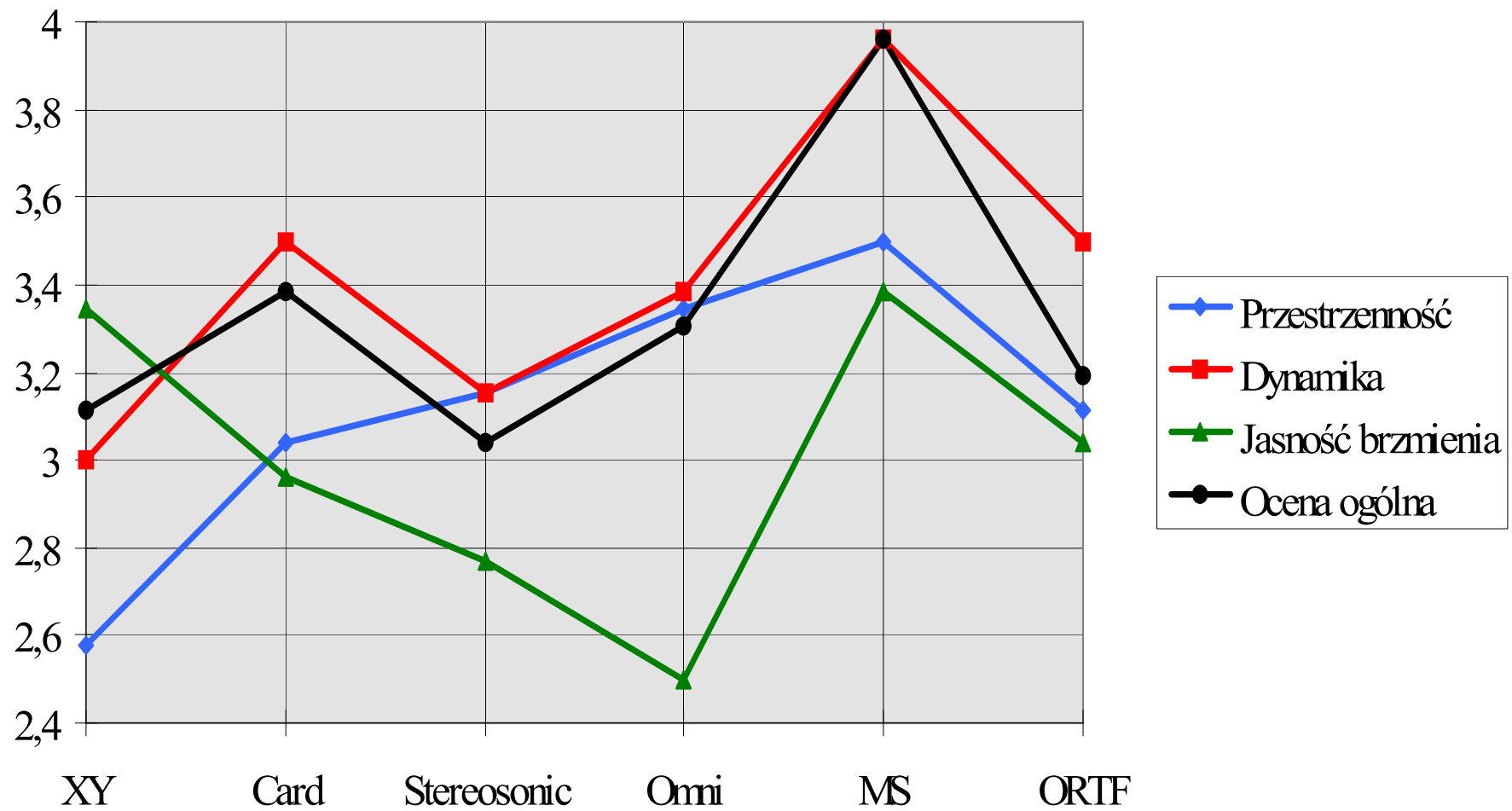
Ocena ogólna



# Analiza wyników



# Analiza wyników





# Analiza wyników

liczba parametrow [5]: 7

3.580	3.380	3.540	3.040	3.500	2.960	3.380	3.460	3.230	3.190
3.120	3.500	3.040	3.190	3.270	3.350	3.230	3.150	3.150	2.770
3.040	3.380	3.380	2.920	3.350	3.380	2.500	3.310	3.190	3.350
3.230	3.500	3.960	3.380	3.960	3.460	3.230	3.190	3.120	3.500
3.040	3.190								

liczba obiektow [6]:

