

3 Hangjelek forráskódolása az MPEG-rendszerben

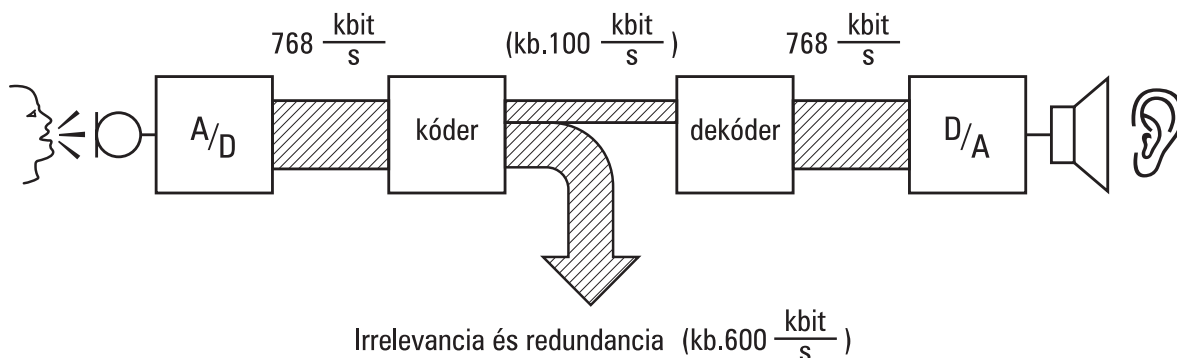
A digitális televízióhoz természetesen hozzátartozik a digitális kísérőhang átvitele is, amely jobb minőséget biztosít, mint az analóg tv FM-hangátviteli rendszere. Hogy a digitális tv sávszélessége elfogadható értékű legyen, megfelelően kialakított adatszőkkentési módszer alkalmazására van szükség a hangjel vonatkozásában is. A digitális kísérőhang átvitelének továbbá olyannak kell lennie, hogy a programszolgáltató szabadon választhasson a különféle adatsebességek és ezzel együtt a hangjel különféle minőségi fokozatai között. A hangcsatornának biztosítani kell két kísérőhangos, sztereo-, ill. surround-üzemmódokat is. Emellett mindezeket a követelményeket kompatibilisen kell teljesíteni; így minden vevő képes a hangjelet dekódolni, függetlenül a választott üzemmódtól és a kódolás adatsebességétől.

A felsorolt követelményeket az ISO/IEC 11172-3 (MPEG-Audio) és az ISO/IEC 13813-3 (MPEG-2 Audio) nemzetközi MPEG-szabványok rögzítik. A DVB-Projektben hozott döntések nyomán a digitális tv kísérőhangjának kódolása is ezen eljárások szerint valósul meg.

3.1 Az adatszőkkentés alapjai

A 3.1. ábra az adatszőkkentés információtviteli csatorna tömbvázlatát mutatja. A hírforrásból származó analóg jelet mintavételezzük, kvantáljuk, majd kódoljuk. Ezzel az analóg jel digitális jelfolyammá alakul. Ha a stúdiótechnikában szokásos 48 kHz-es mintavételezési frekvenciát alkalmazzuk 16 bites kvantálással, akkor egy monocsatornára 768 kbit/s-os adatsebesség adódik. Ahhoz, hogy ezt a jelet lehetőség szerint kis ráfordítással továbbíthassuk, az adatsebességet egy ún. forráskódolóban kisebb értékre kell csökkentenünk. A vevőoldali dekódoló a csökkentett adatsebességű jelet ismét visszaalakítja PCM-jellé, amit digitál-analóg átalakítás követ. Az analóg jel ezután már a hangszóróhoz vezethető.

A forráskódolás művészete abban van, hogy az adatszőkkentés úgy valósuljon meg, hogy az az információ érzékelésekor – hangkódolás esetén a hang hallgatása során – ne legyen észrevehető, ill. a hangminőségnek az adatsebesség csökkentése miatti csorbulása lehetőleg kevésbé legyen zavaró.



3.1. ábra
Adatszőkkentés elve hangátvitelnél

A híradástechnikai irodalomban e téren két alapvető lehetőséget különböztetnek meg: a redundancia-, ill. az irrelevanciacsökkentést.

Redundancia alatt bizonyos jelrészeknek az előre jósolhatóságát értjük. A redundanciacsökkentés a jelforrás statisztikai tulajdonságairól kíván meg ismereteket. A redundáns jelrészek a kódnél elhagyhatók, ill. a dekódnél a jel azokkal kiegészíthető. Így az eredeti jelalakot maradéktalanul visszkapjuk. Hangjelek esetében, ameddig nem mesterségesen előállított hangjelekről van szó („generic audio”), ilyen statisztikai feltételek alig vannak a jelben, úgyhogy tisztán redundanciaredukció segítségével nagy kompressziós arányok nem érhetők el.

Az irrelevanciacsökkentés ezzel szemben az információt befogadó személy véges érzékelőképességét használja ki az adatok csökkentésére. Hangjelek esetében a kóderben azokat a jelrészeket elhagyjuk, amelyeket az emberi fül korlátozott amplitúdó-, idő- és spektrális felbontóképessége folytán úgysem tud észlelni. Az irrelevanciacsökkentés, ellentétben a redundanciacsökkentéssel, megfordíthatatlan folyamat. Mégis, mivel ez a csökkentés az érzékelő számára nem észrevehető, szubjektíve az irrelevanciaredukció nem okoz romlást a jel minőségében. Mivel az irrelevanciaredukció az információt befogadó korlátozott érzékelőképességét használja ki, ezért a hangadatok illetően komprimálása az emberi hallásérzékelés törvényszerűségeinek alapos ismeretét követeli meg.

3.2 Pszichoakusztikai alapok

Először az állandó vagy hosszabb ideig tartó hangok érzékelésével foglalkozunk, amelyeknél az időbeni felépítés hatása az észrevehetőségben nem állapítható meg. Ezek azok a hangesemények (hangjelenségek), amelyek hossza nem rövidebb, mint 200 ms. A különálló szinuszhangok észlelését írja le az un. nyugalmi hallásküszöb és a hangfelület. Ha a hangesemény nem csak egy szinuszből áll, úgy előfordulhat, hogy az egyik hang a másikat elfedi. Hangok észlelése erős hanghatások után időt vesz igénybe, ez a jelenség adatesökkentésnél szintén felhasználható.

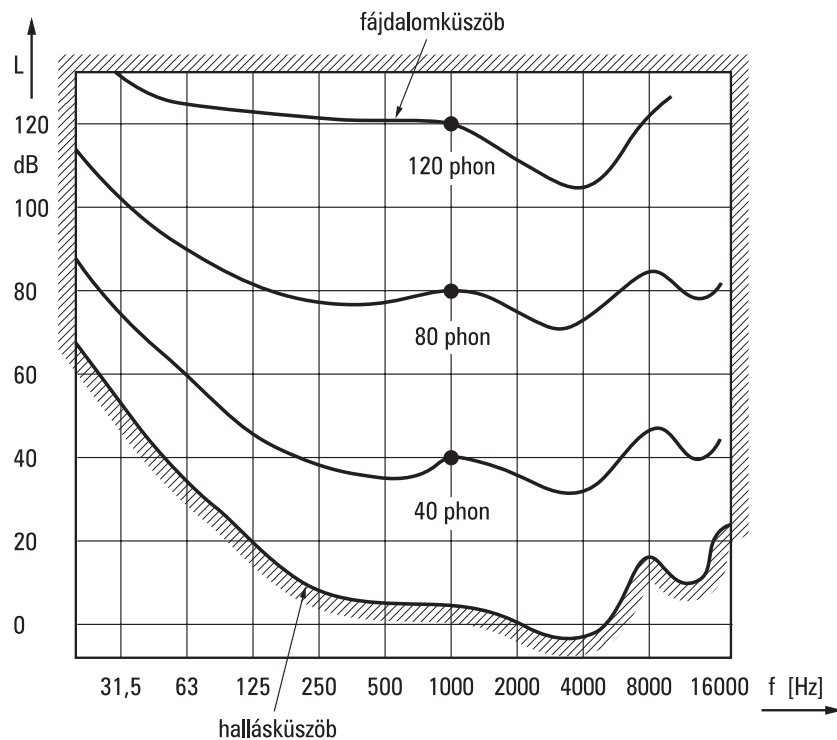
3.2.1 Nyugalmi hallásküszöb és hangfelület

Adott frekvenciájú szinuszhang nyugalmi hallásküszöbe alatt azt a hangnyomásszintet értjük, amelyiket még éppen érzékelni tudunk. A hallásküszöb frekvenciafüggő. Az L hangnyomásszintet a következő összefüggéssel definiáljuk:

$$L = 20 \lg \frac{p}{p_0}, \quad (3.1)$$

ahol p a hangnyomás, és p_0 a vonatkoztatási érték: $p_0 = 2 \cdot 10^{-5} \text{ N/m}^2 = 20 \text{ } \mu\text{Pa}$.

A nyugalmi küszöb meghatározásakor a vizsgálatban résztvevőknek úgy kellett egy szinuszos jelforrás hangszintjét beállítani, hogy a hangot már éppen ne érzékeljék. Ezeket a hangnyomásértékeket azután különböző frekvenciákra is meghatározták. Nagyszámú kísérleti személlyel elvégezve a mérést, és az eredményeket átlagolva, kapták meg az un. hallásküszöböt (3.2. ábra). Mivel a hangjelenség hosszú időn át változatlan, ezért szoktuk nyugalmi hallásküszöbnek is hívni.



3.2. ábra
Hallásgörbék

A hallásküszöbtől 120 dB távolságra helyezkedik el a fájdalomküszöb. A (nyugalmi) hallásküszöb és a fájdalomküszöb közötti területet hívjuk hallásfelületnek is. A 3.2. ábrán ezenkívül még az azonos hangerőérzet görbéi vannak bejelölve. Ezeken a görbéken látható, hogy a hangerőérzet, valamint a hallásküszöb is erősen frekvenciafüggő, ami már arra utal, hogy a hangjelek kódolásakor a spektrum különböző részeit más-más módon kell kezelni.

Azonban a természetben előforduló hangok legtöbbször nem egyedüli szinusztrezgés. Ezért meg kell vizsgálni, hogy milyen hatások jelentkeznek, ha egyszerre több hangból álló hangeseményt hallunk.

