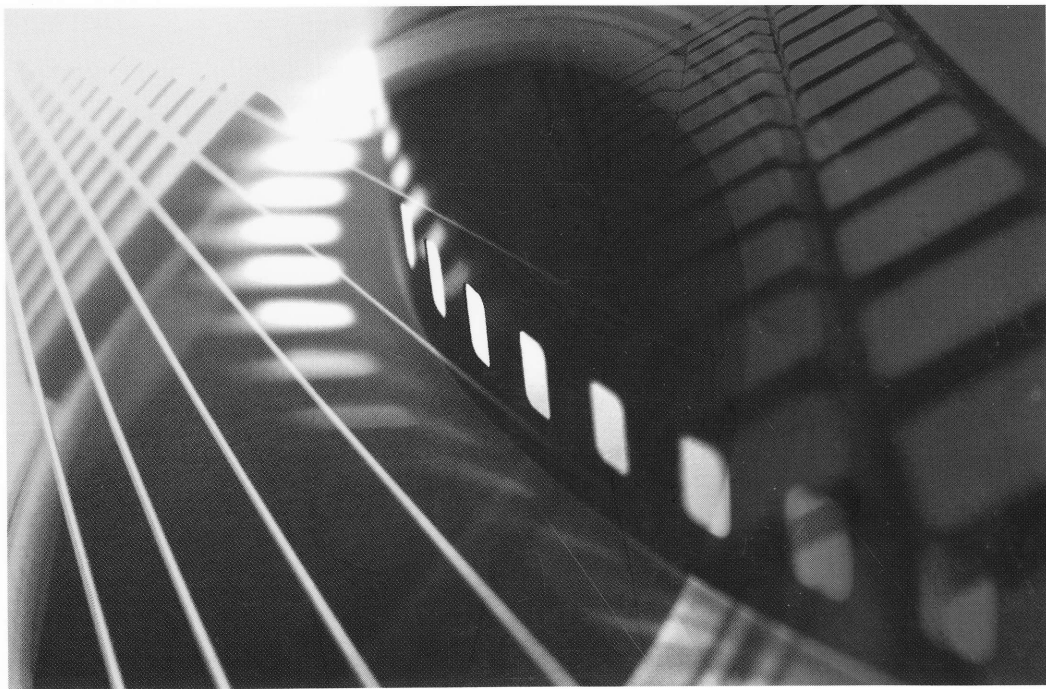
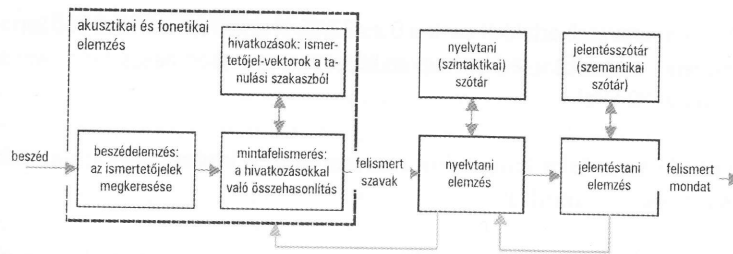


Andreas Holzinger

A multimédia alapjai



VOGEL



3.35 ábra
A beszédfelismerés munkaszakasa.

A hangokat legtöbbször nem lehet egyértelműen egy szóhoz kapcsolni, csak egy listát készíthetünk, hogy melyik szóhoz milyen valószínűséggel tartozik az adott hangsor. A nyelvtani elemzés a korábbi szakaszban elkövetett hibát ki tudja mutatni, és a valószínűségi listát át tudja alakítani.

nyelvtani elemzés

A jelentéstani elemzés is segít a hibák kiküszöbölésében, és az elsőbbségi szabályoknak megfelelően befolyásolja a nyelvtani elemzést.

jelentéstani elemzés

4. A hang megjelenése számítógépes környezetben

Miután a 2. modulban áttekintettük a digitális jelfeldolgozás alapjait, most az ott szerzett ismereteinket alkalmazzuk a hangok esetében, majd kiegészítjük azokat a hangjelek kódolásáról szóló részben. Ezután a legfontosabb hangfájl-formátumokról ejtünk szót.

4.1. Nyers hangadatok

Az analóg, állandó és folyamatos (például hangszerek által kibocsátott) hangjeleket a digitális számítógép nem képes közvetlenül megjeleníteni, ezért a jelet először „le kell tapogatni” (mintavétel), majd mennyiségi mutatókba kell alakítani és kódolni kell (lásd: 2. modul, Jelek és kódolás).

A hangjelek digitalizálásakor három összetevőt nem szabad figyelmen kívül hagynunk:

- a mintavételi frekvenciát (gyakoriságot),
- az amplitúdót,
- a hangcsatornák számát (monó vagy sztereó).

A multimédia területén a következő három szabványos mintavételi frekvencia fordulhat elő: 11,025 kHz, 22,05 kHz, és 44,1 kHz. A zenei CD-ket 44,1 kHz-es mintavételezik. Bár ez az érték első látásra magasan az ember hallástartományja fölé esik, Shannon és Nyquist mintavételi elmélete szerint (lásd a 2. modult) a közvetíthető sáv szélesség csak feleakkora, mint a mintavételi frekvencia.

A stúdióminőség eléréséhez 48 kHz-et használnak, a professzionális DVD-k esetében 96 kHz-et.

44,1 kHz mintavételi értékkel csak a 0 és 22,05 kHz frekvencia közé eső tartományt lehet elérni. Ez a terület nagyon közel van az emberi hallás tartományához (20 Hz–20 kHz).

