

Metody oceny jakości dźwięku

Subiektywne testy odsłuchowe

Wprowadzenie

- Problem: **ocena jakości dźwięku** – urządzenia (np. zestawy głośnikowe), pomieszczenia, algorytmy.
- Metody obiektywne - np. pomiar SNR czy THD+N - nie dają pełnych informacji o jakości sygnału.
- Ważne jest to, jaka jest jakość dźwięku odbieranego **subiektywnie** przez słuchacza.
- Stosuje się **subiektywne testy odsłuchowe**.
- Słuchacze - eksperci odsłuchują testowy zestaw sygnałów i oceniają każdy z nich według podanej skali.
- Jest to ocena subiektywna, ponieważ opiera się na indywidualnym wrażeniu odnoszonym przez słuchacza.

Zastosowania testów subiektywnych

Przykładowe zastosowania testów subiektywnych – ocena:

- jakości sprzętu audio (kolumny głośnikowe, odtwarzacze),
- jakości dźwięku w pomieszczeniu,
- jakości transmisji mowy lub muzyki,
- jakości kodowania mowy/muzyki (np. kompresja stratna),
- wyników działania algorytmu (np. redukcja szumu, rekonstrukcja starych nagrań).

W tych (i podobnych) przypadkach testy subiektywne są jedyną możliwością uzyskania wiarygodnej oceny jakości sygnału dźwiękowego.

Etapy testu subiektywnego

- Przygotowanie testu:
 - procedura testowa,
 - zestaw sygnałów testowych.
- Zgromadzenie grupy słuchaczy, ich weryfikacja.
- Trening – zapoznanie z procedurą testową.
- Testy – każdy słuchacz dokonuje oceny.
- Analiza wyników:
 - weryfikacja słuchaczy (po teście),
 - weryfikacja postawionej tezy: przyjęcie / odrzucenie.
- Opracowanie raportu.

Wiarygodność wyników testu

Czynniki wpływające na wiarygodność wyników testu:

- wiarygodna grupa słuchaczy,
- dobór sygnałów testowych (muszą być reprezentatywne dla badanego przypadku),
- wybór procedury testowej: czas trwania, metoda prezentacji sygnałów, sposób oceniania,
- warunki w pomieszczeniu testowym (brak zakłóceń),
- jakość sprzętu (głośniki, słuchawki),
- usytuowanie głośników i słuchacza w pomieszczeniu,
- poziom głośności prezentowanych sygnałów,
- poprawna analiza statystyczna i weryfikacja wyników.

Dobór grupy słuchaczy

Eksperti czy zwykli słuchacze?

Zwykły słuchacz:

- reprezentuje „statystycznego użytkownika” oraz „docelowego konsumenta”,
- można wykonywać testy preferencji: „A czy B”,
- nie ma doświadczenia w ocenianiu jakości, nie wie jak przydzielać oceny („4,4 czy 4,5?”),
- istnieje duże ryzyko, że wyniki uzyskane od takich słuchaczy zostaną odrzucone na etapie weryfikacji po teście – czyli nie mamy wyników.

Dobór grupy słuchaczy

Eksperci czy zwykli słuchacze?

Ekspert:

- osoba potrafiąca „wychwytywać” nawet niewielkie zniekształcenia i różnice między sygnałami,
- posiada doświadczenie w ocenianiu sygnałów,
- ma prawidłowy słuch (kontrola audiometryczna),
- udziela wiarygodnych i powtarzalnych ocen.

Zaleca się przeprowadzanie testów z **grupą ekspertów**.

Najczęściej wystarcza uzyskać wyniki od **10-20 ekspertów**.

Jeżeli słuchacze nie są ekspertami: co najmniej 20 osób.

Weryfikacja słuchaczy

Weryfikacja – eliminowanie niewiarygodnych słuchaczy.

Przed testem – na podstawie:

- wyników wstępnych testów (np. podczas treningu)
 - wychwycenie różnic zauważalnych dla ekspertów, a nie zauważalnych dla zwykłego słuchacza,
- badań audiometrycznych,
- doświadczenia w testach subiektywnych.

Po teście – na podstawie analizy wyników, np.:

- wyniki znacząco różne od średniej,
- niewiarygodne wyniki, np. nie powtarzalne, lub sygnały zniekształcone oceniane wyżej niż oryginalne.

Skala ocen

Typowa skala ocen do oceny jakości lub zniekształceń:
od **1** (najgorsza jakość) do **5** (najlepsza jakość).

UWAGA: skala jest ciągła. Można dać ocenę np. 4,31.

Poniższe etykiety są tylko punktami odniesienia.

Ocena	Jakość	Zniekształcenia
5	Doskonała	Niestłyszalne
4	Dobra	Stłyszalne, nie dokuczliwe
3	Średnia	Lekko dokuczliwe
2	Słaba	Dokuczliwe
1	Zła	Bardzo dokuczliwe

Skala ocen

Skala ocen **względnych**, stosowana w celu porównania jakości jednego sygnału względem innego. Również jest to ciągła skala.

Ocena	Porównanie jakości
3	Znacznie lepsza
2	Lepsza
1	Nieco lepsza
0	Taka sama
-1	Nieco gorsza
-2	Gorsza
-3	Znacznie gorsza

Faza treningu

Przed właściwym testem wykonuje się **trening**.

- **Objaśnienie** procedury testowej (co ma być oceniane, jak będą prezentowane sygnały, w jaki sposób mają być oceniane, itp.) – najlepiej bezpośrednio (ustnie), może być też pisemna lub odtwarzana instrukcja.
- **Demonstracja** procedury testowej – przykładowe sygnały (nie wykorzystane później w teście, ale prezentowane w ten sam sposób). Można wskazać które sygnały są oryginalne, a które zniekształcone. Można też wykorzystać fazę treningu do **weryfikacji** słuchacza.

Konieczna jest przerwa między fazą treningu a zasadniczym testem.