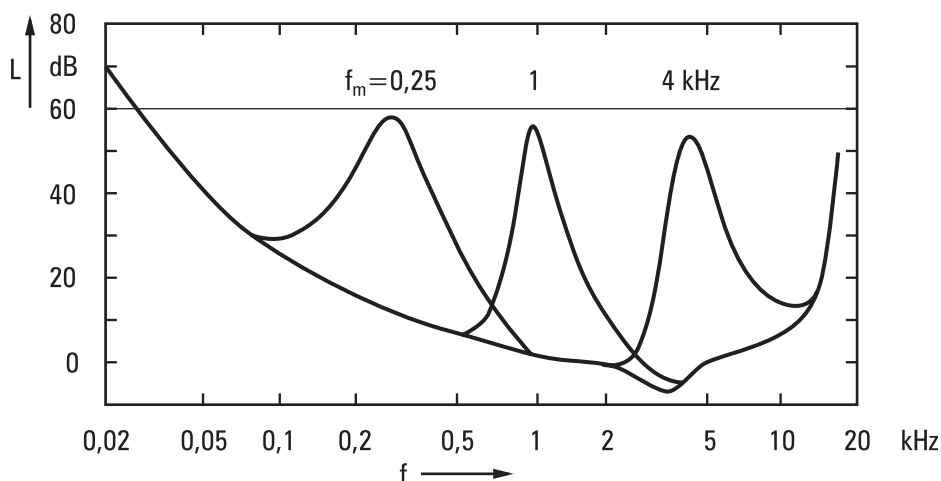
3.2.2 Elfedés

Az elfedés a mindennapi életből jól ismert jelenség. Így az a hangesemény, pl. zene, amely nyugodt csendes környezetben jól hallható, zavaró hanghatások mellett, pl. légalapács működése közben nem érzékelhető. Ezt a jelenséget hívjuk elfedésnek. Ahhoz, hogy a hasznos hang mégis érzékelhető legyen a hangszintjét jelentősen meg kell növelni. Az elfedési jelenség mennyiségi leírását szolgálja az ún. együttes hallásküszöb, amelyik azt adja meg, hogy adott zavaró hang mellett mekkora a vizsgáló hang szintje akkor, amikor az még éppen hallható.

3.2.2.1 Folytonos hangok okozta elfedés

Keskenysávú zaj jelenlétében a hallásküszöb szinuszos hangokra úgy határozható meg, hogy a vizsgálatban résztvevő személyeknek a szinuszos hang szintjét addig kell csökkenteniük, hogy az éppen ne legyen hallható. Ezt a vizsgálatot több személlyel 20 Hz és 20 KHz között minden frekvenciára elvégezték. A kiírtlagolt szintértékek adják azután az ún. együttes hallásküszöböt. A 3.3. ábra az együttes hallásküszöböt tünteti fel 60 dB

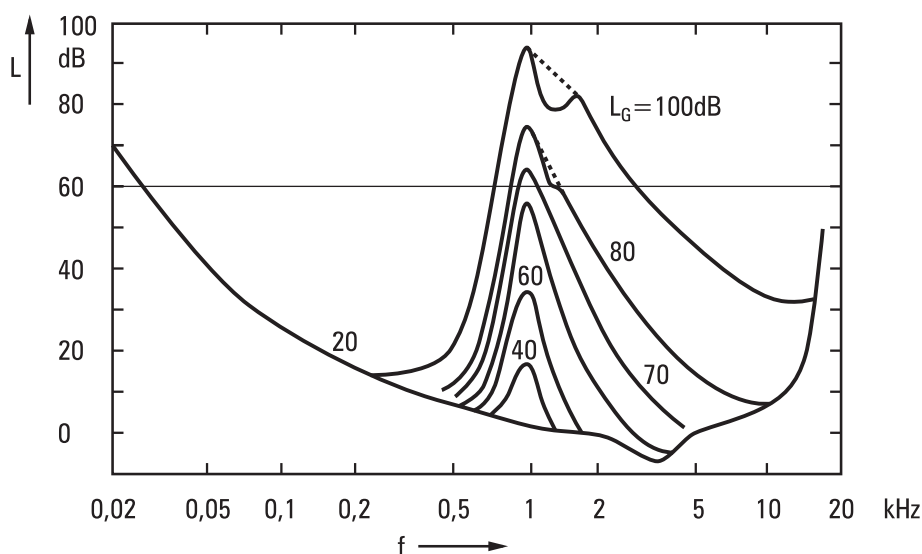
szintű keskenysávú maszkoló zajokkal felvéve, melyek közepes frekvenciája 250 Hz, 1 kHz és 4 kHz volt. A zajok sávszélessége pedig: 100 Hz, 160 Hz, ill. 700 Hz. Az összes hangot, amelyeknek a szintje az együttes hallásküszöb alatt van, a keskenysávú zaj elfedi, és így azok az emberi hallás számára érzékelhetetlenné válnak. Látható, hogy az együttes hallásküszöb görbéinek alakja erősen frekvenciafüggő, továbbá hogy a maximumok távolsága a 60 dB-es vonaltól növekvő frekvenciák irányában nő.



3.3. ábra

Különböző közepes frekvenciájú 60 dB-es szintű keskenysávú zajokhoz tartozó hallásküszöbgörbék

A maszkológörbe alakját nem csak a frekvencia határozza meg egyedül, hanem a szint is. A 3.4. ábra ezt érzékelteti. Ez esetben a maszkoló keskenysávú zaj közepes frekvenciája 1 kHz, sávszélessége 160 Hz, azonban a szintje, L_G más-más. A maszkoló frekvenciája és szintje mellett egy harmadik jellemzőnek is szerepe van, ez pedig a maszkoló tonalitása. Egy tonális (a hangnemben meghatározó) maszkolóhang a 3.4. ábrától eltérő görbéket eredményez.

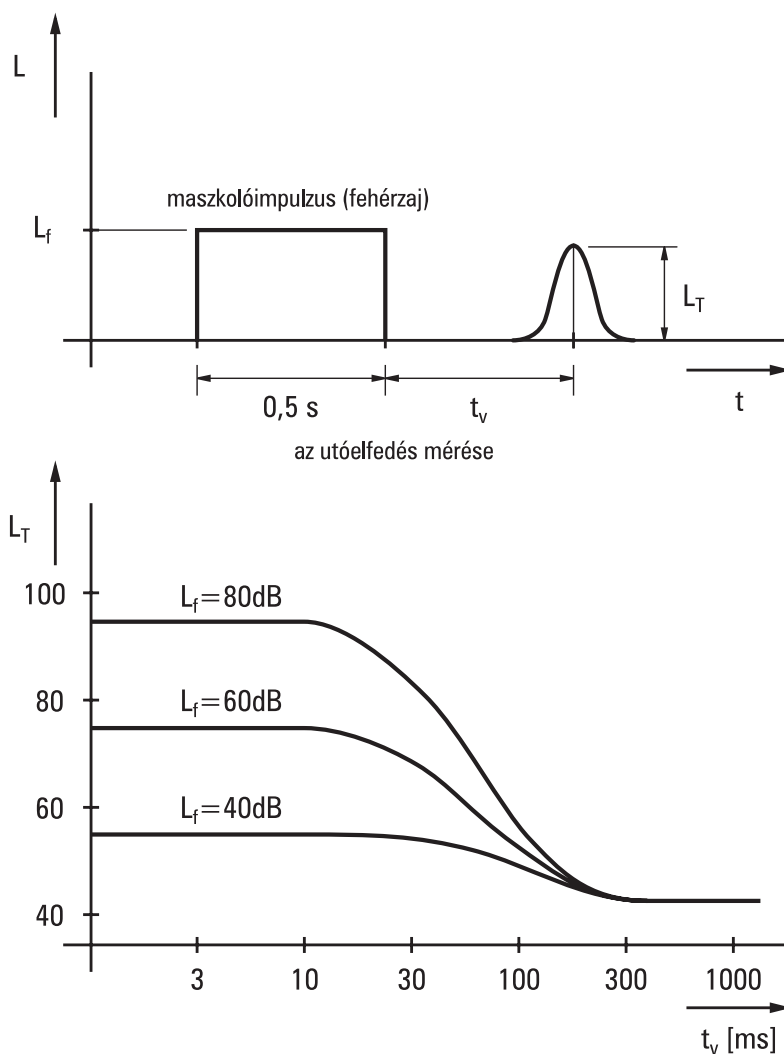


3.4. ábra

Különböző szintű keskenysávú maszkoló zajok hallásküszöbgörbéi

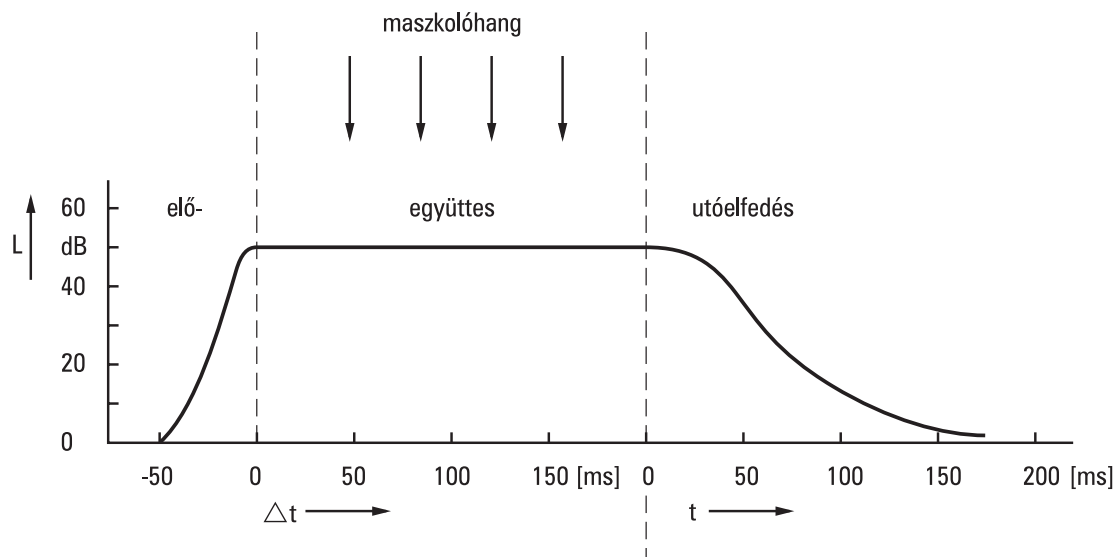
3.2.2.2 Időfüggő elfedési jelenségek

A nagyon rövid ideig tartó hanghatások is okoznak elfedést. A hatás vizsgálata ez esetben is kísérleti személyek bevonásával történt. A maszkolóimpulzus 0,5 s időtartamú, L_f szintű, Gauss-eloszlású, fehérzaj volt. Ezt a maszkolóimpulzust t_v idő után egy 20 μ s időtartamú Gauss-impulzus követte (3.5. ábra). A kísérleti személyeknek most az volt a feladata, hogy a Gauss-impulzus L_T szintjét úgy állítsák be, hogy az már éppen ne legyen hallható.



