

# Objektív beszédminősítés

Fegyó Tibor  
fegyo@tmit.bme.hu

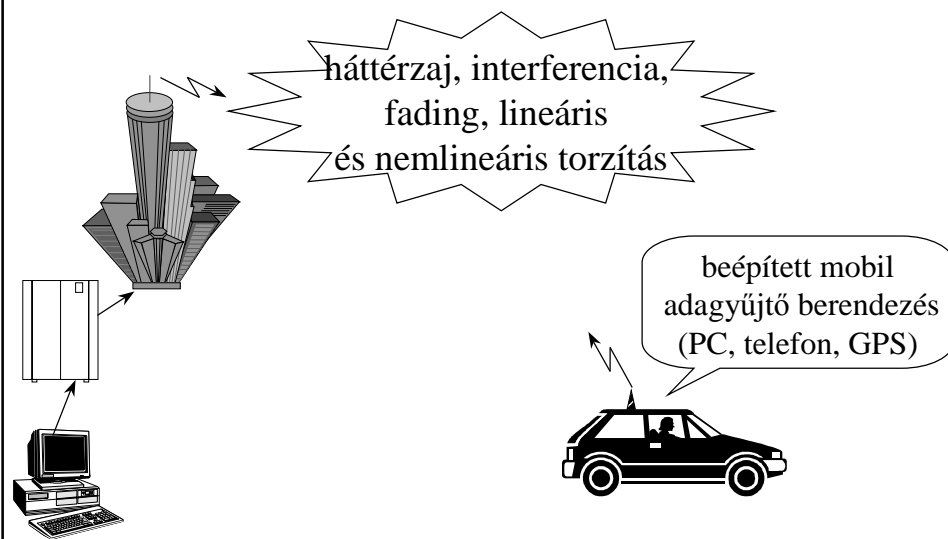
## Bevezető kérdések

- Mi a [beszéd] minőség [a beszédkommunikációban]?
- Mi befolyásolja a minőséget?
- Lehet-e mérni szubjektív / objektív módon a minőséget?
- Mennyire megbízhatóak az értékelések, avagy mindig ugyanúgy értékelünk?
- Mi alapján értékeljük a beszédet?
- Mi rontja el a beszéd minőségét?
- Lehet-e befolyásolni / javítani a beszéd minőségét?

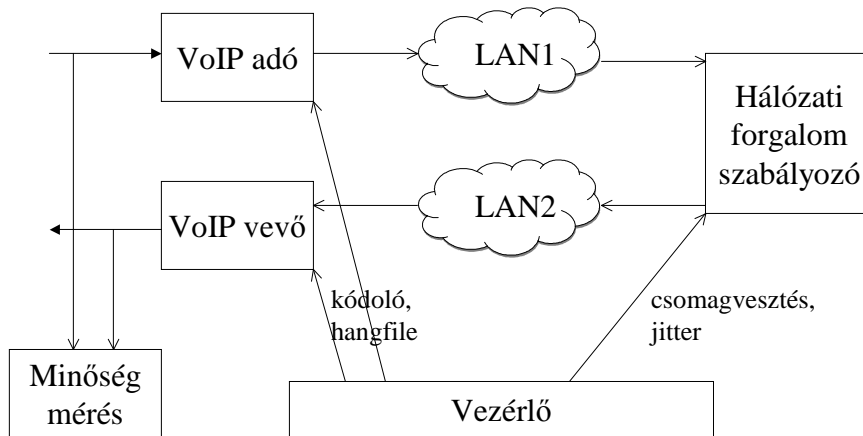
## Alkalmazási példák

- beszédminőség rendszeres mérése telefon csatornán
- nagyobb rekonstrukció után a minőség változásának kimutatása
- országos minőségterkép készítése

