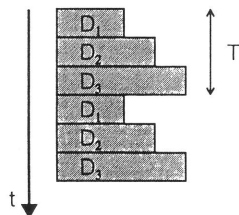


2-4 ábra

Szigorúan egyöntetű adatfolyam. Minden csomag adatmennyisége azonos

Ez a jellegzetesség tipikus nemtömörített audio- és videoátvitelnél. Példa erre egy kamera által szolgáltatott adatfolyam teljes képkódolással. Hasonló példa egy audio CD által szolgáltatott audioszekvencia.

- Ha az adatmennyiség (időben) periodikusan változik, akkor az adatfolyamot **közel egyöntetűnek** (*weakly regular*) nevezzük.



2-5 ábra

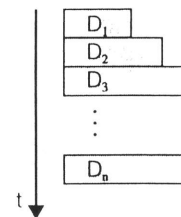
Gyengén egyöntetű adatfolyam. A csomagok adatmennyisége periodikusan változik

Néhány videotömörítő eljárás a következő elveket alkalmazza: az egyes teljes képeket, mint egészet kódolják és tömörítik. Az ily módon létrejött információegység az adatfolyamon belül viszonylag nagy adatsomagot állít elő. Ily módon a korlátozott hosszúságú adatsomagokat átvivő átviteli rétegeket figyelmen kívül hagyják. Ezeket a csomagokat aztán periodikusan, pl. 2 másodpercenként továbbítják. Minden ily módon kódolt és tömörített video egyesképénél csak az egymást követő képek közötti különbséget viszik át, mint információt.

Példaként említhető az MPEG (1. 5.7. fejezet, 105. oldal). Itt az I-képek tömörített egyesképek, ellenben a P- és B-képek képkülönbségként vannak figyelembe véve, ami lényegesen csökkent adatmennyiséghez vezet. Az egyes I, P és B tömörített csomagokhoz nincs meghatározott mennyiségű bithányad

hozzárendelve. Átlagosan azonban az I:P:B viszony az így létrejövő adatmennyiségekben az egyes képekre 10:1:2. Ezáltal egy ilyen adatfolyam hosszú idejű átlagban **közel egyöntetűnek** nevezhető.

- Az **adatfolyamok nem egyöntetűek**, ha az adatmennyiség nem állandó, se nem periodikus változású (l. 2-6 ábra). Az átvitel és a feldolgozás itt bonyolultabb, mint az előző változatokban.



2-6 ábra

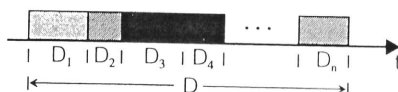
Nem egyöntetű adatfolyam. A csomagok adatmennyisége sem azonos, sem periodikusan változó

Olyan tömörítő eljárásoknál, melyeknél változó bithányadú adatfolyam keletkezik, az egyes információegységek hosszát (melyek az aktuális egyes képekhez tartoznak) a megelőző képre vonatkozó képtartalom változás határozza meg. A létrejövő információegységek hossza ezek után legtöbbször a videoszekvenciáktól függ és az adatfolyam szabálytalan.

### Összefüggés az egymást követő csomagok között

A harmadik jellegzetesség a folyamatoságra, illetve az egymás utáni csomagok közötti összefüggésre vonatkozik. Vajon az egymás utáni csomagokat közvetlen egymást követően viszik át, vagy szünetek keletkeznek a csomagok között? Ez az erőforrás kihasználásának fogalmával írható le. A hálózat egy ilyen erőforrás.

- A 2-7 ábra a 34. oldalon egy **összefüggő információtovábbítást** mutat. Az **összes csomagot** egymást követően, minden szünet nélkül viszik át. A szükséges többletinformáció, az alapadatokkal együtt van figyelembe véve. Ilyenek például a rétegfüggetlen hibafelismerő kódok. Ily módon a szóban forgó erőforrás 100%-ig kihasznált.

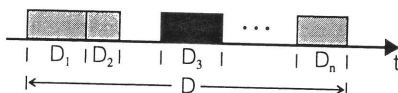


2-7 ábra

Összefüggő adatfolyam. A csomagokat közbülső szünet nélkül viszik át

Az összefüggő adatfolyam maximális forgalmat enged meg és ezáltal az erőforrás maximális kihasználását teszi lehetővé. Példa erre egy ISDN-kapcsolat B csatornája 64 kbit/s audioátvitelnél.

- Az **összefüggő adatfolyam** nagyobb csatornkapacitású vonalon való átvitele az egyes csomagok között szünetet hoz létre. Minden olyan adatfolyamot, melyben az adatsomagok között szünetek vannak, *össze nem függőnek* nevezünk. Példa erre a 2-8 ábra. E tekintetben nincs jelentősége annak, hogy minden csomag között van-e szünet és a szünetek hossza változik-e?



2-8 ábra

Nem összefüggő adatfolyam. A csomagok között szünetek vannak

Egy DVI-PLV eljárás szerint kódolt, átlagosan 1,2 Mbit/s-os adatfolyam továbbítása FDDI-hálózaton keresztül elkerülhetetlenül szüneteket hoz létre a csomagok között. Ezáltal itt egy nem összefüggő adatfolyam keletkezett.

A következő **példák**ban a fentebb megadott jellegzetességeket röviden értelmezzük:

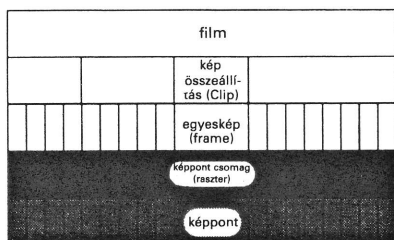
Egy kamera PAL szabvány szerinti videojelét vesszük és számítógéppel digitálizáljuk. Semmilyen tömörítést sem alkalmazunk. Az így létrejövő adatfolyam szigorúan periodikus, szigorúan egyöntetű és összefüggő, amint az a 32. oldal 2-4 ábráján látható. A csomagok között nincsenek szünetek. Ezután ezt a DVI-RTV eljárással kombinálva az Action-Media-II kártya segítségével tömörítjük. Az így keletkező adatfolyam (hosszú idejű átlagban) most gyengén periodikus, közel egyöntetű és egy 16 Mbit/s-os Token-Ring vonalon továbbítva nem összefüggő lesz.

## 2.6. Információegységek

Folyamatos (és időfüggő) média egyedi információegységek időbeli sorozatából áll. A következőkben a protokoll adategységekhez hasonlóan (*protocol data unit PDU*) az ilyen információegységeket *logikai adategységnek* (*logical data unit LDU*) nevezzük. Az információ- és adatmennyiség jelentése valamely LDU-ban nagyon különböző lehet:

### Példák

1. Tekintsük első példaként Joseph Haydn *A medve* szimfóniáját. Ez négy tételből áll, vivace assai, allegretto, menuet und finale vivace. Minden tétel önmagában zárt rész ebben a kompozícióban. Különböző hangszerekre hangjegyek sorozatából áll. Ezek a hangjegyek egy digitális rendszerben mintavételi értékeknek felelnek meg. Ezáltal semmilyen tömörítésnek nem kell lennie a rendszerben. Hajtsunk végre egy lineáris karakterisztikájú PCM kódolást. A CD-DA minőségű rögzítéshez 44 100 mintavétel szükséges másodpercenként, melyet csatornánként 16 biten kódolunk. Egy CD-n ezeket 1/75 sec időtartamú szakaszokba fogunk össze. Ezek után LDU-nak az egész művet, a tételket, az egyes hangjegyeket, az 1/75 sec-ban összefogott mintavételi értékeket vagy akár az egyes biteket is tekinthetjük. A mindenkori alkalmazástól függ, hogy mit kell LDU-nak tekinteni. Némely műveletet az egész művön, mint LDU-n végrehajthatunk. Más műveletek a legkisebb önálló jelentésű egységre vonatkozathatók (ebben az esetben a hangjegyekre). A digitális jelfeldolgozás a mintavételi értékekre, mint LDU-ra vonatkozik.
2. Egy példa a 36. oldalon a 2-9 ábrán bemutatott videoszekvencia: a tömörítetlen videoszekvencia egyes jelenetekből tevődik össze (Clips), melyek mindegyike egy jelenetet mutat be. Egy ilyen jelenet egyes képek sorozatából áll. Az egyes kép különböző tartományokra, pl.  $16 \times 16$  képpontú részekre bontható folytatólagosan. Egyetlen képpontnak lehet luminancia és krominancia értéke.  
Az egyes kép tehát nem az egyetlen lehetséges LDU egy mozgóképszekvenciában, egy jelenet vagy egy képpont szintén lehet LDU. Egy MPEG vagy VDI szerinti kódolt videoadatfolyamban az egyes kép-sorozatok redundanciái, a képek közötti tömörítéssel, az adatmennyiség csökkentésére használhatók ki. A legkisebb zárt, önálló jelentésű egységek itt az egyes kép szekvenciák.



2-9 ábra

Egy mozgóképszekvencia aláosztásai (Logikai adategységek, LDU-k)

Az aláosztásnak ez a jelensége bemutatja egy audio- vagy videoadatfolyamnak a hierarchikus felbontását. Mint legáltalánosabb megnevezés és legátfogóbb információegység a példáinkban a szimfónia és a film szerepelt. Ily módon tekinthetjük ezeket akár *zárt*, akár *nyitott* LDU-nak. A zárt LDU-k meghatározott hosszúságúak. Ezek többnyire tárolt szekvenciák. Egy adatfolyam, mint nyitott LDU, előre nem ismert hosszúságú. Például ilyen adatfolyamot szolgáltat egy kamera vagy egy mikrofon egy számítógép számára.

Ezen alapvető, a multimédia rendszerek sajátosságaira vonatkozó megfontolások után a következő fejezetben az audioadatokat vizsgáljuk meg közelebbről. Ezen belül a hangsúlyt a beszédfeldolgozásra helyezzük.

### 3. Audiotechnika

Az akusztika a hangfolyamatokat tárgyalja folyadékokban, gázokban és szilárd testekben. Ennek során az akusztikus jelátalakítókat, az akusztikus hullámok terjedési törvényszerűségeit, a hangsugárzást, az akusztikai mérés technikát, valamint a fiziológiai és pszichológiai akusztikát vesszük tekintetbe [Moln68a, Skud71a]. Az audiotechnika az akusztikus jelek feldolgozásával foglalkozik, melyeket az ember észlelni képes. A lényeges munkaterületek itt a magnetofonon, illetve a DAT jelförzítőn (Digital Audio-Tape) történő tárolás, a beszédfeldolgozás és a kódolás.

Audioadatoknak (és más média adatainak) optikai lemezen való tárolását a 6. fejezetben a 125. oldalon („Optikai tárolók”) mutatjuk be, mivel ez a technológia korlátozódik az audiojelekre. Az audio- és a videojelek tömörítésével külön fejezetben foglalkozunk a 6. fejezet 125. oldalán („Adattömörítés”), mert gyakran használt módszereket alkalmaznak az információ kódolására a különböző médianál is így az összefüggések jobban értelmezhetők. A legtöbb multimédia-alkalmazás az audio médiumot zene és/vagy beszéd formájában használja. A beszédkommunikációknak különösen az elosztott multimédia-alkalmazásokban van a legnagyobb jelentősége. Ezért itt a MIDI mellett a beszéd médium lényeges összetevőire összpontosítunk: beszédszintézis, beszédfelismerés és beszédtovábbítás [Flan72a, Flan72b, Shau90a, Mang92a]. Mindenekelőtt néhány olyan megjegyzést állítunk fel, melyek minden audiojelre érvényesek.

