

Veszteséges audiotömörítési eljárások vizsgálata

Mérési útmutató

Multimédia rendszerek és szolgáltatások laboratórium I.

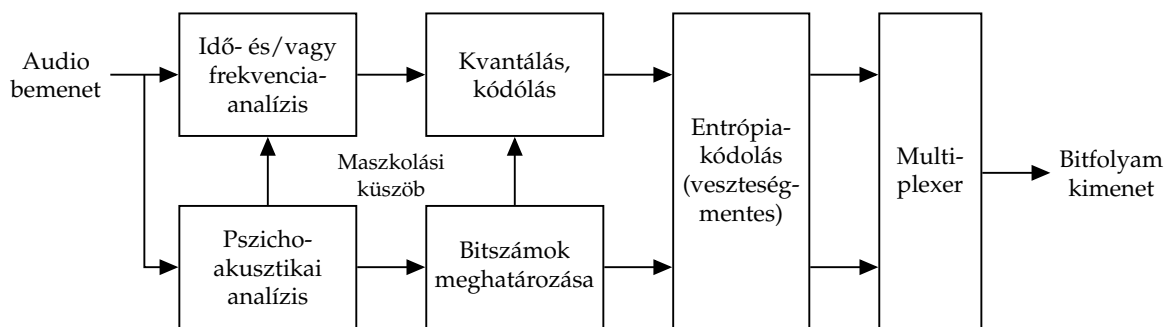
Rucz Péter
rucz@hit.bme.hu

2018.

1. Bevezetés

A veszteséges audiotömörítési eljárások a hordozható lejátszók elterjedésével nagy teret hódítottak. Ennek természetesen a tömörítési eljárások hatékonysága, illetve az ilyen módon tömörített hanganyagok jó minősége az oka. Az ilyen tömörítési eljárások hatékonyságát szemlélteti a következő néhány adat. Egy átlagos minőségű MP3 bitfolyam bitsebessége 128 kbit/s. Ez 48 kHz-es mintavételi frekvencián mintánként átlagosan mindössze 1,33 bitet jelent. Az audio CD-k 16-bit-es mintáihoz hasonlítva ez tizenkétszeres tömörítési arány, nagyjából azonos hangminőség mellett. Ilyen mértékű tömörítési arány csak pszichoakusztikai alapú veszteséges eljárásokkal érhető el. A pszichoakusztikai veszteséges tömörítő eljárások lényege, hogy az audiojelből olyan információkat hagy el, melyek nem okoznak érzeti különbséget. Vagyis azokat a részleteket hagyja el a jelből, melyeket a hallgató amúgy sem hallana meg. Annak érdekében, hogy megállapítsuk, melyek azok a komponensek, melyeket fülünk nem érzékel, a jel pszichoakusztikai elemzése szükséges.

Az 1. ábra egy általános veszteséges tömörítési eljárás modelljét mutatja be. Az audio bemenet egyszerre kerül feldolgozásra a kódoláshoz használt analízismodulban és a pszichoakusztikai modulban. Ebben a lépésben a pszichoakusztikai modul már szolgáltat információkat az analízis modul részére, pl. a kódoláshoz használt időablakok kiválasztását illetően. Jelen mérés során kiemelt jelentőséggel bír a pszichoakusztikai analízis modul, melyet részletesen a 3. szakaszban vizsgálunk meg. Egyelőre megelégszünk annyival, hogy a pszichoakusztikai modul feladata, hogy az egyes – frekvenciatartománybeli – mintákhoz bitszámokat, vagy kívánt bitszámokat rendeljen. Az alacsony bitszám és így a magas tömörítési arány eléréséhez az idő- és frekvenciatartománybeli maszkolási jelenségeket használhatjuk ki. A kódoló a mintákat transzformálja (pl. diszkrét koszinusz-transzformációval), majd a kiszámított bitszámok



1. ábra. Veszteséges audiotömörítő eljárások általános blokkvázlata

alapján újrakvantálja. Az újrakvantált minták veszteségmentes entrópiakódolás (pl. Huffman-kódolás) után kerülnek a végső kimeneti bitfolyamba. A végső bitfolyam a mintákon kívül a visszafejtéshez vagy éppen szinkronizációhoz szükséges mellékadatokat is kell, hogy tartalmazzon. A kódolással kapcsolatban leggyakrabban megfogalmazott követelmény a bitfolyam maximális bitsebessége (pl. 192 kbit/s). A bitsebesség tartása érdekében a jobb minőségű kódoló eljárások iteratívan (analízis–szintézis iteráció) határozzák meg az egyes mintákhoz rendelhető bitek számát az adott bitsebesség mellett elérhető maximális minőségre (pl. minél kisebb torzítás) törekedve.

Ebben a mérésben egy veszteséges tömörítéshez használható pszichoakusztikai modellt fogunk megvizsgálni, amely a frekvenciatartománybeli egyidejű maszkolás jelenségét (lásd a 2. szakaszt) használja ki. Alkalmazásunkban a modellt egy egyszerűsített újramintavételezési eljárásban fogjuk alkalmazni. Ahhoz, hogy a modell és a tömörítő eljárás minden lépése áttekinthető legyen a laborfoglalkozás alkalmával, a veszteséges tömörítéshez kevésbé kapcsolódó nemlineáris kvantálási módszerrel és a kvantált minták entrópiakódolásával nem fogunk foglalkozni.