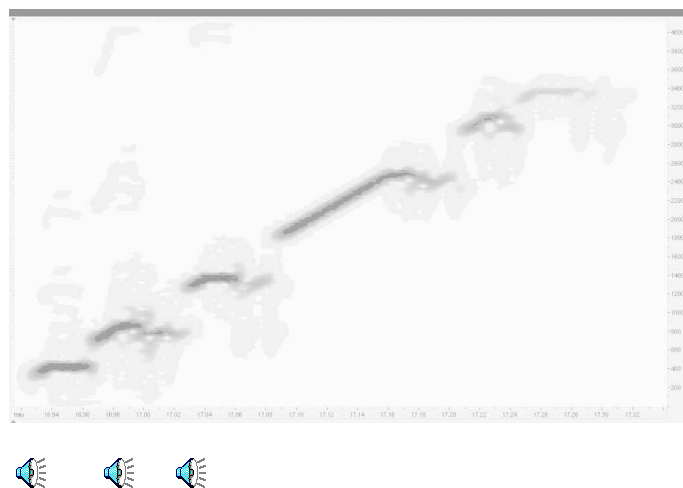
## Példa hangminták gyűjtésére mobil hálózat



## Példa hangminták gyűjtésére VoIP hálózat



## Példa torzított felvételekre



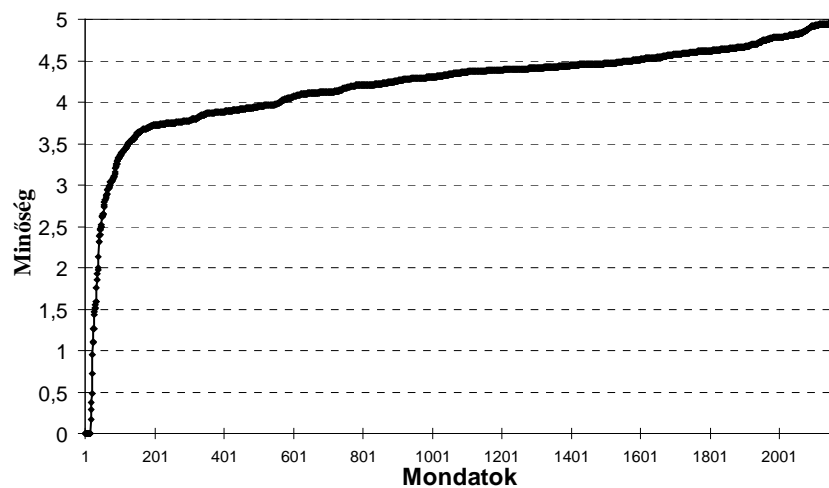
## Szubjektív beszédminősítés

- Szabvány ITU-T P.800
  - abszolút értékelés előre definiált skála alapján (ACR)
  - ‘jelenség’ észlelési tesztek
  - romlás megfigyelése eredetihez képest (DCR)
  - referencia rendszerrel összehasonltás
    - pl. MNRU (Modulated Noise Reference Units P.81)
    - ‘jól’ meghatározott referenciák
- MOS (mean opinion score) sok ember véleményének átlaga

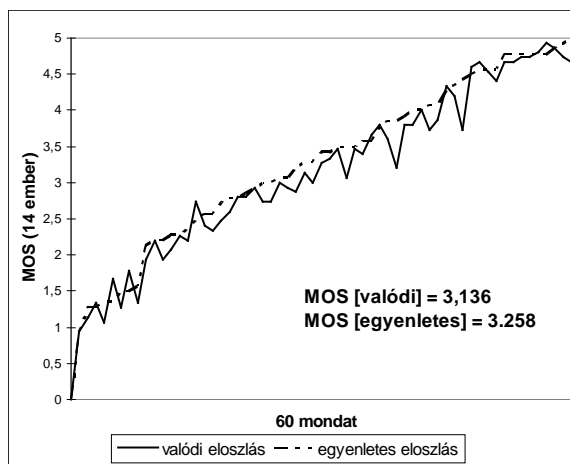
## Az abszolút értékelés problémája

- A kiválasztott tesztalmaz minőség eloszlásától függ a az egyes felvételek értékelése, mivel ugyanaz a felvétel jobb minőségű környezetben gyengébb osztályzatot kap, mint rosszabb minőségű környezetben.
- Egyenletes tesztalmazt könnyebb összeállítani, viszont a valóságban több a jó minőségű felvétel.