#### Alapok

Az akusztikus hullámok, gázokban és szilárd testekben a hangfolyamatok különböző frekvenciatartományokra oszthatók:

- **Infrahangok:** 0-tól 20 Hz
- **Hallható hangok:** 20 Hz-től 20 kHz-ig
- **Ultrahangok:** 20 kHz-től 1 GHz-ig
- **Hyperhangok:** 1 GHz-től 10 THz-ig

A multimédia rendszerekben csak a hallható hangok tartományának van jelentősége. Az akusztikus jelek, nyomáshullám-ingadozások a hallható hangok tartományában vannak, melyeket emberek és gépek gerjesztenek és érzékelnek. Minden felvétel valamilyen átalakítóval jön létre; az akusztikus információ felvételére az emberi fül szolgál. Ebben a technikában érzékelőként mikrofonokat alkalmaznak. Akusztikus jeleket az ember többnyire a hangszálaival hoz létre. Néha azonban másképpen is előállítunk hangot, például tapsolással. A műszaki gyakorlatban ehhez átalakítóként hangszórót használnak.

Az **audiomédium** úgy tekinthető, mint a hallható tartományba eső **akusztikus jelek összessége**. A *beszéd* az emberi beszédszervek által előállított jel. Ezeket gépi úton is lehet utánozni. A *zenei jelek* a 20 Hz és 20 kHz közötti tartományba esnek. A beszéden és a zenén kívül ebben a frekvenciatartományban szintén megkülönböztetünk *zajokat*, mint egyéb audiojeleket. A zaj egy nem meghatározott célú hangjelenség. Ezt a röviden tárgyalt elhatárolást nem szabad dogmatikusan fel fogni, csupán kiindulási pontnak kell tekinteni. Definiálhatnánk példának okáért az érthetetlen emberi beszédet is zajnak.

### 3.1 Zene – MIDI

A számítógéppel összefüggésben a beszéd mellett a zene játssza a legfontosabb szerepet. Különösen a MIDI (**Music Instrument Digital Interface**) kifejlesztését kell figyelembe venni, melyet a zenei ipar 1983 óta alkalmaz. Ez az interfész definíció két elektronikus zenei eszköz között (és két számítógép között) kódolt zenei jelek átvitelét teszi lehetővé.

Itt a tényleges mintavételi értékek helyett eszközfüggő jelmegjelenítéseket alkalmaznak. A kódolás többek között az eszköz megnevezését, a hangjel kezdetét és végét, az alaphangfrekvenciát és a hangerőt tartalmazza. A MIDI 10 oktávon keresztüli kódolást tesz lehetővé, ami 128 hangnak felel meg. Ha például egy zenész leüt egy zongorabillentyűt, akkor a hang kezdete és a leütés erőssége átvivődik. Ha a billentyűt elengedi, ismét átvivődik egy ennek megfelelő jel. Ez az eljárás egy 10 perc hosszú zenevisszaadáshoz kb. 200 Kbyte MIDI-adatot vesz igénybe, ami egy olyan hangrend, amely lényegesen alatta marad a 44 kHz-es mintavételi ismétlődésű jelnek megfelelő adatmennyiségnek. A MIDI a zenei adatok legkompaktabb előállítását, ami természetű hangvisszaadást tesz lehetővé.

#### MIDI

A MIDI segítségével egy számítógép az egyes hangszereket megszólaltathatja. **Másfelől** a számítógép ezen interfész révén kódolt zenei adatokat tud venni, tárolni és **tovább** feldolgozni. A MIDI-terminológiában egy billentyűzet segítségével lehet **ezeket** az adatokat generálni és egy szintetizátorral visszaadni. Egy sorbaállító **eze- ket** az adatokat átmenetileg tárolni tudja és adott esetben képes ezeket megváltoz- títani. **Egy** multimédia rendszerben a sorbaállító a számítógéppel valósítható meg.

#### A billentyűzet

A billentyűzet a zongora klaviatúrájának felel meg. Segítségével a zenész a legkül- **önözőbb** hangszereken képes játszani (a játékot imitálni). Eközben a billentyű **nyomásának** sebességével és gyorsulásával a hangerősség befolyásolható. Ezen- **ként** 5 oktávban 61 billentyűnek kell lennie. A billentyűzet ily módon a zenész játé- **kat** MIDI-adatokká változtatja.

#### A szintetizátor

A szintetizátor olyan akusztikus jelek előállítására szolgál, melyek MIDI-adatok **formájában** jönnek létre. Ez belülről a legkülönözőbb formában jöhet létre: **a jelek** előre tárolva lehetnek. Hangvisszaadáskor a mintavételi értékek egy digitál- **is** átalakító segítségével akusztikai jelekké alakulnak. Egészében véve az **egyes** hangok időtartománybeli összekapcsolása valósul meg. Más megoldásokban **az akusztikus** jeleket szintetikus úton állítják elő.

#### A sorbaállító

**Eredetileg** egy sorbaállító az a feladata, hogy az előállított MIDI-adatokat **először** és egy későbbi időpontban visszaadja. Ma egy sorbaállító egy zenemű szer- **kesztője**. Itt az adatok megfelelő módon megváltoztathatók. Az előállításához több **lehetőség** áll rendelkezésre: a legismertebb előállítást és feldolgozási lehetőséget **maguk** a hangjegyek adják. Itt a képernyőn egy kottalap jelenik meg, amely megfe- **lel** a zenei darabnak. A sorbaállító ezeket a hangjegyeket MIDI-adatokká alakítja **le**. **Egy** nagyon technológia-orientált előállítást jelent az egyes MIDI-adatok köz- **vetlen** előállítása. Ehhez egy-egy csatornát állítanak elő: a felhasználó a megvalósi- **tandó** jeleket elhelyezi az időtengelyen.

Gyakran a billentyűzetből indulnak ki: a bal szélén, mint Y tengelyen, külön- **öző** billentyűk helyezkednek el. Az idő mint az X tengely jelenik meg. Egy nyo- **mógomb** lenyomásának időtartama ebben a diagramban, mint egy vonás jelenik **meg**. Egy pontosabb forma magában foglalja a mindenkor leütési erősség kódolá- **sát is**, mint a vonás amplitúdóját.

### MIDI-szabvány

A MIDI-szabvány szerint 16 **csatornán** vihetők át jelek. Minden felhasznált csatornához hozzá van rendelve egy szintetizátor. A zenei adatok, melyek ezen a csatornán folynak, a szintetizátor segítségével a megfelelő hangszereken visszaadhatók. Ily módon egyidejűleg **több hangszer** a különböző hangokon szólaltatható meg egy MIDI-csatlakozáson keresztül, mivel az egyes csatornához különböző hangszerek vannak hozzárendelve. A MIDI-definíciók 128 hangszert, beleértve a zajhatásokat (mint telefon és helikopter) egyértelmű számokkal jelölnék meg. Példaképpen a 0 számmal a zongorát, *acoustic grand piano*, 12-vel a marimbát, 40-nel a hegedűt és 73-mal a fuvolát jelölik.

Egyes hangszereken egyszerre csak egy **hang** játszható. Erre példa a fuvola. Más hangszereken egyidejűleg több hang is előállítható. Egyszerre játszható több hang például az orgonán. Az egy csatornán belül egyidejűleg játszható hangok maximuma lényeges jellemzője egy szintetizátornak. Szokásos érték a 3 és a 16 közötti hang csatornánként.

### Gyártóspecifikus kódolás

Az általánosan érvényes kódolás mellett a MIDI a gyártóspecifikus kódolás lehetőségét is magában foglalja. Ezt a gyártó által adott egyértelmű *System-Exclusive* jelzés vezeti be és egy *End-of Exclusive* jelzés zárja le. A **MIDI-Clock** jel segítségével a vevő az adó üteméhez tudja magát szinkronizálni. Erre a célra negyedhangonként 24 jelölés vihető át. Másik lehetséges megoldásként az **SMPTE** (*Society of Motion Picture and Television Engineers*) időkód vihető át. Az SMPTE egy *Óra: Perc: Másodperc: Képszám* formátumot definiál például 30 kép/s-ra. Ez a 30 kép/s sebességű információátvitel mindenesetre átlépné a MIDI-re érvényes sávzélességet. Ezért megállapodásszerűen létezik egy **MIDI time code**, mely nem viszi át a teljes időbeli ábrázolást minden egyes képpel együtt.

Egy MIDI-adatok kezelését tartalmazó multimédia alkalmazás előállítja a hangszer és a csatornák közötti kapcsolatokat is. Ezenkívül lehetővé teszi ez az alkalmazás a szintetizátoron való azonnali megszólltatást. A MIDI az adatok időben helyes sorrendű visszaadását követeli a korrekt visszaadáshoz. Itt érvényes a multimédia rendszereknél az ismert valósídió követelmény.

## 3.2 Beszédjelek

A beszédet az ember vagy gép hozhatja létre.

### Gépek

**Gép útton szintetizált beszéd** hozható létre, ahol a beszéd alkalmasint viszonylag természetellenesen hangzik. A beszéd felismerés rendszerint különböző összehangolt úttján történik. Ily módon manapság mintegy 20 000 szót tartalmazó beszéd felismerő rendszer lehetséges. Zavaró hatású a beszéddialektus, az érzelmi-  
g erősen befolyásolt kiejtés és a környezeti zaj. Mind ez ideig és várhatóan a belát-  
tás jövőn belül mindig lesz egy számottevő egyértelműségi zavar az emberi agy és  
számítógépek teljesítményű számítógépek teljesítménye között.

### Az ember

**Az ember** (Más élőlények, mint pl. a denevérek, szintén használnak akusztikus jeleket információközlésre. Ezeket a továbbiakban nem tárgyaljuk.) a beszéd szerveit és működését nem tudatos módon használja. Ez megkövetel egy nagyon tudatos illesztést a beszédalkotó beszélő személyhez és beszédmódjához. A különböző dialektusok és dialektusok színezett kifejezőmódok ellenére a beszéd megérthető. Az agy képes arra, hogy jól megkülönböztesse a beszédet a zavaró hatásoktól. Ehhez hozzájárul a beszéd felismerésével való hallás is. Az egy füllel történő zavarkiszűrés az ember számára sokkal nehezebb. Az emberi agyban a beszéd felismerési hibák nem tudatos javítása érdekében a tartalom ismeretében a nyelvtani szabályok, a fonetikus és lexikális szókincs birtokában. Az emberi beszédképzés leírása és modelljei például az [1] hivatkozásban találhatók.

Az emberi felismerés szubjektív hangmagasságon történik, ami nem arányos a frekvenciával. A mély hangoknál két, 1:2 frekvenciaviszonyú, oktávnyira lévő hang a meghatározó. A magasabb hangoknál úgy találták, hogy ez a frekvenciaviszony kisebb.

Megvizsgálható az azonos hangerősségű frekvencia függvényében. A fül a 400 Hz és 6000 Hz közötti frekvenciatartományban a legérzékenyebb. Ezeket Fletcher és Munson által felvett görbék azt mutatják, hogy az emberi hallás az alacsony és a nagyon magas frekvenciák tartományában lényegesen érzéketlenebb.

A beszédjeleknek két alapvető sajátosságuk van, amely a beszédjel-feldolgozásban jól használható:

A zöngés beszédjelek meghatározott időintervallumon belül közel periodikus jellegű oszcillációt mutatnak. Ezért ezeket a jeleket kvázistacionáriusnak tekinthetjük a kb. 30 milliszekundumos időintervallumban.

Egyes beszédhangok spektruma jellegzetesen kiemelt, amiből többnyire 5 frekvencia lényeges. Ezeket a frekvenciákat formánsoknak nevezik. Ezek a formánsok a beszédképző út rezonanciái miatt jönnek létre.

### 3.3 Beszédkiejtés

A beszédkiejtés a **beszéd mechanikai előállításával** foglalkozik. A beszédkiejtés már régóta kutatott terület [Endr84a, Fell85a]: Helmholtz már a 19. század közepén leutánozta a beszédképző csatornát csatolt mechanikai rezonátorokkal. Ily módon beszédhangokat lehetett előállítani. Dudley 1940-ben előállította az első beszéd szintetizátort a mechanikus rezgőkörök villamos leutánzásával.

