

Veszteséges, érzeti alapú audio forráskódolási eljárások vizsgálata

MPEG Audio kódolók

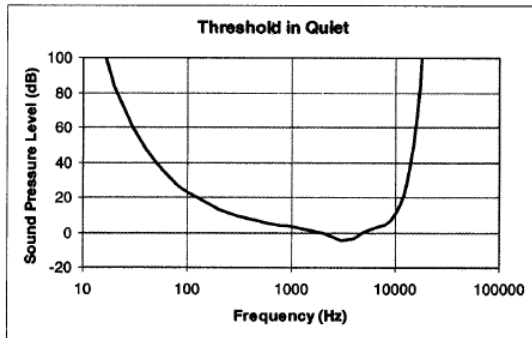
BME - Hálózati Rendszerek és Szolgáltatások Tanszék

Bevezetés

Tartalom

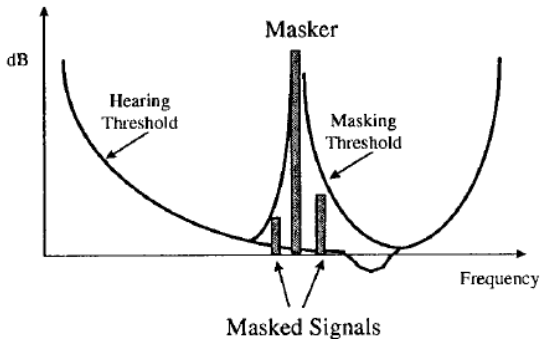
- ▶ Kódolási elvek
- ▶ Pszichoakusztikus (érzeti) alapelvek
- ▶ Kódoló implementációk rendszertechnikai áttekintése
- ▶ Pszichoakusztikus kódolások objektív elemzése