## Példa a minőség eloszlására egy valódi rendszer esetén



## A szubjektív vélemény változása a környezet függvényében



## Szubjektív minősítés szerepe

- pontos módszer, bár az eredmény függ a kísérlet körülményeitől
- tömeges mérésre nem alkalmas, mert drága és lassú
- az objektív minősítés kiindulási alapja, kalibrálásához szükséges

## Objektív (gépi) minősítés

- feladata az MOS lehető legjobb közelítése
- a hatékonyság mértékei:

- maximális korreláció MOS-el

$$\frac{1/n * \sum \{(x_i - m_x)(y_i - m_y)\}}{\sigma_x \sigma_y}$$

- minimális négyzetes eltérés MOS-tól

$$\frac{1}{n} * \sum (x_i - y_i)^2$$

## Felhasználási területek

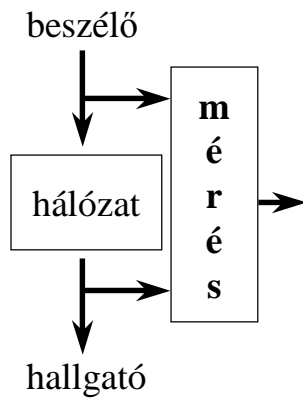
- Analóg beszédátvitel
  - vezetékes telefonok
  - mobil telefonok
- Digitális beszéd átvitel
  - kis sebességű kódolók, tömörítők (PSQM/ITU P.861, PESQ/ITU P.862)
    - tervezésük során folyamatos kontrollt biztosít
  - GSM kódoló (PSQM +)
    - kiválasztása során nagy mennyiségű szubjektív és objektív tesztet végeztek el
  - IP, ATM (csomag kapcsolt) beszéd átvitel, pl.: VoIP

## Minőség vs. sávzélesség

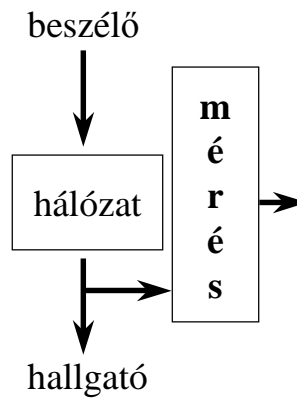
- Beszéd
  - Lineáris kódolás: 8kHz 16bit : 128 kbps
  - Telefon (nemlineáris: A/μ law) 8kHz 8bit (12bit) : 64kbps
  - VoIP (LPC, CELP...) 4-64 kbps
- Zene
  - CD minőség: 44,1kHz, 16bit (stereo) : 1411 kbps
  - MP3: 64-128-256-384 kbps
  - Realaudio: 32kbps
- Késleltetés ingadozásra (jitter) érzékeny a realtime média, függetlenül a sávzélességtől.