Procedura testowa

Test może być przeprowadzany w różny sposób.

- Wszyscy słuchacze oceniają jednocześnie:
 - o ile warunki i sposób przeprowadzania testu na to pozwalają,
 - uwaga na warunki – np. miejsce w sali może wpływać na ocenę.
- Każdy słuchacz ocenia osobno. Jedna ciągła sesja na słuchacza:
 - znacznie dłuższy czas testów,
 - warunki muszą być jednakowe,
 - słuchacze mogą wykonać test w dostępnym terminie.

Procedura testowa

- Czas trwania sesji testowej nie powinien być zbyt długi (zmęczenie słuchacza).
- Maksymalny czas trwania sesji: **20–30 minut**.
- Maksymalnie **10–15 ocenianych sygnałów** w sesji.
- Jeżeli potrzebna jest więcej niż jedna sesja: przerwa między sesjami musi trwać przynajmniej tyle, co jedna sesja.
- Dla testów wykonywanych indywidualnie: kolejność prezentowanych sygnałów powinna być losowa (kolejność sygnałów nie powinna mieć wpływu na ich ocenę).

Prezentacja sekwencji testowych

Sygnały testowe mogą być prezentowane w sposób:

- **automatyczny** (narzucony z góry)
 - każda oceniana sekwencja powinna być powtórzona,
 - konieczne przerwy na dokonanie i wpisanie oceny,
 - metoda stosowana zwykle przy grupie słuchaczy oceniającej sygnały w tym samym czasie.
- **interaktywny**
 - słuchacz sam decyduje o liczbie powtórzeń,
 - najczęściej „badamy” tą metodą jednego słuchacza w danej chwili,
 - dłuższy test, ale wyniki mogą być bardziej wiarygodne.

Prezentacja sekwencji testowych

Sposoby prezentacji sygnałów testowych i ich oceny:

- pojedynczy sygnał – ocena jego jakości (problem punktu odniesienia, więc należy unikać takiej metody),
- porównanie parami – ocena różnicy między dwoma sygnałami,
- porównanie sygnału ocenianego z referencyjnym,
- porównania z wykorzystaniem ukrytego sygnału referencyjnego (np. test ABC),
- porównanie zestawu ocenianych sygnałów z sygnałem referencyjnym (np. test MUSHRA).

Przygotowanie sekwencji testowych

- **Sygnal referencyjny** (*reference*) – sygnał „oryginalny” (nie zniekształcony). Punkt odniesienia dla oceny badanych sygnałów. Np. „czysty” sygnał z płyty CD.
- Sygnały testowe (oceniane) muszą być tak dobrane, aby możliwe było uzyskanie żądanej informacji.
- Nie można dobierać sygnałów testowych w sposób „tendencyjny”, aby udowodnić tezę.
- Pojedynczy sygnał nie może trwać dłużej niż **10-20 sekund**.
- Sygnały (zwłaszcza muzyka i mowa) nie mogą się gwałtownie urywać (musi być np. wyciszenie).
- Zwykle wybiera się fragment ze środka utworu.

Przygotowanie sekwencji testowych

- Między sygnałem referencyjnym a ocenianym: przerwa ok. 0,5–1 sekundy.
- Między powtórzeniami sekwencji „referencyjny – oceniane”: dłuższa przerwa, ok. 1–1,5 sekundy.
- Jeżeli prezentowany jest ten sam sygnał z różnymi zniekształceniami – nie pod rząd.
- Kolejność sekwencji testowych powinna być losowa, różna dla kolejnych słuchaczy (eliminacja wpływu zmęczenia na wyniki testu).
- „Treść” sygnałów powinna być neutralna, tak aby preferencje słuchacza (np. gatunek muzyczny) nie wpływały na ocenę.

Oceniwane atrybuty

- Ocenie podlegają parametry sygnału – **atomybuty**.
- Podstawowy atrybut:
 - **jakość** sygnału (*basic audio quality*), lub
 - **zniekształcenia** (*distortions*).
- Inne atrybuty mogą być definiowane w zależności od testu. Np. test pomieszczenia – oceniane atrybuty mogą dotyczyć: zrozumiałości mowy, klarowności sygnału, sceny stereofonicznej, zawartości pogłosu, itp.
- Nie należy przesadzać z liczbą atrybutów – utrudnia to słuchaczom ocenę.

Testy oceny bezwzględnej

Testy oceny bezwzględnej:

- prezentujemy sygnały testowe,
- słuchacz musi ocenić jakość sygnału,
- słuchacz musi sam znaleźć „punkt odniesienia”,
- np. ocena jakości dźwięku w pomieszczeniu,
- udział ekspertów bezwzględnie wymagany, nie uzyskamy dobrych wyników od zwykłych słuchaczy – nie potrafią oni przydzielać ocen nie mając podanego sygnału odniesienia,
- takie podejście stosujemy w ostateczności, jeżeli nie można zastosować innej metody.

Testy oceny względnej

Ocena **względna**:

- prezentowane są pary: sygnał referencyjny i oceniany,
- oceniana jest różnica między sygnałem referencyjnym a ocenianym (zniekształcenie),
- prostsze do oceny – sygnał odniesienia jest podany wprost,
- nadaje się do testów porównawczych: „A czy B?”,
- można stosować dla zwykłych słuchaczy („kolegów studentów”), ale trzeba zwiększyć liczbę słuchaczy i dokładnie sprawdzić ich wiarygodność.

Wykrywanie zniekształceń - test AB

- **Zniekształcenie** (*impairment*) – różnica między sygnałem referencyjnym a ocenianym.
- Test AB: prezentujemy słuchaczom sygnał referencyjny, a następnie dwa sygnały: A i B.
- Jeden z sygnałów A/B jest powtórzonym sygnałem referencyjnym, drugi – sygnałem ze zniekształceniami. Wybór jest zwykle losowy.
- Zadaniem słuchacza jest wskazanie sygnału referencyjnego: A czy B.
- Można powtórzyć każdą sekwencję, odwracając kolejność A/B. W ten sposób sprawdzamy czy słuchacz daje powtarzalne oceny.