A beszédjelek képzésénél lényeges követelmény ezen jelek *valósídejű* megjelenítése. Egy beszédképző rendszer ezért például hosszabb feldolgozási idő nélkül alakít át egy szöveget beszéddé.

Néhány alkalmazás csak egy korlátozott szókészletet igényel, erre példa az időmegadási szolgáltatás. A legtöbb alkalmazás azonban nagyon nagy, ill. korlátozás nélküli szókészletet használ.

A kiadott **szövegnek érthetőnek kell lennie és természetesen kell hangzania**. Az érthetőség alapkövetelmény, a természetes hangzás pedig a felhasználók részéről nagyobb elfogadottságot eredményez. A beszédkiejtés gyakran érthető, természetes hangzás nélkül, tehát a természetes hangzás nem jelent feltétlenül érthetőséget.

A további tárgyaláshoz néhány fogalomra van szükség:

- **Beszédalapfrekvenciának** a beszéd legalacsonyabb periódusú jelrészét nevezzük. Ez a zöngés hangoknál jön létre.
- **Fonéma** definíciószerűen a legkisebb jelentésszerű, de önálló jelentést nem hordozó beszédegység. A német nyelvben mintegy 40 fonéma különböztethető meg.
- **Allophonének** nevezünk egy fonémavariációt, melyet hangzásbeli környezetük függvénye határoz meg.
- **Morphema** a neve a legkisebb még jelentést hordozó beszédegységnek. Ezért a ház egy morphema, a hízalás nem.
- A **zöngés hang** képzése a hangszálak segítségével történik. Példák erre az M, W és L hangok. A zöngés hangok kiejtése erősen függ a beszélő személytől.
- A **zöngétlen hangok** képzésekor a hangszálak nem működnek. Az F és az S hangok zöngétlenek. A zöngétlen hangok viszonylag függetlenek a beszélő személytől.

Egzakt módon megkülönböztethetők:

- A magánhangzók (pl. az *a* a van szóban)
- A zöngés mássalhangzók (pl. az *m* a magas szóban)

- A frikativ zöngés mássalhangzók (pl. a *w* a wolfram szóban)
- A frikativ zöngétlen mássalhangzók (pl. az *s* a sör szóban)
- Az explosiv mássalhangzók (pl. a *d* a dög szóban)
- Az affrikativ mássalhangzók (pl. a *p* a pók szóban)

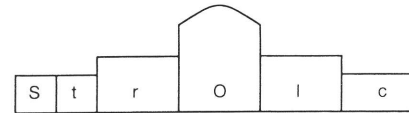
#### Reprodukáló beszédvisszaadás

Az **valósítja** meg a legegyszerűbb beszédkiejtési eljárást. A beszédet előzőleg egy **memória** hozza létre és ezt rögzítik. A beszédvisszaadás időpontjában ezt a tárolt **beszédet** játsszák vissza.

**Itt mindig** felismerhető a beszélő személye. A lejátszáskor egy meghatározott **szókészlet** vagy egy meghatározott számú mondat áll rendelkezésre jó **minőségben**. A beszéd PCM kódolt (l. 3.5 fejezet, 49. oldal) formában és minőségben **tárolható**. További adattömörítési eljárások alkalmazhatók függetlenül a **beszédtípus** sajátosságaitól (l. 3.4 fejezet, 46. oldal és 5. fejezet, 73. oldal).

#### Hangláncolás az időtartományban

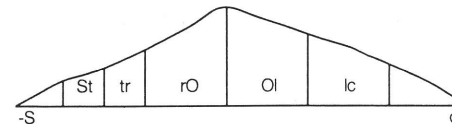
A **beszédkiejtés** a hangok időbeli egymáshoz kapcsolásával is megvalósítható. Ekkor **egyes** beszédegységeket mint építőelemeket tesszük össze, ahol egy ilyen **komponenst** különböző síkokon hozhatunk létre.



3-1 ábra

Az időtartományban hangláncolt phonem [Endr84a]

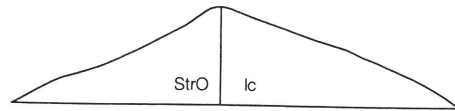
A **legegyszerűbb** esetben a fonémákat tekintjük beszédegységnek. A 3-1 ábrán a **„Storch”** szó egyes fonémáit mutatjuk be. Lehetséges, hogy korlátozott számú **fonémával** egy korlátozott szókészletet állítsunk elő. Különösen kritikus az egyes **fonémák** közötti átmenet. Ezért egy jobb közelítésben már figyelembe veszik a **fonémák** környezetét, az allophonokat. Azonban még kb. 230 allophon **alkalmazásával** sem lehet megnyugtató megoldást találni az átmenetek problémáira.



3-2 ábra

Az időtartományban hangláncolt diphon [Endr84a]

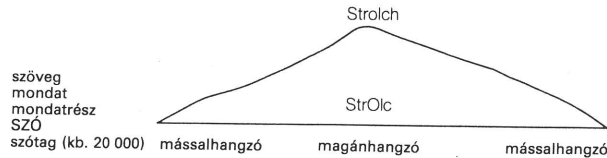
Minden fonéma összevonható egy diphonná. A 3-2 ábrában a 43. oldalon a „Strolch” szónak egymáshoz rendelt diphonok segítségével való előállítását mutatjuk be. Összesen mintegy 1400 diphon definiálható.



3-3 ábra

Az időtartományban hangláncolt fél szótag [Endr84a]

A beszédegységek átmenetei problematikájának kérdését még tovább tompítva képezhetünk fél szótagokat. A beszédet, ahogyan ezt a 3-3 ábrán bemutatjuk, ezen feltevés alapján egy sor fél szótagból állítjuk elő. Hozzávetőlegesen 20 000 német szótagot lehet hangláncolással az időtartományban használni.



3-4 ábra

Az időtartományban hangláncolt szó [Endr84a]

Egy szó legjobb kiejtése az egész szó tárolásával érhető el. Ekkor a beszédszintézisnek ahhoz a fajtájához jutunk, amikor a visszaadásnál előre tárolt szavak szekvenciáit alkalmazzuk.

Minden esetben az egymást követő hangzasegységek átmenetei jelentik a fő problémát. Ezt a hatást *koartikulációnak* nevezzük. A koartikuláció a több hangzás között fennálló kétoldalú kölcsönhatás. (A hangzás tulajdonságai ugyanis lényegesen függenek a környezeti hanghatásoktól.) Ezek a tulajdonságok az emberi beszéd elkerülhetetlen velejárói és a beszédszervek tehetetlenségén alapulnak.

Ezekon túlmenően a beszédkiejtés *prosódijája* is figyelembe veendő. Ez egy mondat hangsúlyozását és hanglejtését jelenti. Egy kérdőmondat kiejtése például erősen különbözik egy egyszerű megállapítástól. A prosódia tehát függ a beszéd szemantikájától és a hangláncolásnál az időtartományban járulékosan figyelembe kell venni.

### Hangláncolás a frekvenciatartományban

A beszédkiejtés a frekvenciatartománybeli hangláncolással, pl. az ún. *formánsképzéssel* is megvalósítható. A formánsok a beszédjel spektrumának energiakoncentrációi. Ennél az eljárásnál a beszédképző utat szűrőkkel képezik le. Jellemző értékek ekkor a szűrők középfrekvenciái és a sávszélességek. Gerjesztésként egy beszédalapfrekvenciájú impulzust választanak, amely minden zöngés hangra érvényes. A zöngétlen hangoknál evvel ellentétben egy zajgenerátorból származó jelet vesznek.

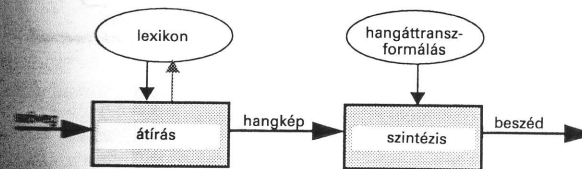
Az egyes beszédelemek, mint például a fonémák, a formánsok jellemző értékekkel vannak definiálva. Itt hasonló problémák állnak elő, mint az időtartománybeli hangláncolásnál. A koartikulációnál említett átmenetek itt is nagyon kritikus pontok. A proszódia is figyelembe kell venni járulékosan. Az új, hangspecifikus dolgozó eljárások az idő- és a frekvenciatartománybeli szintézist kombinálják [Fie92a]. Az első eredmények elsősorban a frikativ és az explosiv hangzás megvalósításában mutatkoznak.

Az emberi beszédképző út legeredményesebben egy többpólusú létraszűrővel képezhető le. Ily módon az első négy vagy öt formáns korrekt módon modellezhető. A hangszálakat zajgenerátorral és egy hanggenerátorral szimulálják. Ez a *Linear-Predictive-Coding eljárás* a [Schr85a] hivatkozás szerint nagyon hasonló a szintézishez.

Egy másik realizálási lehetőséget nyújtanak a *csőmodellek*. Itt a hangképző utat egy egyszerűsített mechanikai csőmodellel közelítik és hullámdigitális technikával (villamos analógia) képezik le.

### Szószintézis

A szószintézis segítségével egy adott szöveg akusztikus jellé transzformálható [Fie92a]. A 3-5 ábra egy ilyen rendszer komponenseit mutatja.



A szószintézis komponensei időtartománybeli hangláncolással

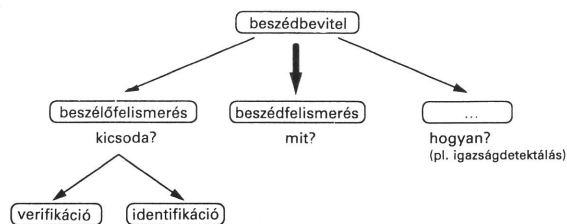
Az első lépésben egy átírás következik be, a szöveg átalakítása a megfelelő hangképbé. Ebben a lépésben a legtöbb eljárás egy lexikonnal dolgozik, amely szavak nagy csoportjából vagy csupán szótagokból és hangcsoportokból áll. Egy ilyen könyvtár előállítására igen nagy ráfordítást követel, azonban ez a felhasználó interaktív ellenőrzésével folyamatosan javítható: a felhasználó felismeri az alkalmazott forma hiányosságait az átírásnál. A kiejtést manuálisan javítja, a tapasztalatait a lexikon részévé válnak. Ennek során egyéni megvalósítás éppúgy jól elképzelhető, mint egy közösen használt lexikon.

A második lépésben a hangkép akusztikus jellel történő átalakítása történik meg. Itt történhet egy egymásra láncolás az idő- vagy a frekvenciatartományban. Míg az első lépésben szinte kizárólag szoftver megoldást alkalmaznak, addig a második lépésben jelprocesszorokat és dedikált processzorokat használnak.

A koartikuláció és a prosódia problémája mellett ügyelni kell itt a *kétértelmű kiejtésre*. Egy kijelentés sokszor csak a tartalom többi részének ismeretében értelmezhető, ez tehát szemantikafüggő. Példa erre a (német) Wachstube szó. Ez akár Wach-Stube, akár Wachs-Tube bontásban értelmezhető. Ez két teljesen különböző kiejtést hordoz magában. Ez a problémakör csak a tartalmi környezet járulékos ismeretében oldható fel.

### 3.4 Beszédbevitel

A beszédbevitel az ábrában bemutatott kérdésfeltevési esetekkel foglalkozik:



3-6 ábra

A beszédbevitel tématerületei

- Az emberi beszéd a beszélőre jellemző sajátosságokkal rendelkezik. A beszédbevitel ezáltal a **beszélő személy felismerését** is szolgálhatja [OSha86a]. A számítógépnek a *beszédlenyomat*, egy akusztikus ujjlenyomat alapján fel kell ismernie a beszélőt.