# Minősítési eljárások

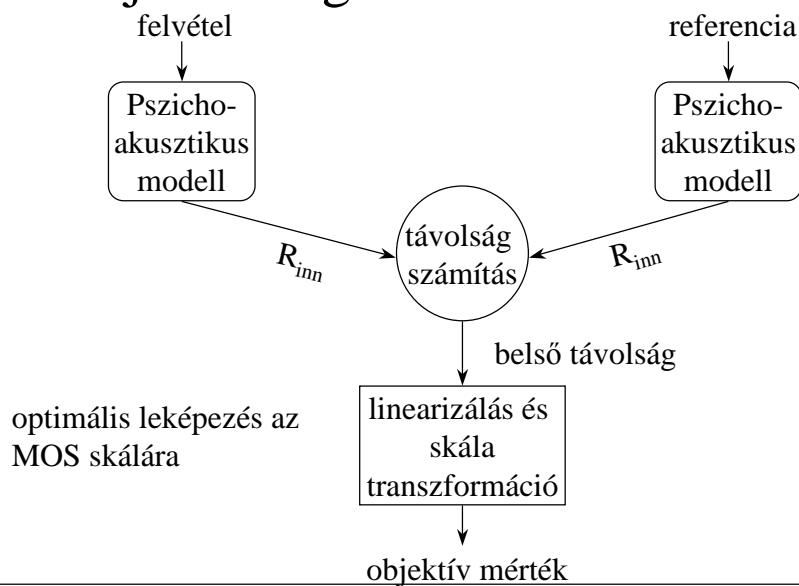
## Összehasonlítás alapú mérés



## Abszolút besorolás (beszéd --- mérőjel)



# A jelfeldolgozás általános menete





## Pszicho-akusztikus modell

- Idő - frekvencia leképezés
  - Keretekre vágás
    - rövid idejű (15 - 50 ms), átlapolódó (50%) keretek
  - Ablakozás
  - Fourier transzformáció
- Pszicho-akusztikus érzeti modellezés
  - Az emberi hallás modellezésén alapul, célja a hallható különbségek kiemelése, és a nem észlelhetőek elnyomása
  - monoton legyen a kapcsolat a belső távolság és az MOS között

## Pszicho-akusztikus modell elemei

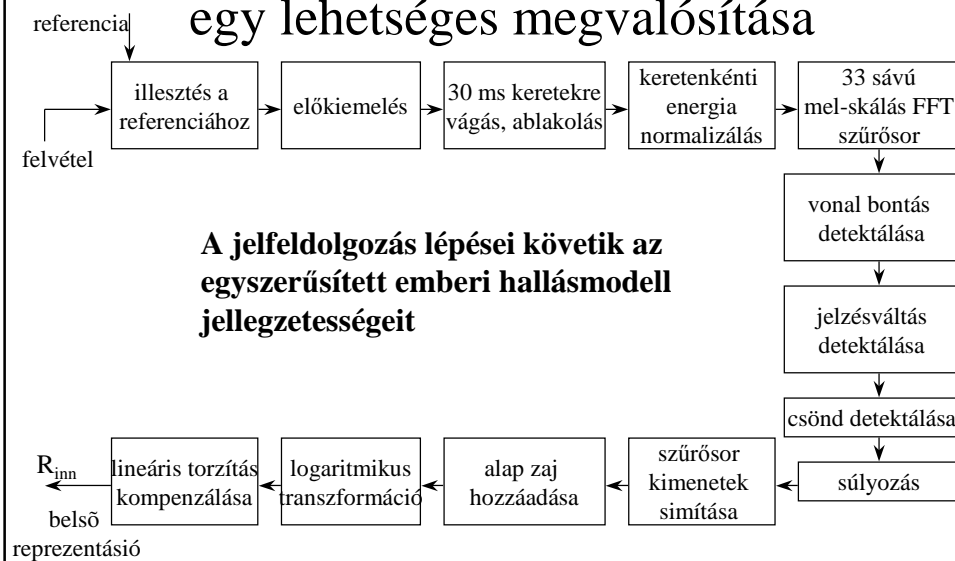
- Transzformálás az érzeti tartományra
  - nemlineáris frekvencia skálák (mel, bark...)
- Frekvencia elfedés
  - közeli frekvenciák esetén az erősebbik elnyomja a gyengébbet
- Időbeli elfedés
  - egymás utáni rövid impulzusokat egynek hallunk
  - egy erős hang elnyomja a környező gyengébbeket
- Pszicho-akusztikus hangosság
  - jel energia és hangosság kapcsolata nemlineáris

## Problémák a szabványtól eltérő alkalmazási területen (NMT)

Hagyományos (analóg méréseknél, ill. kis bitsebességű kódolóknál ajánlott) módszerek nem használhatók, mert:

- a csatorna nem stacionárius, nehezen modellezhető
- az eredeti hanganyag nem használható referenciaként
- a mérések nem reprodukálhatók az időben változó környezet miatt
- a jelzésváltás a hangcsatornán belül történik, de ezt nem kell figyelembe venni