Abszolút hallásküszőb



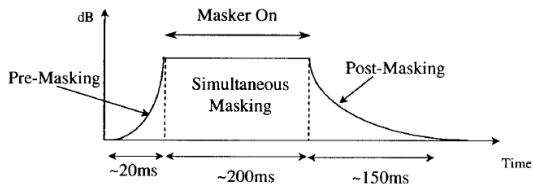
ábra: Abszolút hallásküszőb

Egyidejű (frekvenciatartománybeli) maszkolás



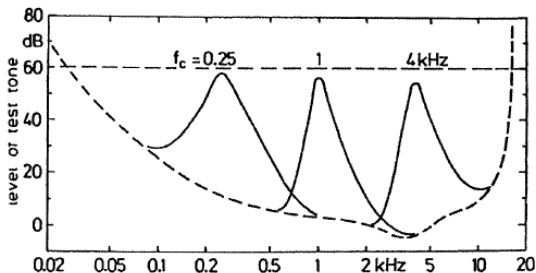
ábra: Egyidejű maszkolás

Időbeli maszkolás



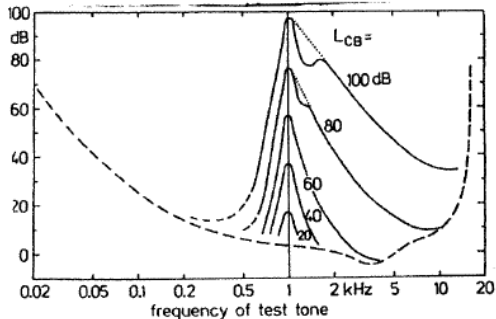
ábra: Elő- és utóelfedés

Egyidejű (frekvenciatartománybeli) maszkolás



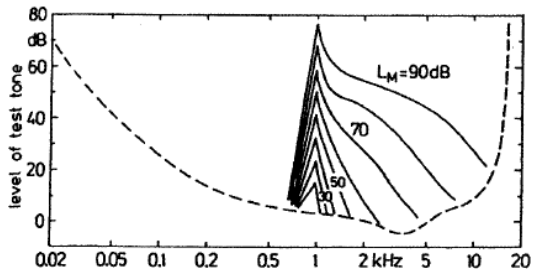
ábra: Példák

Egyidejű (frekvenciatartománybeli) maszkolás



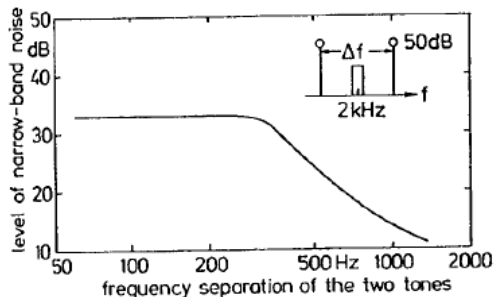
ábra: Példák

Kritikus sávok



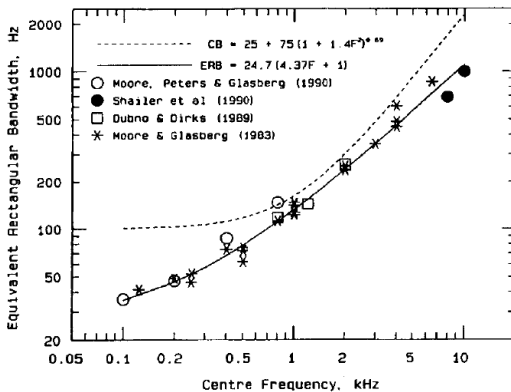
ábra: példák

Kritikus sáv szélességek



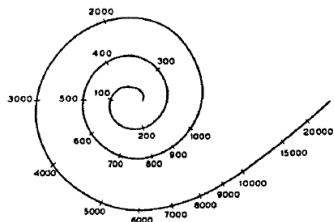
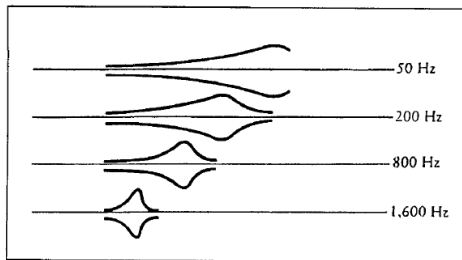
ábra: Maszkolási küszöb két tonális maszkoló távolságának függvényében

Kritikus sáv szélességek

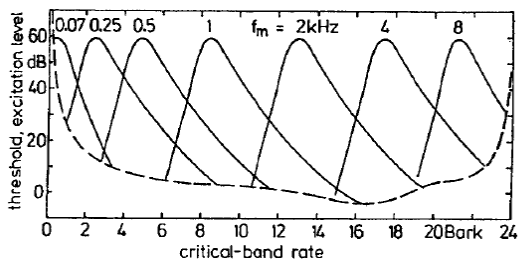


ábra: Ekvivalens sáv szélesség modellek

A hangmagasság érzékelése a csiga mentén



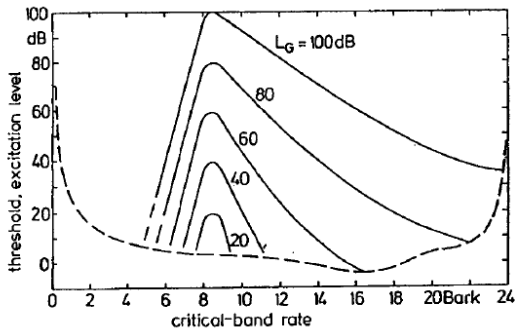
Kritikus sávok



ábra: Maszkolási görbék a Bark skála függvényében

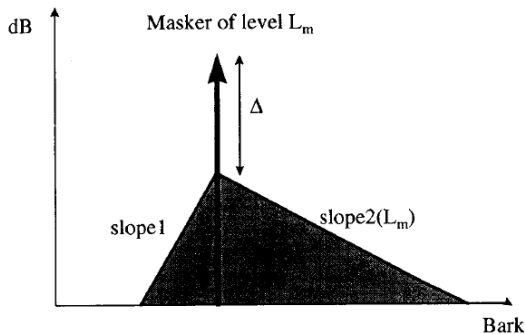
$$z[\text{Bark}] = 13 \arctan(0.00076f[\text{Hz}]) + 3.5 \arctan((f[\text{Hz}]/7500)^2)$$

Kritikus sávok



ábra: Maszkolási görbék a maszkoló szintjének függvényében

Egyszerűsített maszkolási modell



$$\Delta_{\text{tone masking noise}} = 14.5 + z \text{ dB}$$

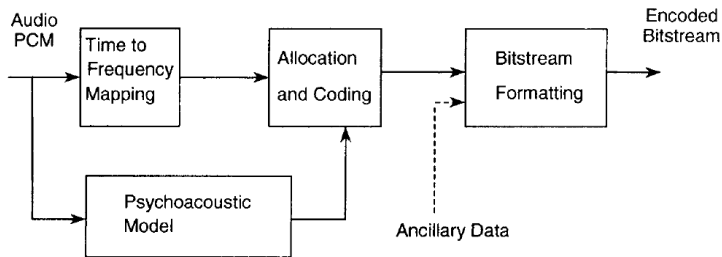
$$\Delta_{\text{noise masking tone}} = C \text{ dB}$$