Ekkor valójában egy identifikáció történik. Egy beszédjel adott és ennek alapján a rendszer felismeri a beszédhez tartozó személyt. Ez alkalmazható a bűnüldözésben.

Más feladatot jelent a verifikáció. Itt mind a beszédminta, mind a beszélő személy ismert. A rendszernek azt kell megállapítania, hogy a beszéd valóban **emlő a személytől** származik. Ezáltal valósítható meg egy hozzáférési jogszabály.

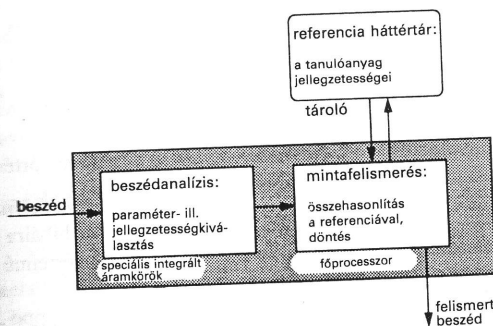
A beszédbevitel központi problémája azonban mégiscsak a **beszéd tartalmának a felismerése**. Valamely beszéd szekvencia alapján az ennek megfelelő szöveget kell előállítani. Ezt alkalmazzák a beszédvezérelt írógépnél, egy fordítórendszerrel vagy részben a testi fogyatékosok munkahelyein.

A beszédbevitel egy további alkalmazási területe egy beszédminta **hogyan** megállapítani. Itt például egy igazmondás vizsgálót képzelhetünk el alkalmazásként.

A médiatranszformációk mellett a beszéd felismerésnek igen nagy jelentősége van. A beszéd felismerő rendszer legfontosabb minőségi jellemzője az a valószínűség  $\leq 1$ , hogy az **egyes szavak** korrekt felismerésére jellemző. Egy szó mindig csak meghatározott valószínűséggel ismerhető fel. Ebben szerepet játszik a környezeti zaj, a beszéd akusztika és maga a beszélő a maga fizikai és pszichikai állapotával.

Példaként említsük meg egy  $p = 0,95$  egyedi szófelismerési hányadú rendkívül pontos rendszert. Ez 5% hibásan felismert szót jelent. Egy három szóból álló mondatban a helyes bevitel valószínűsége  $p = 0,95 \times 0,95 \times 0,95 = 0,86$ -ra csökken.

Ez a kis példa azt kívánja megvilágítani, hogy egy beszéd felismerő rendszernek **milyen nagy egyedi szófelismerési valószínűséggel** kell rendelkeznie.



A beszéd felismerés elve: a feladatoknak rendszerkomponensekhez történő rendelése az „adat- jellegzetesség kiválasztás útján” alapul szerint



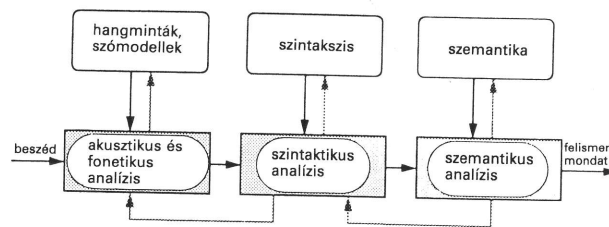
A 3–7 ábra a 47. oldalon egy ilyen rendszer koncepcionális összetevőit mutatja be vázlatosan.

Az elv az, hogy az egyes beszédelemek speciális jellemző ismérveit összehasonlítják egy korábban készített jellemző tulajdonságú készlettel. Ehhez először mindig az aktuális beszédszekvencia ismérveit minősítik. Ezek után következik a meglévő referenciákkal való összehasonlítás az aktuális beszédszekvenciákban való előfordulások megállapítására. Az azonosított beszéd ekkor már mint paraméterezett szekvencia tárolható, továbbítható vagy feldolgozható.

A konkrét megvalósítások a tulajdonságok megállapítására leggyakrabban dedikált alkatrészeket vagy jelprocesszorokat alkalmaznak. Az összehasonlítást és a döntést rendszerint a rendszer főprocesszora végzi. A lexikont a referenciatulajdonságokkal együtt legtöbbször a számítógép háttértárolója tárolja.

A konkrét eljárások a jellegzetes ismérvek definíciójában különböznek. A 47. oldal 3–7 ábráján bemutatott elvek különböző ismérvekre vonatkoztatva is többször alkalmazhatóak.

A 47. oldalon lévő 3–7 ábrán bemutatott beszédfelismerési elveket a 3–8 ábra szerint vázolt három szakaszra bonthatjuk.



3–8 ábra

A beszédfelismerés komponensei. Idő- és frekvenciatartománybeli eljárások

1. A mindenkori eljárások jellegzetes ismertetőjeleire vonatkozó módszer alapján a 47. oldal 3–7 ábrájának elveit először a hangmintákra és/vagy a szómodellekre alkalmazzuk. Itt egy **akusztikus és fonetikus analízis** történik.
2. A második lépésben az eddigiekben meghatározott beszédelemeket szintaktikai analízisnek vetjük alá. Ezáltal a korábbi vizsgálat hibáira derülhet fény. Gyakran előfordul, hogy az első lépésben nem lehet egyértelmű döntést hozni, ekkor a **szintaktikai analízis** további támogatást nyújt.
3. A harmadik lépésben történik meg az eddig felismert beszéd **szemantikai analízise**. Itt a korábbi döntések hibái felismerhetők és a további analízáló eljárásokkal ezeket együttesen ki lehet küszöbölni. Ez a lépés, mesterséges intelli-

gencia és a neurális hálózatok jelenlegi fejlettségi szintje mellett nem könnyen alkalmazható.

**Hangzajok** legtöbbször az ismérveknek az idő- és/vagy a frekvenciatartományban vizsgálataival dolgoznak. Ezen kritériumok számára a beszédelemek (formánsok, hangzajok stb.) éppúgy rendelkezésre állanak, mint a beszédvisszaadásnál.

**Hangzajok problémája** a teremakusztika a jelen lévő zajok miatt. Itt ezenkívül a hangzajokról és tárgyakról frekvenciafüggően visszavert hanghullámoknak a primer hangzajokkal való kölcsönhatása is létrejön.

**Meghatározottnak** kell lenniük továbbá a szóhatároknak. Itt arra kell ügyelni, hogy a szomszédos szavak gyakran egybefolyhatnak.

**A beszédegységeknek** az alkalmazott mintákkal való összehasonlításához időnormát kell hozzárendelnünk. Ugyanazon szavak ugyanis röviden és hosszánként is kiejthetők. Az időtengely azonban nem nyújtható vagy zsugorítható egyszerűen, hiszen a méretváltoztató tényező nem arányos az összédtartamokkal. Előfordulnak rövid és hosszú zöngétlen hangzások (pl. s, sz). Az egyes hangok különböző mértékben nyújthatók és részben van egy olyan minimális időtartam, ami feltétlenül szükséges a megértésükhöz.

### Beszédőfüggetlen és beszédőfüggő beszédbeviteli rendszerek

Egy **beszédőfüggetlen** rendszer ugyanolyan megbízhatóság mellett lényegesen kevesebb szó felismerésre képes, mint egy, a beszélőhöz kötött rendszer. Ezért minden **beszédőhöz kötött** rendszer előzőleg „begyakoroltatandó”. Ilyenkor legtöbbször előre adott beszédmintákat kell utánamondani. Abból indulunk ki, hogy sok mai **beszédőfüggetlen** rendszer kevesebb mint félórás előzetes gyakorlásra van szükség. A beszélőhöz **kötött** rendszerek kb. 20 000 szó felismerésére képesek. Beszédőfüggetlen rendszerek nagyságrendileg 100 szó felismerésére alkalmasak. Ezeket a számokat csak **átlagos** becsléseknek szabad tekintenünk. Konkrét összehasonlításnál a peremfeltételeket nagyon pontosan kell ismernünk. (Vajon a méréseket süketszobában végezték-e? Alkalmazkodott-e a beszélő a rendszerhez, például az időnormák egyeztetésére stb. érdekében?)

## 35 Beszédátvitel

A **beszédátvitel** területe a **beszédjelek** lehetőleg **hatékony kódolásával** foglalkozik, **azaz** a céllal, hogy lehetőleg kis átviteli sebességre legyen szükség. Emellett ne álljon **elő** számottevő minőségromlás. A további részleteket a kódolásról és a tömörítésről az 5. fejezetben a 73. oldalon tárgyaljuk („Adattömörítés”). Itt csupán

néhány olyan elvet tárgyalunk, melyek szoros összefüggésben állnak a beszédbevitellel és a beszédkiejtéssel.

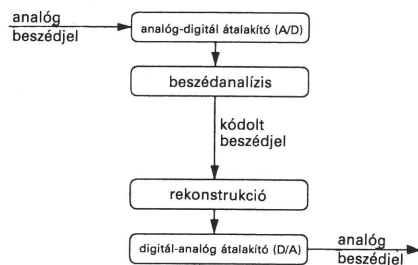
### Jelalak kódolás

A kódolásnak ez a fajtája nem vesz tekintetbe semmilyen beszédspecifikus sajátosságot és paramétert. Ezáltal az audiojel leghatékonyabb kódolására törekszik. A stereo audiojeleknél szokásos PCM-eljárás, amelyet a legmagasabb követelményű CD-technológiában alkalmaznak, a következő adatsebességű:

$$\text{Adatsebesség} = 2 \times 44\,100/\text{s} \times \frac{16 \text{ bit}}{8 \text{ bit/byte}} = 176\,400 \text{ byte/s.}$$

A távbeszélő minőség ezzel az 1 378 125 Kbit/s adatsebességgel szemben csupán 64 Kbit/s adatsebességet követel meg. A *delta kód modulációval (DPCM, Delta Code Pulse Modulation)* legalább ilyen minőség érhető el 56 Kbit/s adatsebességgel. Az *adaptív kód modulációval (ADPCM)* további javulás érhető el 32 Kbit/s-ra.

### Forráskódolás



3-9 ábra

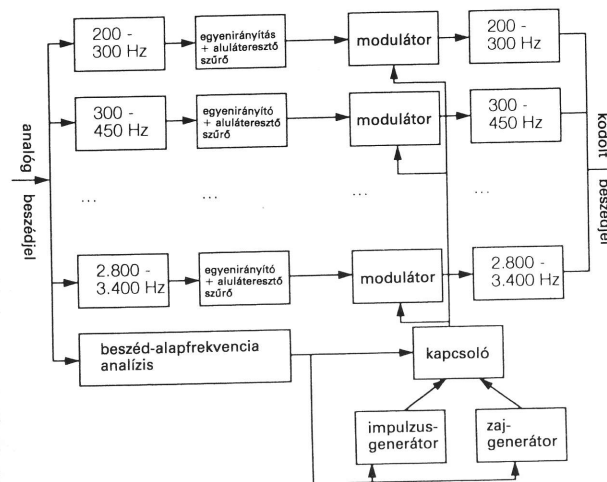
Forráskódolás, paraméterezett rendszerek. Egy beszédátviteli rendszer komponensei

Paraméterezett rendszerek forráskódolást alkalmaznak. Itt a beszédspecifikus sajátosságokat adatszőkítésre használják. A következőkben ismertetésre kerülő csatorna vocoder egy ilyen eljárásra mutat példát, melynek elvét ábrázoltuk a 3-9 ábrán.

A csatorna vocoder a *subband kódolás* kiterjesztése. Itt a beszédanalízis a jelet egy sor frekvenciasávra bontja, mivel abból a feltételezésből indul ki, hogy a beszédre mindig csak bizonyos frekvenciamaximum a jellemző. A 3-10 ábra az 51. oldalon egy ilyen sávzélesség-felosztásra mutat példát. Járulékosan kihasználják a zöngés és a zöngétlen hangok közötti különbségeket. A zöngétlen hangokat zajgenerátorral gerjesztik. A zöngés hangokat impulzusok megfelelő sorozatával

gerjesztik. Ehhez az impulzusok ismétlődési frekvenciája pontosan megfelel az előzőleg kimért beszéd alappfrekvenciáknak. Egy csatorna vocoder segítségével mintegy 3 Kbit/s adatsebesség valósítható meg. A minőség azonban nem minden esetben kielégítő.