Ocena zniekształceń - test ABC

- Test **ABC (BS.1116)** służy do oceny zniekształceń, przede wszystkim gdy stopień zniekształceń jest niewielki.
- Prezentujemy słuchaczom trójki sygnałów:
 - **A**: zawsze sygnał referencyjny (nie oceniany),
 - **B**: pierwszy sygnał oceniany względem A,
 - **C**: drugi sygnał oceniany względem A.
- Jeden z sygnałów: B lub C, jest powtórzonym sygnałem referencyjnym, ale słuchacz nie wie który z nich.
- Słuchacz ocenia sygnały B i C.
- Przy indywidualnym teście: słuchacz może odsłuchiwać sygnały dowolną liczbę razy.

Ocena zniekształceń - test ABC

- Spodziewamy się, że sygnał referencyjny uzyska najlepszą ocenę (brak zniekształceń).
- Oceny przydzielane sygnałom referencyjnym mogą być wykorzystywane do weryfikacji słuchaczy po teście.
- Jeżeli słuchacz uporczywie „słyszy” różnice między referencyjnym B lub C a referencyjnym A, to znaczy, że nie jest wiarygodny i należy go wykluczyć z analizy.
- Można też powtórzyć każdy przykład (lub wybrane), odwracając kolejność B/C. W ten sposób sprawdzimy, czy słuchacz udziela powtarzalnych odpowiedzi.

Ocena zniekształceń - test MUSHRA

- Metoda ABC zawodzi w przypadku silnie zniekształconych sygnałów – wszystkie oceny są niskie, a my chcemy zbadać różnice między kilkoma badanymi przypadkami.
- W tym przypadku stosuje się test **MUSHRA (BS.1534)**.
- Słuchacz odsłuchuje grupy sygnałów – może porównywać między sobą sygnały o różnym stopniu zniekształcenia.
- Metoda stosowana dla pojedynczych słuchaczy.
- Słuchacz ma do dyspozycji:
 - sygnał referencyjny,
 - grupę sygnałów, które ma ocenić.

Ocena zniekształceń - test MUSHRA

Grupa sygnałów ocenianych przez słuchacza zawiera:

- sygnał **referencyjny** (ukryty), powinien uzyskać **najlepszą** (maksymalną) ocenę,
- **oceniane** sygnały (różne stopnie zniekształceń),
- punkty **zaczepienia** (*anchors*): sztucznie zniekształcone sygnały (typowo: filtracja dolnoprzepustowa 3,5 kHz i 7 kHz); te sygnały powinny uzyskać **najniższe** oceny.

Słuchacz ocenia **każdy** sygnał na skali od **0** (najgorsza jakość) do **100** (najlepsza jakość), z krokiem 1.

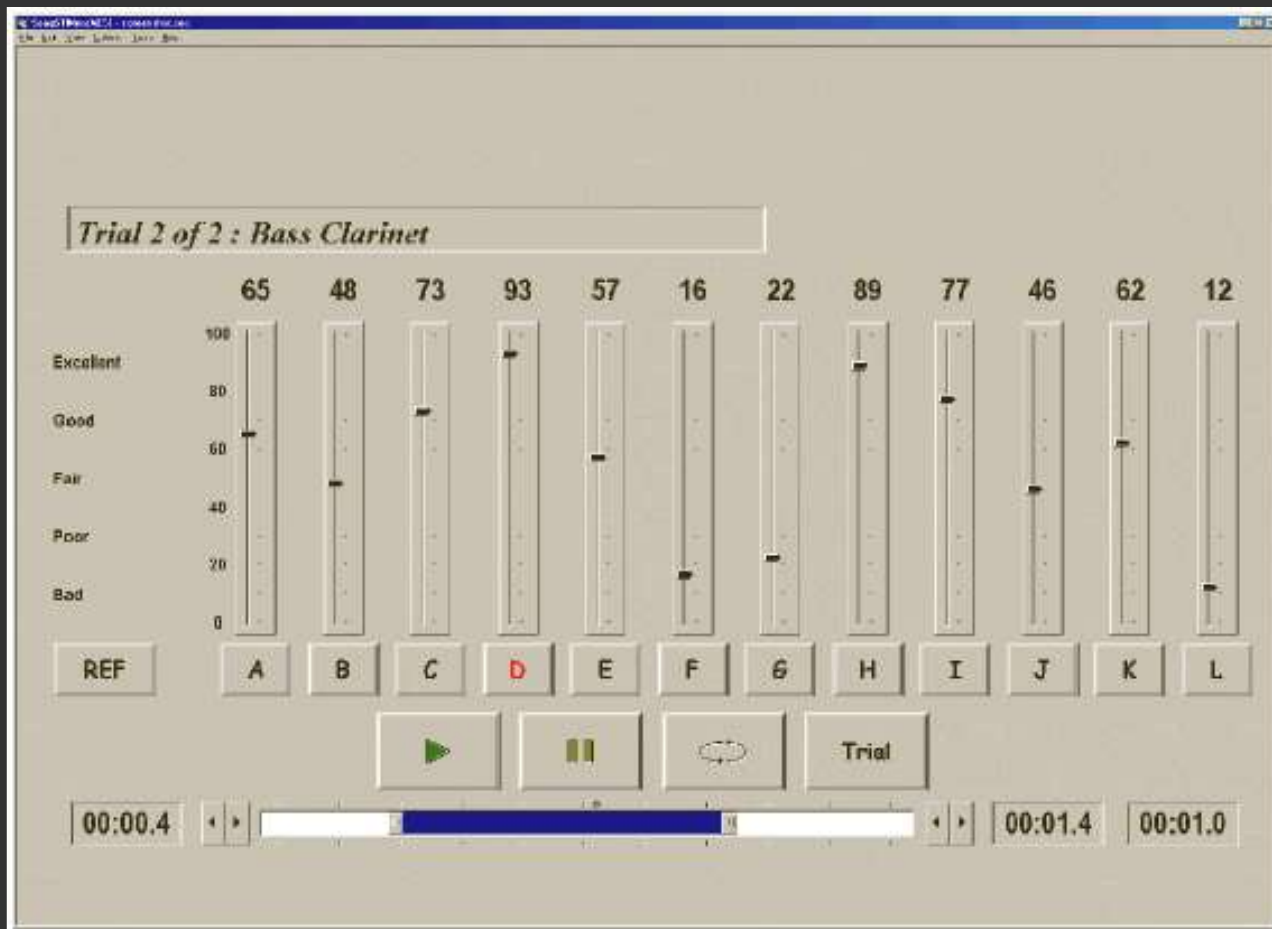
Jeżeli słuchacz nie potrafi wskazać sygnału referencyjnego lub punktów zaczepienia – nie jest wiarygodny.