ábra: Egyszerűsített maszkolási modell

Spreading function, pl:

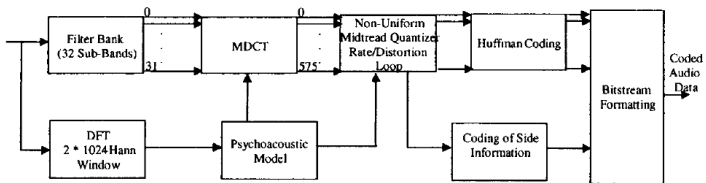
$$10 \log_{10}[B(dz)] = 15.81 + 7.5(dz + 0.474) - 17.5[1 + (dz + 0.474)^2]^{1/2}$$

Általános kódolási modell



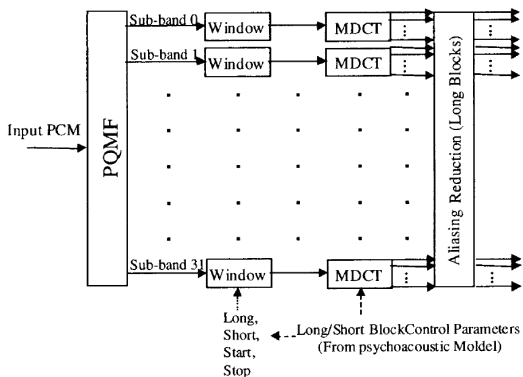
ábra: Általános audio kódoló blokkvázlat