További eljárások – a jövő európai mobilhálózat követelményeiből kiindulva – 13 Kbit/s adatsebességet alkalmaznak. Manapság egy további csökkentéssel 6 Kbit/s elérése a cél. Ekkor is el kell érni azt a minőséget, mint a nem tömörített 64 Kbit/s adatsebességű jelnél.

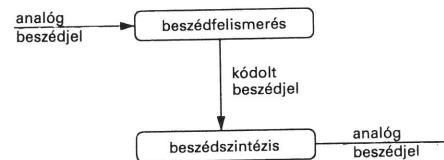


3-10 ábra

Egy csatorna vocoder beszédanalízis komponensei

### Felismerő szintézis eljárások

Kísérletek folynak az irányban is, hogy a tisztán felismerő szintézis eljárással a beszédjel átviteli sebességét számottevően csökkentsék.

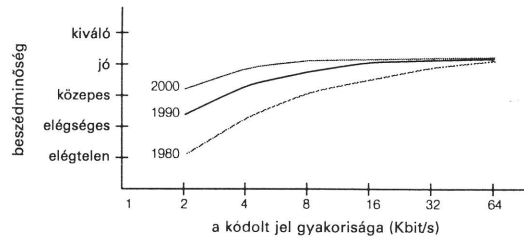


3-11 ábra

Felismerő-szintézis rendszer. Egy beszédátviteli rendszer komponensei

Itt egyfelől beszédanalízis, másfelől a beszéd-szintézis rekonstrukciója történik meg. A beszédelemeknek még csupán a jellegzetes ismertetőjeleit viszik át. Ezek például lehetnek az egyes formánsokhoz tartozó digitális szűrők közepes frekvenciái és sávzélességei. Ezekkel a módszerekkel egész 50 Bit/s adatsebességig lehetett menni. A reprodukált beszéd minősége és a felismerési hibaarány ma még nem elfogadható.

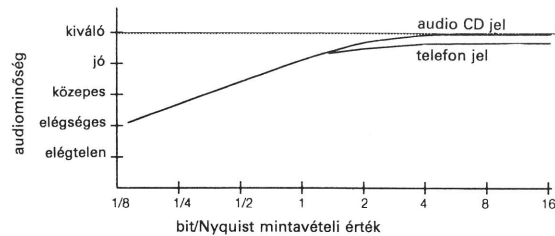
### Elérhető minőség



3-12 ábra

Tömörített beszéddel elérhető minőség a tömörített jel adatsebességének függvényében [FLAN92a]

Az alapvető kérdésfelvetés a beszéd és audioátvitelnél a multimédia rendszerekben az, hogy milyen adatsebességgel érhető el előre megadott minőség. E tekintetben nagyon érdekes a Flanagan által közzétett függvénykapcsolat, ami a 3-12 ábrán látható. Abból indulunk ki, hogy a telefon minőség 8 Kbit/s-nál alacsonyabb adatsebességgel elérhető.



3-13 ábra

Tömörített audiojellel elérhető minőség a mintavételi értékekhez rendelt bitszám függvényében a lehetséges minimális mintavételi gyakoriságra vonatkoztatva (Nyquist gyakoriság), [FLAN92a] szerint

A 3-13 ábrán az 52. oldalon az audiominőség látható a mintavételekhez tartozó bitszám függvényében. Itt például egy kiváló CD minőséget biztosító 16 bites mintavételi érték egy 2 bites mintavételi értékre csökkenthető. Ez másképpen azt jelenti, hogy az eredeti adatsebességnek csak  $\frac{1}{8}$  része szükséges. 1991-ben például az Audio Process Technology Ltd. cég már bemutatott egy rendszert, melyben 4 bit/mintavétellel CD minőséget tudtak elérni. A szerző 15 különböző típusú zene-műből kivágott, tömörített és nem tömörített részletet hasonlított össze. Még egy nagyon jó minőségű berendezésen történő visszajátszásnál sem lehetett szubjektív különbséget megállapítani.

## 4. Videotechnika

Az audioteknika mellett a video- illetve televíziótechnika képezi a multimédia rendszerek alapját. E fejezetben néhány erről a területről származó fejlesztést mutatunk be, amely a *video* médium alapvető megértése szempontjából jelentős. A további részletek tekintetében utalunk a televíziótechnika területén fellelhető alapvető szakirodalomra. (Lásd a [Joha92a] hivatkozást.) A [MICR92a] hivatkozásban például néhány korszerű hardver megoldás is fellelhető.

A jelenlegi és a jövőbeli televíziótechnikában, beleértve HDTV-t is, a megjelenítés az emberi szem sajátosságaiból indul ki. A tárgyalás a Németországban érvényes eljárásokat és értékeket tekinti kiindulásnak és a további módszereknél ezekkel a paraméterekkel való kapcsolatokra hivatkozik. Végül tárgyalni fogjuk a digitális televíziótechnikát az ennél fellépő adatsebességgel együtt, és a számítógép szerepét a videotechnikában.

### 4.1 Az emberi szem sajátosságai

Az álló- és mozgóképek emberi érzékelésének szerve a szem. A szem belső sajátosságai az idegrendszeri *feldolgoásával* együtt meghatározzák a videorendszerek néhány koncepcionális peremfeltételét.

Egy kép egyes pontjai csak a kép nagyságától és a szemlélőnek a képtől való távolságától függően láthatók. A szem **helyfüggő felbontóképessége** meghatározza ezáltal a kép nagyságát, beleértve a vízszintes és függőleges képpontok számát.

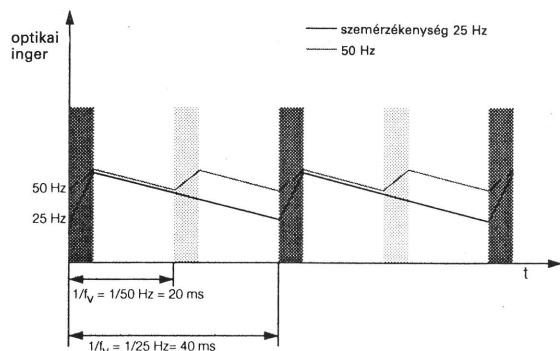
A szem mint **optikai aluláteresztő** is felfogható; a magas frekvenciák az éleken és nem az összemosódó felületekben jelennek meg. A szem a változásokat csak egy meghatározott határig tudja megkülönböztetni. Ezt a hatást az ember egy egyszerű kis kísérlettel is tudja igazolni: *pislogással*, illetve a *szem becsukogatásával* a levágási frekvencia lecsökken, ezért csak a világos és sötét foltok maradnak felismerhetők.

További jellemző a **mozgóképfelbontás határa**. A folyamatos akusztikus nyomáshullámmal ellentétben az egyes képek diszkrét szekvenciáit folyamatos szekvenciaként vesszük tudomásul. Körülbelül 16 kép/s-tól áll elő a folyamatos mozgás érzete. A mozgóképfelbontás hatását sajnos nem tudjuk közvetlenül alkalmazni, mert a szem az optikai ingereket csak tökéletlenül tudja tárolni.

Túl alacsony képváltásfrekvencián a fényességérzet periodikus ingadozása áll elő, ez az ún. **villódzás**. A határérték itt 50 Hz felett van. Emiatt a villódzásmentes mozgóképhez viszonylag magas képváltásfrekvenciát kell alkalmaznunk. Mind a filmtechnikában, mind a televízióanaltechnikai fogások segítségével ennél alacsonyabb frekvenciákon tudunk dolgozni.

A filmnél például a 16 kép/s képváltás nagyon zavaró, ha semmilyen technikai fogást nem alkalmazunk. Ennek a villódzásnak az elkerülése érdekében a filmnél egyetlen kép vetítését kétszer megszakítják, így módon  $3 \times 16 = 48$  Hz képráfrissítési frekvencia jön létre.

A televízióanaltechnikai legutóbbi időkig egy meglehetősen drága képtárolási módszerrel tudták a villódzási jelenséget kiküszöbölni. A tárolóban a képváltási frekvenciával frissítették a tartalmat, ami a mozgóképfelbontási frekvencia felett volt (pl. 25 Hz). A képernyőn a képadatok nagyobb frekvenciával jelentek meg (pl. 70 Hz), ami a villódzási hatást kiküszöbölte és ami megfelel a jó minőségű számítógépek 70 Hz-es képváltási frekvenciájának.



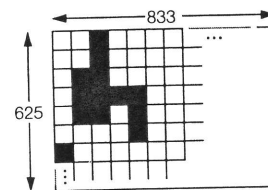
4-1 ábra  
Villódzási hatás: a szemben keletkező inger 25 Hz-es és 50 Hz-es képsémlelőadási frekvenciánál

A televízióanaltechnikai teljes kép két sorreltolással egymásba illesztett félképből tehető össze. Az egyes félképeket egymást követően egy sorreltolási eljárással viszik át. A teljes kép átvitele 25 Hz, illetve 30 Hz-cel történik, miáltal a félképek átvitelére

$2 \times 25 = 50$  Hz, illetve  $2 \times 30 = 60$  Hz adódik. Az 56. oldalon 4-1 ábra erre mutat példát. A szemérzékenység 25 Hz-es képváltásnál lényegesen gyorsabban csökken (szaggatott vonal), mint 50 Hz-nél.

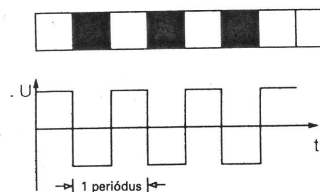
## 4.2 Fekete-fehér televízió

A mai televízióban a **szélesség-magasság** arány 4:3. A kép a német és az európai szabvány szerint 625 sorból áll. Ebből a soronkénti képpontok számára  $4/3 \times 625 \approx 833$  adódik (lásd 4-2 ábra).



4-2 ábra  
Egy mozgóképfelbontása. A szélesség-magasság arány 4:3

Az ilyen fekete-fehér jel sávszélességét első közelítésben az egy soron belüli **fekete-fehér-fekete** hirtelen átmenetekből határozhatjuk meg. Egy ilyen sor, amint ezt a 4-3 ábra mutatja, maximum  $833/2$  periódust tartalmaz.



4-3 ábra  
A televíziójel sávszélessége. A maximális frekvencia

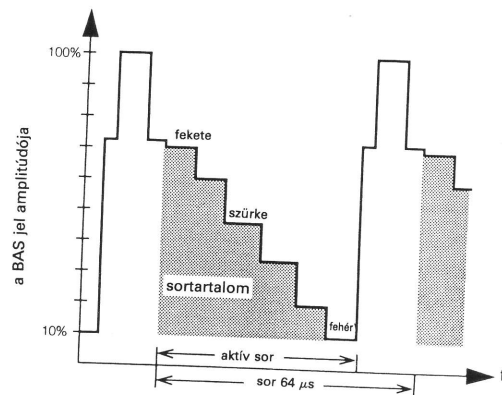
A soronkénti képpontok tényleges száma az eltérítési időből ( $64 \mu\text{s}$ ) és a mintavételi frekvenciából ( $13,5 \text{ MHz}$ ) számítható. Ebből  $864$  képpont/sor adódik [CCIR91a]. Ily módon a maximális videosáv szélesség elméleti értéke a következőképpen adódik.

$$864/2 \text{ periódus/sor} \times 625 \text{ sor/kép} \times 25 \text{ kép/másodperc} = 6,75 \text{ MHz.}$$

Az  $5,5 \text{ MHz}$  sáv szélesség azonban teljesen kielégítő minőséget ad. Gyakorta az  $5 \text{ MHz}$  szintén szokásos érték. A Kell-tényezőt szintén figyelembe kell venni. A  $625$  soros felbontás ellenére ez nem használható ki teljesen, a lebegési hatás nem elfogadható képet ad. Ezt a tényezőt kísérletileg lehet meghatározni és eredetileg értéke  $0,64$  (időnként megadnak  $0,67$  értéket is).