Ocena zniekształceń - test MUSHRA

- Test MUSHRA przeprowadza się w trybie indywidualnym.
- Słuchacz powinien mieć możliwość odtworzenia każdego sygnału w ramach grupy dowolną liczbę razy.
- Sekwencja grup sygnałów oraz kolejność sygnałów w grupie powinny być losowe (nie powtarzać się dla kolejnych słuchaczy).
- Grupa powinna zawierać nie więcej niż 15 sygnałów, wliczając referencyjny i punkty zaczepienia.
- Test może składać się z kilku grup. Należy stosować ogólne zalecenia co do czasu trwania sesji testowej.

Test MUSHRA

Przykładowy panel służący do oceny sygnałów. Słuchacz nadaje oceny za pomocą suwaków.



Analiza wyników testów

- Mamy wyniki dla grup sygnałów, np. dla referencyjnych i dla badanych.
- Chcemy wykazać, że są różnice ocen w tych grupach, albo że nie ma różnic (zależnie od celu testu).
- Możemy policzyć **wartość średnią** i **odchylenie standardowe** ocen w obu grupach.
- Widzimy, że są pewne różnice między grupami.
- Pytanie: **czy te różnice są istotne?**
- Na to pytanie odpowiadają **testy statystyczne**.
- Wykonujemy testy dla **zmiennych niezależnych** (*independent variables*).

Podstawy testów statystycznych

- Każdy typ testu statystycznego ma zdefiniowaną **hipotezę zerową H_0** , którą testuje. Np.: „nie ma istotnych różnic między wynikami w dwóch grupach”.
- Wynikiem testu są wartości:
 - **wynik** testu (liczba specyficzna dla testu),
 - **przedział ufności** (*confidence interval*, CI): przedział, w którym mieści się zadany odsetek wyników (typowo: 95%),
 - **wartość p** (*p-value*): prawdopodobieństwo wystąpienia H_0 .

Podstawy testów statystycznych

- Dla wartości p przyjmuje się **poziom istotności** (*significance level*), typowo: 0,05 (czasem 0,01).
- Jeżeli $p > 0.05$: nie możemy odrzucić H_0 (pozostaje ona w mocy).
- Jeżeli $p < 0.05$: odrzucamy H_0 i przyjmujemy **hipotezę alternatywną H_1** , zdefiniowaną przez typ testu.
- Zależnie od naszego wyboru, H_1 może być:
 - **dwustronna** (*two-tailed*), np. „są istotne różnice między grupami”,
 - **jednostronna** (*one-tailed*), np. „wyniki w grupie pierwszej są istotnie wyższe” (**prawostronna**) lub „są niższe” (**lewostronna**).

Test dwumianowy

- Rozpatrzmy przykład testu AB. Słuchacze mają wskazać który z sygnałów – A czy B – jest referencyjny.
- Wyniki tworzą rozkład dwumianowy: poprawne (1) lub niepoprawne (0) wskazanie referencji.
- Stosujemy **test dwumianowy** (*binomial test*).
- H_0 : prawdopodobieństwo poprawnego wyniku (1) jest równe 50%, czyli wyniki są czysto losowe.
- Wychodzi nam $p = 0,0024$.
- Mamy $p < 0,05$, więc odrzucamy H_0 i przyjmujemy H_1 .
- Jeżeli wykonamy test dla prawostronnej H_1 , to przyjmujemy H_1 : „prawdopodobieństwo wskazania referencji jest większe niż 50%”.

Porównanie ocen w dwóch grupach

- Załóżmy, że chcemy porównać oceny w dwóch grupach, np. sygnały referencyjne i oceniane w teście ABC.
- Wybór testu statystycznego będzie zależał od tego, czy wyniki (próbki) pochodzą z **rozkładu normalnego** (gaussowskiego).
- Jeżeli tak, możemy stosować **testy parametryczne**.
- Jeżeli nie, musimy stosować **testy nieparametryczne**, które mają mniejszą moc niż parametryczne (co znaczy, że ich wyniki są mniej wiarygodne, zwłaszcza gdy p jest bliskie 0,05).
- Nie należy stosować testów parametrycznych bez sprawdzenia czy rozkład jest normalny.

Test Shapiro-Wilka

- Istnieje kilka testów statystycznych sprawdzających czy wartości pochodzą z rozkładu normalnego.
- Test **Shapiro-Wilka** jest jednym z nich.
- H_0 : **wartości pochodzą z rozkładu normalnego.**
- Jeżeli $p > 0,05$: możemy pozostawić H_0 i stosować testy parametryczne.
- Jeżeli $p < 0.05$, to niestety, ale musimy odrzucić H_0 i pozostają nam testy nieparametryczne.