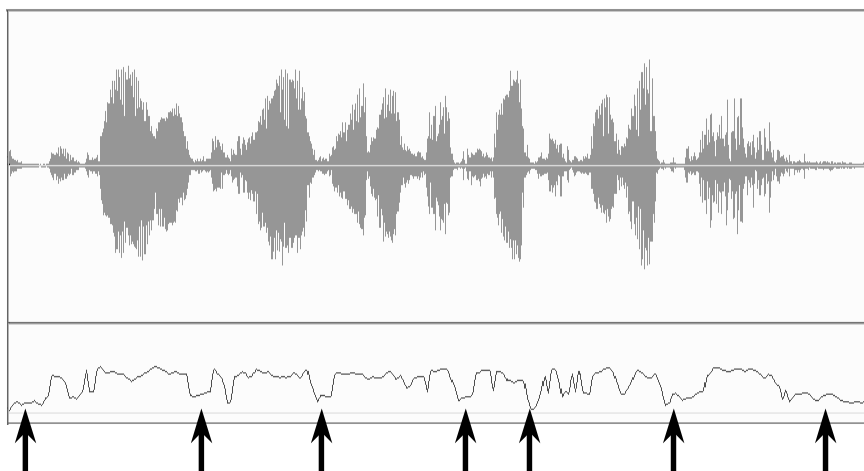
## A pszicho-akusztikus modell egy lehetséges megvalósítása



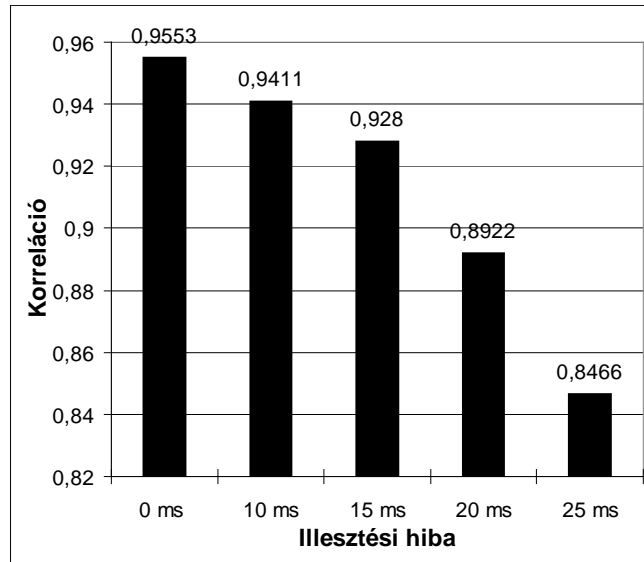
## Egy alkalmazás során felmerült speciális jelenségek

- **Jelzészváltás:** A hangúton belül történik. Mivel a bázis-állomás kezdeményezi, így a kezdete hallható a felvételek során. 🗣️
- **Csönd detekció:** A csendes részeken az additív zaj relatív nagyobb spektrális torzítást okoz, amit kompenzálni kell.
- **Áthallás:** Speciális kategóriája a torzításnak, de csak az egyéb zajokkal azonos módon tudjuk kezelni.