2. Néhány pszichoakusztikai fogalom

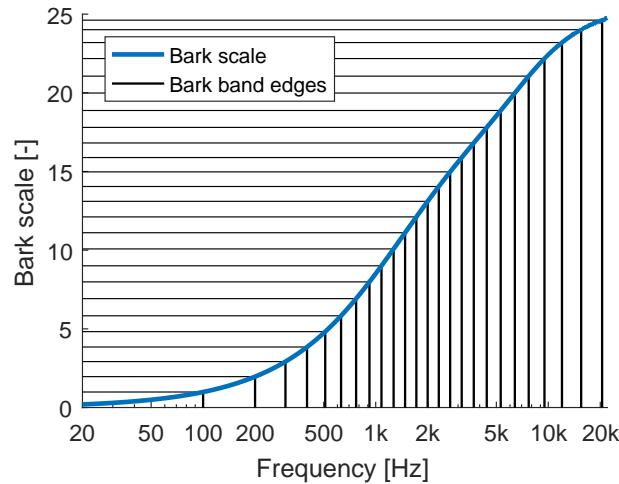
2.1. Tonális és atonális komponensek

Azokat a hangokat, melyekhez érzett hangmagasság társítható, tonális hangoknak, több komponensből álló hangok esetén az ilyen komponenseket pedig tonális komponenseknek nevezzük. Ezzel szemben azokat a hangokat vagy komponenseket, melyekhez nem társítható érzett hangmagasság atonálisnak mondjuk. Tonális például egy fúvós vagy vonós hangszer tartott hangja, egy megütött xilofonrúd, harang vagy megpengetett húr hangja. Atonális például egy szélessávú zaj, egy megütött cintányér, vagy egy erősen tompított és torzított megpengetett húr hangja. Komplex hangokban, például egy repülőgép által lesugárzott zajban, vagy egy olyan zeneműben, melyben egyszerre különféle hangszerek is megszólalnak, egyidejűleg vannak jelen tonális és atonális komponensek. A tonális és atonális komponensek megkülönböztetése azért fontos a pszichoakusztikai analízis szempontjából, mert ezek jelentősen eltérő mértékben maszkolják a velük egyidejűleg megszólaló egyéb komponenseket.

2.2. A Bark-skála és a kritikus sáv szélesség

A Bark-skála egy olyan frekvenciaskála, melyen egyenlő távolságok azonos észlelt hangmagasság-különbségeknek felelnek meg. Nevét Heinrich Barkhausen német fizikusról kapta. A Bark-skála igazodik a Harvey Fletcher által definiált kritikus sáv szélességekhez. Fletcher kísérleteiben azt vizsgálta, hogy különböző sáv szélességűre megszűrt fehér zaj mennyire zavarja a sáv közepi harmonikus jel észlelését. Azt találta, hogy létezik egy frekvenciafüggő ún. kritikus sáv szélesség, amely felett a zavarás sáv szélességének növelése már nem növeli a zavaró hatást. A kritikus sáv szélesség a Bark-skálán minden frekvencián 1 Bark. A Bark-skálát a 2. ábra szemlélteti. Megfigyelhető, hogy kisfrekvencián a skála közel lineáris, míg az 500 Hz-nél nagyobb frekvenciákon logaritmikus menetű.

Az audiotömörítésben a Bark-skálát a maszkolási görbe számításánál használjuk ki. Egyrészt egy maszkoló komponens (*masker*) maszkoló hatásának erőssége a maszolt (*maskee*) komponensre jelentősen függ a komponensek frekvenciabeli távolságától. Ezt a függést a Bark-skála fölött szokás megadni. Másrészt a kvantáláshoz használandó bitszámokat is érdemes a szabványos kritikus sávokban megadni. (A szabványosított kritikus sávok frekvenciahatárait mutatják a 2. ábra függőleges vonalai.)



2. ábra. A Bark-skála a hallható frekvenciatartományban

2.3. Maszkolás a frekvenciatartományban

Ismert jelenség, hogy egy megszólaló hang „zavarja” a vele egyidejűleg észlelt további hangok érzékelését. Amennyiben az egyszerre észlelt hangok frekvenciái egymáshoz közeliak, a zavaró hang teljesen elfedheti, vagyis maszkolhatja a másik komponensét. Ezt nevezzük frekvenciatartománybeli vagy egyidejű maszkolásnak, a zavaró és az elfedett részleteket pedig rendre a maszkoló (*masker*) és maszkolt (*maskee*) komponenseknek. Az effektust a belső fülben található bazális membrán (*basilar membrane*) viselkedése okozza. Különböző frekvenciájú gerjesztések a térben változó merevségű membrán különböző részeit hozzák rezgésbe, azonban egy harmonikus gerjesztés a membránon nem pontszerű, hanem kiterjedt felületet gerjeszt meg, így a keskenysávú gerjesztéshez egy szélesebb sávú válasz tartozik. A frekvenciasáv kiszélesedését nevezzük *spreading*-nek, a mértékét leíró függvényt pedig *spreading function*-nek.