Test t-Studenta

- Test **t-Studenta** (*t-test*) dla zmiennych niezależnych porównuje **wartości średnie** w dwóch grupach.
- Jest to test **parametryczny**: wyniki muszą pochodzić z rozkładu normalnego.
- Można go zastosować np. do wyników testu ABC.
- H_0 : **średnie w obu grupach są identyczne** (nie ma istotnych różnic wyników między grupami).
- Jeżeli odrzucamy H_0 ($p < 0,05$), to przyjmujemy H_1 :
 - dwustronną (wyniki w dwóch grupach są różne),
 - prawostronną (wyniki w 1. grupie są wyższe), lub
 - lewostronną (wyniki w 1. grupie są niższe).

Test U Manna-Whitneya

- Test **U Manna-Whitneya** jest nieparametrycznym odpowiednikiem testu t-Studenta.
- Nie warto go stosować, jeżeli wyniki pochodzą z rozkładu normalnego (ma mniejszą moc).
- Również można go zastosować do wyników testu ABC.
- H_0 : **rozkład wartości w obu grupach jest identyczny** (nie ma istotnych różnic wyników między grupami).
- Jeżeli odrzucamy H_0 ($p < 0,05$), to przyjmujemy H_1 :
 - dwustronną (wyniki w dwóch grupach są różne),
 - prawostronną (wyniki w 1. grupie są wyższe), lub
 - lewostronną (wyniki w 1. grupie są niższe).

Testy dla wielu grup wyników

- Jeżeli chcemy porównać kilka grup wyników (np. z testu MUSHRA), musimy sięgnąć po testy wielu zmiennych.
- Test **ANOVA** jest testem **parametrycznym**. Wymaga rozkładu normalnego i **równości wariancji** dla wszystkich badanych grup.
- Test **Kruskala-Wallisa** jest testem nieparametrycznym.
- H_0 : **brak istotnych różnic między grupami** (ANOVA: jednakowe średnie; K-W: jednakowe mediany).
- Jeżeli odrzucimy H_0 ($p < 0,05$), to przyjmujemy H_1 (tylko dwustronną): różnice między grupami są istotne.

Testy post hoc

- Odrzucenie H_0 w teście ANOVA lub K-W daje tylko informację, że są różnice między grupami.
- Nie wiemy jednak jaki jest charakter tych różnic, np. czy wszystkie grupy się różnią między sobą, czy tylko jedna grupa jest różna od pozostałych.
- Nie należy wykonywać testów t-Studenta lub M-W dla par „każdy z każdym” – nie da to poprawnych wyników.
- Do tego celu służą specjalne *testy post hoc*, np. *test Dunna*.
- Testy *post hoc* mają małą moc statystyczną – wyniki należy traktować z pewną rezerwą.

Narzędzia do analizy statystycznej

- Profesjonalne oprogramowanie do analiz statystycznych, np. *Statistica*, SPSS, PAST.
- Język *Python* i procedury z biblioteki *scipy.stats*,
- Język skryptowy *R*, opracowany z myślą o analizach statystycznych,
- *MATLAB* z pakietem *Statistics and Machine Learning Toolbox*.

Raport z testów subiektywnych

Przykładowa zawartość raportu:

- cel przeprowadzenia testu,
- charakterystyka słuchaczy i sygnałów testowych,
- dokładny opis procedury testowej, warunków,
- opis sposobu analizy wyników,
- końcowe wyniki (w formie wykresów i tabel),
- dokładniejszy opis wyników (jeżeli chcemy zamieścić szczegółowe analizy, powinny być one w załączniku),
- wnioski wynikające z uzyskanych wyników.

Raport z testów subiektywnych

Opracowanie wyników analizy statystycznej powinno zawierać:

- liczbę osób uwzględnionych w analizie,
- wartości średnie i odchylenia standardowe dla poszczególnych osób oraz sygnałów,
- typ przeprowadzonego testu statystycznego,
- przyjętą hipotezę (np. „stwierdzono istotnie wyższe oceny dla badanych sygnałów względem referencyjnych”),
- wynik testu, przedział ufności, (obowiązkowo) wartość p ,
- jeżeli p jest bardzo małą liczbą, wystarczy napisać np. $p < 0,01$.

Wykres pudełkowy

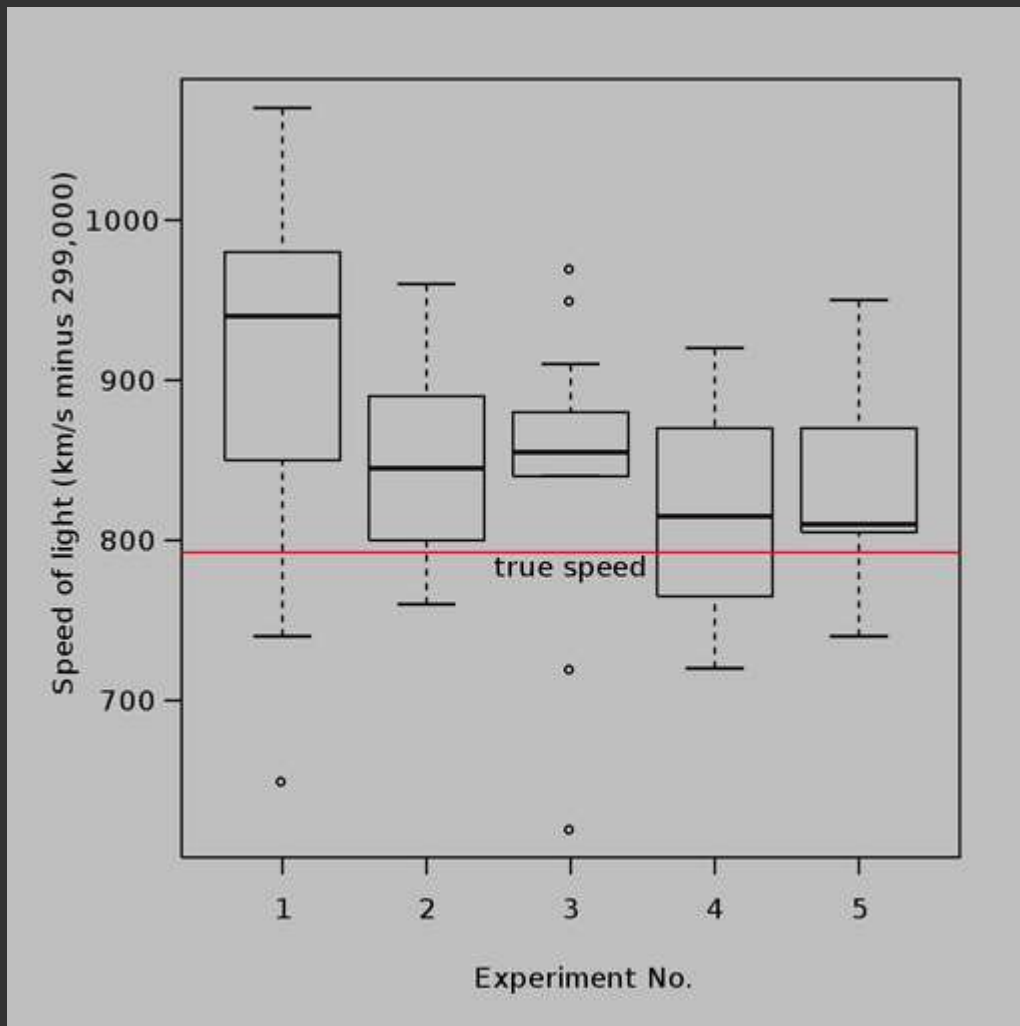
Wykres pudełkowy (*box plot*, „*box & whiskers*”) przedstawia graficznie rozkład wartości zmiennej.