Fontos kérdést vet fel a hangfájlok mérete, ennek következtében az adattömörítés is (lásd a 4. modult).

Az adatállomány méretét a következő egyenlet segítségével számíthatjuk ki:

$$F = c \cdot s \cdot t \cdot b$$

- F fájl méret,
 c csatornák száma (monó $c = 1$, sztereó $c = 2$),
 s mintavételi sebesség Hz-ben,
 t idő másodpercben,
 b bájtok száma (8 bit = 1 bájt; 16 bit = 2 bájt).

A következő ábra a legfontosabb lehetőségeket mutatja:

3.36 ábra
 A tömörítés fontosságát a hangfájlok mérete mutatja.

Minőség	b	c	s (kHz)	F ($t=60s$)
Stúdió	2	4	48,00	23 MB
CD (ISO 10149)	2	2	44,10	10,5 MB
Magas minőségű beszéd	2	1	44,10	5,25 MB
PC-minőség	1	2	44,10	5,25 MB
MM szabvány	2	2	22,05	5,25 MB
Belépő szintű PC	1	2	22,05	2,5 MB
Alacsony minőségű zene	1	2	11,00	1,3 MB
Alacsony minőségű beszéd	1	1	11,00	650 KB

A 3.36 ábra hatásosan ábrázolja a hangfájlok méreteit, ezért a folyamatosan növekvő merevlemez-kapacitás ellenére sem lehet elhanyagolni az adattömörítésre irányuló eljárásokat, mint amilyenek a következők:

- „Downsampling”, vagyis a mintavételi frekvencia csökkentése 22 kHz-re vagy 11 kHz-re,
- a felbontás csökkentése 1 bájtra,
- a sztereó hangról való lemondás,
- adattömörítés (például ADPCM, MPEG).

4.2. Digitális hangfelvétel, PCM

Ahogy a 2. modulban tárgyaltuk, a digitális hangfelvételek esetében az impulzuskód-moduláció elnevezés onnan származik, hogy az analóg hangjeleket egységnyi időcsomagokra (impulzusokra) osztjuk, majd digitalizáljuk. E módszer kialakításakor a mágnesszalagon kívül nem létezett elérhető áron más felvételre alkalmas eszköz, ezért a PCM hangjeleket képmagnóval vették fel, másodpercenként 50, illetve 60 kép ritmusban, PAL vagy NTSC rendszerben, hagyományos **videoszalagra**. Ezt a módszert vették át PCM-1630 szabványként a zenei CD-k.

A DAT felvétel szintén a PCM kódoláson alapul. Számítógépes környezetben a WAV fájlformátum szintén PCM kódolású.

Figyelem: mivel a PCM felvételek mintavételi „mélységüket” és mintavételi sebességüket illetően különbözhetnek, közvetlen digitális adatcsere esetén ügyelni kell rá, hogy rendelkezünk megfelelő átalakítóval!

A PCM formátumot a telekommunikáció területén az ISDN technológiánál alkalmazzák, ahol a mintavételi frekvencia 8 kHz, a felbontás pedig 8 bit, ami által másodpercenként 64 kbit vagy 8 Kb-át adatáramlás érhető el. A beszéd így nagyon jól érthető, de a mai igények már meghaladják ezeket az értékeket.

Mint már említettük, a PCM alkalmazására további példa a WAV formátum, amivel gyakran találkozhatunk a számítógépek világában. Bár ebben a formátumban elvileg bármekkora mintavételi érték és bitmélység tárolható, a mintavételi frekvencia általában 44,1 kHz, a felbontás pedig 16 bit. Ha ezeket a lehetőségeket teljes mértékig kihasználjuk, nagyon jó minőségű felvételek készítésére van módunk. Az ehhez szükséges adatsebesség és tárolási kapacitás óriási: egy sztereó jelhez másodpercenként 172 Kb-át szükséges.

4.3. Digitális hangfelvétel, ADPCM

Ha a bitsebességet DPCM-mel csökkentjük (lásd a 2. modult), az a beszédhangok esetében elkésztető eredményt ad. Ennek oka a beszédhangok tulajdonságaiban keresendő. Megvizsgáltunk néhány hangjelet és megállapítottuk, hogy hullámuk hangonként teljesen különböző. Ebből arra következtethetünk, hogy minél jobban alkalmazkodunk az új hullámhoz annak megjelenéséhez, annál jobb becslést érhetünk el.

ITU-T G.721 ajánlása szerinti, az általában másodpercenként 64 kbit-es hangjelek (kódolás μ -law, a-law vagy lineáris formában) másodpercenként 32 bit-esre való tömörítésére használt eljárás neve ADPCM (Adaptive Differential Pulse Code Modulation).

Az ADPCM eljárás a szomszédos jelek hasonlóságán (nem azonosságán!) alapul. Az eljárás során a korábbi jellefutás alapján, előrejelzési együtthatók segítségével próbáljuk kikövetkeztetni a következő jelet.

A gyakorlatban ez az alkalmazkodás úgy néz ki, hogy egyforma hosszú időközönként (ami általában 20 ms) újra kiszámoljuk az előrejelzési együtthatót.

Az eljárás erről az állandó alkalmazkodásról kapta a nevét (alkalmazkodó különbség alapú PCM = ADPCM).