MPEG-I Layer 3 (mp3) kódoló



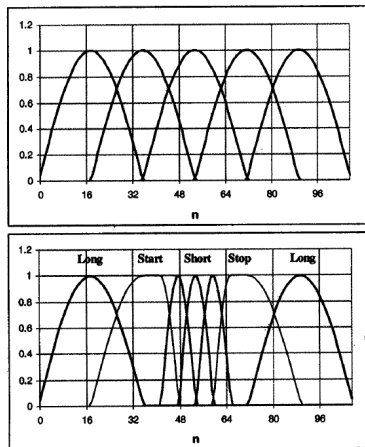
ábra: MP3 kódoló

MPEG-I Layer 3 (mp3) kódoló



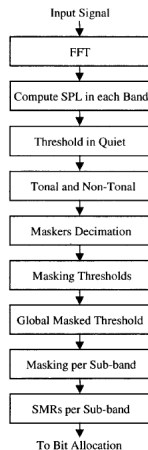
ábra: MP3 kódoló

MP3 kódoló



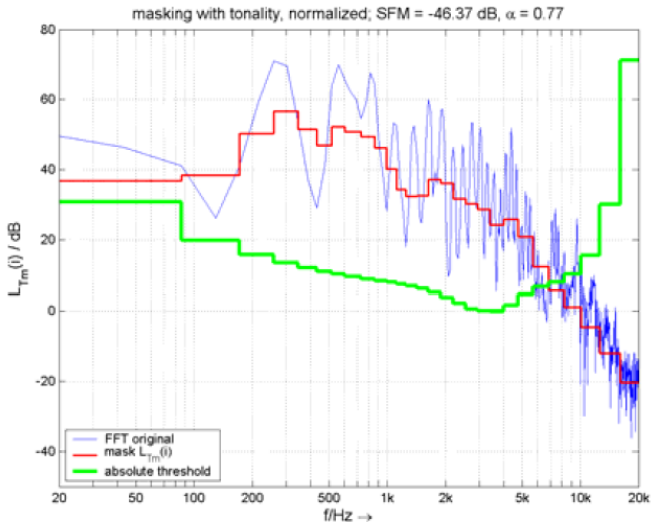
ábra: Ablakfüggvény típusok

MP3 kódoló

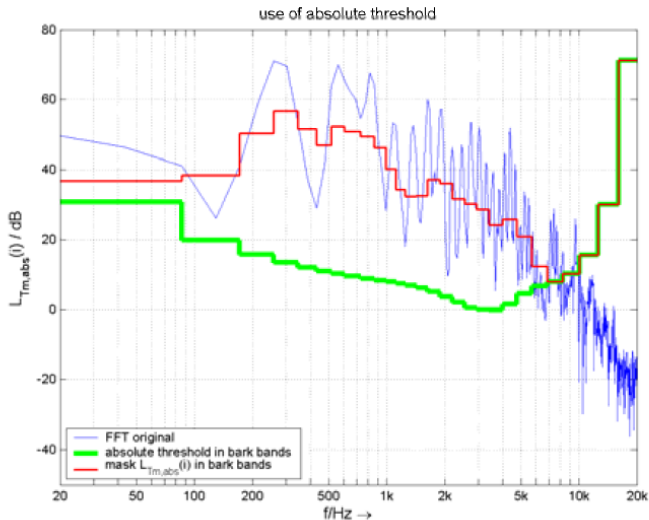


ábra: MPEG-I Hallásmodell

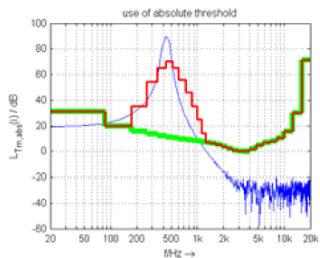
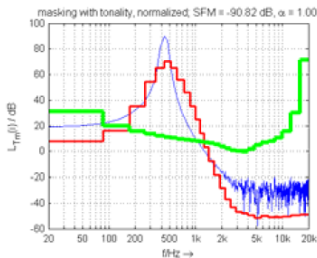
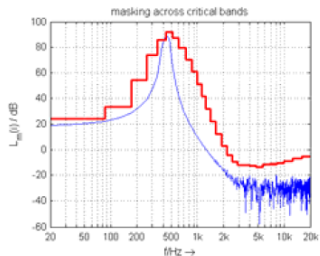
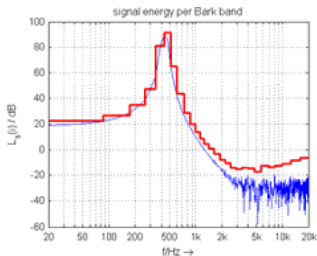
Példák



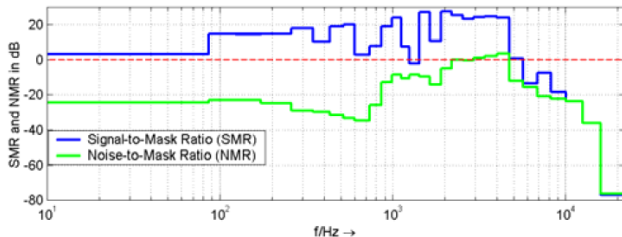
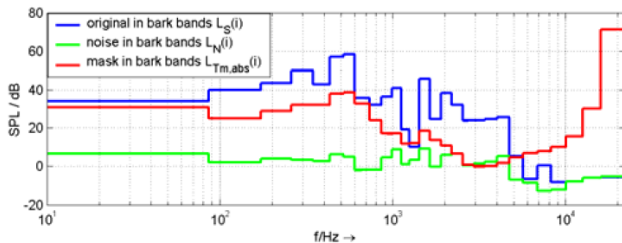
Példák



Példák

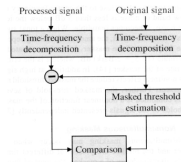


Példák

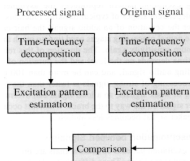


Objektív metrikák alapelvei

- ▶ **NMR vizsgálat (Noise-To-Mask ratio):** A referencia hangminta és a kódolt hangminta közötti különbségjel (hibajel) összehasonlítása a referencia hangmintából számított maszkolási küszöbvel. A hibajel jellemzően a veszteséges újrakvantálásból származik, és így a hibajel energiájának jelentős részét a kvantálási zaj adja. A hibajel spektrumát összehasonlítva a maszkolási küszöbvel, becsülhető, hogy a kvantálási zaj mennyire hallható.
- ▶ **Hallásmodell alapú értékelés:** A teljes hallójárat "jelfeldolgozási" folyamatainak modellezése, a fülben lévő alaphártyán létrejövő gerjesztés-ingerület kapcsolat modellezése. A teljes hallásfolyamat jelfeldolgozási modelljét alkalmazva mind a referencia, mind a kódolt hangminta értékelhető válik a két hangminta között észlelt különbség mértéke.