Emiatt a maximális függőleges felbontásra

$$625 \times 0,64 = 400 \text{ sor adódik.}$$



4-4 ábra

A képjel, a képkiváltási jel és a szinkronjel (a német rövidítés BAS). Különböző szürkességi fokozatok előállítására egy sorban

A képcsőben minden képet soronként, balról jobbra haladva rajzolnak fel. Minden sor végét fel kell ismerni és a sugár visszafutása alatt a képcsövet el kell sötétíteni (Gondoljuk meg, hogy a képcső a digitális technika mai állása mellett még mindig nagy jelentőséggel bír,) amit a 4-4 ábrán bemutatott, a vízszintes eltérítésben előállított kihagyásokkal lehet elérni.

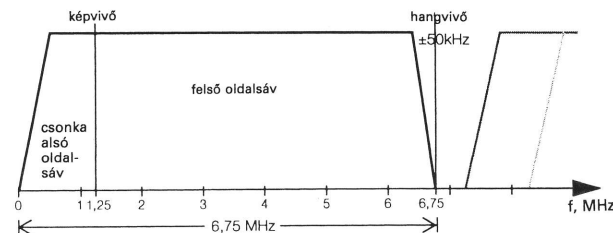
Minden sor lefutására maximum

$$40 \text{ ms}/625 \text{ sor} = 64 \mu\text{s} \text{ áll rendelkezésre.}$$

Ily módon a sorfrekvenciára

$$1/64 \mu\text{s} = 15\,625 \text{ Hz adódik.}$$

A függőleges kioltó szünetek a félképek között találhatóak. Ezeknél szintén sötét lesz a képcső, miközben a sugár a képcső tetejére visszajuthat. Az adó és a vevő félképegyüttfutásához és sorogyüttfutásához szinkronizáló jelek szükségesek.



4-5 ábra

A BAS jel spektruma

Ezekből a kép-, kioltó- és szinkronjelekből áll elő a (németül) a BAS névvel jelölt összetett jel, melyet az 58. oldalon a 4-4 ábra szemléltet. A fényességet (fényssűrűséget) általában *luminanciának* nevezik. Amint a 4-4 ábra az 58. oldalon mutatja, a feszültség az aktív soron belül a luminanciával arányos.

Egy ilyen kb.  $5,5 \text{ MHz}$  sáv szélességű jel amplitúdómodulációjához  $11 \text{ MHz}$  vezetési sáv szélességre lenne szükség. Féloldalsáv amplitúdómoduláció alkalmazásával az alsó oldalsávot csak csonkítva visszük át (lásd a 4-5 ábrát). Ily módon a vevőben a szinkronizációhoz szükséges képvivőt is átvisszük. Az analóg televízióban a hangot egy CCIR szabvány szerinti frekvenciamodulációval állítjuk elő, melynek vivőfrekvenciája a képvivő felett található  $5,5 \text{ MHz}$  távolságra.

### 4.3 Színes televízió

A szürkességfokozatokon kívül a mai televízióban a kép színinformációt is hordoz. A különböző színinformációk kódolása akár külön jelben, akár egy, a luminanciával közös jelben történhet.

### Komponenskódolás

A komponenskódolás a videoinformáció alkatelemeit egymástól elkülönítve veszi figyelembe. Így itt pl. a szinkronjelet külön, járulékosan viszik át.

Egy **RGB jel** a vörös, a zöld és a kék színjeleket külön-külön tartalmazza. Minden más szín ezen három elsődleges szín kombinációjaként, az additív színkeverés elve alapján állítható elő. A vörös (R), a zöld (G) és a kék (B) színek úgy vannak szabványosítva, hogy  $R + G + B = 1$  éppen a semleges fehér színt adja.

Az egyes színek elkülönítése helyett adódik a lehetőség, hogy a fénysűrűséget (luminancia) és a színinformációt (krominancia) különválasszuk. A luminanciát ugyanis a fekete-fehér jellel való kompatibilitás miatt mindig át kell vinni.

Az **YUV jel** a következőképpen tevődik össze:

$$Y = 0.30R + 0.59G + 0.11B$$

$$U = (B - Y) \times 0.493$$

$$V = (R - Y) \times 0.877$$

Hasonló kódolással áll össze az **YIQ jel** az NTSC rendszerben:

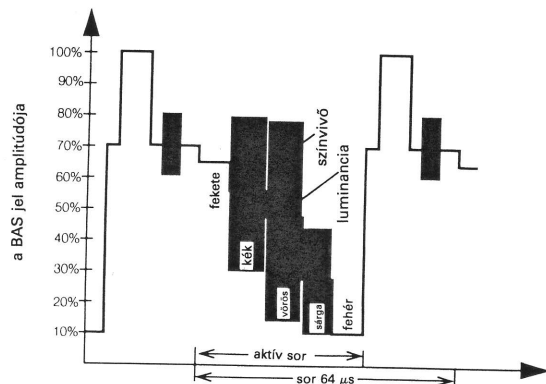
$$Y = 0.30R + 0.59G + 0.11B$$

$$I = 0.60R - 0.28G - 0.32B$$

$$Q = 0.21R - 0.52G + 0.31B$$

### FBAS jel

A komponenskódolás alternatívájaként az **FBAS** jel mindezeket az információkat egyetlen jellel foglalja össze. Az egyes összetevőket (RGB, YUV, ill. YIQ) egyetlen jellel kell kombinálni. Ekkor mindig a fénysűrűség-információból és a krominancia különbség jelekből indulunk ki. A jelek összetevésénél azonban a krominancia jelek a luminancia jelekre zavarólag hatnak. A televíziótechnikában ezért olyan modulációs eljárásokat alkalmaznak, melyek egymásra lehetőleg csekély kölcsönhatást fejtenek ki.



4-6 ábra

A színjel, a képjel, a képkioltási jel és a szinkronjel, az egy soron belüli színek megjelenítésével együtt

### NTSC

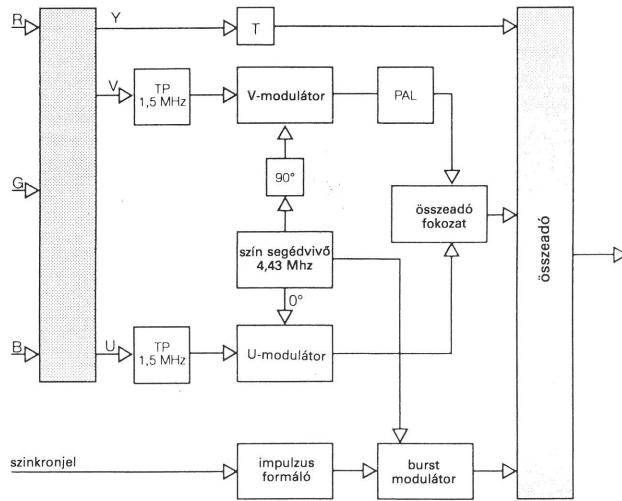
Az NTSC a **National Television Systems Committee** rövidítése és a legrégebbi, ennek következtében a legelterjedtebb televíziós szabvány. Az USA-ból származik. Itt színvivőként a kb. 4,429 MHz-es vagy a kb. 3,57 MHz-es frekvenciát használják. Az NTSC kvadratúra-amplitúdómodulációt használ elnyomott színsegédvívvel és kb. 30 Hz-es kép váltás frekvenciával dolgozik; a kép 525 sorból áll.

### SECAM

A SECAM a **Sequential Couleur avec Memoire** rövidítése és túlnyomórészt Franciaországban, valamint Kelet-Európában alkalmazzák. Az NTSC-vel és a PAL-lal ellentétben frekvenciamoduláción alapul. A SECAM a PAL-hoz hasonlóan 25 Hz-es képfrekvenciát alkalmaz és minden kép 625 sorból áll.

### PAL

A **Phase Alternating Line** (PAL) rendszert 1963-ban W. Bruch javasolta a Telefunken-nél. E rendszert Nyugat-Európa nagy részén alkalmazzák. A 4-7 ábra a 62. oldalon a képelőállítás folyamatát mutatja be vázlatosan.



4-7 ábra

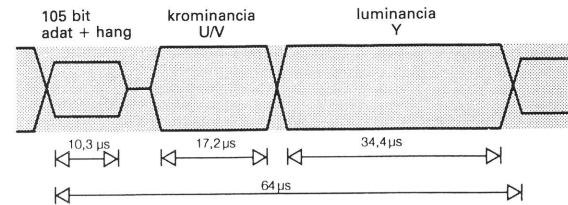
Színes jel előállítása a PAL eljárásnál: az RGB jeltől az FBAS jelig a PAL szabvány szerint

A színsegédvív az FBAS jel spektrumában a képvivőtől 4,43361875 MHz távolságra található. Az alapelv egy kvadrátúra-amplitúdómoduláció. A színsegédvívöt az U szintkülönbségi jellel közvetlenül megszorozzák.

A 90 fokos fáziseltolású színsegédvívöt ezek után a V szintkülönbségi jellel szorozzák meg. Ezután a két eredményt összeadják. Ez így egy reguláris kvadrátúra-moduláció lenne. Magához a tényleges kvadrátúramodulációhoz képest a modulált V jel minden második sorának fázisát megfordítják, miáltal a fázishiba csökken.

#### 4.4 D2-MAC

A D2-MAC (*Doubinary Multiplexed Analogue Components*) egy közbülső fokozat a mai televízió és a HDTV között (l. a 4.5 fejezet 63. oldal). Ezt a közbülső megoldást Németországban, mint a PAL-rendszer követőjét már bevezették. Elfogadottsága azonban nem éri el a várt szintet.



4-8 ábra

D2-MAC időmultiplex eljárás

A D2-MAC-ban az egyes komponenseket időmultiplex módon viszik át. A 4-8 ábrában a mozgókép egy sorának felosztása látható. A rendelkezésre álló 64 µs-ból külön-külön nettó 34,4 µs jut a luminanciára, 17,2 µs a krominanciára és 10,3 µs a hangra és az adatokra.

Miként a PAL-rendszerben, itt is 625 sorból áll a kép. Ebből 574 sor látszik, míg a PAL-nál 575 látható. A képváltásfrekvencia és a sor egymásbafejtés ugyanolyan, mint a PAL-szabványnál. Mind a 4:3, mind a 16:9 szélesség-magasság arányt támogatja.

Az audio és az adatjel egy kétállapotú bináris eljárással van kódolva. A digitális hang és a járulékos adatok számára összesen  $105 \text{ bit}/64 \mu\text{s} = 1,64 \text{ Mbit/s}$  áll rendelkezésre. Az audio különböző üzemmódokban működtethető. Két igen jó minőségű sztereójelből 8 kevésbé jó csatornáig kódolható.