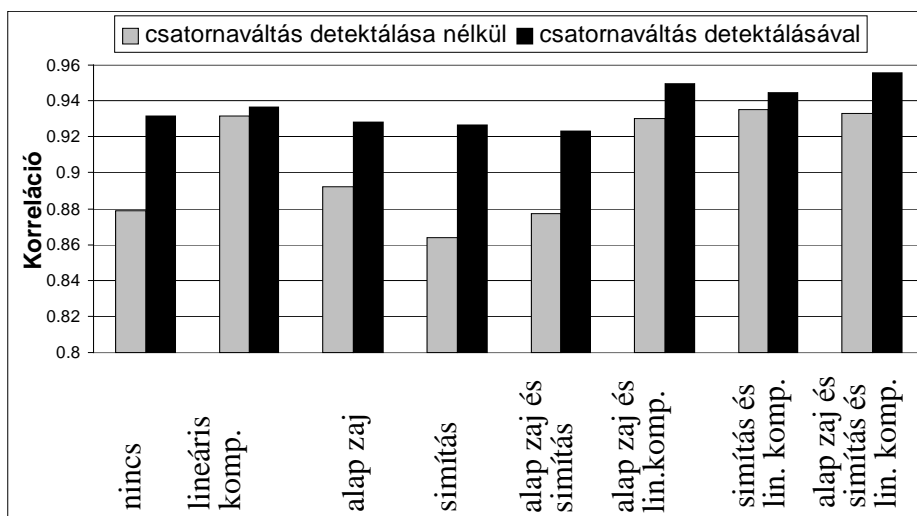
## Keretenkénti minőséggörbe



## Az illesztés hatása a korrelációra



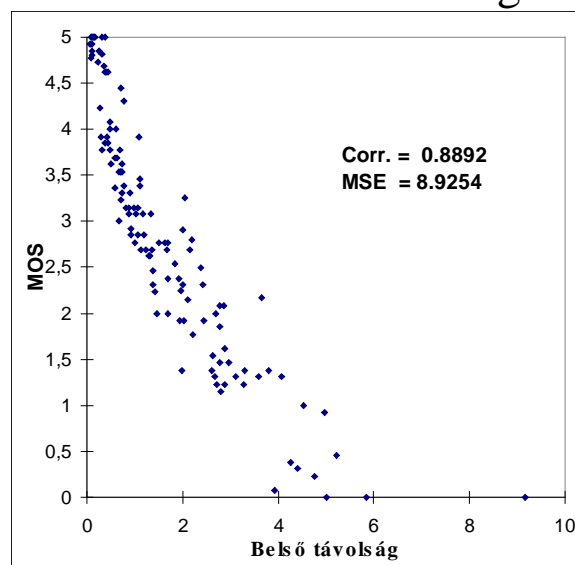
## A pszichoakusztikus modell elemeinek hatása a korrelációra



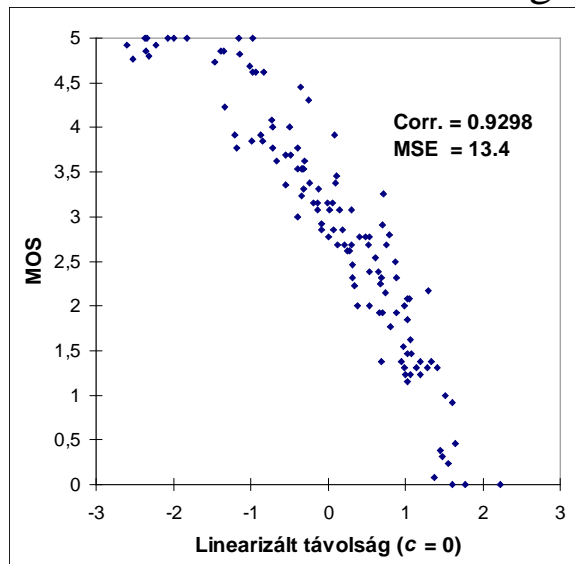
## Linearizálás és skála transzformáció

- Belső távolság = (pl.) keretenkénti euklideszi távolság átlaga 🗣️
- A belső távolság monoton, de nemlineáris módon követi az MOS-t
  - $D_{lin} = \log( D_{inn} + c ) / \log$ : kísérletezéssel megállapított fv/
- A lineáris távolság skálája nem egyezik az MOS skálával
  - $Q_{obj} = a D_{lin} + b$
- $Q_{obj}$  értéke  $D_{lin}=0$  esetén 5.0 kell legyen, ehhez a  $c$  konstans iteratív módon lehet meghatározni

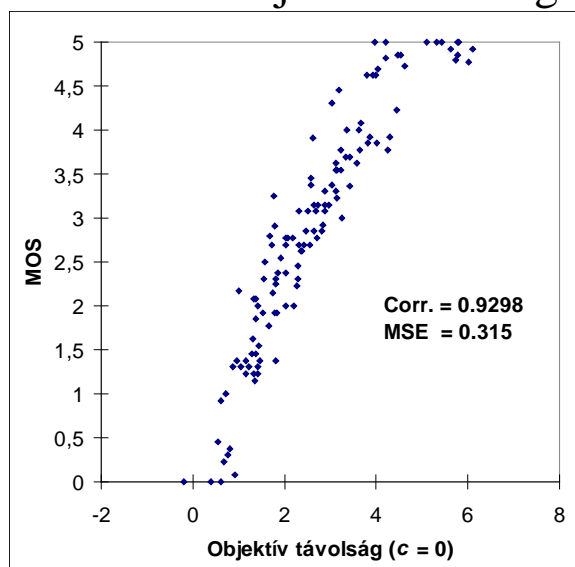
## A kísérleti eredmények szórása MOS vs. belső távolság