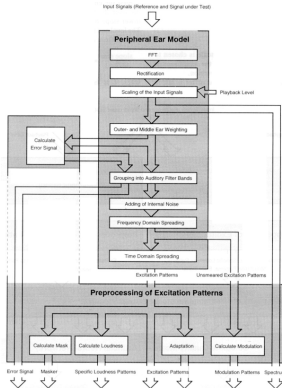
ábra: NMR vizsgálat



ábra: Hallásmodell alapú értékelés

PEAQ model

Egyesíti az NMR és a hallásmodell alapú értékelést



ábra: PEAQ modell

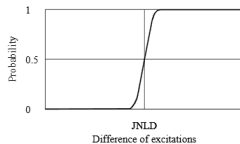
PEAQ model

Egyesíti az NMR és a hallásmodell alapú értékelést

- ▶ Az NMR és a hallásmodell kimeneti változóit egy kognitív modell egyesíti egy neurális háló alkalmazásával. A neurális háló súlyozótényezőit kísérleti úton határozták meg, hogy a kimeneti változók súlyozásával a kapott objektív metrika a lehető legjobban illeszkedjen a szubjektív vizsgálatok (páros összehasonlítások) eredményeihez.
- ▶ A kognitív modell a következő jelenségeket próbálja figyelembe venni: pl. ismert hangok (pl. beszéd, zongorahang) minőségromlására érzékenyebbek vagyunk (előzetes tanulás hatása), mint ismeretlen hangokéra (pl. szintetikus hangminták)
- ▶ Ismert, már korábban tapasztalt torzulások (pl. ismert zajok, ismert zajjelenségek) okozta minőségromlás jobban észlelhetőek, mint ismeretlen zajok, ismeretlen torzulási termékek
- ▶ A hozzáadott zaj, illetve torzítási termékek, zavaróbbak, mintha valami hiányzik a hangból.
- ▶ Lineáris torzulások (pl. frekvenciaszelektív csillapítás, sáv szélesség korlátozás) kevésbé zavaróak mint a nemlineáris torzulások

PEAQ model

- ▶ A modell figyelembe veszi a különböző amplitúdójú hibák detektálási valószínűségét is
- ▶ A referencia és a kódolt hang közötti különbség észlelése szintje (JNLD - Just Noticeable Difference) szintfüggő
- ▶ 20 dB SPL esetén, JNLD: 0.75 dB
- ▶ 80 dB SPL esetén, JNLD: 0.2 dB
- ▶ A referencia és a kódolt hang közötti hibajel és az adott szinthez tartozó JNLD függvényében a kódolási hiba észlelésének valószínűsége változik. A kognitív modell ezt a hibaészlelési valószínűséget is figyelembe veszi az értékelésnél



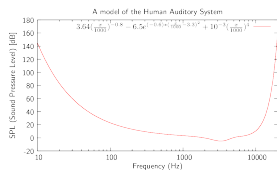
ábra: Észlelési valószínűség a JNLD függvényében

Hallásmodell alapú értékelés - modell elemei

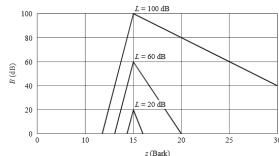
▶ Külső fül modell:

- ▶ **Abszolút hallásküszöb**, mely a külső fül és a középfül átviteli függvényének eredője
- ▶ **Belső zaj**, mely a fülön belüli vérkeringésből, és spontán (gerjesztés nélküli) idegi aktivitásból származik

- ▶ **Csiga (cochlea) modell:** A csiga járatában a hang különböző frekvenciakomponensei állóhullámokat generálva különböző helyeken ingerlik a csiga hallósejtjeit (szőr sejtek). Ennek végeredménye a hangmagasság-érzékelés.
- ▶ **Magasabb szintű feldolgozás modellje.** Kognitív modell, neurális háló alapú modell



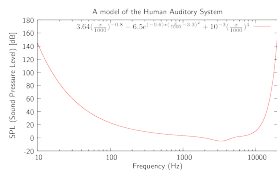
ábra: Abszolút hallásküszöb



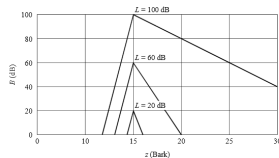
ábra: Gerjesztési modell szinuszos gerjesztés esetén

