Vízjelek hanganyagban

BME - TMIT
VITMA378 - Médiabiztonság
feher.gabor@tmit.bme.hu

Hanganyagok

- Tömörítetlen
  - CD anyagok
- Tömörített
  - MD lejátszók, MP3 lejátszók
    - Zune, iPod, …
- A tömörítés alapja
  - Veszteségmentes:
    - A hanghullámot máshogy írjuk le, becslést alkalmazunk
    - Pl.: DPCM
  - Veszteséges: Az emberi fül nem hall meg minden ‘hangjegyet’
    - Pl: MP3

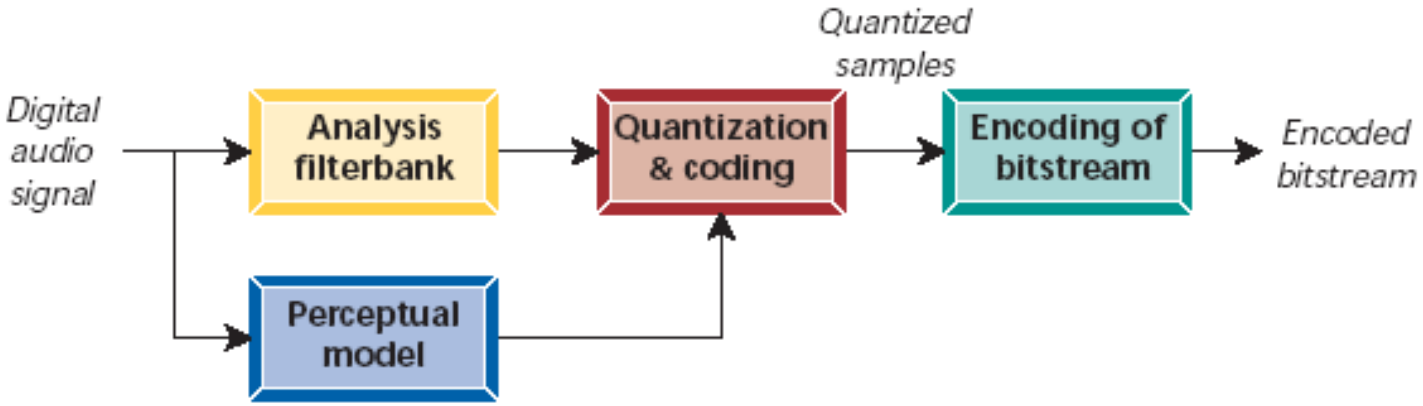
MPEG Layer 3

- Csatornák
  - Mono, Dual, Stereo, Joint-stereo
- Különböző MPEG verziók
  - MPEG-1
    - 32, 44.1, 48 kHz
  - MPEG-2
    - 16, 22.05, 24, 32, 44.1, 48 kHz
  - MPEG-2.5 (Fraunhoffer IIS)
    - 8, 11.05, 12, 16, 22.05, 24, 32, 44.1, 48 kHz

- MPEG hangkódolás 3 réteg
- Layer I
  - 32 frekvenciatartomány
  - Pschyo-akkusztikus modell
  - Kvantálás
  - Több mint 128kbps/csatorna
- Layer II
  - Layer I +
  - Skálázási faktor
  - Cél: 128kbps/csatorna
- Layer III
  - Layer II +
  - Újabb (hibrid) szűrés
  - Entrópia kódolás
  - Cél: 64kbps/csatorna