3.5. ábra
Különböző szintű maszkolóimpulzusok utóelfedésének
hatása a hallásküszöbre (L_T).

A maszkolózáj megszűnése után az elfedési hatás még mintegy 10 ms-ig ugyanúgy érvényesül, mint közvetlenül a megszűnés pillanatában. Csak ezután kezd az együttes hallásküszöb csökkenni, és kb. 200 ms múlva éri el a nyugalmi állapotnak megfelelő értéket. Érdekes, hogy az elfedés nem csak a maszkolóhang alatt vagy után lép fel, hanem akkor is, amikor a Gauss-impulzus azt megelőzi. Ezt a hatást az irodalomban előelfedésnek hívják, amely azonban csak akkor lép fel, ha Gauss-impulzus és a maszkolóimpulzus között 20 ms-nál nincs hosszabb idő. A 3.6. ábra az összes elfedési hatást együtt érzékelteti.



3.6. ábra
Elfedési hatások összefoglaló ábrázolása

A maszkolóhanggal időben egybeeső elfedési jelenségeket szimultán elfedésnek hívjuk. Az előelfedési hatás sokkal rövidebb az utóelfedésnél, 5 ms-mal a maszkolóhang előtt már csökkenni kezd, ahogyan az a 3.6. ábrán is látható.

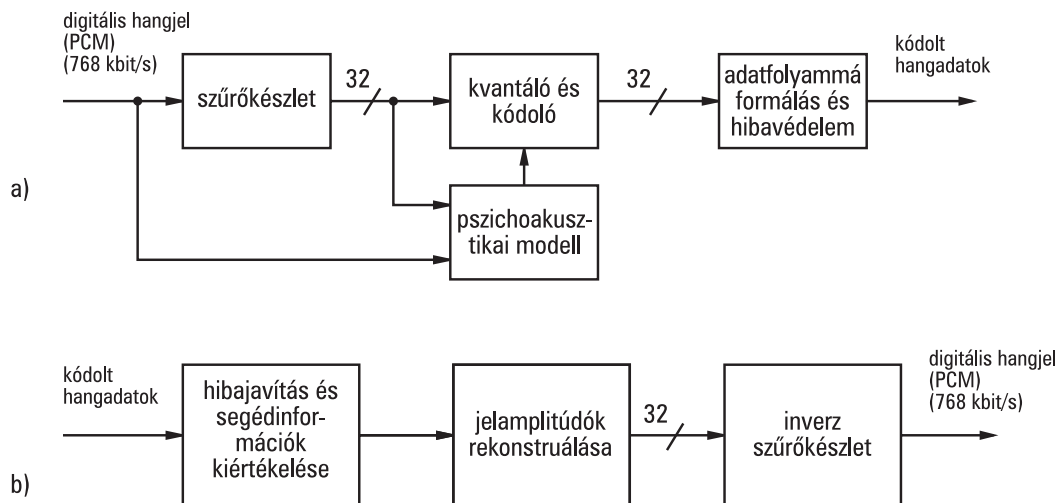
3.3 Hangjelek forráskódolása az elfedési jelenségek hasznosításával

A 3.3 szakaszban a hangjelek forráskódolásának az ISO 11172 és az ISO 13818 MPEG-szabványok szerinti megvalósítását ismerhetjük meg. Az MPEG-hangkódoló alapfelépítésének bemutatása után az úgynevezett 1-es rétegű (Layer 1) komprimálási mód részletezése következik, amit azzal a céllal fejlesztettek ki, hogy a kódér megvalósítása minél egyszerűbb legyen. Ezt követi a 2-es réteg (Layer 2) ismertetése, amely a Layer 1-nél összetettebb felépítésű azért, hogy a hangminőség romlása nélkül nagyobb adattömörítés legyen elérhető. A teljesség kedvéért röviden az MPEG-szabvány harmadik rétegével is foglalkozunk, amely a Layer 2-nél is nagyobb tömörítést tesz lehetővé. A Layer 3-at azonban a digitális tv-műsorszórásban nem szándékozzák alkalmazni, használata sokkal inkább az internetes hangátvitelben (MP-3) van elterjedőben. Az MPEG-szabvány kifejlesztésénél gondoltak arra is, hogy a hangjelek dekódolása sokkal kisebb ráfordítást igényeljen, mint a kódolás, ami a műsorszóró rendszereknél igen nagy jelentőségű, mert a végberendezéseknek gazdasági szempontok miatt egyszerűbbeknek kell lenni, hiszen ezekből nagyon sokat készítenek.

Ezen a helyen kell rámutatni arra, hogy a szabványban csak a dekódolás és az adatfolyam leírása rögzített, mert így biztosítható a különböző gyártók által előállított készülékek közötti kompatibilitás. A kódér leírása csupán informatív jellegű, amitől az MPEG-kóderek gyártói el is térhetnek.

3.3.1 Az MPEG-rendszerű kódolási eljárás alapfelépítése

Az MPEG-rendszerű hangkódér alapstruktúrája a 3.7. ábra szerinti.



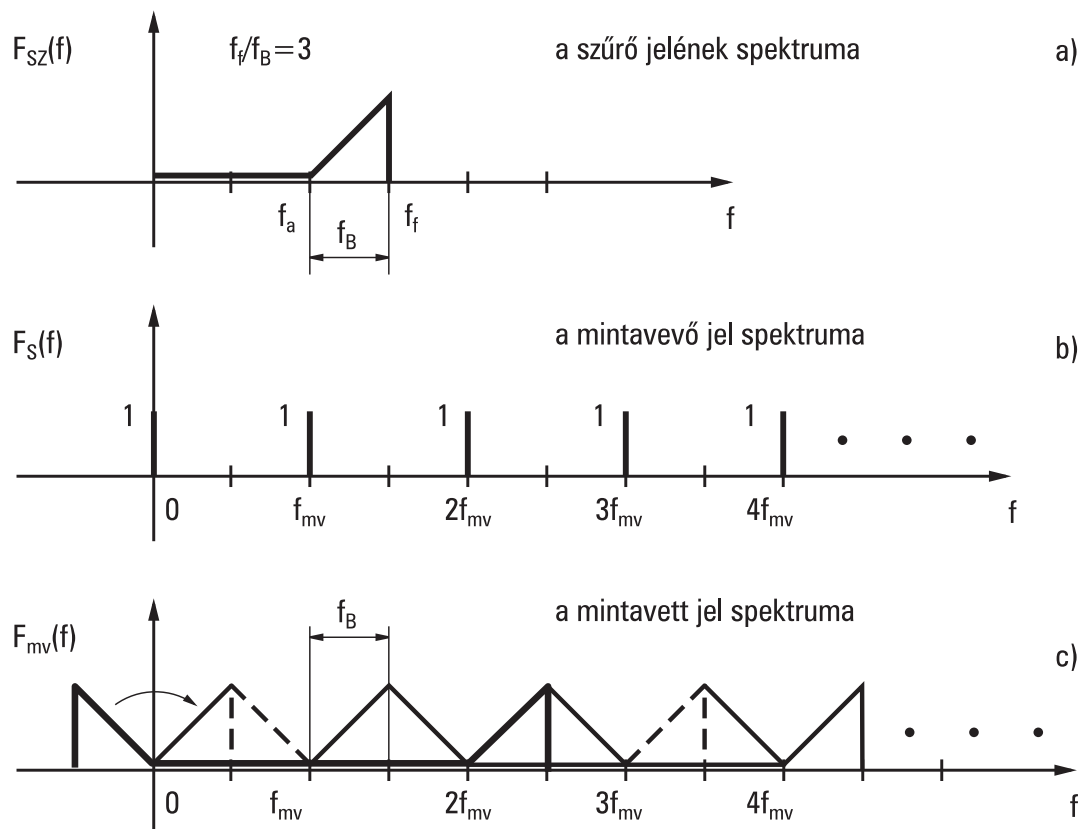
3.7. ábra
Adatcsökkentés hangkódolás elve az MPEG-szabványban;
a) kódoló, b) dekódoló

A kódolandó digitalizált PCM-hangjel először egy szűrőkészletbe jut, amely azt 32 egyforma sávszélességű sávra bontja. Egyidejűleg mind a 32 csatornában a jelet alulmintavételezik, azaz a mintavételezési frekvencia a digitalizáláshoz alkalmazott mintavételi frekvencia 32-ed része. A szűrőkészletet a kvantáló követi, amely mindegyik részsávban a jelet úgy kvantálja, hogy a lépcsők száma lehetőleg kevés legyen, de a kvantálási zaj még mindig a maszkolási küszöb alatt maradjon, hogy az emberi hallás számára észrevehetetlen legyen. A megengedhető kvantálás meghatározásakor természetesen a pszichoakusztikai törvényszerűségekből kell kiindulni. Ezért a kapcsolat megfelelő része, amely a 3.7. ábrán mint pszichoakusztikai modell van feltüntetve, kiértékeli a bemenőjelet, utána kiszámítja minden egyes részsávra az elfedési jelenség figyelembevételével a megengedhető kvantálást. Ezután a kódoló a kvantált mintákat és néhány más segédadatot, amelyek a dekóder számára a hangjel visszaállításához szükségesek, szabványos bitfolyammá alakítja, és hibavédelemmel látja el.

A dekóderben először az átviteli hibák korrekciója valósul meg, ezt követi a járulékos adatok és a hangjel mintáinak helyreállítása. A rekonstruált 32 részsáv egy inverz szűrőkészletbe jut, amely ezeket újra egy egységes frekvenciasávra rendezi. Ezt a fajta adattömörítési eljárást hívjuk részsávú (subband) kódolásnak.