Külső fül modell - csiga

- Csiga (cochlea) modell:** A csiga tulajdonképpen egy nemlineáris frekvencia-hely (a gerjesztés helye a csigán belül) leképzést valósít meg. Ennek egyik elterjedt modellje a Bark skála, mely a 20 Hz-15 kHz tartományt 24 nem átfedő tartományra bontja. a csiga szőrsejtjei idegi aktivitással válaszolnak az alaphártya rezgéseire. Szinuszos gerjesztésre az idegi aktivitás a csiga mentén háromszög jelleggörbéjű. Tehát a szinuszos hang frekvenciája alatti, és a fölötti frekvenciáknak megfelelő helyek is gerjesztett állapotba kerülnek. A hang frekvenciája alatt a gerjesztés jelleggörbéje 27 dB/Bark, a fölött -5 dB/Bark-tól -30dB/Barkig terjed a hangnyomás szintjétől függően.



ábra: Abszolút hallásküszöb



ábra: Gerjesztési modell szinuszos gerjesztés esetén

Külső fül modell - csiga

- Hallásküszöb

$$A(f)[dB] = -0.6 \times 3.64(f[kHz])^{-0.8} + 6.5 \times \exp^{-0.6(f[kHz]-3.3)^2} - 10^{-3}(f[kHz])^4$$

- Frekvencia (kHz - Bark-skála) leképzés

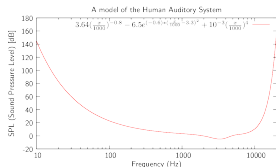
$$f[kHz] = 650 \sinh\left(\frac{z[Bark]}{7}\right)$$

Hozzáadott zaj:

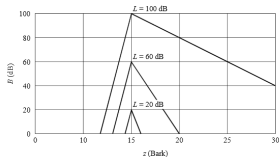
$$N[dB] = 0.4 \times 3.64(f[kHz])^{-0.8}$$

- Frekvencia elmosás (gerjztési modell)
alsó: 24 dB/Bark felső:

$$SLOPE[dB/Bark] = -24 - \frac{230Hz}{f_c} + 0.2L[dB]$$



ábra: Abszolút hallásküszöb



ábra: Gerjztési modell szinuszos gerjztés esetén

Külső fül modell - csiga

- ▶ Sávok energiáinak nem-lineáris összegzése

$$E_k = \left(\sum_i E_{i,k}^{0.4} \right)^{1/0.4}$$

Sávok energiájának időbeli elmosása (időbeli maszkolás)

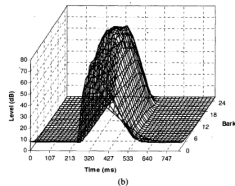
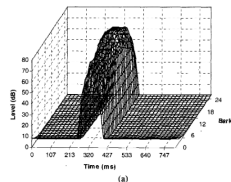
$$\tau(f_c) = \tau_{min} + \frac{100\text{Hz}}{f_c} (\tau_{100} - \tau_{min})$$

ahol $\tau_{min} = 8\text{ms}$ és $\tau_{100} = 30\text{ms}$.
Exponenciális súlyozás a keretek energiáira:

$$E_f[k, n] = aE_f[k, n-1] + (1-a)E[k, n]$$

ahol

$$a = \exp \frac{4}{187.5\tau}$$



ábra: Fehérzaj burst (a) időbeli elmosás nélkül (b) időbeli elmosással

Spektrális burkoló változásának (moduláció) figyelembevétele

Spektrális burkoló modulációja:

$$E_{\Delta}(f, t) = \int_{t'=0}^t \left| \frac{dE(f, t')}{dt} \right| \exp^{(t'-t)/\tau(f)} dt'$$

$$\bar{E}(f, t) = \int_{t'=0}^t |E(f, t')| \exp^{(t'-t)/\tau(f)} dt'$$

ebből a moduláció:

$$\text{mod}(f, t) = \frac{E_{\Delta}(f, t)}{1 + (1/c_1)\bar{E}(f, t)}$$

Modulációs különbség a referencia és a kódolt hangminta között:

$$\text{moddiff}(f, t) = w \frac{|\text{mod}_{proc}(f, t) - \text{mod}_{orig}(f, t)|}{|\text{mod}_{orig}(f, t)|}$$