MP3 algoritmus

- Filterbank
- Perceptual model
- Quantization & coding
- Encoding

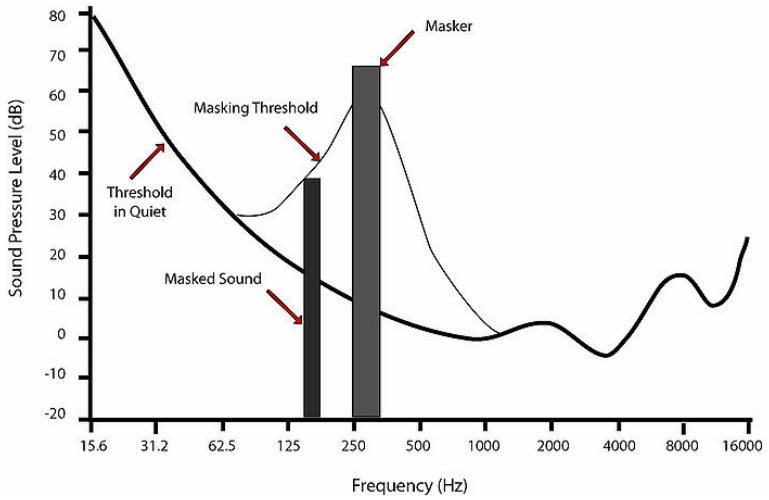


HAS

- Human Auditory System
  - Az ember hallását jellemzi
    - Az emberi fül nem képes minden hangot felfogni
    - Tipikusan 20-20000 Hz között hallunk
      - Leginkább érzékeny a 2kHz-5kHz tartomány
      - 16 kHz felett már a felnőttek nem hallanak jól
    - Az emberi agy is befolyásolja, hogy mit hallunk
      - Pszichoakusztika
      - Masking
      - Haas effektus

Masking

- Egy teljesen jól hallható hang bizonyos egy másik hang hatására hallhatatlan lesz számunka
  - A hangosabb jel elnyomja a közeli halkabb jelet.



  - Történhet egyszerre, de időben eltolva is

"Phantom" fundamentals

- Egy hangot akkor is hallhatunk, ha nincs is ilyen forrás, akkor ha a felharmonikusai jelen vannak
  - f –et hallhatunk, ha halljuk 2f, 3f, 4f, stb.-t
- Olcsóbb hangfalat lehet gyártani

Haas effektus

- Két közel azonos hang forrása, amely max 30-40 ms késleltetéssel érkezik, nem megkülönböztethető, az első érkező számít
  - A második jel akár erősebb is lehet bizonyos keretek között
- Színházakban így működik az erősítés

MPEG pschyo-akkusztikus modell

- Az ember számára hallhatatlan információk eltávolítása
  - Masking (Az egyik közeli jel elnyomja a másikat)
    - Közeli frekvencián az erősebb jel győz
    - A magas frekvenciákat könnyebb eltakarni
    - Elő- (5-20ms) és utó (50-200ms) takarás
  - Az elnyomott jel kevesebb biten van ábrázolva

MP3 frame tömörítés

- Egy mp3 frame 1152 mintát tartalmaz (kompatibilitás miatt ennyi)
- A frame fel van osztva 32 frekvenciatartományra (filterbank), ezek tovább vannak osztva 18 adaptív tartományra
- Módosított DCT és FFT segítségével meghatározzák az egyes tartományok spektrális intenzitását
- A tartományok külön vannak kvantálva
- Kvantálás után entrópia kódolás

Adatrejtés - tömörítetlen hang

- A pschyo-akkusztikus modell segítségével adatrejtés a hangállományban
  - Emberi fül számára hallhatatlan
  - DE: tömörítés hatására eltűnhet!

Adatrejtés - MPEG hangállomány

- Adatrejtés illesztése
  - Skálázási faktorok módosítása (±1)
  - Minták módosítása (±1)
  - Nagyobb változások már hallható változást okoznak
  - Visszaállíthatatlan!
- Szükség van az eredeti állományra is

Adatrejtés használata

- Szteganográfia – titkos üzenetek
- Vízjelek
  - A zenemű megjelölése szerzői jog védelme szempontjából
    - „Ujjlenyomat” – az mű azonosítása
      - Csak az arra jogosult lejátszók játszhatják le, pl.: on-line zeneletöltés
    - Nyomonkövetés
      - Adott zenemű kiszivárogtatójának azonosítása

Ujjlenyomat

- Az ujjlenyomat (fingerprint) egy speciális vízjel, amely a művet azonosítja
  - Inkább tekinthető egy hash algoritmusnak, amely a tartalomra nézve készül, nem bitről bitre
    - További hasonlóság, hogy az ujjlenyomat is nagyon kevés ütközést (vagy hibát) engedhet meg
- Az ujjlenyomat nem módosítja a tartalmat!
- Hang ujjlenyomat esetén cél lehet, hogy CD művet azonosítsunk akár mobiltelefonon keresztül is!