- Krawędzie pudełka: pierwszy kwartył Q1 (25% wartości jest mniejszych niż Q1) i trzeci kwartył Q3 (75% wartości jest mniejszych niż Q3).
- Linia w środku pudełka: mediana Q2 (wartość środkowa).
- „Wąsy” przedstawiają minimalną i maksymalną wartość.
- Wartości ewidentnie odstające (*outliers*) pokazuje się jako punkty i nie wlicza się ich do wąsów.
- Na jednym wykresie można pokazać kilka zmiennych.

Matplotlib: funkcja *boxplot*.

Wykres pudełkowy

Źródło: *Wikipedia* (wykres pudełkowy).



Problemy związane z testami subiektywnymi

Badanie jakości sygnałów za pomocą testów subiektywnych napotyka na szereg trudności:

- konieczność zgromadzenia grupy ekspertów,
- długi czas badania (i koszt),
- zmęczenie słuchaczy i czynniki indywidualne wpływające na wyniki,
- konieczność analizy statystycznej wyników w celu uzyskania wiarygodnych ocen,
- problem powtarzalności wyników.

Obiektywizacja testów subiektywnych

- Obiektywizacja polega na zastąpieniu grupy słuchaczy przez program komputerowy: „test zobjektywizowany”.
- **Algorytm perceptualny** ocenia jakość sygnału w taki sposób, w jaki robi to człowiek.
- Założenie: wysoka zgodność (korelacja) wyników testu „komputerowego” z wynikami testów odsłuchowych.
- Zalety:
 - skrócenie czasu testów,
 - wyeliminowanie konieczności pracy z grupą słuchaczy,
 - powtarzalność wyników.
- Przykład oprogramowania: *Opticom Opera*.