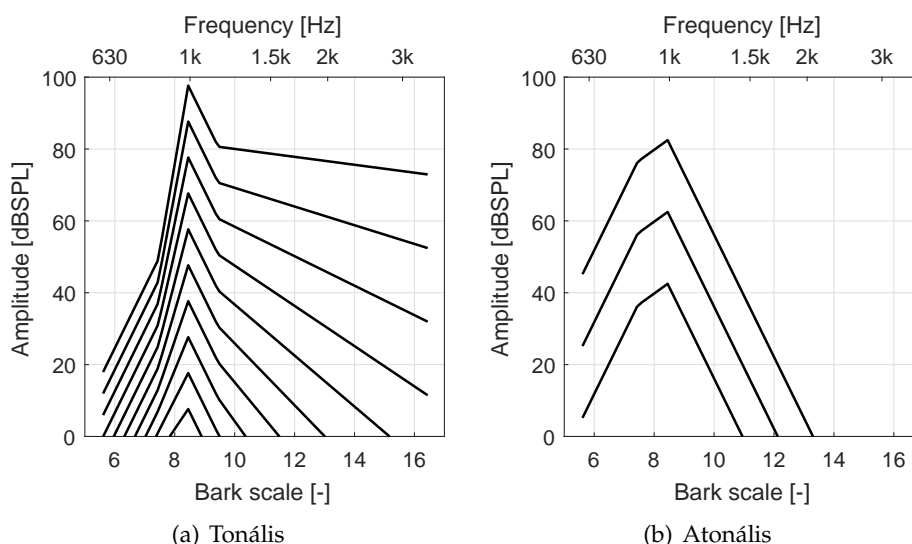
Az egyidejű maszkolás erőssége függ a maszkoló jel jellegétől (tonális vagy atonális), amplitúdójától és a maszkolt és maszkoló komponensek frekvenciáinak távolságától. Egy harmonikus és egy zaj jel közelített maszkológörbéit mutatja a 3. ábra. Figyeljük meg, hogy a tonális esetben a görbe jellege jelentősen függ a jel amplitúdójától, míg az atonális esetben a görbék alakja a jelszinttől független. Érdekes még megfigyelni, hogy a görbék jelentősen aszimmetrikusak. A nagyobb amplitúdójú tonális komponensek jóval nagyobb területet maszkolnak, különösen a maszkoló komponensnél nagyobb frekvenciák felé, ezt mutatja, hogy a maszkológörbék lefutásának meredeksége csökken a maszkoló jel amplitúdójának növelésével. A maszkoló görbéket minden esetben jól közelítik a Bark-skála fölött szakaszonként lineáris függvények.

Az egyidejű maszkolás a legjelentősebb pszichoakusztikai hatás a veszteséges audiotömörítés szempontjából. A maszkolt komponensek detektálásával és elhagyásával jelentős kódolási nyereség érhető el.

2.4. Maszkolás az időtartományban

Egy audiojel nem pusztán a jel megszólalásával egyidejűleg, hanem az után is kifejti maszkoló hatást. Ezt a jelenséget előre-maszkolásnak (*forward masking*) nevezzük. Az előre-maszkolás időtartama jelentősen függ a maszkoló komponens frekvenciájától, azzal nagyjából fordítottan arányos. A hatás időtartama kisfrekvenciás jelek esetén elérheti akár a 120 ms-ot is, erőssége pedig az idővel exponenciálisan csökken. Mivel ez az időtartam az analízismodulokban használt időszelvények hosszának akár többszöröse is lehet, a hatást a legtöbb veszteséges tömörítőeljárás is kihasználja.

Érdekes megemlíteni a hátra-maszkolásnak (*backward masking*) nevezett érdekes jelenséget is, vagyis azt a hatást, amikor egy később megszólaló hangosabb hang képes maszkolni egy



3. ábra. Különböző amplitúdójú tonális (a) és atonális (b) komponensek maszkolási görbéje a frekvenciasávban. A tonális eset egy 1 kHz frekvenciájú harmonikus jel, míg az atonális eset egy ugyanilyen középfrekvenciájú keskenysávú zaj maszkolási görbéit mutatja.

korábbi, halkabb hangot. A jelenség magyarázata az, hogy az idegpályán gyorsabban terjed a nagyobb amplitúdójú inger, mint a kisebb amplitúdójú. A hátra-maszkolás időtartama a maszkoló jel amplitúdójától függően 2 ms és 5 ms közötti. Mivel ez az időtartam rövidebb, mint a legtöbb pszichoakusztikai analízismodulban használt szeletek időtartama, így ezt a hatást a veszteséges audiotömörítő eljárások többsége nem használja ki.

3. Pszichoakusztikai analízis

Ebben a szakaszban egy veszteséges audiotömörítéshez használható pszichoakusztikai analízismodul működését tekintjük át. Az itt bemutatott model az MPEG szabványban – ISO/IEC 11172-3 (MPEG-1, layer 1) – mint ajánlás szerepel, „1-es pszichoakusztikai model” néven. A modell részletesebb leírása megtalálható pl. az [1] könyvben. Emellett természetesen másfajta analízismodulok és számítási módok is léteznek, a mérés céljaira ez a modell az egyszerűsége miatt megfelelő.