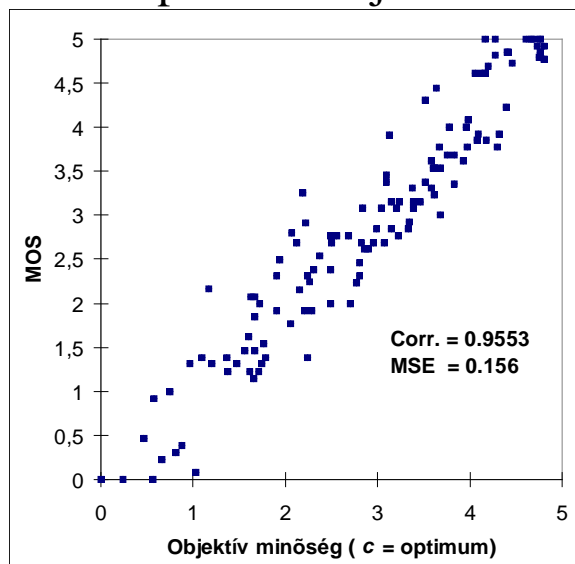
## A kísérleti eredmények szórása MOS vs. Lineáris távolság



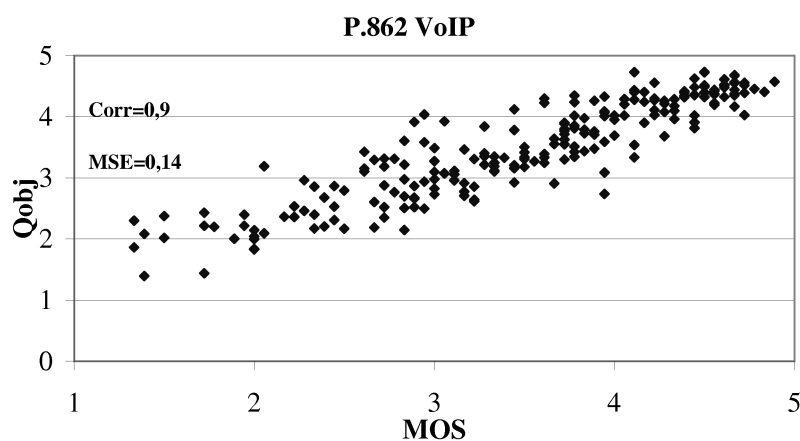
## A kísérleti eredmények szórása MOS vs. Objektív távolság



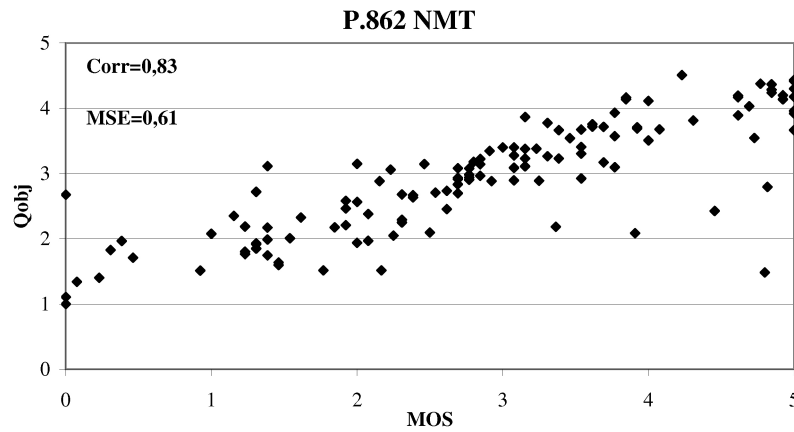
## A kísérleti eredmények szórása MOS vs. Optimális objektív távolság