Itt az értékek kedvezőbbek: 8 bitből 4 bit, illetve 16 bitből akár 12 bit is megtekinthető. Egy 8 bit felbontású, 8 kHz-es mintavételi frekvenciájú rendszer bitsebessége 32 kbit/s.

Számos hangkártya támogatja a WAV fájlok tárolását ADPCM formátumban. Ha ismét abból indulunk ki, hogy van egy 16 bit felbontású, 44,1 kHz-es rendszerünk, a sztereó jel adatsebessége 43 Kbajt/s lesz. Az eredeti PCM-mel összehasonlítva az adatsebesség negyedére csökken, bár az ADPCM esetében a minőség nem olyan jó, mint a PCM-nél.

**Minőségcsökkenés
az ADPCM-mel**

ADPCM kódoló

A kódoló alapvetően a visszaalakítóhoz (dekóderhez) hasonlóan működik, amikor a jel alakulását „megjósolja”. Először a logaritmikus bemeneti értékeket változtatja lineárisra (uniform PCM). Az így újra számszerűsített értékeket egy – az adott kódoló-megvalósításnak – megfelelő eltolással súlyozzuk. Ezt az eltolást használjuk arra, hogy az egységes PCM értéket el tudjuk helyezni.

Így az ADPCM kódoló alkalmazkodik a jel jellegzetességeihez azáltal, hogy változtatja a „számszerűsítési” és előrejelzési lépcsőfokok számát. A kimeneti jeleket csupán mint különbségeket közvetíti. Ezért ahhoz, hogy az esetleg eltérő megvalósítású számszerűsítő és előrejelző egység összeférhetőségét biztosítsuk, ezeket a mennyiségeket az ADPCM kódolású kimeneti jelekkel együtt kódoljuk.

Alkalmazkodó viselkedés

A következőkben az IMA algoritmust mutatjuk be. A többi algoritmus a G.721, illetve a G.723 ajánlás, valamint a CD-I szabvány (Compact Disc Interactive Audio Compression Algorithm) írja le.

IMA algoritmus

**IMA: Interactive Multimedia
Association**

Ezen algoritmus segítségével 4:1 arányú tömörítésre nyílik lehetőség. Gyorsasága az előrejelző egyszerűségéből következik.

Az előrejelző érték egész egyszerűen az előző hangminta, ezért ezt a lépést „nem alkalmazkodónak” is nevezik.

Az alkalmazkodás a számszerűsítő részben történik: az éppen számszerűsített jel számszerűsítési lépéseinek számát jelölő három bitet besoroljuk indexként egy táblázatba. Az ott feltüntetett értéket levonjuk, ez újfent indexként jelenik meg egy következő táblában, ezt pedig tároljuk a következő lépéshez.

A második táblában szereplő érték a következő minta számszerűsítési lépéseinek száma. Az eljárás teljes mértékben megfordítható, így nincs szükség további információkra a létrejött bitfolyamon belül.

Az ADPCM eljárás emellett lehetőséget biztosít a hibajavításra is: amennyiben a tömörített adatfolyamon belül bithibát észlel, az egyes mintát csak az adott számszerűsítési lépésen belül „hamisítja”.

ADPCM visszaalakító

A visszaalakító (dekóder) fordított alkalmazkodó számszerűsítőt (Inverse Adaptive Quantizer) alkalmaz, így teremt meg a különbségjelet. A mintát azután e különbségjel és az alkalmazkodó előrejelző (Adaptive Predictor) segítségével állítja elő és így változtatja az egyenjelet hagyományos PCM jellé.

Az ADPCM-et a telekommunikációban (ISDN) is előszeretettel alkalmazzák, mivel kiválóan alkalmas beszéd közvetítésére. Ebben az esetben a tömörítési módszereket a telekommunikációban alkalmazott frekvenciatartományokhoz igazítják.

ADPCM-et alkalmaznak a Sun számítógépek hangeszközeiben is. Az ADPCM-et és az ADPCM-et alkalmazó programokat más rendszerekre is átvitték, így a mai számítógépeken az ilyen anyagok mindenféle hardvereszköz bevonása nélkül kódolhatók és visszaalakíthatók.

Az ADPCM alkalmas beszéd közvetítésére

4.4. A GSM szabvány

A GSM (Global System for Mobile Communication, globális mobilkommunikációs rendszer) Európa-szerte alkalmazott (digitális) szabvány rádiótelefonok számára, ami lehetővé teszi a nemzetközi barangolást („roaming”) a hálózatban.

A GSM szabvány az alappfrekvenciát 900 MHz-ben (például D-hálózat) határozza meg. A GSM 1800 (például E-hálózat) a GSM szabványon alapul, de az 1,8 GHz frekvenciatartományba tartozik.

A GSM hálózatba különböző nyelvi és adatszolgáltatások is tartoznak, amelyek azonban különböző mértékben szabványosítottak.

ETSI = Európai Telekommunikációs Szabványok Intézete, Nizza

Az összeurópai mobiltelefon-előírásokat az ETSI GSM 06.10 foglalja össze.

A GSM időosztásos (time multiplexing) digitális átvitel, átviteli egységenként nyolc csatornával, 200 kbit/s sávzélességgel és 13 kbit/s beszédkódolással.

A GSM rendszer az ETSI ajánlása a beszéd tömörítésére. Ezt használják a D-telefonhálózatban. A GSM rendszer csak jelentős szoftverberuházással alkalmazható.

A GSM kódolás alapja a lineáris előrejelző kódolás (Linear Prediction Coding) matematikailag összetett, és csak nagyon leegyszerűsítve ábrázolható (résztelesebb információ az irodalomjegyzékben, példák a [W18]-ban).