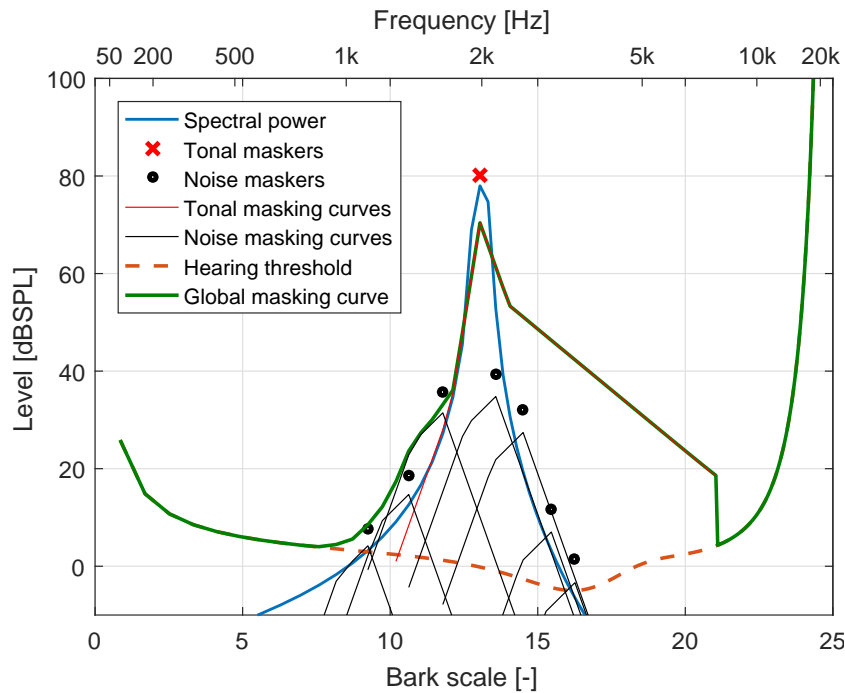
Az analízismodul által legfontosabb megállapítandó jellemző a 4. ábrán bemutatott frekvenciatartománybeli maszkolási görbe, mely az egyszerre megszólaló komponensek egyidejű maszkoló hatásait összesíti. Ez a frekvenciafüggő érték ugyanis az a szint, amit a maszkoló komponensek elnyomnak, tehát az ennél a küszöbszintnél kisebb szintek nem érzékelhetők. Ezt kihasználva csökkenthető az átvitt bitek száma, hiszen elegendő sávonként úgy megválasztani a bitszámot, hogy a kvantálásból adódó kvantálási zaj¹ az átvitt jelkomponensek még éppen maszkolják, azaz a kvantálási zaj a maszkolási küszöbszint alatt maradjon.

A 4. ábrán egy 2 kHz frekvenciájú harmonikus jeltől számított maszkolási görbe (*global masking curve*) látható, a számítás lépéseit is mutatva. A számítás lépései az alábbiak:

1. Teljesítményspektrum számítása

Az analízis első lépéseként a jel egy szakaszának (más szavakkal szeletének vagy ablakának) spektrális teljesítményét (*spectral power*) határozzuk meg. A spektrális szivárgás csökkentése érdekében a jelet egy kiválasztott ablakfüggvénnyel (pl. Hann-ablak) súlyozzuk. Mivel a frekvenciatartománybeli maszkolási hatások függenek a jel abszolút szintjétől, ezért a maszkolási jelenségek számításához a dimenziótlan felvételhez egy névleges

¹Helyesebben kvantálási torzítást, hiszen a kvantálásból adódó hibajel függ a kvantált jeltől.



4. ábra. Keskenysávú jel maszkolási görbéje

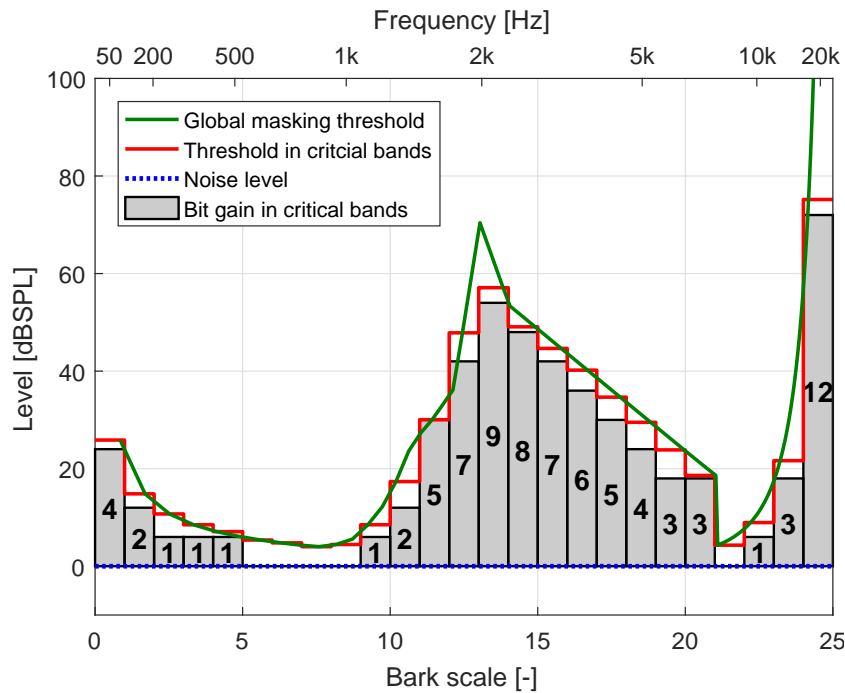
lejátszási szintet kell rendelnünk. Kódoláskor a lejátszási szint ismeretlen, így a névleges szintnek egy *worst case* szintnek kell lennie. Ezt a szintet úgy rögzítjük, hogy a 4 kHz sávközepű sávban, amelyben a fül érzékenysége maximális, egy 1 bit amplitúdójú harmonikus jel 0 dB SPL szintűnek feleljen meg egy 16 bit-en diszkrétizált minta esetén. Így egy teljes kivezérlésű jelhez a 90 dB SPL szint tartozik.