Az egyes rétegek főbb jellemzőinek áttekintése előtt vizsgáljuk meg, hogy a 3.8. a) ábra szerinti $f_f - f_a = f_B$ sávszélességű jelből, $f_a \neq 0$ esetén, mekkora mintavételi frekvenciával kell mintákat venni ahhoz, hogy a mintákból az eredeti f_B sávszélességű jel (esetünkben a részsáv jele) visszanyerhető legyen. Természetesen a felvázolt spektrumot tekinthetnénk egy $0 \dots f_f$ frekvenciatartományú jel spektrumának is, ahol a $0 \dots f_a$ tartományban nincs spektrumösszetevő. Ebben a jelben a legnagyobb frekvencia f_f , és így az un. alapsávi, azaz a $0 \dots f_f$ frekvenciatartományú jelekre vonatkozó, jól ismert mintavételi tétel értelmében a mintavételi frekvenciának legalább $2f_f$ -nek kellene lennie. Látni fogjuk, hogy ennél kisebb mintavételi frekvencia is alkalmazható, amit már abból is gyanítani lehet, hogy a spektrum egy $0 \dots f_B$ frekvenciatartományú alapsávi jelnek a frekvenciatengely mentén f_a -val való eltolásával is létrehozható. A $0 \dots f_B$ frekvenciatartományú jelre pedig $f_{mv} \geq 2f_B$, ami kisebb $2f_f$ -nél.

Ezek után vizsgáljuk részletesebben annak a részsávú jelnek a mintavételezését, amelyre teljesül, hogy $f_f / f_B = m$ egész szám, és a sávhatárokon nincsenek diszkrét spektrumvonalak. Legyen $f_{mv} = 2f_B$.

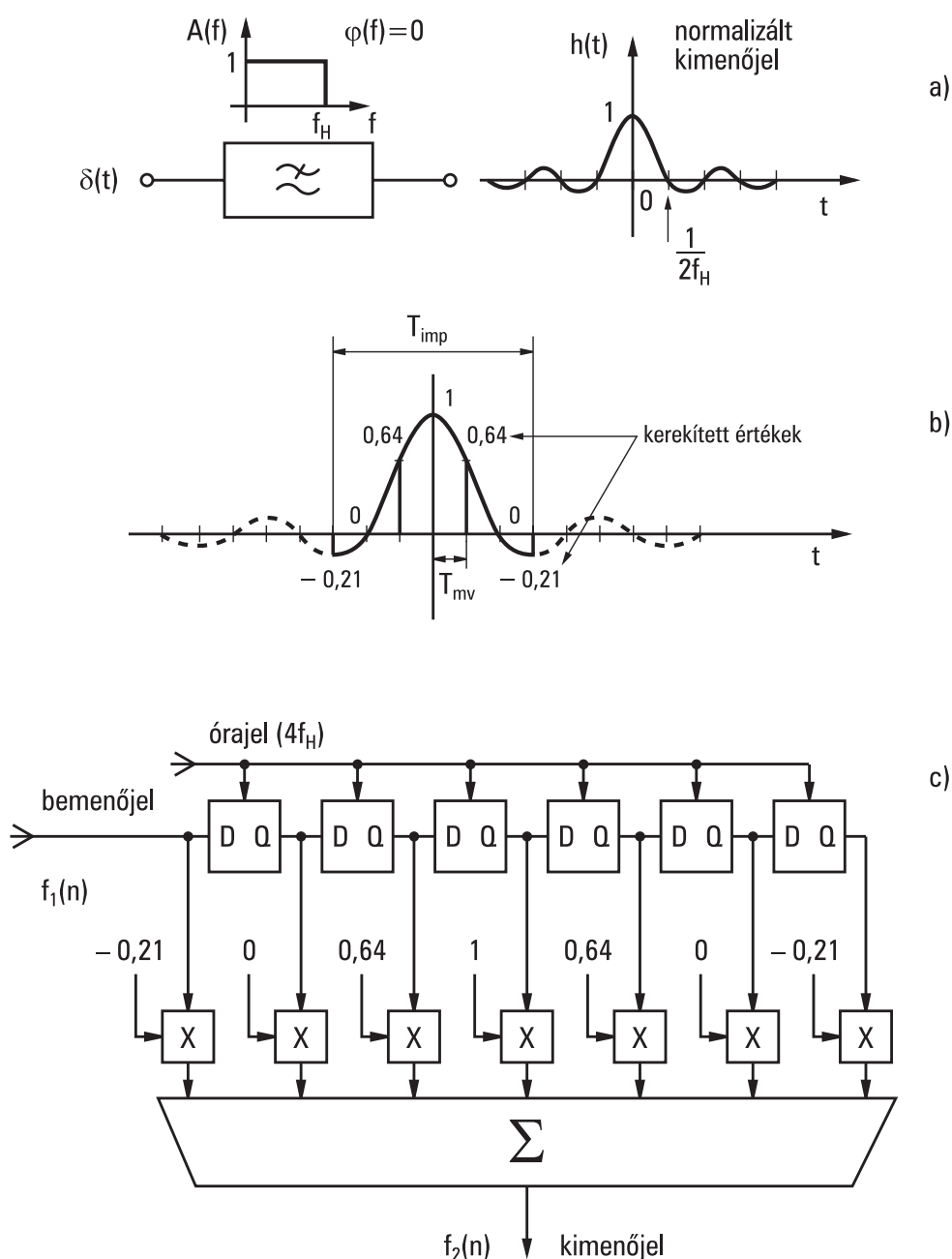


3.8. ábra
Sávszűrő jelének mintavételezése

A 3.8. b) ábrán a mintavevő jel spektruma látható, ahol mindegyik spektrumvonalnak az időtartományban egy szinuszos rezgés felel meg. Mivel a mintavételezés a szűrő jelének és a mintavevő jelnek a szorzását jelenti, ezért mindegyik szinuszjel szorozódik a szűrő jelével, azaz két oldalsávós amplitúdómoduláció jön létre, ahol az egyes vivőfrekvenciák a mintavételi frekvencia egészszámú többszörösei, a szűrő jele pedig a moduláló jel. A c) ábrán az f_{mv} -hez tartozó AM-DSB jelet vastag vonal jelöli, a $2f_{mv}$ -hez tartozó szaggatott. A c) ábrán jól látható, hogy egy $f_a \dots f_f$ frekvenciasávú szűrővel az eredeti spektrum kiszűrhető, azaz a szűrő kimenetén a jel visszanyerhető. Ha a szükséges feltételek nem teljesülnek, akkor a mintavett jel spektrumában átlapolódások jönnek létre, amely lehetetlenné teszi az eredeti spektrum kiszűrését, ill. a jel rekonstruálását.

Az MPEG-szabványban a részsávú kódolás valamennyi szűrője 750 Hz sávszélességű, így valamennyi részsávra teljesül az $f_f / f_B = m$ feltétel, a mintavételezés valamennyi sávban 1,5 kHz-es mintavételi frekvenciával megvalósítható. A szaknyelvben ezt hívják alulmintavételezésnek (subsampling).

Természetesen a részsávok szűrése és alulmintavételezése is digitálisan valósul meg. Az első részsáv jelét egy 0...750 Hz frekvenciasávú aluláteresztő szűrő kimenetén kapjuk meg. Jelöljük a 3.9. a) ábra szerinti ideális, egyelőre még analóg aluláteresztő szűrő felső határfrekvenciáját f_H -val.

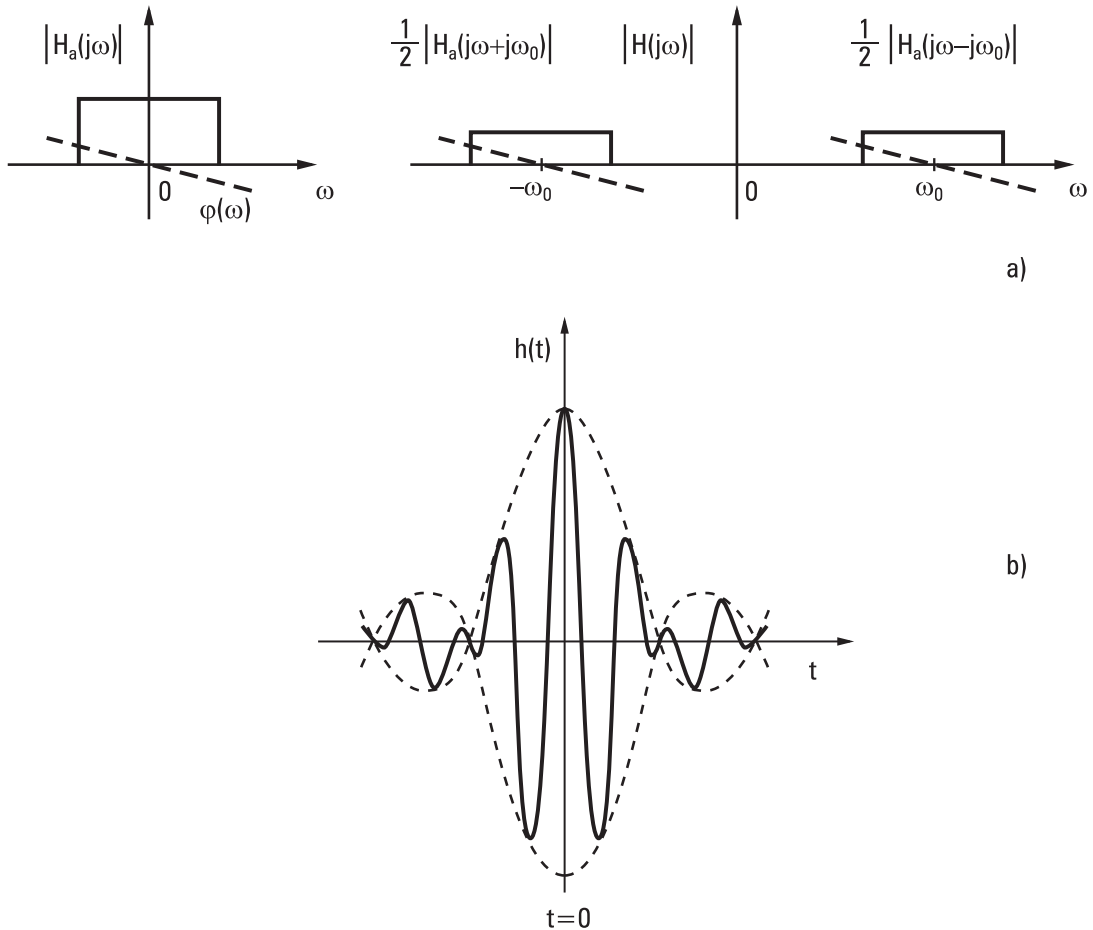


3.9. ábra
Digitális aluláteresztő szűrő súlyfüggvénye és struktúrávázlata

A szűrőt $\delta(t)$ -val gerjesztve, a kimenetén a $\sin x/x$ függvény alakú válaszjel, a szűrő súlyfüggvénye jelenik meg. A kimenőjel f_H -nál nagyobb frekvenciájú komponenseket nem tartalmaz, ezért ha a mintavételi frekvencia $2f_H$ -nál nagyobb, akkor a mintavételi tétel értelmében a jel a mintákból rekonstruálható. Legyen $f_{mv} = 4f_H$. Ebben az esetben a mintavételezett kimenőjel a 3.9. b) ábra szerinti. Annak érdekében, hogy ne kelljen végtelen sok mintát figyelembe venni, a kis értékűeket hagyjuk el. Ez a jel kis torzulását okozza csak, hatásában olyan, mintha az amplitúdókarakterisztikát változtattuk volna meg kismértékben, hogy a szűrő válaszjele éppen a csonkított impulzus legyen. Az a) és a b) ábrán is látjuk, hogy a szűrő nem kauzális, mert a kimenőjel már létezik, amikor a bemeneten a gerjesztés még 0. A gyakorlati megvalósításnál ezért be kell vezetni egy késleltetési időt, amely fele a megcsonkított $\sin x/x$ impulzus T_{imp} idejének. Így a már kauzális válaszfüggvényt a 3.9. c) ábra szerinti elrendezés szolgáltatja. Az MPEG-