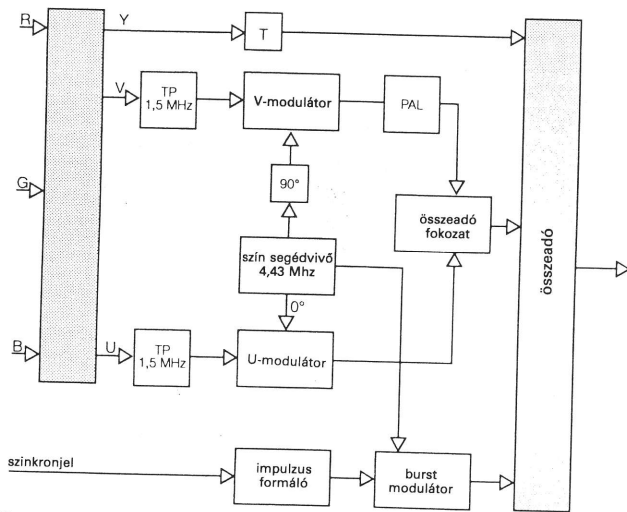
#### 4.5 HDTV

A jövőbeli televíziótechnika a *High Definition Television (HDTV)* elvén fog alapulni. A következőkben az eredetileg európai rendszerből kiindulva néhány jellemző példát fogunk bemutatni. Az átviteli eljárás a jelenleg kedvelt európai HDTV-rendszerben ugyanúgy van jelölve; *HD-MAC* (High Definition Multiplexed Analogue Components) [VHCa89a].

A HD-MAC-et egy sor új követelmény alapján specifikálták:

- A képnagyság és a felbontás definíciója alapjául egy kb. 2–3 m-es néző távolságot vettek fel. Ezáltal a vízszintes látószög kb. 45°. A szélesség-magasság arány 16:9. Ebből több mint 1100 aktív sor adódik, soronként 1900 képponttal. A  $2 \times 625 = 1250$  sorból legalább 1150-nek aktívnak kell lennie.





4-7 ábra

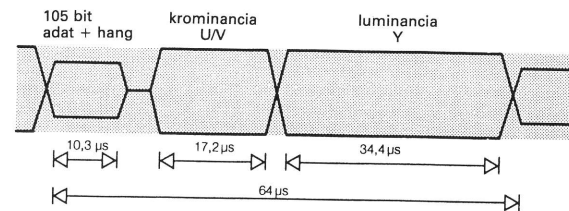
Színes jel előállítás a PAL eljárásnál: az RGB jeltől az FBAS jelig a PAL szabvány szerint

A színsegédvív az FBAS jel spektrumában a képvivőtől 4,43361875 MHz távolságra található. Az alapelv egy kvadrátúra-amplitúdómoduláció. A színsegédvívöt az U szintkülönbségi jellel közvetlenül megszorozzák.

A 90 fokos fáziseltolású színsegédvívöt ezek után a V színelkülönbségi jellel szorozzák meg. Ezután a két eredményt összeadják. Ez így egy reguláris kvadrátúra-moduláció lenne. Magához a tényleges kvadrátúramodulációhoz képest a modulált V jel minden második sorának fázisát megfordítják, miáltal a fázishiba csökken.

#### 4.4 D2-MAC

A D2-MAC (*Doubinary Multiplexed Analogue Components*) egy közbülső fokozat a mai televízió és a HDTV között (l. a 4.5 fejezet 63. oldal). Ezt a közbülső megoldást Németországban, mint a PAL-rendszer követőjét már bevezették. Elfogadottsága azonban nem éri el a várt szintet.



4-8 ábra

D2-MAC időmultiplex eljárás

A D2-MAC-ban az egyes komponenseket időmultiplex módon viszik át. A 4-8 ábrában a mozgókép egy sorának felosztása látható. A rendelkezésre álló 64  $\mu\text{s}$ -ból külön-külön nettó 34,4  $\mu\text{s}$  jut a luminanciára, 17,2  $\mu\text{s}$  a krominanciára és 10,3  $\mu\text{s}$  a hangra és az adatokra.

Miként a PAL-rendszerben, itt is 625 sorból áll a kép. Ebből 574 sor látszik, míg a PAL-nál 575 látható. A képváltásfrekvencia és a sor egymásbafűzés ugyanolyan, mint a PAL-szabványnál. Mind a 4:3, mind a 16:9 szélesség-magasság arányt támogatja.

Az audio és az adatjel egy kétállapotú bináris eljárással van kódolva. A digitális hang és a járulékos adatok számára összesen  $105 \text{ bit}/64 \mu\text{s} = 1,64 \text{ Mbit/s}$  áll rendelkezésre. Az audio különböző üzemmódokban működtethető. Két igen jó minőségű sztereójelből 8 kevésbé jó csatornáig kódolható.

#### 4.5 HDTV

A jövőbeli televíziótechnika a *High Definition Television (HDTV)* elvén fog alapulni. A következőkben az eredetileg európai rendszerből kiindulva néhány jellemző példát fogunk bemutatni. Az átviteli eljárás a jelenleg kedvelt európai HDTV-rendszerben ugyanúgy van jelölve; *HD-MAC* (High Definition Multiplexed Analogue Components) [VHCa89a].

A HD-MAC-et egy sor új követelmény alapján specifikálták:

- A képnagyság és a felbontás definíciója alapjául egy kb. 2-3 m-es néző távolságot vettek fel. Ezáltal a vízszintes látószög kb. 45°. A szélesség-magasság arány 16:9. Ebből több mint 1100 aktív sor adódik, soronként 1900 képponttal.  $2 \times 625 = 1250$  sorból legalább 1150-nek aktívnak kell lennie.

- A képélességnek a kétszer annyi sor és a seregymásbafűzés miatt lényegesen jobbnak kell lennie. Ebből a luminancia sáv szélességre több mint 30 MHz adódik.
- A kontraszt átfogásnak is javulnia kell. A felvételnél 40-ről 100-ra, a visszaját-zásakor 300-ra.
- A mozgásfelbontás legalább 16 Hz-et követel meg. Ezt a HDTV másodper-cenként 50 különböző félképpel tudja megvalósítani.
- A villódzás nagyobb felületeknél kritikusabb, mint kisebbeknél. Ebből egy kb. 100 Hz-es képismétlődési frekvencia adódik a vevőben.
- A hangminőségnek is lényegesen javulnia kell a jelenlegi televízióhoz képest. Itt a CD-k sztereo hangminőségére törekszenek. Ezenkívül több hangcsatorna szükséges a többnyelvű közvetítéshez.
- Kívánatos a meglévő rendszerekkel való kompatibilitás (PAL, D2-MAC). Ez valamilyen konverterrel is elérhető.

A multimédia rendszereknél az egyik legfontosabb paraméter az **adatsebesség**. E tekintetben a CCIR [CCIR82]-szabvánnyal összevetve a következő összehasonlí-tást tehetjük (l. a 4.6 fejezet 67. oldal):

A sorok száma a jelenlegi 625-ről a duplájára, 1250-re növekszik. A képpontok száma soronként első közelítésben a következőképpen növekszik

$$1250 \times 16/9 > = 2222,22 \text{ képpont}$$

szemben a **PAL jellel, ahol**

$$625 \times 4/3 = 833,33 \text{ képpont.}$$

A CCIR szerint exakt módon 32  $\mu$ s-os soridő és 72 MHz-es mintavételi frekvencia mellett

$$2304 \text{ képpont/sor (amiből 1920 képpont látszik)}$$

szemben a

$$864 \text{ képpont/sor értékkel.}$$

Ebből összességében a következő tényező adódik:

$$2 \times 2 \times \frac{16/9}{4/3} = \frac{2304}{864} = 5,33$$

Ebből a HDTV adatfolyamra a következő adatsebesség adódik:

$$216 \times 10^6 \text{ bit/s} \times 5,33 = 1,152 \times 10^9 \text{ bit/s}$$

A kompressziós eljárásokkal számottevő minőségromlás nélkül az adatsebesség **216 Mbit/s-ra** csökkenthető a jelszakaszokra osztással.

Az adásnak a stúdióban való előkészítésekor a tényleges adást egy nyolcszoro-sra terjedő keverésnek kell megelőznie. A kódolás hibái ezen stúdióelőkészítés miatt láthatóvá válnak. Itt az ún. **Contribution-Quality-nek** (stúdiótól stúdióig, mintavételi helytől a stúdióig) kb. 140 MHz-nek kell lennie ennek elkerülésére. A **TV-Distribution-Quality-ra** a 34 Mbit/s érték a szabványosítás küszöbén áll. A D2-MAC-hoz és a PAL-hoz való kompatibilitás elérése érdekében egy **kétfoko-sú mintavételi sémát** alkalmaznak. A valójában nagy felbontóképességű HDTV **2304 sorból** minden 625 sort csökkentett teljesképváltással visznek át. (A **HD-MAC vevőben** ellenben **mind a teljes felbontás, mind a teljesképváltás megmutat-kozik**). Ez a sorszámcsoökkentés és a félképváltási frekvencia felezése egyszerűbbé teszi ezen jelek a PAL és a D2-MAC jelekre történő konvertálását.

A HD-MAC önmagában megkülönbözteti a stúdió, a gyártási, az átviteli és a rep-likációs szabványt [Scho98].

#### Az elő szabvány

Ez egy mindenek fölé rendelt szabványról van szó, mely lehetővé teszi, hogy ebből **mindenből** az összes egyéb HD-MAC jelet és a más szabványú TV jeleket egyszerűen **alkalmazhatunk**. Ezért ezt **High Definition Progressive-nek (HDP)** nevezik. Itt egy kép-**2304 sorból** áll 50 Hz-es teljesképváltási frekvenciával. Ekkor ún. progresszív rasz-**ter** alkalmaznak (nincs sorvisszafutás). Ez egyszerűsíti a szabványkonvertálást.

A luminancia mintavételi frekvencia 1,44 MHz.

A krominancia mintavételi frekvencia egyenként  $2 \times 5,33 \times 6,75 \text{ MHz} = 72 \text{ MHz}$ .

Ebből adódik a 4:2:2 kódolásnál a következő adatsebesség:

$$(144 \text{ MHz} + 72 \text{ MHz} + 72 \text{ MHz}) \times 8 \text{ bit} = 2,304 \times 10^9 \text{ bit/s}$$

#### A gyártási szabvány

A gyártási szabvány különböző lehetőségeket ír le, a stúdió szabványtól és a végbe-**vezetés** szabványig történő egyszerűbb levezetésére. Ezt **High Definition Quin-****telem-nek (HDQ)** nevezik. A gyártási szabvány nem ír elő sorvisszafutást és egye-**ben** lényegileg megegyezik az átviteli szabvánnyal.

### Az átviteli szabvány

Az átviteli szabványnak, a *High Definition Interlace*-nek (**HDI**), a vevő és az átviteli út szempontjából van jelentősége. Itt 1250 sor 50 Hz félképváltás frekvenciával kerül kódolásra sorvisszafutásnál.

A luminancia sáv szélesség  $5,33 \times 13,5 \text{ MHz} = 72 \text{ MHz}$ .

A krominancia sáv szélessége  $5,33 \times 6,75 \text{ MHz} \times 36 \text{ MHz}$  adódik.

Az átviteli szabványban 4:2:2 szerinti kódolásnál az adatsebességre a következő adódik.

$$(72 \text{ MHz} + 36 \text{ MHz} + 36 \text{ MHz}) \times 8 \text{ bit} = 1,52 \times 10^9 \text{ bit/s}$$

### A reprodukciós szabvány

A reprodukciós szabvány a vevőben történő képvisszaadást írja le. Itt egy digitális képmémlelési tároló által villódzásmentes visszaadás valósítható meg 50 Hz-es teljesképváltási frekvencia mellett.

### A HDTV-rendszerek fejlettségi szintje

Ma világszerte három különböző HDTV-rendszeren dolgoznak:

- Eddig az európai **HD-MAC** eljárást ismertették ebben a fejezetben. Egy EUREKA programot hoztak létre EU 95 néven. Ez a program 1986-ban kezdődött 35 európai gyártó, televíziós társaság és kutatóhely részvételével. Fontosnak tartottak egy bizonyos szintű kompatibilitást. Ennek ellenére a meglévő PAL-rendszerben történő visszaadáshoz konverterre volt szükség.
- Az első HDTV-fejlesztések Japánban történtek. Ez a **MUSE** nevű eljárás nem volt kompatibilis a meglévő televíziós szabványokhoz. Ez a rendszer 1125 képporral és 60 Hz-es képfrekvenciával dolgozik. 1992-ben Japánban ezzel a rendszerrel már napi 8 órás televíziós program létezett [Sawa90].
- Az amerikai törekvések egy **NTSC**-vel kompatibilis rendszer létrehozására irányulnak. Itt 1050 sorral 59,94 Hz képváltással egy teljesen digitalizált rendszeren dolgoznak.