2. Maszkoló komponensek meghatározása

- A spektrális teljesítmény alapján először tonális maszkoló (*tonal masker*) komponenseket keresünk a jelben. A tonális komponenseket a spektrális teljesítmény lokális maximumainál feltételezzük, azzal a módosítással, hogy a maximumértékek megkezdésénél nem csak a szomszédos frekvenciavonalhoz tartozó értékeket, hanem egy tágabb környezetből való kiemelkedést veszünk figyelembe. A tonális komponenseket a 4. ábrán a piros keresztjek jelölik, az amplitúdókat a teljesítményspektrum csúcserképei és az ezzel szomszédos frekvenciavonalakon felvett értékek összegeként kapjuk.
- Minden kritikus sávban zaj maszkolókat (*noise maskers*) is megállapítunk. A zaj maszkolókat a kritikus sávok középfrekvenciájára vesszük fel, az amplitúdójukat pedig a sáv összteljesítményéből számítjuk, elhagyva a tonális komponensek teljesítményét, ha voltak ilyenek a sávban.
- Ezután a kiszámított maszkoló komponenseket rendszerezzük, melynek során egy 1 Bark szélességű ablakkal végighaladva a frekvenciasávban, az ablakon belüli legerősebb maszkoló komponenszt tartjuk csak meg, megjelölve, hogy az tonális vagy zaj jellegű maszkoló volt-e. Így minden kritikus sávban egy-egy maszkoló komponenszt kapunk.

3. Maszkolási görbe felvétele

- A maszkoló komponensekhez kiszámítjuk a maszkoló görbéket, minden egyes maszkoló komponenshez egy-egy egyedi maszkolási görbét (*individual masking curve*) fel-



5. ábra. Bitszám-hozzárendelés a zaj-maszk arány alapján

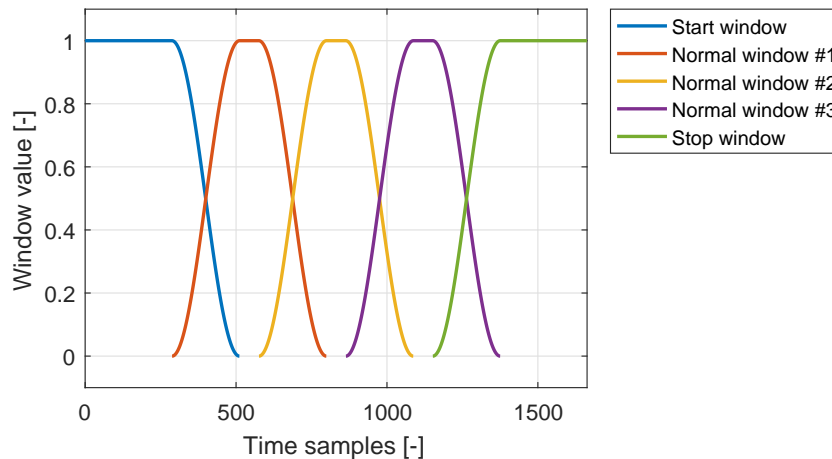
véve. Ezeket a maszkolási görbéket mutatják a 4. ábra vékony piros (tonális) és fekete (zaj) vonalai.

- (b) Az egyedi maszkológörbék és az emberi hallásküszöb (4. ábra szaggatot barna vonala) figyelembe vételével alakul ki a teljes vizsgált frekvenciatartományt lefedő globális maszkolási határ (*global masking threshold*), melyet a 4. ábra zöld vonala mutat.

4. Bitszám-hozzárendelés

A globális maszkolási határ alapján rendelhetünk bitszámokat az egyes kritikus sávokhoz, melynek menetét az 5. ábra szemlélteti. Egy adott frekvenciasávban a minták bitszámának csökkentése növeli az ebbe a sávba eső kvantálási zaj teljesítményét. A bitszámok mindaddig csökkenthetők, amíg az ilyen módon megnövekedett kvantálási zaj még a maszkológörbe alatt marad. Így sávonként a maszkolási határgörbe minimumának (5. ábra piros vonala, *threshold in critical bands*) és az eredeti jel zajsztijének (5. ábra szaggatot kék vonala, *noise level*) különbsége határozza meg, hogy sávonként mennyire emelhető meg a zajszt. Ezt az arányt zaj-maszk aránynak (*noise to mask ratio*, NMR) nevezzük. Ahogy az 5. ábrán is látható, 0 dB SPL zajszt esetén a zaj-maszk arány a sávonkénti maszkolási minimumszinttel egyezik meg. Ebből a maximális bitszám-csökkenés sávonként egyszerűen meghatározható, mivel egy bit elhagyása $20 \log_{10} 2 \approx 6$ dB-lel növeli a kvantálási zaj szintjét.

Egy hanganyag (pl. zenei mű vagy beszéd) spektruma természetesen időben változó. Ezért a frekvencia-analízist nem a teljes anyagon, hanem annak kis szeletein (*frame*, *slice*, *window*) végezzük el és a bitszám-hozzárendelést is szeletről szeletre számítjuk ki. Az egymást követő szeletek természetesen hatnak egymásra, például az időbeli maszkolás jelenségének következményeként, ezeket a hatásokat azonban az egyszerűsített modellünkben nem vesszük figyelembe.