szabvány szűrőinél a 3.9. c) ábra szűrőjével ellentétben az úgynevezett szűrőegyütthatók száma nem 7, hanem 512, ami biztosítja az amplitúdókarakterisztika oldaléleinek nagy meredekségét, ill. a jó sávészérválasztást.

A részsávú kódolás sávszűrőinek átviteli karakterisztikája származtatható az un. prototípus- vagy vonatkozási aluláteresztő átviteli karakterisztikájából a 3.10. a) ábra szerint a frekvenciatengelyen való eltolással.



3.10. ábra
Sávszűrő átviteli karakterisztikája és súlyfüggvénye

Így a sávszűrő átviteli karakterisztikája:

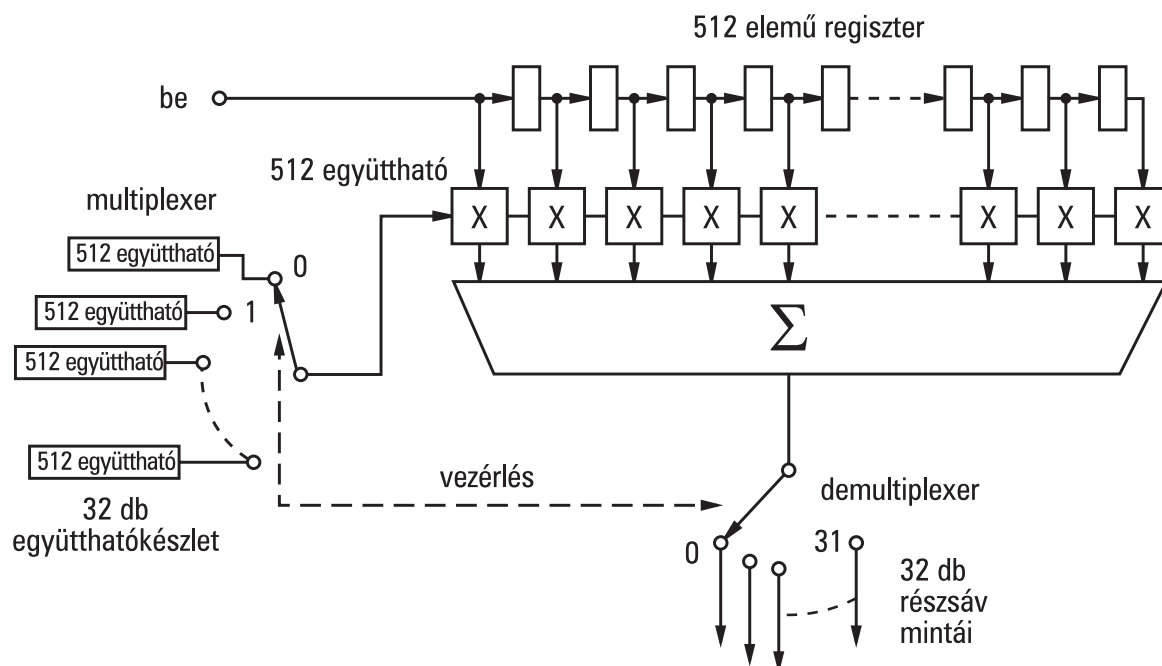
$$H(j\omega) = \frac{1}{2}H_a(j\omega + j\omega_0) + \frac{1}{2}H_a(j\omega - j\omega_0), \quad (3.2)$$

ahol $H_a(j\omega)$ a prototípus-aluláteresztő átviteli karakterisztikája, amelynek sávészélessége csak 375 Hz. Inverz Fourier-transzformációval megkapjuk a szűrő súlyfüggvényét. A csillapítási tétel alkalmazásával:

$$h(t) = \frac{1}{2}h_a(t)e^{-j\omega_0 t} + \frac{1}{2}h_a(t)e^{j\omega_0 t} = h_a(t)\cos\omega_0 t, \quad (3.3)$$

azaz a sávszűrő súlyfüggvénye a prototípus-aluláteresztőszűrő $h_a(t)$ súlyfüggvényével amplitúdóban modulált szinuszrezgés (3.10. b) ábra).

Miután ismerjük a súlyfüggvényeket, amelyek csak ω_0 -ban különböznek, az aluláteresztő mintájára meghatározhatók a szűrőegyütthatók.

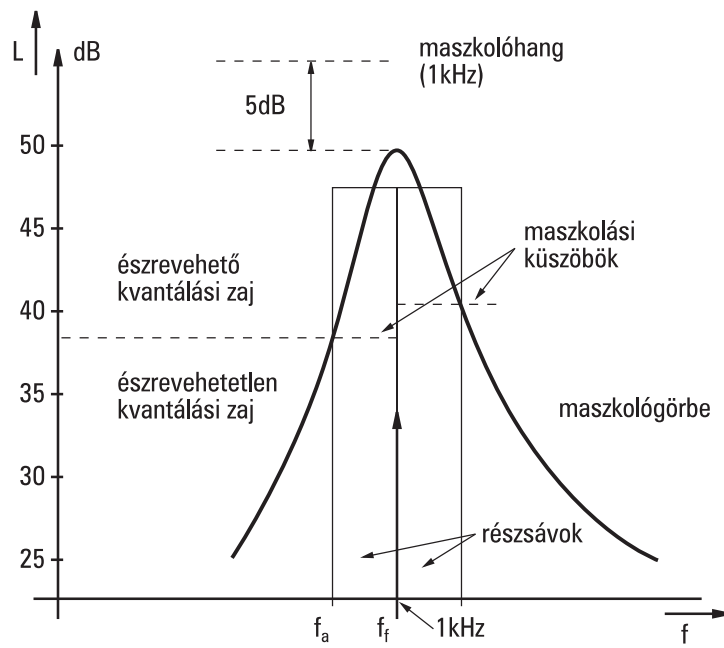


3.11. ábra
32 részsávós polifázisú szűrő felépítése

A multiplexerrel és demultiplexerrel kiegészített szűrőelrendezés az un. polifázisú szűrő (3.11. ábra), amely természetesen az első, az aluláteresztő részsáv jelét is meghatározza.

A szűrő tehát 32 fázisban számítja ki a részsávok mintáit. Az első 512 szűrőegyüttható az első részsáv szűrőjét realizálja. Az 512 szorzás eredményének összege adja az első részsáv első mintáját. A 2. részsáv együtthatóit megkapjuk, ha az un. prototípusszűrő együtthatóit megszorozzuk $\cos \omega_{02} nT_{mv}$ -vel. A 2. fázisban ezek szorozódnak meg az 512 mintával, hogy összegezés után a 2. részsáv első mintáját adják. Valamennyi részsáv első mintájának hasonló kiszámítása után a regiszter tartalma 32 új mintával eltolódik, hogy kezdődhessen a részsávok második mintájának kiszámítása. Tulajdonképpen egy 32 egyenletből álló lineáris egyenletrendszerbe helyettesítettük be a PCM-mintáknak megfelelő független változókat, hogy meghatározzuk a szűrők jeleit. A dekóderben a feladat fordított: ott a szűrők jeléből kell, az egyenletrendszert megoldva, a PCM-mintákat kiszámítani. Mivel az ismeretlenek száma 32, megegyezik az egyenletek számával, feladat elvégezhető. A műveleti blokk neve inverz szűrőkészlet.

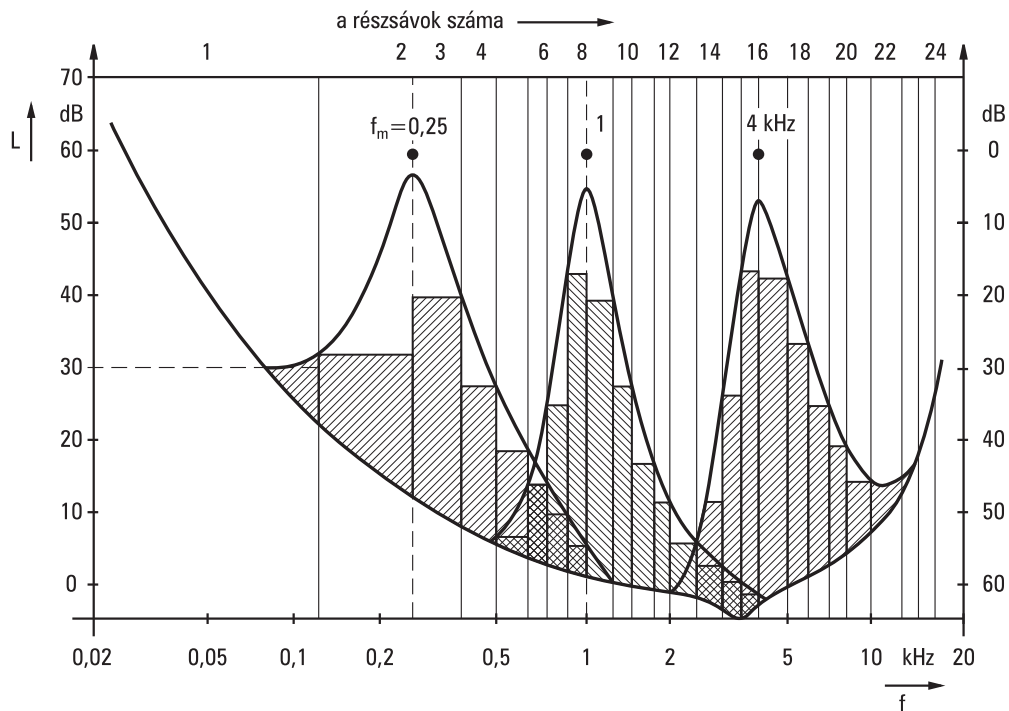
Ezek után vizsgáljuk a kódolás további részleteit. A részsávú kódolással elérhető kompressziós arány lényegében az egyes részsávokban alkalmazott kvantálási lépcső függvénye. Ennek meghatározása a következők szerint történik. Tételezzük fel, hogy a maszkoló egy 1 kHz-es szinuszos hang (3.12. ábra). A legkritikusabb helyzet akkor áll elő, ha ez a hang a részsáv felső sávhatárán van, mivel a maszkológörbe a kisebb frekvenciák felől meredekebben változik. Ez esetben a legkisebb az adott részsávban az elfedési hatás. A maszkológörbe és az alsó sávhatár metszéspontja határozza meg azt a szintet, amely alatt ebben a részsávban az összes hang észrevehetetlen, úgyhogy ezeket a dekóderoldalon nem szükséges visszaadni,