SREDNIA :

3.390	3.320	3.217	3.213	3.498	2.948	3.345
-------	-------	-------	-------	-------	-------	-------

DYSPERSJA:

0.129	0.065	0.180	0.159	0.241	0.270	0.295
-------	-------	-------	-------	-------	-------	-------

MACIERZ KOWARIANCJI:

0.017	-0.002	0.009	-0.016	-0.008	-0.008	-0.017
-0.002	0.004	0.001	0.004	-0.001	-0.006	0.006
0.009	0.001	0.032	-0.015	0.006	0.023	0.006
-0.016	0.004	-0.015	0.025	0.023	0.009	0.036
-0.008	-0.001	0.006	0.023	0.058	0.052	0.066
-0.008	-0.006	0.023	0.009	0.052	0.073	0.051
-0.017	0.006	0.006	0.036	0.066	0.051	0.087

# Analiza wyników

MACIERZ KORELACJI:

1.000	-0.227	0.404	-0.785	-0.261	-0.217	-0.442
-0.227	1.000	0.103	0.344	-0.042	-0.335	0.324
0.404	0.103	1.000	-0.515	0.142	0.465	0.107
-0.785	0.344	-0.515	1.000	0.597	0.212	0.758
-0.261	-0.042	0.142	0.597	1.000	0.807	0.924
-0.217	-0.335	0.465	0.212	0.807	1.000	0.645
-0.442	0.324	0.107	0.758	0.924	0.645	1.000

MACIERZ STAT. t:

-----	-0.465	0.883	-2.536	-0.541	-0.445	-0.985
-0.465	-----	0.207	0.732	-0.083	-0.711	0.686
0.883	0.207	-----	-1.202	0.288	1.049	0.215
-2.536	0.732	-1.202	-----	1.490	0.434	2.324
-0.541	-0.083	0.288	1.490	-----	2.738	4.820
-0.445	-0.711	1.049	0.434	2.738	-----	1.687
-0.985	0.686	0.215	2.324	4.820	1.687	-----

# Analiza wyników w oparciu o teorię zbiorów rozmytych

---

- Cel stosowania tego rodzaju analizy:
  - uzyskanie oceny ogólnej dla każdego systemu w oparciu o oceniane parametry
  - zbadanie stopnia wpływu poszczególnych parametrów na końcową ocenę ogólną

# Procedura analizy przykładowego systemu (XY)

1. Zestawienie ilości poszczególnych ocen przyznanych każdemu parametrowi (dla danego systemu):

Ocena \ liczba głosów \ Parametr	5	4	3	2	1
Szerokość bazy	4	8	11	3	0
Ciągłość bazy	1	3	8	11	3
Przejrzystość	0	9	12	4	1
Przestrzenność	0	4	8	13	1
Dynamika	0	4	18	4	0
Jasność brzmienia	0	12	11	3	0

# Procedura analizy przykładowego systemu (XY)

## 2. Normalizacja i ujęcie wyników w formie macierzy:

$$R_{XY} =$$

0.154	0.308	0.423	0.115	0
0.038	0.115	0.308	0.423	0.115
0	0.346	0.462	0.154	0.038
0	0.154	0.308	0.5	0.038
0	0.154	0.692	0.154	0
0	0.462	0.423	0.115	0