6. ábra. Átlapoló ablakfüggvények

4. Transzformáció és újrakvantálás

Az előző szakaszban láthattuk, hogy a frekvenciatartománybeli vagy egyidejű maszkolás kihasználásával lehetőségünk van egy hangminta bitszámának csökkentésére, és így a hanganyag tömörítésére. Azonban az elemzés a frekvenciatartományban ad meg bitszámokat, míg a hangminta időbeli minták sorozata. Ezért, hogy a tömörítés megvalósítható legyen, a mintákat a frekvenciatartományba transzformáljuk és ezek bitszámát csökkentjük, vagyis a frekvenciatartománybeli mintákat kvantáljuk újra és tároljuk el a fájlban. A minták visszaállítása az időtartományba már a dekódoló egység feladata.

A transzformáció módszere többféle lehet, az MP3 például a diszkrét koszinusz transzformáció (*discrete cosine transform*, DCT) egyik változatát használja. De megvalósítható a transzformáció FFT-vel vagy wavelet transzformációval is. A DCT előnye az FFT-hez képest, hogy valós mintákból valós frekvenciatartománybeli mintákat állít elő, így a laborfoglalkozáson is ezt a transzformációt alkalmazzuk.

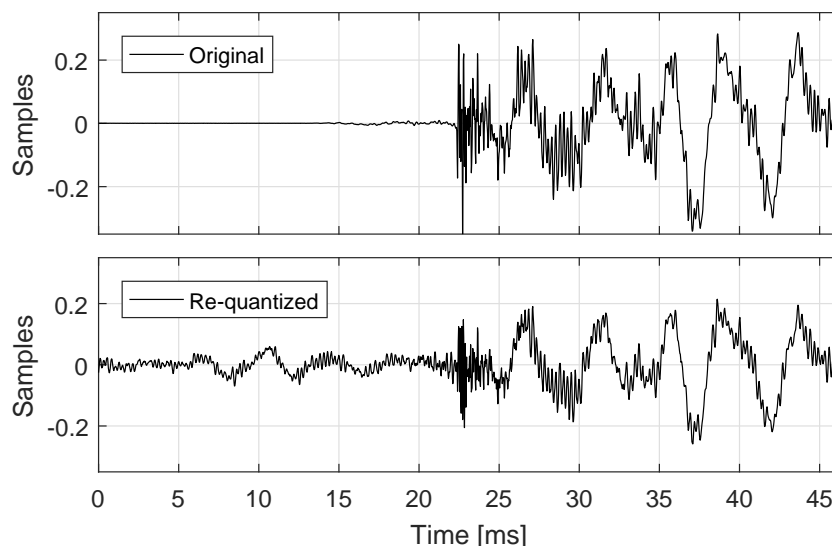
4.1. Az ablakozás szükségessége

Azzal, hogy a frekvenciatartományban egy sávban csökkentjük a bitszámot, az időtartományban a teljes szelvényben elterülő kvantálási zajt kapunk. Mivel ezek a zajok szeletről szeletre változnak, az újrakvantált és időtartományba visszatranszformált szeleteket egymás után téve a szeletek határán ugrásokat kapnánk, az összefűzött szeleteket meghallgatva pedig „pattogást” hallanánk a szelethatárokon. Ennek elkerülése végett az eredeti jelből átlapoló szeleteket szükséges venni, majd az átlapoló szeleteket kvantálni újra (ez természetesen többlet eltárolandó információt is jelent), végül pedig dekódoláskor a megfelelő súlyozó ablakfüggvények alkalmazásával sima átmenetet képezni.

Időben átlapoló ablakfüggvényeket mutat be a 6. ábra. Az ábrán minden ablak egységnyi szélességű, az átfedésnél pedig emelt koszinuszos karakterisztikájúak. Látható továbbá, hogy az ablakfüggvények összege minden mintára az 1 értéket adja. A normál ablakok mindkét szélükön fel- illetve lefutnak, a start és stop ablakok viszont az egyik szélükön nem, így utóbbiakat a felvétel elején és végén alkalmazhatjuk.

4.2. Változó ablakhosszak

A 7. ábra azt a jelenséget mutatja be, amikor egy gyors tranziens előtt az ablakban egyenletesen széteső kvantálási zaj a tranziens előtt is megjelenik. A hatást az okozza, hogy a pszichikus analízis a teljes ablak spektruma alapján számítja ki a bitszámokat, így, ha jelentős



7. ábra. Kvantálási zaj szétterülése a teljes ablakban (*pre-echo*)

amplitúdókülönbségek vannak az ablakon belül, akkor az eredetileg „halkabb” részben akár már hallható is lehet a szétterült kvantálási zaj. Mivel a kvantálási torzítás erős összefüggésben áll a kvantált jellel, így a tranziens egyfajta torzított „visszhangja” jelenik meg a tranziens előtt. Ezért az effektust *pre-echo*-nak (elő-visszhang) is nevezik. Gyors lecsengés esetén a probléma nem jelentős, mivel ekkor az időtartománybeli maszkolás segít elnyomni a „visszhangot”.

A jelenség elkerülése végett a pszichoakusztikai analízismodul feladata a gyors tranziensek megtalálása az audiojelben. Az ilyen tranzienseknél célszerű a kódoláshoz használt szeletek méretét lokálisan csökkenteni. Az MP3 például az eredetileg 576 mintából álló szeleteket három rövidebb, egyenként 192 minta méretű ablakra bontja fel a tranziens szakaszokban.