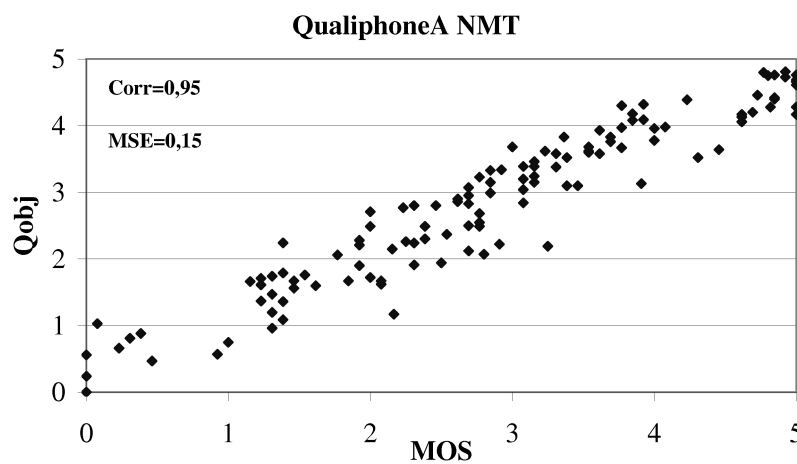
## A kísérleti eredmények szórása Illeszkedő eljárással VoIP mérés



## A kísérleti eredmények szórása Nem illeszkedő eljárással NMT mérés

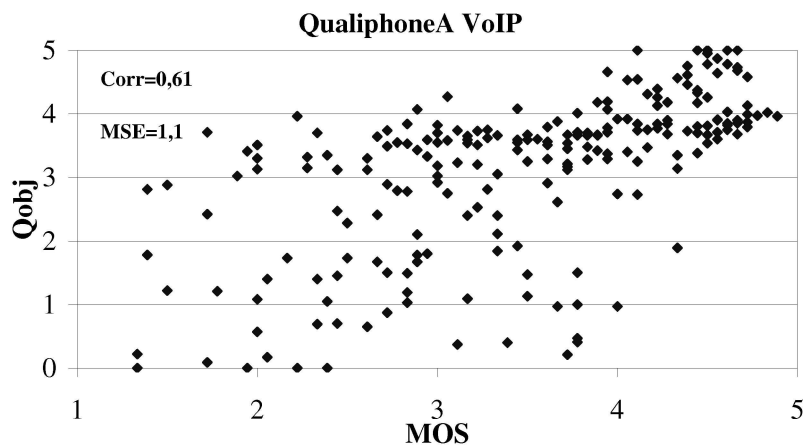


## A kísérleti eredmények szórása Illeszkedő eljárással NMT mérés





## A kísérleti eredmények szórása Nem illeszkedő eljárással VoIP mérés



## Aktív vs. passzív mérések

- Aktív mérések
  - + Irányított mérések, adott/ vett jel ismert
  - + Pontos minőségmérés lehetséges
  - Plusz terhelést jelent a hálózaton, különösen tömeges mérések esetén
  - Hozzá kell férni a végpontokhoz
  - PESQ eljárást is adaptálni kell speciális megoldások esetén
- Passzív mérések
  - + Nincs szükség plusz terhelésre
  - Hogyan mérjük ismeretlen jel minőségét?
  - + Elegendő a központi elemekhez hozzáférni

## Passzív mérési módszer

- Ötlet: beszédfelismerés a zajban nehezebb
- Jellemzők ismert minta esetén
  - Mondat felismerési valószínűsége
  - Többi mondat illeszkedési valószínűsége
  - Általános beszédmodell(ek) illeszkedési valószínűsége
- Jellemzők ismeretlen minta esetén
  - Általános beszédmodell illeszkedési valószínűsége
- Mindkét megoldás hasonlóan gyenge a PESQ-hoz képest

## Felhasználás 2

- Minőség mérés – rendszer tesztelés
- Adat LAN-hoz adódó VoIP forgalom esetén előzetes minőségmérés
- Összekapcsolt rendszerek esetén egymás ellenőrzése (VoIP – VoIP, PSTN – VoIP, GSM – VoIP)
- Mesterségesen előállított beszéd minőségének mérése
- Rádió adások minőségének mérése

## Továbblépések

- Audio jelek minőségének mérése
- Képek minőségének mérése (tömörítés esetén)
- Videó jelek minőségének mérése