Hang ujjlenyomat

- Felhasználás
  - Megfigyelés: reklámok, zeneszámok automatikus mérése
  - Kapcsolt tartalom megjelenítése
  - Másolásvédelem, megosztás
  - Automatikus zenegyűjtemény rendezés

Ujjlenyomat megvalósítása

- Követelmények
  - Robusztus – az azonosítás megmarad, ha a tartalom torzul (pl. tömörítés)
  - Megbízhatóság
  - Méret - Az ujjlenyomat mérete (bit/s vagy bit/zeneszám)
  - Felbontás – Mennyi információra van szükség, hogy az azonosítást elvégezzük
    - Függ az alkalmazástól is. Rendelkezésre állhat az egész zeneszám, de lehet, hogy csak egy része érhető el aktuálisan
  - Visszakeresési idő, skálázhatóság

- Néha a paraméterek összefüggenek.
  - Pl.: kisebb felbontás -> kisebb megbízhatóság
    Nagy megbízhatóság -> nagy keresési idő

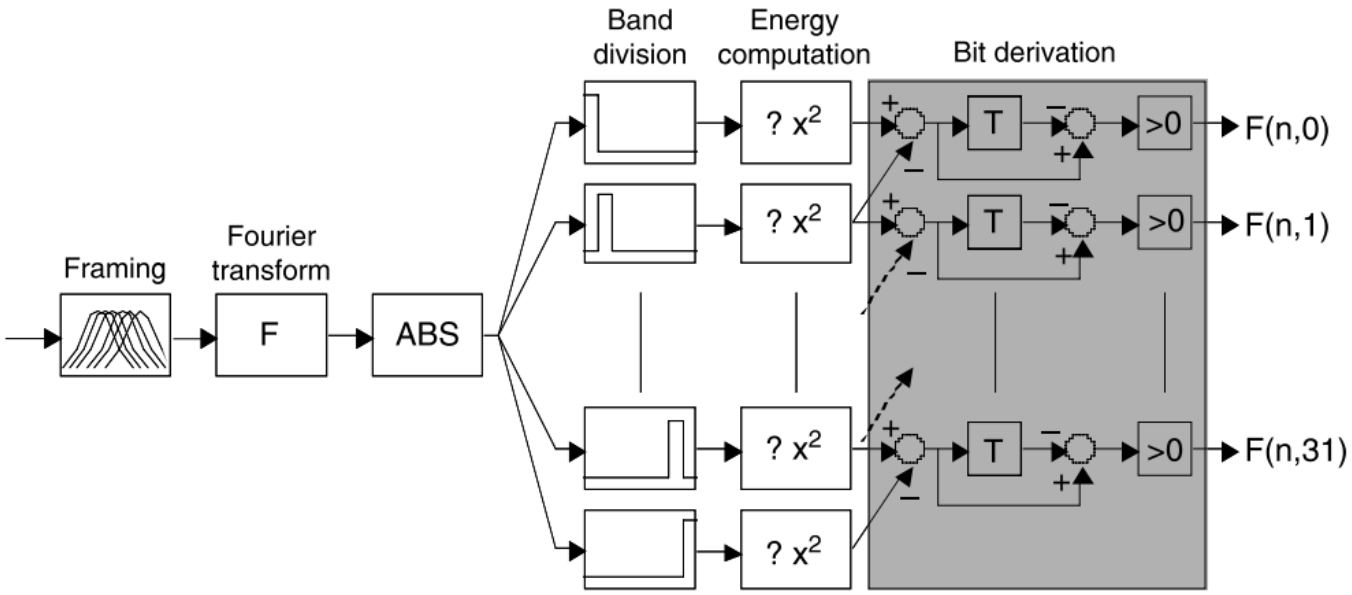
PHILIPS hang ujjlenyomat

- Nemsematikus vizsgálat – ahogy halljuk a zenét (Sematikus: műfaj, BPM, …)
- Bináris ujjlenyomatok tárolása
  - A hasonlóság a Hamming távolságon alapszik
- 3 másodperces (11.6 ms x 256) felbontás
  - Elég a robusztussághoz
  - Nem feltételezi, hogy a mű egészben rendelkezésre áll
    - A támadó például levághatja a mű elejét/végét
  - Megfigyelésnél nincs szükség sok információra (hallgatás/letöltés)

Ujjlenyomat meghatározása

- Alap folyamat:
  - Keretekre osztás
  - Keretekben tulajdonságok meghatározása
    - Pl. hang esetén: Fourier együtthatók, Mel-frequency cepstral együtthatók (MFCCs), spektrális tulajdonságok, Linear Predictive Coding (LPC) együtthatók, stb…
  - Al-ujjlenyomat (subfingerprint) meghatározása a tulajdonságok alapján
  - Több al-ujjlenyonat (fingerprintblock) egymás után azonosítja a művet

PHILIPS ujjlenyomat



- 32 bites subfingerprint miden intervallumban
  - 1 intervallum: 11.6 ms
- Fingerprintblock: 256 egymás utáni subfingerprint

PHILIPS ujjlenyomat 2.