3.12. ábra
Maszkolási küszöb meghatározása

ill. a kvantálás ebben a részsávban olyan durva lehet, hogy a kvantálási zaj szintje ezt az értéket (szaggatott vonal) elérheti.

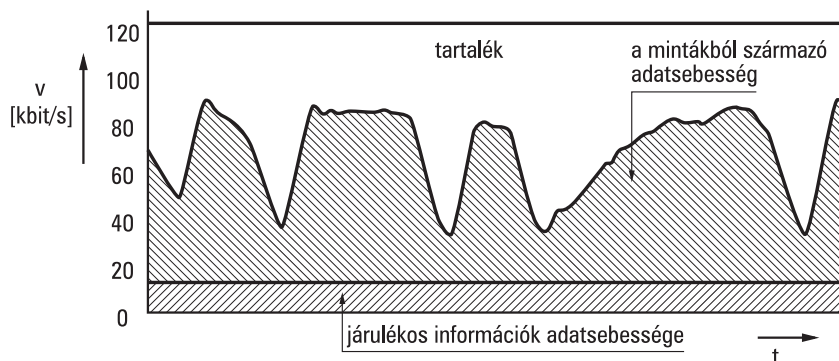
Többnyire azonban nem csak egyetlen maszkolóhang létezik, hiszen pl. a beszéd vagy zene sokkal összetettebb hangjelenség. A 3.13. ábra három szinuszos hanghoz tartozó három maszkolót tüntet fel.



3.13. ábra
Amplitúdóspektrum az elfedési görbékkel és a megengedhető kvantálási zajjal

Vannak olyan részsávok, amelyekben két maszkolóhang is elfedést okozhat.

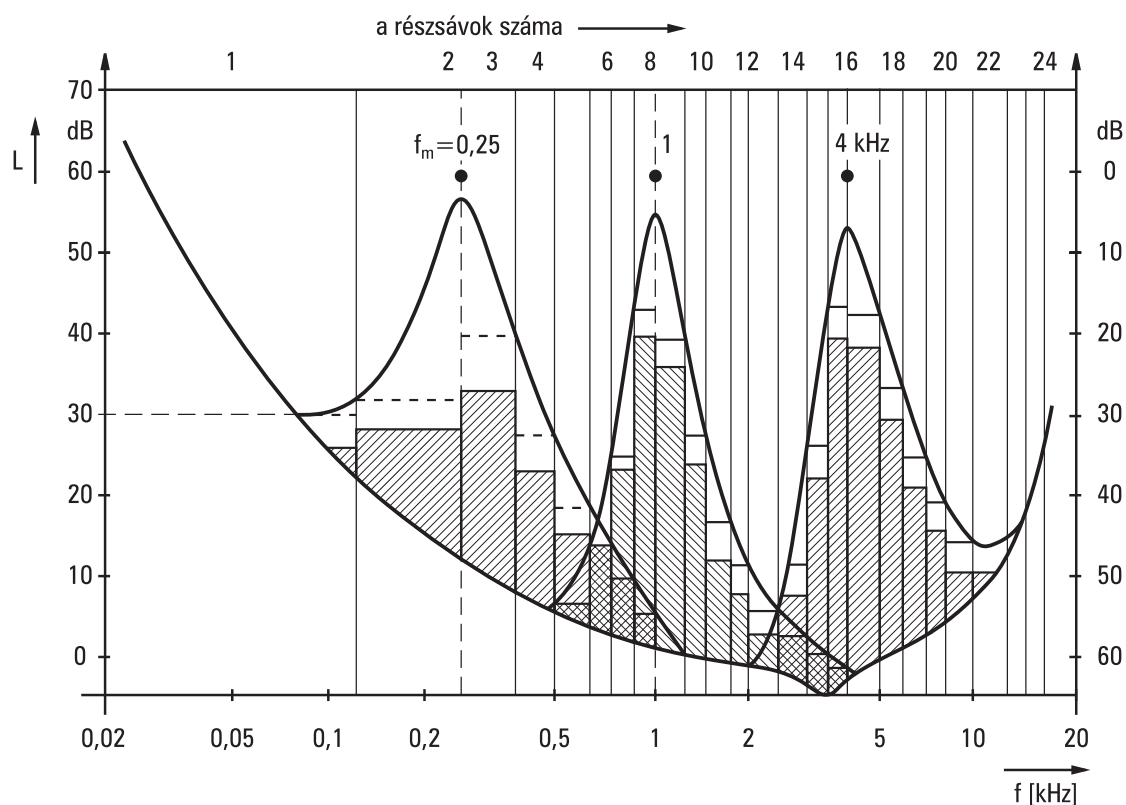
A 3.13. ábrán jól látható, hogy a maszkolási küszöbök értéke nagymértékben függ a elfedést okozó jel amplitúdójától és spektrális megoszlásától. A megengedhető kvantálási zaj azonban közvetett módon meghatározza a jel átviteléhez szükséges adatsebességet, ezért az az elfedés mértékének függvényében ingadozni fog (3.14. ábra).



3.14. ábra
Dinamikus adatsebességtartalék

A 3.14. ábrán v a szükséges adatsebességet jelöli. A járulékos információk a részsávok rekonstruálását és a dekóder vezérlését biztosítják. A járulékos információk adatsebessége közel állandó, ellentétben a kvantált mintavételi értékek átviteléhez szükséges adatsebességgel. Mivel a legtöbb átviteli csatornát állandó adatsebességre tervezik, időfüggő adatsebesség esetén a kapacitás egy része kihasználatlan maradna. Ezért a kvantálási szintek számát állandóan olyan mértékben meg kell növelni, hogy a csatorna kapacitása teljes mértékben ki legyen használva. Ezzel elérhető, hogy a 3.13. ábra szerinti maximális értékű kvantálási zaj alá kerülhet a pillanatnyi kvantálási zaj.

Ha a jelben a kvantálási zaj eléri a maszkolási görbét, úgy mint a 3.13. ábrán, akkor ezzel a jellel már semmilyen további jelfeldolgozás nem végezhető el anélkül, hogy a kvantálási zaj hallhatóvá ne válna. Ez már egy egyszerű korrektoráramkörre is érvényes. Ha pl. 5 kHz-nél a frekvenciamenetben kiemelést végeznénk, akkor a kvantálási zaj túllépné a maszkolási görbét. Ugyanez érvényes akkor is, ha az egyszer már kódolt jelet újra kódolnánk. Hogy a jel mégis korlátozottan ismét feldolgozható legyen, a maszkolási görbe és a kvantálási zaj között a 3.15. ábrán látható módon bizonyos távolságot kell hagyni.



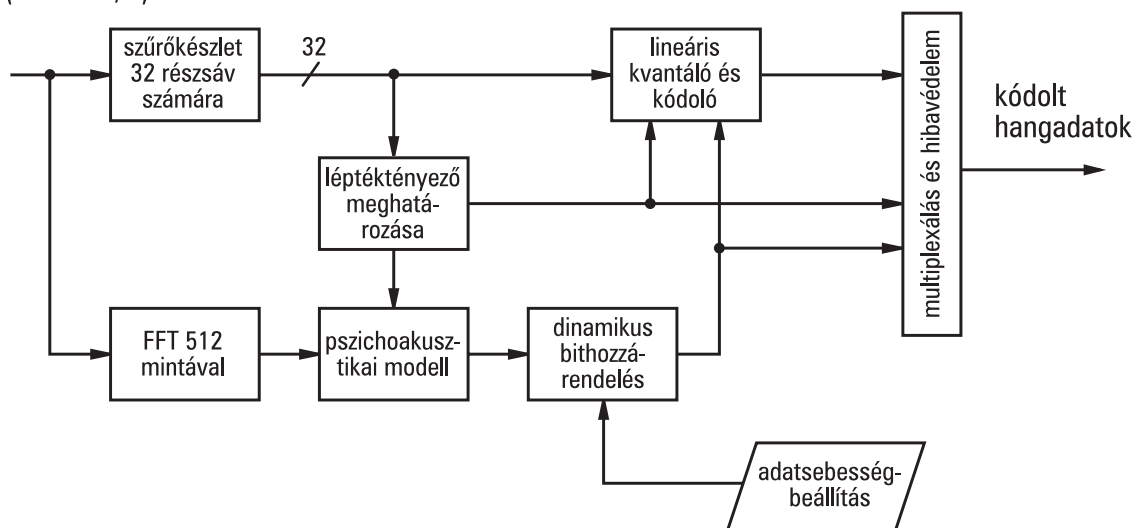
3.15. ábra
Adatsebességtartalék kihasználása a maszkolási küszöb és a kvantálási zaj
közötti távolság maximalizálására

Egy iteratív folyamat során a maszkolási görbe és a kvantálási zaj közötti távolság addig növekszik, amíg a rendelkezésre álló csatorna maximális adatsebessége teljesen ki nem lesz használva. Következésképpen a jelen végezhető utófeldolgozás lehetősége annál nagyobb, minél nagyobb az adatsebesség.