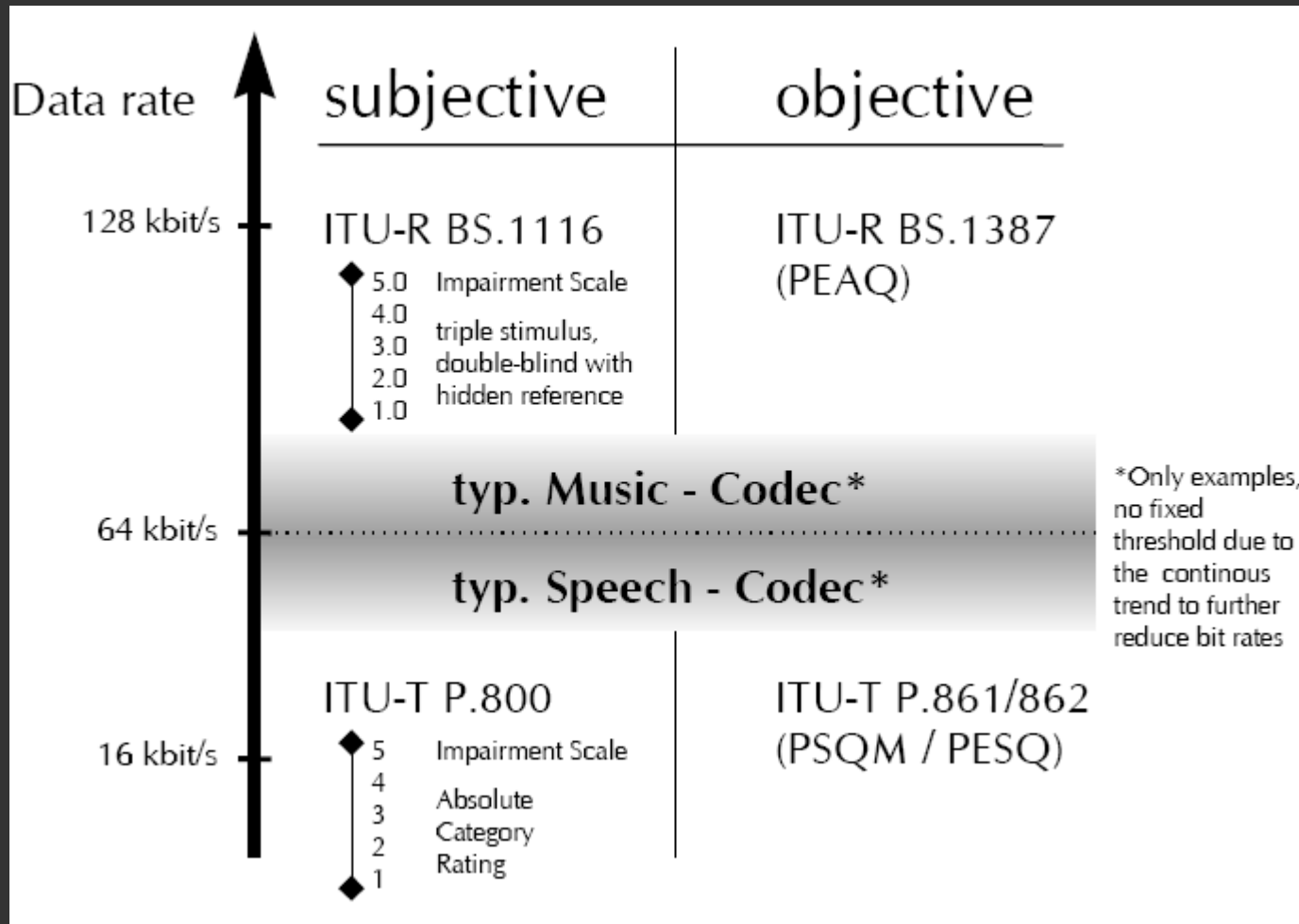
Komputerowe testy subiektywne

Najważniejsze zobiektywizowane testy używane do oceny jakości dźwięku:

- *PESQ – Perceptual Evaluation of Speech Quality*
– badanie jakości sygnału mowy, uwzględnia pakietową transmisję danych;
- *PEAQ – Perceptual Evaluation of Audio Quality*
– badanie jakości sygnałów szerokopasmowych (muzycznych).

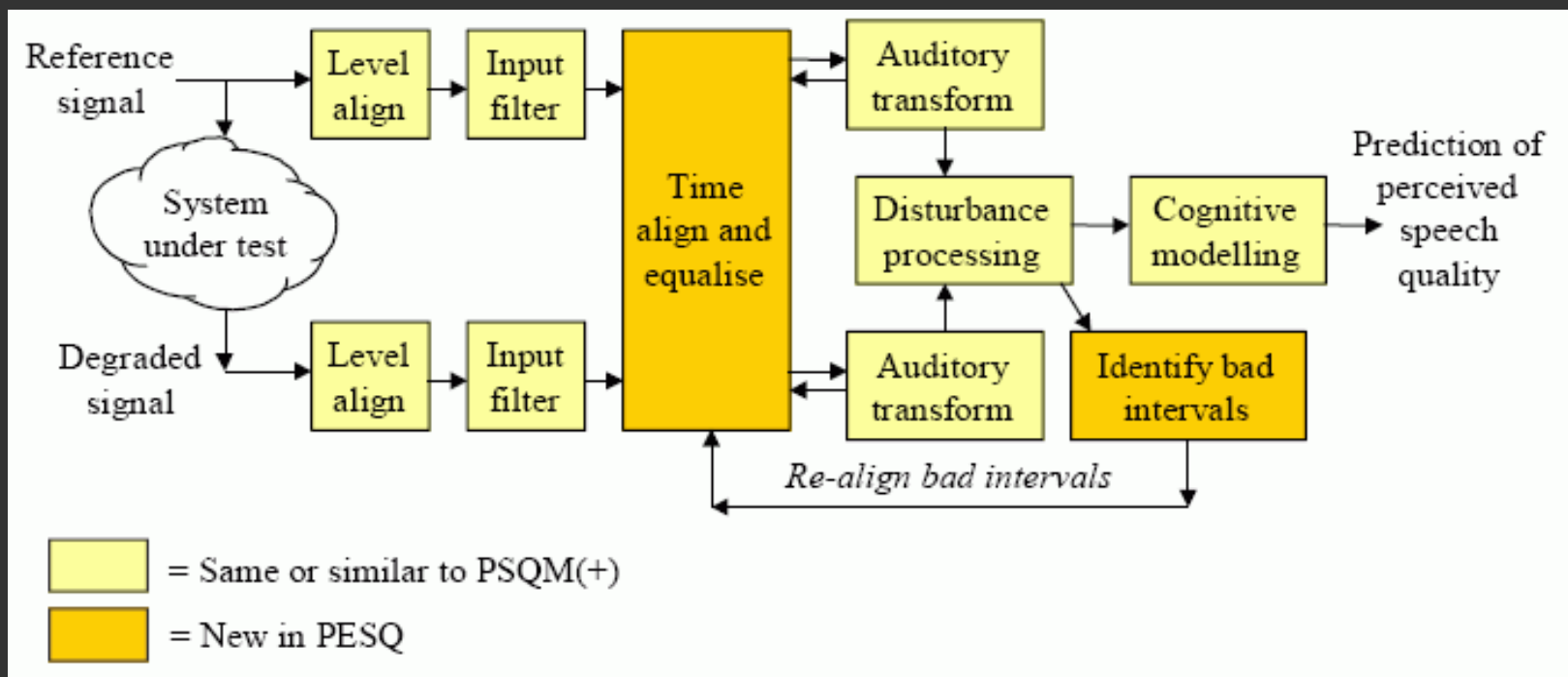
Komputerowe testy subiektywne

Przykład: testowanie kodeków muzyki i mowy.



PESQ

Test PESQ służy do badania jakości mowy, a szczególnie zniekształceń w wyniku transmisji i kodowania (kompresji). Uwzględnia zmienne opóźnienie transmisji (*jitter*), np. w sieciach VoIP.



PESQ

- Wynik testu PESQ: miara **MOS** (*Mean Opinion Score*) w skali od **1.0** (najgorsza jakość) do **4.5** (najlepsza jakość).
- Najlepsza jakość w skali MOS sięga wartości 5.0, ale taki wynik nie jest nigdy osiągnięty w analizie statystycznej wyników testów subiektywnych.