LPC = lineáris előrejelző kódolás

Az LPC a zöngés hangokat alacsonyabb rezgésszámmal írja le. Az egyes mintavételi értékek lineáris függésben állnak és korábbi értékek súlyozott összegei alapján jelzik őket előre.

Számos hangkártya támogatja a WAV fájlok tárolását GSM formátumban. Ha ismét egy 16 bit felbontású, 44,1 kHz-es rendszert veszünk, abban a sztereó jel adatsebessége 18 Kbajt/s.

A GSM minősége mindenestre jelentősen gyengébb, mint a PCM-é vagy az ADPCM-é.

4.5. μ -Law kódolás

G.711 (CCITT) μ -law és a-law

A μ -law-átalakítás olyan hangtömörítési módszer, amely a CCITT (Comité Consultatif International de Télégraphique et Téléphonique, Nemzetközi Telekommunikációs Bizottság) G.711-es ajánlásában már 1969-ben szerepelt, és nemzetközi szabvánnyá is vált.

A kiindulási alap a logaritmikus átváltás és egy 13 bites átalakító felbontás, amely az átalakítás után 8 bitre csökken. Maga az átalakítás az ajánlásban előírt „13-ról 8 bitre” és „8-ról 13 bitre” táblázatok segítségével történik.

Az átalakítás ilyen módjával a kódolást az emberi fül logaritmikus hallóképességéhez igazítják.

Az alacsonyabb kódolási felbontás minőségében megfelel a nagyobb számszerűsítési felbontásnak.

Az átalakítás matematikailag a következőképpen írható le:

$$y = 255 - \ln(1+x) \text{ minden } x \text{ esetén } = 0$$

$$y = 127 - \ln(1+x) \text{ minden } x \text{ esetén } < 0$$

ahol x mindig egy 0 és 1 közötti szabványos jelérték, vagyis az értékek 0 és 255 között mozognak.

A μ -law kódolást ISDN-en történő átvitelre Észak-Amerikában és Japánban használják. Egyes átviteli csatornák a nullák folyamatos átvitelét nem engedik, mert akkor elvész a bitáramlás.

μ -law Észak-Amerikában
és Japánban, a-law Európában

A G.711-nek megfelelő a-law kódolást Európában az ISDN telefonhálózatban alkalmazzák, ahol 13-ról 8 bitre átalakító táblázatokat használnak és -128 és 127 közötti értékeket.

ISDN kapcsolatnál csak egyszeri kódolás és visszaalakítás történik, vagyis nincs értelme a 8 kHz esetében 8 bites korlátozásnak. A hardver- és szoftvergyártók azonban mégis ragaszkodnak ehhez, mivel

- a 8 bites PCM kodekek beépített elemekként elérhetők,
- az ADPCM az ISDN része lesz,
- a meglévő hardver- és szoftvereszközöket (digitális telefonközpontok, meglévő hardver, telepített szoftver) nem lehet más adatszélesítéshez és mintanagysághoz igazítani.

Az azonban meg kell jegyezni, hogy a G.711-es formátum nem megfelelő magas minőségű hangtárolásra, hangfeldolgozásra és hangátvitelre.

A μ -law nem megfelelő a magas minőségű hangadatok számára

A meglévő formátum megváltoztatása valószínűtlen, a 8:13-as tömörítési arány pedig egész egyszerűen nem elegendő annyi adat számára, amennyi például egy zenei CD-n található.

Mivel azonban a beszéd vagy adatáram továbbítására rendelkezésre álló csatornák áteresztőképessége korlátozott, ez a formátum mégis gyakran használatos (például ISDN-en átvitt hangjelek kódolására vagy hangátvitelre H.261 vagy SMP használata esetén).

5. Hangfájlok

5.1. Áttekintés

Az idők folyamán az éppen rendelkezésre álló számítógéprendszereknek megfelelően számos hangfájl-formátumot dolgoztak ki, ám ezek közül *csak néhány* sikerült elterjednie, és azok közül is alig néhány felel meg az Internet által támasztott követelményeknek.

A hangfájlok körében az átvitel módja szerint megkülönböztetünk *statikus állományokat*, amelyeket a számítógépen tárolnak, valamint *áramló*

adatállományokat, amelyek átvitele egymás után történik, de lejátszásuk valósidejű („élő”), ami különösen az Internet szempontjából fontos (lásd a 6. modult).

Mielőtt az egyes hangfájl-formátumokat részletesen tárgyalnánk, egy táblázatos összefoglalást mutatunk be a legfontosabb formátumokról, hogy kellő áttekintést nyerjünk:

3.37 ábra
A legfontosabb hangfájl-formátumok áttekintése.