3.3.2 1-es rétegű kódolás

Az alapelvek után vizsgáljuk meg részletesebben az MPEG-kódolót és -dekódolót. A 3.16. ábra az MPEG 1-es rétegű hangkódolás tömbvázlatát tünteti fel. A szűrőkészlet a jelet 32 egyenlő sávszélességű részsávra osztja. A részsávokban a jel mintavételezése is megvalósul 1,5 kHz-es mintavételi frekvenciával.

digitális hangjel (PCM)
(768 kbit/s)



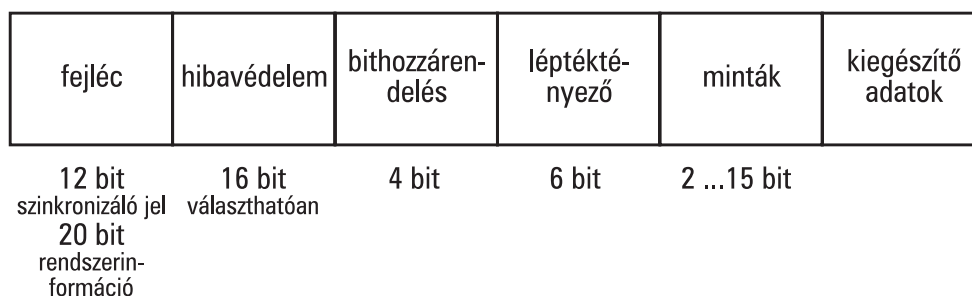
3.16. ábra
MPEG-szabványú 1-es rétegű kódoló tömbvázlata monojelre

Ezután a kódér minden részsávban folyamatosan 12 mintából álló blokkokat képez, amelyekben a legnagyobb minta a léptéktényező. Ezeket 6 bites felbontással szintén kvantálja, ill. kódolja, ami által 2^6 , azaz 64 hangerőfokozat különböztethető meg. Minden hangerőfokozat 2 dB-es szintnek felel meg, így az eredő dinamikatartomány kb. 120 dB. A „lineáris kvantáló és kódoló” elvégzi a 12 mintának a léptéktényezővel való osztását, az un. normálást, a minták lineáris kvantálását és kódolását. A „pszichoakusztikai modell” meghatározza a maszkolási küszöböt és ezáltal a szükséges kvantálási lépcsők számát. Mind a 12 minta egyformán kvantálódik, ami megengedhető, mivel a maszkolás hatása az elő- és utóelfedés miatt az egész, a mintavételi frekvenciától függően 8-tól 12 ms-ig tartó blokkra kiterjed. (A blokk időtartama 48 kHz-es mintavételi frekvenciánál és részsávonként 32-szeres alulmintavételezésnél: $12 \cdot 32 / (48 \cdot 10^3) = 8$ ms.)

A részsávokra történő felbontással párhuzamosan megvalósul a jelnek 512 mintavételi értékkel számított Fourier-transzformációja is, hogy a jel spektruma még nagyobb felbontással álljon rendelkezésre az alacsony frekvenciák tartományában, ahol a részsávú felbontás alapján végzett analízis nem ad optimális eredményt. A spektrum vizsgálata kiterjed a helyi maximumok meghatározására és annak eldöntésére, hogy ezek a jelnek tonális vagy atonális komponensei. Erre azért van szükség, mert egy adott jelnél a tonalitás – ahogyan erről már szó volt – erőteljesen befolyásolja a maszkolási görbe alakját. Tehát a részsávok maximális hangszintjéből meghatározhatók a maszkolási görbék, ill. kvantálási lépcsők. Az előre megadott adatsebesség és a maszkolási görbe alapján a kvantálási szintek száma annyi lesz, hogy a távolság a maszkolási görbe és a kvantálási zaj között maximális legyen.

Az átviendő multiplex jel a részsávok kvantált mintáinak kódjai mellett tartalmazza a dekódoláshoz szükséges információkat, azaz a léptéktényezőt és a dinamikus bitkiosztást, amely minden mintához megadja a felhasznált bitek számát. Mindezeket az adatokat a fejléc előzi meg, amely a dekóder vezérléséhez szükséges információkat tartalmazza. A 3.17. ábrán az adatfolyam felépítését láthatjuk az 1-es rétegű MPEG-kódér kimenetén, monojel kódolásakor.

adatkeret 384 PCM-mintával (48 kHz mintavételi frekvenciánál 8 ms)



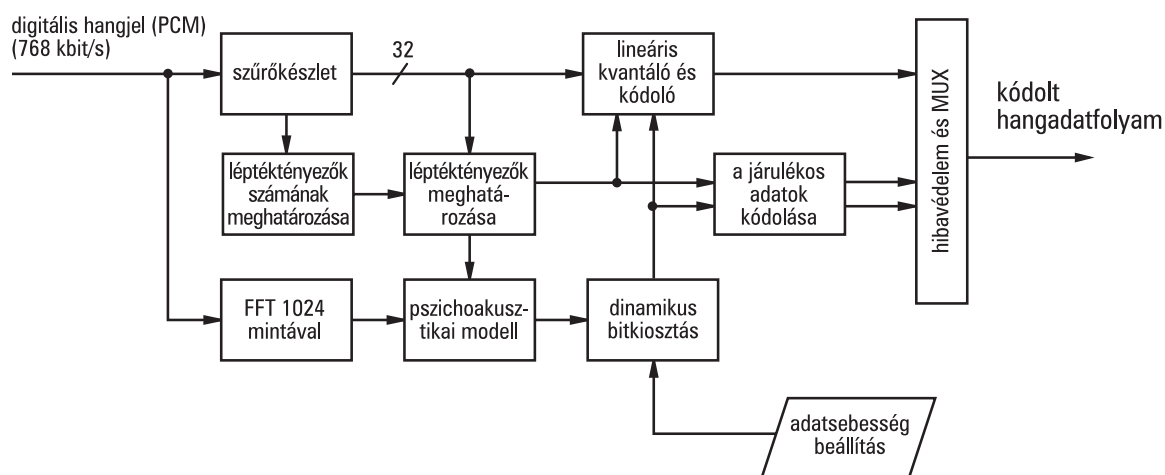
3.17. ábra

1-es rétegű adatfolyam egy keretének felépítése

A kiegészítő adatok az adatkeret végén lehetővé teszik pl. „surround-sound” (térbeli hangzás) megvalósítását.

3.3.3 2-es rétegű kódolás

Az 1-es rétegű kódolással ellentétben a 2-es rétegű kódolónál (3.18. ábra) a blokkok 36 PCM-mintából állnak, ami 48 kHz-es mintavételi frekvenciánál a hangjel 24 ms-os szakaszának felel meg. Ebben az esetben előfordulhat, hogy egyetlen léptéktényező nem elég részsávonként, mivel az elfedés hatása maximálisan csak 20 ms-ig tart.



3.18. ábra

MPEG-szabvány 2-es rétegű kódolójának tömbvázlata monojelre

Ha a jelben sok változás van, részsávonként és blokkonként kettő vagy három léptéktényezőre is szükség lehet. Ha ellenben kevés a változás, egy is elegendő.

A kódolóban külön egység vizsgálja, hogy hány léptéktényezőre van szükség. Ezt követően megtörténik a blokk maximumainak, azaz a léptéktényezőknél a meghatározása. A léptéktényező mellett a dekóder számára közölni kell azok számát is. A maszkoló tonálisának meghatározása a 2-es rétegben 1024 értékből történik a nagyobb spektrális felbontás érdekében.

A kvantálási lépcső meghatározása és az adatfolyam formálása az 1-es rétegnek megfelelő, viszont a 2-es rétegben csak az alsóbb részsávokban valósul meg a mintáknak 2-től 15-ig terjedő bitszámmal való megadása. A felsőbb részsávokban a jel energiája kisebb, ezért a

minták ábrázolási lehetősége korlátozott. Így 50 kbit/s adatsebesség felett és 48 kHz-es mintavételi frekvenciánál a 23-tól 26-ig terjedő részsávokban csak 0, 3, 5 vagy 65535 kvantálási szint lehetséges, ami hozzájárul az adatsebesség csökkenéséhez. Mivel csak 4 különböző kvantálás megengedett, ezért a bitkijelöléshez elegendő mindössze két bit. A 27-től 31-ig terjedő részsávokat 48 kHz-es mintavételi frekvencia esetén egyáltalán nem visszük át, mivel az e sávokba eső frekvenciák 20 kHz feletti, tehát egyébként sem hallhatóak. A 3.19. ábrán a 2-es rétegű adatfolyam egy keretét láthatjuk. Annyiban különbözik az 1-es rétegtől, hogy 1152 PCM –mintából épül fel, valamint tartalmazza az alkalmazott léptéktényezők számát is.

adatkeret 1152 PCM-mintával (24 ms 48 kHz-es mintavételi frekvenciánál)

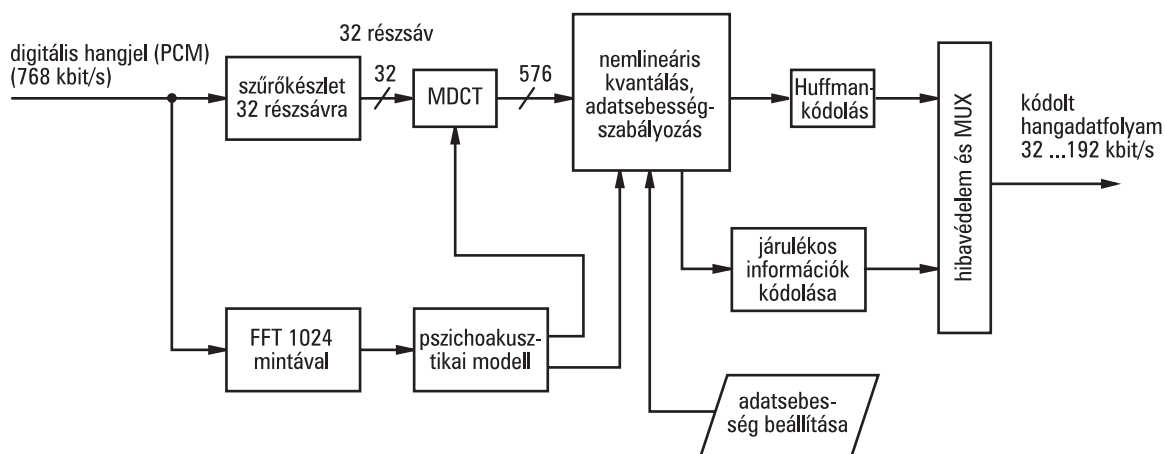
fejléc	hibavédelem	bithozzárendelés	léptéktényezők száma	léptéktényező	kódolt minták	kiegészítő adatok
12 bit szinkronizáló jel 20 bit rendszerinformáció	16 bit választhatóan	4 bit az alsó 3 bit a középső 2 bit a felső részsávokban	2 bit	6 bit	2 ...15 bit	

3.19. ábra
2-es rétegű adatfolyam egy keretének felépítése

3.3.4 3-as rétegű kódolás

A teljesség kedvéért röviden tekintsük át a 3-as rétegű kódolást, amit azzal a céllal fejlesztettek ki, hogy a 2-es rétegnél is nagyobb adattömörítést valósítsanak meg vele egy bonyolultabb jelfeldolgozási folyamat eredményeként.