4.3. Kódolási nyereség becslése

Pusztán az újrakvantáláshoz használt bitszámok alapján nem állapítható meg a teljes kódolási nyereség. Ennek több oka is van. Egyrészt, az újrakvantáláshoz használt bitszámok időablakról időablakra változnak, így a minták értékei mellett az egyes minták bitszámát is el kell tárolnunk valamilyen módon, ami plusz információt jelent. Emiatt is érdemes a bitszámot nem mintáról mintára változtatni, hanem egy adott frekvenciasávban (pl. kritikus sávonként) állandónak változtatni, így az ábrázoláshoz használt bitszámokat elég sávonként tárolni. Másrészt, az újrakvantálás után a mintákon veszteségmentes entrópiakódolást (pl. Huffman-kódolást az MP3 kódoló esetén) is alkalmazunk, ami jelentősen csökkenti a végül tárolandó bitek számát. Emellett pedig egy adott formátumra kódolt fájl a mintákon kívül más információkat is kell, hogy tartalmazzon, melyek további tárolandó adatokként jelennek meg; ilyenek lehetnek például a szinkronizációs keretek, információk a formátumról, meta-adatok a hangfelvételtől, stb.

Ezek ellenére is érdemes tájékoztató jelleggel megbecsülni csak az újrakvantálásból adódó kódolási nyereséget. Ez legegyszerűbben a frekvenciatartománybeli minták bitszámainak összege és az eredeti bitszám és mintaszám szorzatának összevetésével kapjuk. Valamivel pontosabb becslést kapunk, ha még figyelembe vesszük azt, hogy a bitszámok csökkentése után sok frekvenciatartománybeli minta zérus értékű lesz. Ezeket az entrópiakódolás során érdemes egyetlen biten kódolni, akár a többi értékhez tartozó kódszavak hosszának egy bittel való növelése árán is.

4.4. Kiértékelés

A veszteséges tömörítőeljárások kapcsán érdekes nehézség a kódolt anyag kiértékelése. Mivel a modell pszichoakusztikai jelenségeket használ ki, az újrakvantált hanganyag minősége a szokásos objektív mértékekkel (pl. átlagos négyzetes hiba) nem becsülhető. Így a minőség szubjektív kiértékelésére meghallgatásos tesztek alkalmazni, az objektív kiértékelés pedig egy szabványosított pszichoakusztikai elemzőmodell (PEAQ model – *perceptual evaluation of audio quality* [2]) segítségével lehetséges.

5. A felkészülést segítő feladatok és kérdések

1. Mit jelent a pszichoakusztikai alapú veszteséges audio tömörítés kifejezés?
2. Rajzold fel egy általános veszteséges audiotömörítő eljárás blokkvázlatát!
3. Definiáld a kiritikus sáv szélesség és a Bark-skála fogalmát!
4. Mit nevezünk frekvenciatartománybeli vagy egyidejű maszkolásnak?
5. Mit értünk időtartománybeli maszkoláson?
6. Milyen jellemző(k) alapján tekintünk egy jelet tonálisnak vagy atonálisnak?
7. Miért fontos a tonális és atonális komponensek megkülönböztetése a veszteséges audiotömörítésben?
8. Mely részekből áll elő egy audiojel globális maszkolási görbéje?
9. Mit jelent az NMR (zaj-maszk arány – *Noise to Mask Ratio*), és hogyan használjuk fel ezt az arányt a kódolás bitszámainak kiszámításánál?
10. A veszteséges audiotömörítő eljárások frekvenciatartománybeli minták formájában tárolják a jelet. Ez miért szükséges?
11. Miért szükséges az újrakvantálás során az ablakozás? Mit okoz az ablakozás elhagyása?
12. Mit nevezünk *pre-echo* (elő-visszhang) effektusnak?
13. Milyen módszerrel értékelhető egy veszteségesen tömörített audiominta minősége?
14. Írj Matlab szkriptet, mely előállítja egy $T = 2$ s hosszúságú, $A = 1$ amplitúdójú, $f = 2000$ Hz frekvenciájú szinuszjel mintáit, $f_s = 48$ kHz mintavételi frekvencia mellett!

6. Feladatok

A méréshez előkészített fájlok az alábbi helyről tölthetők le:

http://last.hit.bme.hu/download/fospeclabor1/audio/audio_lab.zip

A mérés összes feladatát Matlab környezetben fogjuk megoldani, így a mérés sikeres elvégzéséhez a Matlab ismerete szükséges. A méréshez kiadott fájlok között a `tools` mappában találhatóak segédfüggvények, az `exercices` mappában pedig a feladatokhoz előkészített fájlok. A további feladatok elvégzéséhez ezeket a kódokat érdemes megismerni, értelmezni. Ebben segítenek a függvényekhez fűzött magyarázatok, melyeket a `help` vagy `doc` parancsokkal lehet megnézni. Az előkészített függvények a fent bemutatott pszichoakusztikai elemzést és a minták frekvenciatartománybeli transzformációjához illetve bitszámcsökkentéséhez használt eszközöket valósítják meg.

1. A környezet beállítása

Futtasd le az `ex_00_setup.m` Matlab szkriptet, mely beállítja a továbbiakban használt függvények elérési útjait. (Sikeres futtatás után ezt a lépést többször nem kell elvégezni.)