Ez a rövid áttekintés azt mutatja, hogy sajnos valószínűleg egységes HDTV-szabvány sem lesz.

## 4.6 A digitális televízió

Az eddigiekben a különböző televíziótechnikákat és hozzájuk tartozó adatsebességeket foglaltuk röviden össze. Ezeknél azonban még nem beszélhetünk a digitális televízióról. Már 1982-ben létrejött a CCIR-nél egy nemzetközi digitális televíziószabvány, amely többek között egy meghatározott mintavételi ajánlást ír le, beleértve a kvantálást is.

### Zárt kódolás

A digitalizálás legegyszerűbb lehetősége a zárt kódolás. Itt az egyetlen jellel összeállított komponenseket együttesen alakítják egy digitális jellel.

Ha a *BAS jelből* indulunk ki, melyet digitalizálunk, akkor egy 5 MHz sáv szélességű jelet nyerünk, melynek adatsebessége

$$2 \times 5 \text{ MHz} \times 8 \text{ bit} = 80 \text{ Mbit/s}$$

Az *FBAS jelek* digitalizálásához ezt a jelet a színsegédvív frekvenciájának négyeszeresével mintavételezzük. Ennek a számnak a színsegédvív egész számú többszörösének kell lennie, mert ekkor a színsegédvív interferencia zavarása nagyon csekély. Ezenkívül ekkor a modulált színsegédvív feldolgozása a dekóderben egyszerűbb. Ebből a következő adatsebességet nyerjük

$$4 \times 4,43 \text{ Mhz} \times 8 \text{ bit} = 141,8 \times 10^6 \text{ bit/s}$$

Ez a zárt kódolás elvileg egyszerűbb, mint a komponensekre bontás és az azt követő kódolás. Van azonban ennek ellenére is néhány **probléma**:

- Előáll egy zavaró **áthatás** a krominancia és a luminancia információk között.
- A színes televíziójelek zárt kódolása **szabványfüggő**. A sorszámkülönbségen és a képváltásfrekvencián kívül létrejön még egy további különbség az egyes szabványok között. Ezen jelek multiplexelt átvitelénél ez a szabványfüggőség szintén zavaró, mert ekkor különbséget kell tenni a különböző digitális televíziószabványok között.
- A luminancia információ fontosabb, mint a krominancia információ, ezért ehhez nagyobb **sáv szélesség** rendelhető. Egy zárt kódolási rendszerben a mintavételi frekvencia nem illeszthető az egyedi komponensek sáv szélességéhez.
- Ennek következtében az **adatsökkenés** nem illeszthető az egyes komponensek sajátosságaihoz.

- A komponenskódolásnál a mintavételi frekvenciát nem kell a színsegédvív frekvenciájához kötni. Lehetne például a sorfrekvenciához kötve. Ezáltal néhány látható **zavar** kiesik.

Ezek a zárt kódolásnál fellépő nehézségek és hátrányok indokolják a *komponenskódolás* pontosabb megtárgyalását.

#### A komponenskódolás

A komponenskódolás az egyes összetevők, vagyis itt a luminancia jel és a krominancia jel külön-külön történő digitalizálásából áll. Ezek aztán ezt követően multiplexelt módon közösen vihetők át. A luminancia jelet (Y) 13,5 MHz frekvenciával mintavételezzük, ez fontosabb, mint a krominancia jel. A krominancia jeleket (R-Y, B-Y) 6,75 MHz frekvenciával mintavételezzük. A digitalizált luminancia jelet és krominancia jelet egyöntetűen 8 bitre kvantáljuk. A komponensek 4:2:2 arányú, különböző sávszélessége miatt a kódolás szintén 4:2:2 aránnyal jellemezhető. Ezekkel a következő adatsebesség adódik:

$$(13,5 \text{ MHz} + 6,75 \text{ MHz} + 6,75 \text{ MHz}) \times 8 \text{ bit} = 216 \times 10^6 \text{ bit/s}$$

Ezáltal soronként 864 mintavétel jön létre a luminanciára (ebből 720 látható) és 432 mintavétel mindegyik krominancia komponensre (amiből 360 látható). Egy kép 575 sorból áll 25 teljesképpel másodpercenként. Probléma a nagy adatsebesség és az a tény, hogy ezek a jelek nem illenek bele a PCM hierarchiába (139,264 Mbit/s és 34,368 Mbit/s...). Ezért aztán különböző *alszabványokat* definiáltak kisebb adatsebességre. Ezeket az alszabványokat egyszerűen a komponensek (13,5 MHz, 6,75 MHz, 6,75 MHz) sávszélességeiből vezették le.

### 4.7 Alszabványok és további adatcsökkentések

Az **1. alszabvány** a luminancia jelnél 11,25 MHz-es (1/2-a a szabványnak), a krominancia jelnél 5,625 MHz-es mintavételezéssel (1/4-a a szabványnak) dolgozik. Ekkor a következő adatsebesség adódik:

$$180 \times 10^6 \text{ bit/s} = 22,5 \times 10^6 \text{ byte/s}$$

A **2. alszabvány** a luminancia jelhez 10,125 MHz (1/4-e a szabványnak), a krominancia jelhez 3,375 MHz (1/8-e a szabványnak) mintavételezési frekvenciát definiál. Ez

a jelsebesség jól illik a 140 MHz-hez és egy nagyon jó kompromisszum. Ekkor a következő adatsebesség adódik:

$$135 \times 10^6 \text{ bit/s} = 16,875 \times 10^6 \text{ byte/s}$$

A **3. alszabvány** a luminancia mintavételt 9,0 MHz-re (1/4-a a szabványnak), a krominancia mintavételt 2,25 MHz-re (1/8-a a szabványnak) specifikálja. Ekkor a következő adatsebesség adódik:

$$108 \times 10^6 \text{ bit/s} = 13,5 \times 10^6 \text{ byte/s}$$

A televíziókép digitalizálásakor további adatredukciós lehetőségek is vannak, amelyeket befejezésül ismertetünk. Ezeket mindig egy a 2. alszabvány szerinti digitalizált jelle kell vonatkoztatni.

Az első lépésben a *mintavételezési hézagok* kihasználása történik. A ténylegesen látható tartományhoz tartozó jel kódolására szorítkozunk:

- A luminancia itt 648 mintavételi értékből áll soronként. Ebből azonban csak 540 látható.
- A krominancia soronként 216 értékkel van digitalizálva. Ebből azonban csak 180 képpont látható.
- A kép 575 látható sorból áll.

Ebből a következő adatsebesség adódik

$$(540 + 180 + 180) \text{ mintavétel/sor} \times 575 \text{ látható sor/kép} = 517\,500 \text{ mintavétel/kép}$$

$$517\,500 \text{ mintavétel/kép} \times 8 \text{ bit/mintavétel} \times 25 \text{ kép/s} = 103,5 \times 10^6 \text{ bit/s}$$

A következő lépésben egy *vertikális krominanciafelbontás redukció* valósítható meg. Ekkor csak minden második sor krominancia különbségi jelét vesszük át. Soronként váltakozva digitalizáljuk a két krominancia különbségi jelet. (R-Y, B-Y, R-Y, ...). Így módon, a 2. alszabvány alapján a következő adatsebesség értékelhető ki:

$$(540 + 90 + 90) \text{ mintavétel/sor} \times 575 \text{ látható sor/kép} = 414\,000 \text{ mintavétel/kép}$$

$$414\,000 \text{ mintavétel/kép} \times 8 \text{ bit/mintavétel} \times 25 \text{ kép/s} = 82,800 \times 10^6 \text{ bit/s}$$

A forráskódolás különböző fajtáit lehet alkalmazni az egyes komponensekre. (A továbbiakra utaló *elmélyültebb tárgyalás található az 5. fejezetben a 73. oldalon.*) Legyen itt csupán egy eredmény megadva:

Egy kereten belüli ADPCM eljárással a kódolás 3 bitre csökkenthető (szemben a 8 bittel) mind a luminancia, mind a krominancia jelnél. Ekkor a 2. alszabvány szerint az alábbi adatsebesség adódik:

$$(540 + 90 + 90) \text{ mintavétel/sor} \times 575 \text{ látható sor/kép} = 414\,000 \text{ mintavétel/kép}$$

$$414\,000 \text{ mintavétel/kép} \times 3 \text{ bit/mintavétel} \times 25 \text{ kép/s} = 31,050 \times 10^6 \text{ bit/s}$$

#### 4.8 Videotechnika a számítógépben

A fejezet befejezéseként egész röviden foglalkozunk a számítógépes videotechnikával. Szinte sohasem alkalmazzák itt a sorvisszafutásos eljárást. A zavaró villódzás miatt rendszerint magasabb képváltási frekvenciát alkalmaznak, pl. 70 Hz-et. A képernyőn a különböző színek előállításához egy ún. *Color Look Up Table*-t használnak (CLUT). Itt egy meghatározott számú (n) szín áll rendelkezésre egy adott időpontban az egész kép színezésére. Ezen színpaletta n színe választható az m színből álló szín térből. Általában  $n \ll m$ .

Példaképpen néhány ismert számítógép-videoszabványt mutatunk be. Ezen rendszerek mindegyike egy sor felbontást és színszám-előállítást támogat. A továbbiakban mindig csak egy üzemmódot magyarázunk meg röviden.

- A *Color Graphics Adapter (CGA)* 320 × 200 felbontású, négy különböző szín egyidejű megjelenítésével. Ebből a képenként szükséges tárhelykapacitás

$$320 \times 200 \text{ képpont} \times \frac{2 \text{ bit/képpont}}{8 \text{ bit/byte}} = 16\,000 \text{ byte}$$

- Az *Enhanced Graphic Adapter (EGA)* 640 × 350 képpont mellett 16 szín egyidejű előállítására alkalmas. Ebből a képenként szükséges tárhelykapacitás

$$640 \times 350 \text{ képpont} \times \frac{4 \text{ bit/képpont}}{8 \text{ bit/byte}} = 112\,000 \text{ byte}$$

- A *Video Graphics Array (VGA)* legtöbbször 640 × 480 képpont felbontású. Itt 256 szín állítható elő egyidejűleg. A monitort analóg RGB vezérli. Ebből a képenként szükséges tárhelykapacitás

$$640 \times 480 \text{ képpont} \times \frac{8 \text{ bit/képpont}}{8 \text{ bit/byte}} = 307\,200 \text{ byte}$$

- A *8514/A Display Adapter Mode* 1024 × 768 felbontás mellett 256 szín megjelenítésére képes. Ebből a képenként szükséges tárhelykapacitás

$$1024 \times 768 \text{ képpont} \times \frac{8 \text{ bit/képpont}}{8 \text{ bit/byte}} = 786\,432 \text{ byte}$$

- Az *Extended Graphics Array (XGA)* 640 × 480 felbontás mellett kb. 65 000 különböző színt tud egyidejűleg visszaadni. Az 1024 × 768 felbontásnál 256 szín jeleníthető meg egyidejűleg. Ily módon a szükséges tárhelykapacitás ugyanaz, mint a 8514/A adapternél.

Minden ebben a fejezetben bemutatott eljárás egy ma még viszonylag nagy adatsebességet igényel. A HDTV a jövő rendszerei számára ezzel egyenértékű követelményt támaszt. Ezáltal szükségessé válik az a követelmény, hogy mozgóképek átviteléhez és tárolásához tömörítési eljárásokat alkalmazzunk.