A kódoló tömbvázlata a 3.20. ábrán látható. A 3-as rétegben a jel spektruma az un. módosított diszkrét koszinusztranszformációval (MDCT) a jel tulajdonságaitól függően különböző, de 576-nál nem nagyobb számú részsávra bontható fel, ami sokkal pontosabb maszkolási görbe, ill. kvantálási lépcső meghatározását teszi lehetővé.

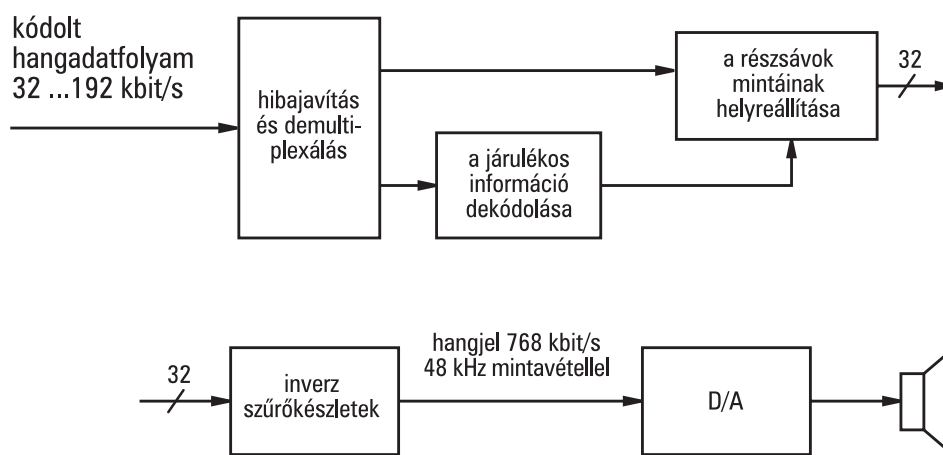


3.20. ábra
3-as rétegű kódoló tömbvázlata (monocsatorna)

A kvantálás nemlineáris, ami további adattömörítést biztosít. A jelben még meglévő redundancia az ún. Hoffman-kódolással csökkenthető, ami a hangminőség romlása nélkül az átviendő adatmennyiség újabb csökkenését eredményezi.

3.3.5 Dekódolás

A dekóderoldalon a hibajavítás után végbemegy az adatfolyam szétválasztása a normált és kvantált hangmintákra és a dekódoláshoz szükséges járulékos információkra. A bithozzárendelés és a léptéktényező kiértékelését követően a 32 részsáv mintái rekonstruálhatók.



3.21. ábra
MPEG-szabványú, 1-es és 2-es rétegű dekóder tömbvázlata
monohang dekódolására

Az inverz szűrőkészlet a részsávokat ismét teljes sávúvá egyesíti. Mivel a dekóderoldalon nem kell pszichoakusztikai modellt kiértékelni, ezért a jelfeldolgozás itt jóval egyszerűbb, aminek különösen a műsorszórásban van nagy jelentősége.

3.3.6 MPEG-hangkódolás paraméterei

Ebben a szakaszban az MPEG-hangkódolás legfontosabb paramétereit foglaljuk össze. A DVB-projekt szerint minden vevőnek rendelkezni kell egy MPEG-kompatibilis hangdekóderrel, amely a szabványban előírt összes üzemmódot támogatja. Tudnia kell monojelet, sztereojelet és két teljesen független hangjelet dekódolni. Egy további üzemmód az ún. joint-sztereó. Az eljárás azt a tényt használja, hogy 2 kHz frekvenciák felett a jel finom részletei nem járulnak hozzá a sztereohatás kialakulásához, hanem csak a burkolója. Ezért 2 kHz felett elég csak az összegjelet kódolni és a két csatorna skálatényezőjét külön-külön átvinni.

Az alkalmazható mintavételi frekvenciák 32 kHz, 44,1 kHz és 48 kHz. A teljes adatsebesség az 1-es rétegnél 32 kbit/s és 384 kbit/s közötti, a 2-es rétegnél 32 kbit/s és 448 kbit/s közötti érték. Ezen határok között a szabványban 14 adatsebesség van meghatározva, amelyeket minden dekódernek tudni kell fogadni.

Továbbá a 2-es rétegre fejlesztett dekódernek az 1-es réteg szerint kódolt jeleket is dekódolni kell. Úgyszintén a hármas rétegű dekódernek tudni kell dekódolni az 1-es és 2-

es réteg szerint kódolt jeleket, azaz kompatibilisnek kell lennie az 1-es és 2-es réteggel. Ilyenkor azt mondjuk, hogy a rendszer felülről kompatibilis.

Az MPEG-kompatibilis dekódernek és így a digitális tv-vevőnek (Integrated Receiver Decoder) a szabványban rögzített összes adatsebességet és üzemmódot támogatnia kell. A programszolgáltatók tehát tág lehetőségekkel rendelkeznek az alkalmazott hangkódolás és ezen keresztül a hangminőség megválasztásában. A 3.1. táblázat a hangminőség és az adatsebesség kapcsolatát adja meg a 2-es rétegű kódolás esetében.

3.1. táblázat

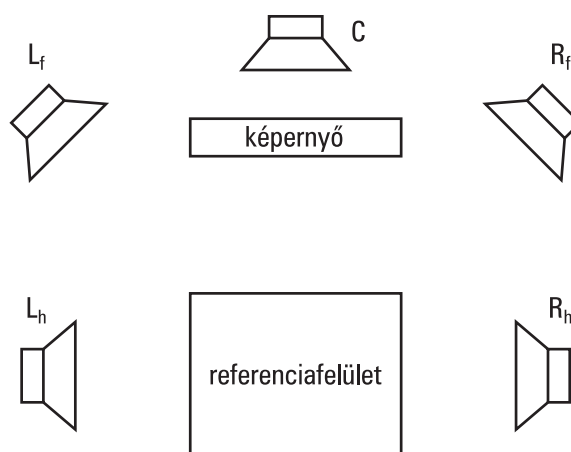
2-es rétegű kódolással átvitt monojel minősége az adatsebesség függvényében

Kódolt monojel adatsebessége [kbit/s]	Tömörítési arány	Minőség
192	4	kiváló
128	6	nagyon jó
96	8	jó (CD)
64	12	kb. FM-hangátvitel
48	16	kb. AM-hangátvitel
32	24	beszédminőség

Az MPEG-hangkódolás az eddigiek mellett lehetővé teszi az 1994-ben nemzetközi szabványként elfogadott surround-hangátvitelt is, amit a következő szakaszban tekintünk át.

3.3.7 MPEG-2 szerinti hangkódolás

A surround-hangátvitel az un. MPEG-2 szerinti hangkódolással valósítható meg. Jellegzetessége, hogy 5 hangszórával hozza létre a térhatású hangzást a 3.22. ábrán látható elrendezésben. A hagyományos sztereóátvitel jobb- és baloldali hangszórói mellett van még egy központi (center) és két oldal-(surround) hangszóró is. Szokás ezért az eljárást 3/2-es sztereónak is nevezni.



3.2 2. ábra
3/2-es sztereó hangszóró-elrendezése

A rendszer kifejlesztésénél nagy hangsúlyt fektettek a kompatibilitásra, azaz egy MPEG-dekódernek tudni kell dekódolni a surround-hangot is. Ez az ún. felülről kompatibilis megoldás lehetővé teszi, hogy a surround-hanggal kisugárzott programot azok a nézők is dekódolhassák, természetesen csak mint sztereoejelet, akiknek nincsen surround-dekódere. A rendszer alulról is kompatibilis, azaz egy MPEG-2 dekóder (surround-dekóder) az MPEG-adatfolyamot is dekódolja. A felülről való kompatibilitás biztosítható, ha az 5 csatornából (L_f , R_f , L_h , R_h és C) lineáris mátrixolással hozzuk létre a két sztereoejelet:

$$L = L_f + 0,71C + 0,71L_h \quad (3.4)$$

$$R = R_f + 0,71C + 0,71R_h \quad (3.5)$$

A kompatibilitás miatt az is szükséges, hogy a surround-jel a 2-es réteg multiplexálásának megfelelően, azaz az adatkeret első részében L és R , a kompatibilis sztereocsatorna jelei, míg a további 3 csatorna információtartalma a keret kiegészítő adatok részében kerüljenek átvitelre.

3.4 Összefoglalás

A digitális tv hangjának adatcsökkentése az MPEG-szabvány szerint valósul meg. Az eljárás a spektrum 32 részsávra bontásán és egy 12 vagy 36 mintából álló blokk képzésén alapul. A blokkok mintáinak kvantálása az emberi hallás elfedési sajátságait figyelembe vevő akusztikus modell alapján valósul meg. A létrejövő kvantálási zajt a jel spektrálisan és időben elfedi, aminek következtében azt nem lehet meghallani. A multiplexált adatfolyamból a dekóder az eredeti hangmintákat rekonstruálja.

A 2-es rétegű kódolással kb. 100 kbit/s adatsebesség mellett a hangminőségben kódolásból származó romlás nem érzékelhető. Ilyenkor azt mondjuk, hogy az átvitel transzparens. Az MPEG-hangdekóder valamennyi üzemmódú és adatsebességű hangjelet dekódolni tudja, ami a felhasználó mindenkori igényének kielégítését teszi lehetővé.

A 3-as rétegű MPEG-hangkódolást azzal a céllal fejlesztették ki, hogy egy összetettebb kódolási folyamat révén a 2-es rétegűnél nagyobb tömörítést érjenek el. A „Joint Stereo” a két sztereocsatorna között meglévő redundanciát hasznosítja, míg a „Surround Sound” a szabvány kompatibilis bővítése.