- Először keretekre bontás történik meg
  - A keretek mérete 0.37s, és „Hanning window” szerint súlyozzuk 31/32 paraméterrel. Ez adja a 11.6 ms hosszt
  - A nagy átfedés miatt az egymás utáni al-ujjlenyomatok lassan változnak -> Ettől robusztusabb lesz

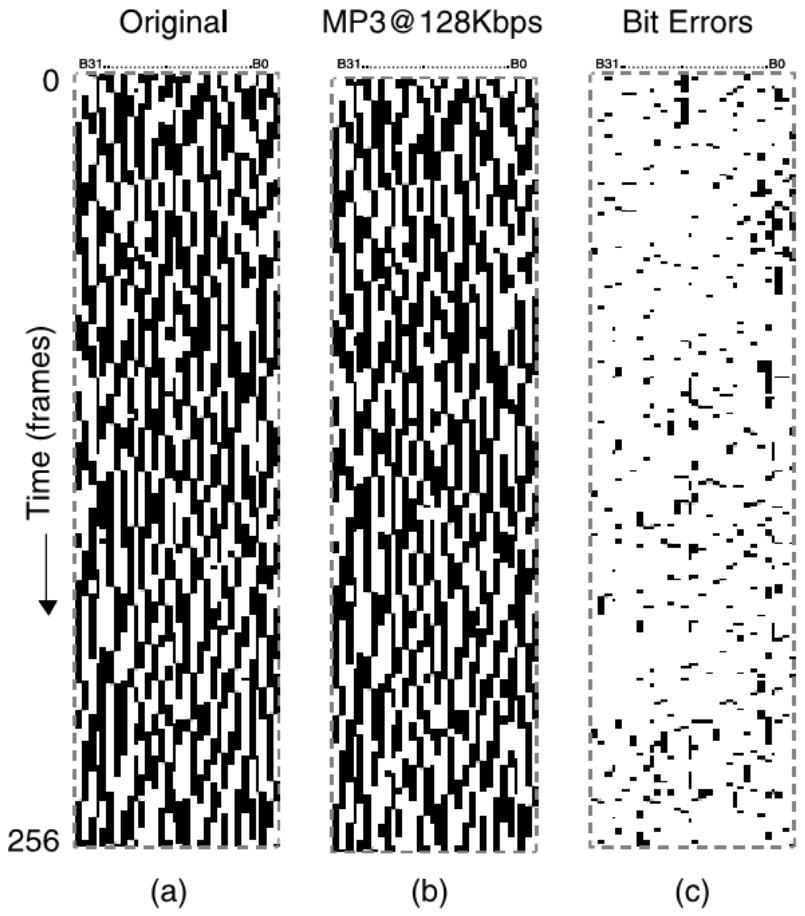
- A keretekben a Fourier spektrum abszolút értéke számít (A fázis információra kevésbé érzékeny az ember)
  - A keret 33 sávra van osztva a 300-2000 Hz tartományban (HAS miatt)

$$F(n,m) = \begin{cases} 1 & \text{if } E(n,m) - E(n,m+1) - (E(n-1,m) - E(n-1,m+1)) > 0 \\ 0 & \text{if } E(n,m) - E(n,m+1) - (E(n-1,m) - E(n-1,m+1)) \leq 0 \end{cases}$$

Ahol E(n,m) az n. keret m sávjának energiája és F(n,m) az m. bit az n. keret al-ujjlenyomatában

Ujjlenyomat minta

- Carl Orff „O Fortuna”
  - Ujjlenyomat blokk (A *c* ábra a különbséget jelöli)
  - 0/1 : fehér/fekete
- A tömörítés miatt nincs pontos egyezés