# Procedura analizy przykładowego systemu (XY)

3. Obliczenie macierzy S:

$$S = R \circ W$$

W - macierz wag poszczególnych parametrów

° - suma logiczna zdefiniowana w logice rozmytej

$R_{XY} =$	0.154	0.308	0.423	0.115	0
	0.038	0.115	0.308	0.423	0.115
	0	0.346	0.462	0.154	0.038
	0	0.154	0.308	0.5	0.038
	0	0.154	0.692	0.154	0
	0	0.462	0.423	0.115	0

# Procedura analizy przykładowego systemu (XY)

## 4. Transpozycja macierzy S:

$$(S)^T = (0.154 \quad 0.462 \quad 0.692 \quad 0.5 \quad 0.115)$$

oraz jej normalizacja :

$$(S')^T = (0.08 \quad 0.24 \quad 0.36 \quad 0.26 \quad 0.06)$$

# Procedura analizy przykładowego systemu (XY)

5. Obliczenie ogólnej oceny końcowej zgodnie z zasadą:

$$\Gamma = s'_1 \cdot 100 + s'_2 \cdot 80 + s'_3 \cdot 60 + s'_4 \cdot 40 + s'_5 \cdot 20$$

Ocena ogólna w przypadku systemu XY wynosi:

$$T = 60.42$$



# Procedura analizy przykładowego systemu (XY)

6. Wytypowanie parametru mającego największy wpływ na ocenę ogólną:c

$R_{XY} :=$	0.154	0.308	0.423	0.115	0
	0.038	0.115	0.308	0.423	0.115
	0	0.346	0.462	0.154	0.038
	0	0.154	0.308	0.5	0.038
	0	0.154	0.692	0.154	0
	0	0.462	0.423	0.115	0

Szerokość bazy  
Ciągłość bazy  
Przejrzystość  
Przestrzenność  
Dynamika  
Jasność  
brzmienia

# Końcowe porównanie wyników

Test porównań parami	Test parametryczny (oceny własne ekspertów)	Test parametryczny (oceny uzyskane w oparciu o "fuzzy logic")
MS	MS	MS
Card	Card	Card
ORTF	Omni	ORTF
Omni	ORTF	Stereosonic
XY	XY	Omni
Stereosonic	Stereosonic	XY