Fájltípus	Kiterjesztés	Eredet	Előny	Hátrány
Apple Interchange fájlformátum	.AIFF	Az Apple rugalmas mintavételi formátuma	Tetszés szerinti mintavételi gyakoriság	Nincs információ a sebességről és időről
SUN-Audio	.AU	A Sun és a NeXT mintavételi formátuma	Java nyelvű programokban gyakori	Rendkívül nagy
Music Instrument Digital Interface	.MID, .MIDI	Hangátírat számítógépes zene számára	Kicsi és rendszerfüggetlen	Szintetikus hangképzés
különböző Amiga-formátumok	.MOD, .NST, .WOW, .FTK, .TT	Az Amiga mintavételi formátumai	Nagyon sok hangzáslétező	Nem túl rugalmas
MPEG Audio Layer I és II	.MPG, .MP2	Moving Picture Experts Group (ISO + IEC)	Jó minőség	Hosszú számítási idő
MPEG Audio Layer III	.MP3	Fraunhofer Institute for Integrated Circuits (IIS), Erlangen (Németország)	Jó minőség az Interneten, pszichoakusztikai tömörítés	Lejátszás előtt visszaalakítás szükséges
Real Audio	.RA, .RAM, .RM	RealNetworks, Seattle, USA	Valódi áramlás	Csak megfelelő lejátszóval játszható le
Wavelet	.WAV	A Microsoft mintavételi formátuma	Széles körben elterjedt „szabvány”	Túl nagy
WMA	.WMA	Windows Media Audio	Alacsonyabb bitsebesség, mint az MP3-nál	Csak megfelelő lejátszóval játszható le

A személyi számítógépek körében leginkább elterjedt multimédiás hangformátumok a Wave és a MIDI (illetve ma már az MPEG Layer III, vagyis az MP3). Közös bennük, hogy a hangzási információt kódoltan tárolják, és lejátszásuk minősége is elég jó, a kódolás módjában azonban különböznek.

Elvileg a Wave fájlok a hangzás alaki lefutását tárolják, míg a MIDI fájlok a zenei tartalmat.

5.2. Wavelet (WAV) fájlok

A WAV rövidítés és egyben kiterjesztés (*fájlnév.wav*) a *Wavelet* („kis hullám”) szó rövidítéséből származik.

Eredetileg a 80-as években fejlesztette ki az IBM és a Microsoft, bináris hangfájlként. Apple számítógépeken is használható, de például a playwav program segítségével Linuxon is lejátszható.

A WAV fájl a hangfelvétel közben a mikrofonnál vagy a hanganyagot tartalmazó CD átjátszásakor érzékelhető hangnyomás „digitális másolata”. A WAV fájlok tehát tetszőleges hangesemény tárolására alkalmasak, vagyis a MIDI fájlokkal ellentétben nem csupán a zenére korlátozódnak.

A WAV formátum (valójában RIFF WAVE) alkalmazkodik a Microsoft által a Windows operációs rendszer számára kifejlesztett RIFF (Resource Interchange Format) formátumhoz.

A RIFF az IFF formátum egy változata, amit azért hoztak létre, mivel a szabványt nem tartották be. Az IFF-fel ellentétben a RIFF-ben az adatokat az Intel processzor „Little Endian” (a legkisebb helyiértékű bájtot a legkisebb memóriacímen tároló számábrázolási forma) bájtsorrendjén alapuló formátumban tárolják. Mindazonáltal a formátum felosztása az IFF formátumával megegyező.

Minden RIFF fájl legfelső darabja (a teljes adatállományt átfogó darab) mindig a RIFF megnevezést viseli, és adattartományának első 4 bájta a fájl típust adja meg. A WAV fájlok belső azonosítója a „Wave”.

Általában azt várjuk egy WAV fájlról, hogy az említett adatokat mindenféle kiegészítő adatok nélkül tartalmazza, PCM formátumban. A szabványos WAV formátum általában két darabból áll (3.38 ábra): a **formátum darab** tartalmazza a hangadatok jellemzőit (formátum típusa, csatornák száma – 1 vagy 2 –, mintavételi gyakoriság, bájtssebesség, mintánkénti bájtt, illetve bit), az **adat darab** pedig magukat a hangadatok, PCM formátumban.

Há nem PCM-et használunk, akkor egy **FACT darab** tartalmazza az alkalmazott bitvétel formátum adatait (3.38 ábra).

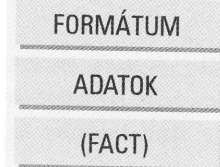
5.3. MIDI

Mielőtt a MIDI-re rátérnénk, három fogalmat kell megkülönböztetnünk.

1. MIDI

- elektromos hangszerek és zenei hatásokat létrehozó eszközök vezérlése céljából szabványosított adatformátum,

RIFF = Resource Interchange Format (erőforráscsere-formátum)



3.38 ábra
A WAV fájl darabjai.

- a fenti szabványnak megfelelő hardver,
- a fenti szabványnak megfelelő zenék gyűjtőneve.

5.3.1. Általánosságok

MIDI: Music Instrument Digital Interface

A zeneiparban már 1983 óta használatos a MIDI (Music Instrument Digital Interface) szabvány, amely lehetővé teszi a számítógépek és elektromos hangszerek közti kódolt zenei jelátvitelt.

E sorrendvezérlő eljárással vagy rövid analóg hangsorokat (például egy zongorabillentyű lenyomását, egy gitárhúr megpendítését) digitalizálunk vagy közvetlenül digitálisan hozzuk létre azokat (például egy MIDI hangszerrel).

A sorrendvezérlő (sequencer) segítségével a MIDI zeneszámok szerkesztővé válhatunk. Segítségével a MIDI zenedarab **hangjegyképlete** vagy **MIDI-eseménytörténete** ábrázolható és szerkeszthető. Megváltoztatható az egyes hangok hangszerhez rendelése vagy akár a hangzásuk is.

A MIDI a zenei adatok legtömörebb ábrázolásmódja, ami a legtermészetesebb lejátszást teszi lehetővé.