Ujjlenyomat kinyerés sebessége

- Mivel úgyis 2 kHz alatt vizsgálunk, ezért először 5 kHz –re újramintavételezik és átalakítják egy hangsávra
  - Egyszerű szűrő is használható, az ujjlenyomat úgyis robusztus
  - A keret így 2048 mintát tartalmaz, minden 64 mintára lesz al-ujjlenyomat értékünk
  - FFT esetén a futási sebesség még megfelelő, akár mobilon is futhat

Mérési eredmények

Processing	Orff	Sinead	Texas	AC/DC
MP3@128Kbps	0.078	0.085	0.081	0.084
MP3@32Kbps	0.174	0.106	0.096	0.133
Real@20Kbps	0.161	0.138	0.159	0.210
GSM	0.160	0.144	0.168	0.181
GSM C/I = 4dB	0.286	0.247	0.316	0.324
All-pass filtering	0.019	0.015	0.018	0.027
Amp. Compr.	0.052	0.070	0.113	0.073
Equalization	0.048	0.045	0.066	0.062
Echo Addition	0.157	0.148	0.139	0.145
Band Pass Filter	0.028	0.025	0.024	0.038
Time Scale +4%	0.202	0.183	0.200	0.206
Time Scale -4%	0.207	0.174	0.190	0.203
Linear Speed +1%	0.172	0.102	0.132	0.238
Linear Speed -1%	0.243	0.142	0.260	0.196
Linear Speed +4%	0.438	0.467	0.355	0.472
Linear Speed -4%	0.464	0.438	0.470	0.431
Noise Addition	0.009	0.011	0.011	0.036
Resampling	0.000	0.000	0.000	0.000
D/A A/D	0.088	0.061	0.111	0.076

BER (Bit Error Rate)

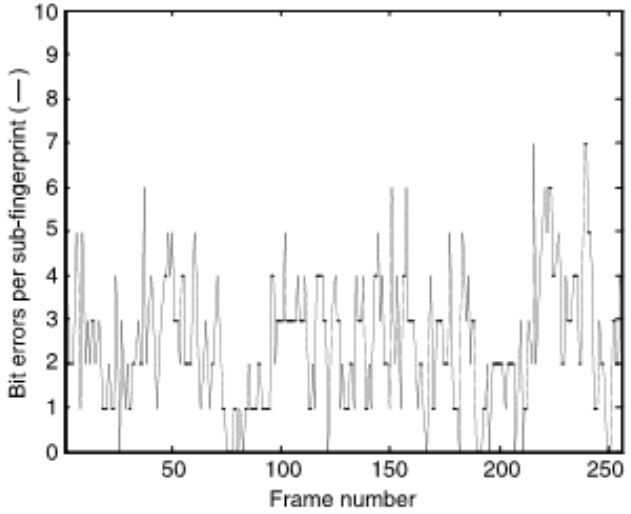
0.35 alatt rendben lehet

Ujjlenyomat visszakeresése

- Adatbázis mérete:
  - Teszt: 10.000 zenemű, átlagosan 5 perc hossz: 250.000.000 al-ujjlenyomat
- Ha minden bit a helyén lenne, akkor egyszerű a keresés
  - De nincs így, a legközelebbit kell megkeresni (legkisebb BER érték)
  - Nyers erő módszere: 200.000 minta/s sebességnél (PC sebessége) kb. 20 perc

Ujjlenyomat visszakeresése 2.

- Mivel a nyers erő nem mindig elfogadható, más is kell
- Megnézhetjük, mi van ha 1 al-ujjlenyomat egyezést biztosra veszünk
- Tapasztalat szerint van ilyen egyezés, akár több is
- LUT: look-up table az al-ujjlenyomatok eléréséhez



Ujjlenyomat visszakeresés 3.

- A LUT mérete hash segítségével csökkenthető. A sebesség így is elfogadható, ütközések után is kb. 800.000x gyorsabb (mérések alapján), mint a brute force
  - Ha nagyon torzulna a tartalom, akkor többet kell keresni: 1 bithibás, 2 bithibás, …, de ez már 3 bithibánál is elfogadhatatlan keresési időt ad!

- Helyette „soft-decoding”, sorba rendezzük a biteket, hogy mennyire biztosan hordozzák az információt, annak alapján, hogy mennyire volt közel a küszöb a bit eldöntésénél. Közeli küszöb: kevésbé bizonyos bit
  - Jelentős gyorsulás érhető el így

Ujjlenyomat visszakeresés 4.