2. Kvantálás hatásának vizsgálata

- Írj egy Matlab szkriptet, amely előállít egy $f = 1$ kHz frekvenciájú, $A = 1/2$ amplitúdójú, $T = 2$ s hosszúságú harmonikus jelet $f_s = 44\,100$ Hz mintavételi frekvenciával. (A mintákat érdemes oszlopvektorként tárolni, a segédfüggvények ilyen formában várják a hangmintákat.) Hallgasd meg a jelet a Matlab `sound` függvényének segítségével. Ábrázold a jel spektrumát, a spektrum számításához használd az átlagolt spektrumot számító `spec_aver` függvényt.
- Szimuláld a 16 bit pontosságú számábrázolást a jel skálázásával. (A b bit-es ábrázolás esetén a kvantálás 2^{b-1} -nel való szorzást, majd kerekítést vagy csonkítást jelent.) Ehhez használhatod `requantize` függvényt. Hasonlítsd össze a kvantált minta spektrumát az eredeti spektrummal. Mit tapasztalsz?
- Kvantáld újra a jelet a `requantize` függvény segítségével 12, 8, 6 illetve 5 bit ábrázolási pontossággal. Hallgasd meg az újrakvantált mintákat a `sound` függvény segítségével. (Ügyelj rá, hogy a `sound` függvény normalizált, vagyis -1 és $+1$ közé eső mintákat vár, így a lejátszáshoz vissza kell skálázni a mintákat.) Vizsgáld meg a kis bitszámokon ábrázolt jelek spektrumát! Milyen jelenség figyelhető meg? Igaz-e, hogy a kvantálásból származó hiba zaj jellegű?

3. Frekvenciatartománybeli maszkolás

- Készíts az előző pontban előállított szkriptről egy másolatot, és használd a 16 bit pontosságú ábrázolást. Vedd a jelnek egy szeletét (ajánlott szeletméret: 512, 1024 vagy 2048 minta), és ábrázold a jel maszkolási görbét. Ehhez használd az előkészített `analyze_slice` függvényt. Figyeld meg a maszkolási görbén a tonális és atonális (zaj) komponenseket és az ezekhez tartozó egyedi maszkológörbét!
- Ábrázold a frekvenciatartományban a zaj-maszk arányt és az újrakvantáláskor elhagyható bitek számát! (Az ábrázolás az `analyze_slice` függvény megfelelő paraméterezésével könnyen megvalósítható.)
- A `requantize_windows` függvény segítségével végezd el a teljes (2 s hosszúságú) minta újrakvantálását. Hasonlítsd össze az eredeti és az újrakvantált jel spektrumát. Megfelel-e az eredmény a várakozásodnak?
- Keverj a harmonikus jelhez az eredeti frekvenciához közeli frekvenciájú (pl. 1 kHz-hez 1,1 kHz), kisebb amplitúdójú jelet. (Az amplitúdók megválasztásakor ügyelj arra, hogy a jel ne legyen túlvezérelve.) Vizsgáld meg, hogy a megválasztott amplitúdók mellett az új komponens a maszkolási görbe alá esik-e. Próbáld meg úgy beállítani a jel amplitúdóját, hogy az a maszkolási határhoz közel essen. Hallgasd meg az eredeti és újrakvantált jeleket és figyeld meg, hogy hallható-e a különbség. Érdemes emellett a jelek spektrumát is vizsgálni. Megegyezik-e a maszkolás számított határa a hallhatóság határával?

4. Hangfelvételek transzformációs kódolása

- Válassz egy szimpatikus minta-felvételt a `samples` könyvtárból. Töltsd be a felvételt az `audioread` függvény segítségével. (Ha a felvétel esetleg többcsatornás, a további analízishez csak az egyik csatornát hagyd meg.)
- Kvantáld újra a felvételt a `requantize_windows` függvény segítségével! Ábrázold egy grafikonon az eredeti és az újrakvantált jelet, illetve ezek különbségét! Hallgasd

meg az eredeti és az újrakvantált jeleket! Hallgasd meg a két jel különbségét is! Mit tapasztalsz a különbségjel meghallgatásakor?

- (c) Ábrázold az újrakvantáláshoz használt bitszámokat az idő és a frekvencia függvényében! (Használd a `pcolor` függvényt.) Számítsd ki a becsült kódolási nyereséget!
- (d) Módosítsd úgy az újrakvantáló függvényt, hogy minden frekvenciatartománybeli mintából 1-gyel vagy 2-vel több bitet hagyjon el az analízisben kiszámított maximumnál! Hallgasd meg az eredményt! Érzékelhető-e a hatás?
- (e) Próbáld ki a transzformációs újrakódolást különböző ablakhosszakkal, illetve az átlapoló minták számának változtatásával! Mi történik zéró átlapolás esetén?
- (f) Végezd el az előbbi vizsgálatok egy részét több hangmintán is! Egy társad segítségével végezz vaktesztet egy hangmintán: megállapítható-e meghallgatásos teszttel, hogy melyik az eredeti és melyik az újrakvantált minta?

5. Időtartománybeli hatások

- (a) Töltsd be a mintafájlok közül a `castanets.wav` hanganyagot! (Ehhez használd az `audioread` függvényt.) Keresd a jelben egy olyan részt, ahol hirtelen tranziens található, pl. a felvétel elején megszólaló gitárnál. Vágj ki egy szeletet a jelből, amelybe beleesik ez a tranziens, majd a `analyze_slice` és `requantize_slice` függvények segítségével kvantáld újra a szeletet. Ábrázold az eredeti és újrakvantált szeleteket! Mit tapasztalsz?
- (b) Vizsgáld meg a jelenséget hosszabb és rövidebb szeletekkel is!

Hivatkozások

- [1] A. Spanias, T. Painter, and V. Atti. *Audio signal processing and coding*. John Wiley & Sons, Inc., New York, 2007.
- [2] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten. PEAQ—The ITU standard for objective measurement of perceived audio quality. *Journal of the Audio Engineering Society*, 48(1–2):3–29, 2000.