A MIDI a hangszerek távvezérlője

Ha igazán pontosan akarnánk fogalmazni, akkor ahelyett hogy azt mondanánk, hogy „a MIDI-t a zene és a számítógép ötvözésére fejlesztették ki”, úgy kéne fogalmazni, hogy „a MIDI a hangszerek távvezérlője”, mert ez ragadja meg a MIDI formátum lényegét, és ebből adódik e fájlok kis mérete is.

Mintavételi értékek helyett a MIDI *hangszerre vonatkozó* ábrázolást tesz lehetővé. A kódolás a következő három adatot tartalmazza:

- hangerő,
- alaphérvencia, és
- a hangszer megnevezése.

A MIDI segítségével 10 oktávnyi hang (128 hangjegy) kódolása lehetséges.

Egy MIDI hangszeren egy zongorabillentyű lenyomásakor a következő dolgok átvitelére kerül sor:

- a hang kezdete,
- a leütési erő és a billentyű elengedése,
- a hang vége.

Egy 10 perces MIDI zeneszám helyigénye körülbelül 200 KB, ami jóval a 44 kHz-es mintavétel mellett keletkező 105 MB (!) adatmennyiség alatt marad.

A MIDI szabvány 128 hangszert határoz meg, és ebben a szabványban (General MIDI Standard) minden hangszerhez tartozik egy szám. Így például a zongorára a 0-val, a trombitára az 56-os számmal hivatkoznak.

zongora = 0
trombita = 56

A MIDI szabványban egyidejűleg 16 úgynevezett csatorna működhet. A hangszerek fajtájuktól függően egy csatornát (ilyen például a trombita) vagy több csatornát (ilyen a gitár) vesznek igénybe. Csatornáként egy-egy szintetizátor szükséges.

5.3.2. Technikai MIDI alapismeretek

A MIDI csatlakozási felülete az RS 232 C-n alapuló ötpólusú DIN-foglalat. Az adatátvitel alapvetően MIDI bájtorként 8 bit szélességben történik, két további vezérlőbittel. A MIDI felületnek a leírás szerint rendelkeznie kell egy UART-tal (Universal Asynchronous Receiver Transmitter), amelynek bemenetét az előírásoknak megfelelően egy optoelektronikus csatoló segítségével galvanikusan választják el a bemeneti foglalatától.

5.3.3. A MIDI előnyei

Különösen a zene területén tapasztalhatjuk a MIDI óriási előnyeit:

- a MIDI fájlok hordozhatók, vagyis különböző operációs rendszereken ugyanazok a MIDI fájlok használhatók,
- a MIDI fájlok viszonylag kicsik,
- a MIDI pontosan megfelel a zeneszerzés klasszikus módszerének (hangjegyek feljegyzése, majd ezek lejátszása különböző hangszereken, azok jellegzetes hangzásvilágával).

Rendszerek felett álló, viszonylag kicsi, zeneszerzésre alakították ki

5.3.4. A MIDI hátrányai

Természetesen hátrányai is vannak a MIDI-nek:

- Ha nincs valóban jó hangszerünk, a hangzás a szintetikus hangképzés miatt nem túl szép (FM szintézis).
- Beszéd vagy zörejek felvételére, például a WAV fájlokkal ellentétben, nem alkalmas.
- A MIDI fájlok különböző rendszereken, az alkalmazott eszközök függvényében másképp hangzanak (míg a WAV fájlok mindig ugyanazt eredményezik)

Szintetikus hangzás, beszédre nem alkalmas, minden rendszeren másképp hangzik

5.4. AU fájlok

Az AU kiterjesztésű hangfájlokat a Sun már 1986-ban kifejlesztette a Sun-3 számítógépek operációs rendszeréhez, ma pedig elsősorban a Java nyelvű programokkal használják.

Az önálló alkalmazásokkal ellentétben a Java kisalkalmazások (appletek) lehetőséget nyújtanak hangkibocsátásra, de ez a lehetőség csak az AU formátum esetében áll fenn. A Sun világából ismert AU formátum 8 bites monó, mintavételi sebessége 8 kHz, μ -law tömörítéssel.

A WAV fájlok Windowson szokásos lejátszását a Java jelenleg sajnos nem támogatja. Mindenesetre vannak olyan ingyenesen hozzáférhető hangszerkesztő eszközök, amelyek lehetővé teszik a különböző formátumok közötti átalakítást (például Chris Craig Goldwave-je).

Az AU fájl lejátszása kétféleképp történhet. Az Applet osztályban a `play` tagfüggvénnyel (metódussal) be lehet tölteni a hangfájlt, és le lehet játszani.

- `public void play (URL url);`
- `public void play (URL url, fajlnev);`

Itt vagy egy hangfájl URL-jét adjuk meg, vagy az URL könyvtárat (elérési utat) és a fájlnevet (lásd a 6. modult).

Az URL könyvtár átadásához általában a `getCodeBase` vagy a `getDocumentBase` tagfüggvényt használjuk. Ezek megadják a könyvtár URL-jét, ahonnan a kisalkalmazás elindult, illetve ahol az adott HTML oldal található.