# MUSHRA

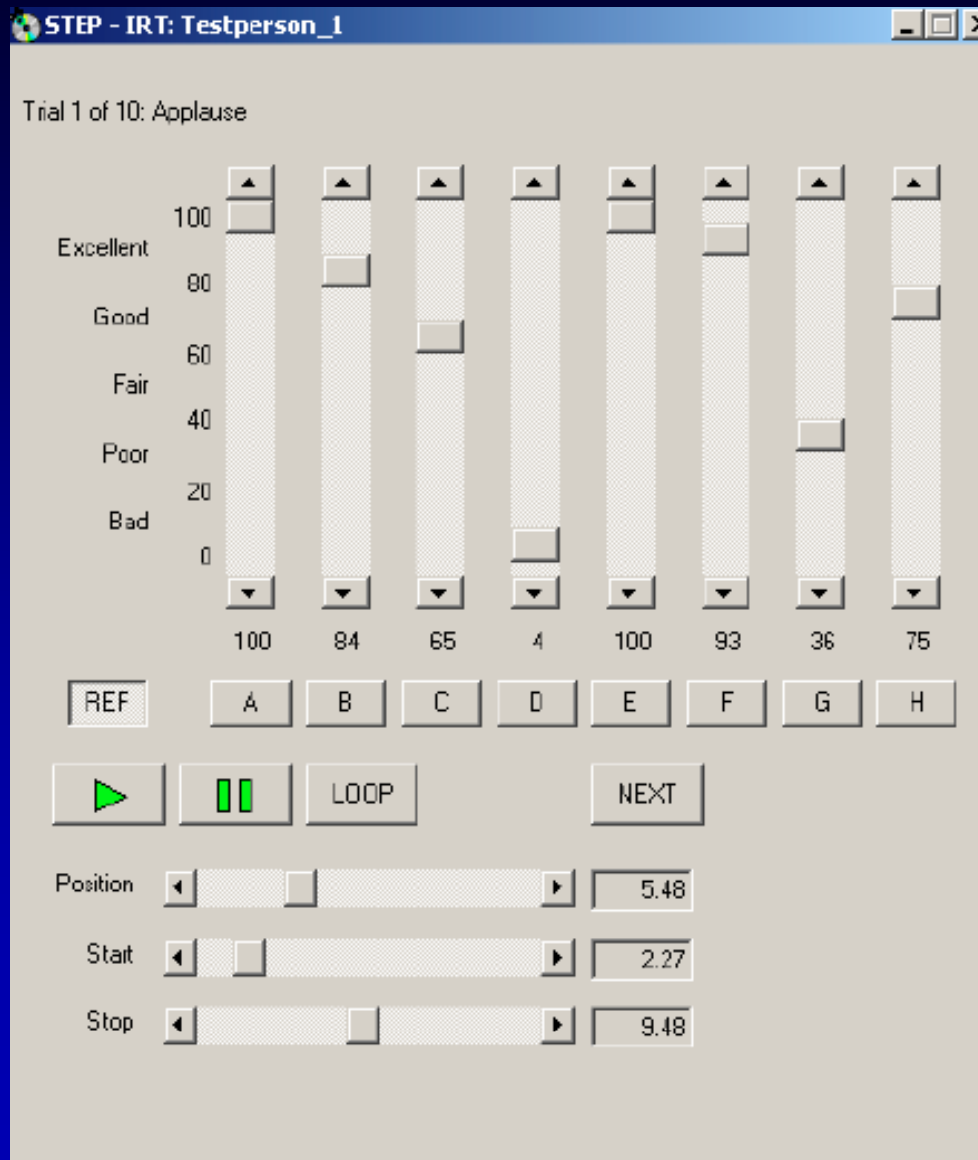
**MUSHRA (ang. *MU*ltiple  
*Stimuli with Hidden Reference and Anchor*) -  
BS.1534.1 -  
stosowany dla średnich i dużych zniekształceń  
sygnału  
zapropozowany przez ITU-R (ang. *International  
Telecommunication Union Radiocommunication*).**

# MUSHRA

**Każdy słuchacz musi przejść wstępną fazę treningową, podczas której zostaje zaznajomiony z procesem testów odsłuchowych.**

**W czasie trwania testów odsłuchowych ustalona liczba osób ma za zadanie porównać i ocenić jakość prezentowanych sygnałów testowych w odniesieniu do sygnału oryginalnego. Każdy uczestnik eksperymentu wyraża swoją opinię w ustalonej wcześniej skali, za pomocą odpowiedniej aplikacji kontrolującej przebieg badania. Następnie indywidualne noty są uśredniane i poddawane obróbce statystycznej, w celu oszacowania całłościowych ocen jakości.**

# MUSHRA



<http://tech.ebu.ch/docs/tech/tech332>  
4.pdf

# MUSHRA

## Anchors

The choice of appropriate anchors is fundamental both for subject rejection and for statistical issues (such as test labs comparison). The MUSHRA methodology was basically developed in order to test stereophonic audio sequences. That is why it needs some adaptation to be used for multichannel audio testing. From those considerations, the choice of the anchors was the following:

- A hidden reference (unprocessed signal)
- A low anchor signal: a filtered version (3.5 kHz low pass) of the unprocessed signal.
- Spatial anchor signal: generated by introducing deliberate crosstalk between the channels, resulting in the distortion of the spatial image. The two first listed anchors are mandatory by the MUSHRA [<http://tech.ebu.ch/docs/tech/tech3324.pdf>]

# MUSHRA

**The following Quality Scale was used:**

- **Excellent**
- **Good**
- **Fair**
- **Poor**
- **Bad**

**The scale is continuous from "Excellent" (100) to "Bad" (0).**