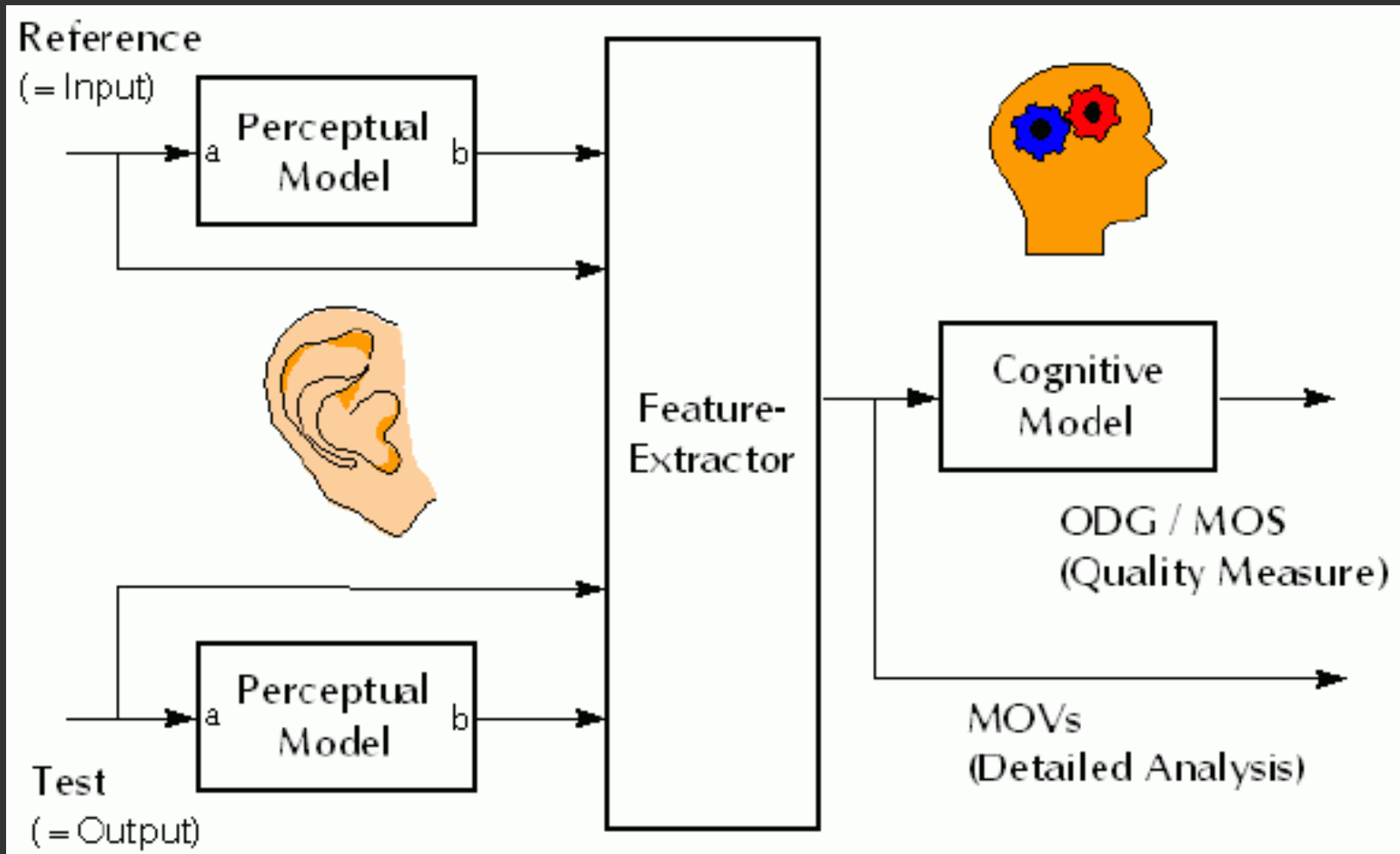
PEAQ

- **PEAQ** - ocena jakości dźwięku szerokopasmowego (np. sygnałów muzycznych).
- Istnieją dwie wersje testu.
 - *PEAQ Basic* - uproszczona wersja, mniej dokładny test, ale analiza może być przeprowadzana w czasie rzeczywistym.
 - *PEAQ Advanced* - wymaga bardziej skomplikowanych obliczeń i dłuższego czasu analizy, ale daje dokładniejsze wyniki
- Struktura obu wersji jest podobna. Wersje różnią się złożonością modelu psychoakustycznego.

PEAQ

- PEAQ porównuje sygnał referencyjny (oryginalny) z sygnałem zniekształconym.
- Na podstawie porównania sygnałów uzyskuje się wektor cech – **MOV (*Model Output Variable*)**. Zmienne MOV opisują poszczególne parametry, które mają wpływ na jakość dźwięku (wielkość zniekształceń, modulacja, itp.).
- **ODG (*Objective Difference Grade*)** – różnica jakości pomiędzy dźwiękiem badanym a referencyjnym, skala od **-4.0** (zniekształcenia bardzo dokuczliwe) do **0.0** (zniekształcenia niesłyszalne).
- **DI (*Distortion Index*)** – inna miara zniekształceń.

PEAQ



Bibliografia

Normy ITU (www.itu.int) dotyczące testów subiektywnych:

- ITU-R BS.1283-1: wykaz standardów dot. testów subiektywnych
- ITU-R BS.1284: ogólne metody subiektywnej oceny jakości dźwięku
- ITU-R BS.1116: ocena małych zniekształceń dźwięku (test ABC)
- ITU-R BS.1534: ocena jakości w systemach kodowania (MUSHRA)
- ITU-R BS.1285: subiektywne testy preselekcyjne
- ITU-R BS.1286: ocena jakości dźwięku z towarzyszącym obrazem
- ITU-T P.800: subiektywna ocena jakości transmitowanego sygnału
- ITU-T P.861: test PSQM
- ITU-T P.862: test PESQ
- ITU-R BS.1387: test PEAQ