- `public URL getCodeBase();`
- `public URL getDocumentBase();`

Ezen eljárás hátránya, hogy a hangfájlt minden híváskor újra be kell tölteni. A másik módszerrel, a `getAudioClip` hívással az `AudioClip` osztályból hívunk egy objektumot, amit tetszőlegesen sokszor lejátszhatunk:

```
public getAudioClip(URL url, fajlnev);
```

Az `AudioClip`-pel rendelkezésünkre állnak a `play`, `loop` és `stop` függvények:

```
public void play();
public void loop();
public void stop();
```

A `play` függvénnyel elindítjuk az előzőleg betöltött hangfájlt, és egyszer lejátszunk; a `loop`-pal is elindíthatjuk, ekkor viszont végteleníve hallgathatjuk a zenét, amit a `stop` függvénnyel állíthatunk meg. Több felvétel egyide-

új lejátszására is van lehetőségünk. Így például egy háttérzenét végtelenítve játszhatunk, anélkül, hogy a kimenetet az esetleg újonnan megszólaló hangok hátrányosan érintenék.

5.5. MPEG

Az MPEG a Moving Picture Experts Group rövidítése (lásd a 2. modult). Az MPEG hang- és képadat-kódolási szabványokat határoz meg (lásd az 5. modul 4.2 fejezetét).

Az MPEG legnagyobb előnye más hangkódolási formátumokkal szemben, hogy az ugyanolyan minőségű MPEG fájlok jóval kisebbek, ami az MPEG kifinomult tömörítési módjának köszönhető.

5.5.1. MPEG 1 for Audio

A hangfelvételt tartalmazó CD-re a hangjelet csatornánként 44,1 kHz mintavételi sebességgel veszik fel, és 16 bites csomagokban tárolják. A mintavételi elmélet szerint így a mintavételi gyakoriság nagyobb, mint a magas minőségű hang sávzélességének kétszerese. Ehhez azonban egy másodpercnyi sztereó hanghoz több mint 1,4 Mbit szükséges.

Az MPEG hangtömörítés a hangfájlok tizenkettő részére való tömörítését is lehetővé teszi érezhető minőségcsökkenés nélkül. Még ha huszonnegyedére csökkentik is őket, a hang sokkal élethűbb, mintha csak a sávzélességet csökkentik (például 44,1 kHz-ről 22,05 kHz-re). Erre az úgynevezett **Perceptual Audio Coding** (érzékelthang-kódolás) segítségével van lehetőségünk, ami azt jelenti, hogy a tömörítés során az emberi fül számára nem hallható elemeket szűrik ki.

5.5.2. Az MPEG Audio kódolás elve

E szabvány célja, hogy CD-minőségű felvételt nyerjünk 1,5 Mbit/s-ról 384 kbit/s-ra csökkentett adatsebesség mellett, és egyben még a 192 kbit/s-tól akár 112 kbit/s-ig tartó tartományba tartozó adatsebességnek is elfogadható minőséget kell eredményeznie.

Az ember érzékenyebben reagál a hang jellegű zavaró tényezőkre, mint a látást érintő zavarokra – vagyis a rövid zúgás vagy pattogás zavaróbb, mint például a villogás.

A kódolás négy módja ismert:

- **Single Channel Coding** (egycsatornás kódolás) monó jelek számára;
- **Dual Channel Coding** (kétszatornás kódolás) például kétnyelvű monó jelek kódolására;

A különböző adatméretek összehasonlítását lásd a 6. modulban

- **Stereo Coding** (sztereó kódolás) sztereó jelek kódolására, ahol is a két csatornát külön kódolják;
- **Joint Stereo Coding** (egyesített sztereó kódolás), ahol kihasználják a két csatorna adatai közti átfedést és felesleget, ami által jobb tömörítési arány érhető el.

A digitális bemeneti jelet az emberi hallás (pszichoakusztika) alapelveinek megfelelő módon 32 egyforma összetevőre osztják (frekvenciacsoportok, frekvenciasáv-szakaszok). Ez az eljárás az idő-frekvenciatartomány átalakítás.

A felesleges jelrészek pszichoakusztikai eltávolítása a hangfájlokból

Az egyes összetevőket az emberi hallásérzékelési tulajdonságoknak megfelelően kódolják. A végső kódolást a három meghatározott réteg (Layer) egyike hajtja végre.

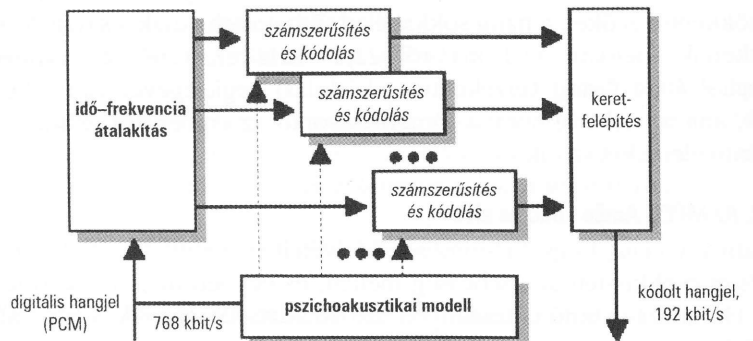
Maszkolás: az egyik jelet átfedi egy másik

Mind a számszerűsítés, mind a kódolás egy maszkolási küszöb bevonásával történik.

A Fourier-transzformáció használatával a jelek frekvenciájuk szerint ábrázolhatók

Ezt a maszkolási küszöböt a pszichoakusztikai modellben minden összetevő számára egyénileg, diszkrét Fourier-transzformációval (DFT) számítják ki. Ennek során megadják a számszerűsítés hibahatárát is, ami azt mutatja meg, hogy mekkora hibaszázalékkal kódolhatunk úgy, hogy az még az emberi fülnek ne legyen észrevehető.

3.39 ábra
Az MPEG hangkódoló működése.



Az MPEG hangformátumnak három növekvő hatékonyságú és tömörítési arányú rétege van, de természetesen ezzel egyidejűleg összetettségük és költségük is növekszik.

- Layer I: („egyszerű” megvalósítás) keretenként 384 mintavételi pont, számszerűsítés előtt egységekre osztás; egységes számszerűsítés,
- Layer II: keret 1152 mintavételi ponttal (24 ms 48 kHz-nél); az egységes számszerűsítés a MUSICAM szabvány szerint,
- Layer III: keret 1152 mintavételi ponttal; módosított DCT; a számszerűsítés nem egységes.

MDCT: módosított diszkrét koszinusz átalakítás

5.6. MP3

Az MPEG Audio Layer III (röviden MP3) olyan hangformátum, amely a hangfájlok tömörítését csekély minőségcsökkenéssel teszi lehetővé. A frekvenciatartomány és a jel-zaj arány lényegében változatlan marad.

Az esetleges veszteségek még jó hi-fi készülékeken sem hallhatók, amennyiben a fájlok megfelelően készültek.

A tömörítési eljárás a „nem hallható” hanginformációk eltávolításán alapul. Sok hang kerül másikkal átfedésbe, másokat fülünk nem képes érzékelni, mivel hallásunk frekvenciatartományán kívül esnek. Az eredetivel összehasonlítva a tömörített adatokat, alig hallhatunk különbséget, ugyanakkor hatalmas adatmennyiséget takarítunk meg. A sztereó zenét a tizenketted részére is zsugoríthatjuk.

Az ilyen nagymértékű tömörítés rengeteg új lehetőséget kínál. Segítségével óriási mennyiségű zenét tárolhatunk egyszerűen, kis helyen – egyetlen CD-n akár 11 albumnyi anyagot!

Az Interneten keresztül sokkal gyorsabban, így jóval olcsóbban küldhetünk MP3 fájlokat, mint az igen elterjedt, de jóval terjedelmesebb WAV fájlokat (lásd a 6. modult).

Ezt a formátumot eredetileg a Fraunhofer Intézet fejlesztette ki az amerikai AC-3 formátum mintájára, a Video CD-k hanganyagának tömörítésére.

Az MP3 formátum további előnye, hogy úgynevezett fejléc nélküli fájlformátum. Az MP3 fájloknak ugyan van fejlécük, ez azonban nem feltétlenül szükséges a fájl lejátszásához. Ez azt eredményezi, hogy ha a fájlból bárhonnán kiszakítunk egy darabot, azt a darabot is lejátszhatjuk. Ezáltal az MP3 lehetőséget kínál az áramló sugárzásra (streaming, közvetítés az Interneten, lásd a 6. modulban) is.

Az MP3 fájlok a Windows Media Player (Médialejátszó), az Apple Quicktime és számos ingyenesen hozzáférhető eszköz és program segítségével lejátszhatók.

Bizonyítható MP3-lejátszóval útközben is hallgathatunk MP3 fájlokat. Mivel eddig az MP3-másolás ellen semmiféle védelem nem ismert, a zeneipar szereplőinek viszonyulása e formátumhoz nagyon ellentmondásos.

Az MP3 a felbontás javítása érdekében MDCT-t (módosított diszkrét koszinusz átalakítást) alkalmaz. Ennek előnye a korábbi eljárásokkal szemben:

- nem jelentkezik elmosódás,

Az MP3 az MPEG Layer III rövidítése, nem pedig az MPEG 3-é, ahogyan gyakran állítják (ilyen formátum valójában nincs is)

Egy egyperces, CD-minőségű, MP3 formátumú zene 1 MB területen tárolható!



3.40 ábra
MP3-lejátszó.

Sync		
ID	Layer	Prot. bit
Bitrate		
Frequency	Pad. bit	Priv. bit
Mode	Mode extension	
Copy	Home	Emphasis
Audio Data		

3.41 ábra
Az MP3 fájl felépítése.

- az egységekre osztás tényezőit sávokra osztják (egy sávba több MDCT-együtthető tartozik és körülbelül az emberi hallástartomány sáv szélességével rendelkeznek), és
- a számszerűsítésből adódó zajok nagymértékben csökkennek.

Az MP3 fájl felépítését a 3.41 ábra mutatja: a fejléc 4 bájt (32 bit) méretű és „sync”-kel kezdődik, amit egy azonosító bit (0 = MPEG-2 és 1 = MPEG-1) követ.

6. Az ember és a számítógép közti felületek

A számítógép hangjelzéseinek vételéhez, feldolgozásához és lejátszásához a számítógépen kívül hangkártyára, hangszóróra, mikrofonra, esetleg egy CD-meghajtóra és a használatához megfelelő szoftverre is szükségünk van. A digitális számítógép segítségével a hangjeleket négyféleképp adhatjuk vissza:

- digitális–analóg átalakító segítségével: a jeleket digitális hangjelekként merevlemezen tárolva,
- szintetizátor segítségével: az információt MIDI fájlban a hangkártyához küldjük, és ott kerül sor a feldolgozására,
- magának a MIDI felületnek a segítségével: az információt MIDI fájlban küldjük egy külső szintetizátorhoz vagy
- CD-meghajtó segítségével.

A hangkártya–hangszóró rendszer ember és számítógép közti felület, ezért a műszaki paraméterek mellett az emberi tényezőt is figyelembe kell venni.

Mivel a hangkártya és a hangszóró egységet képez, a teljes rendszer minőségét a két összetevő közül a rosszabbik fogja meghatározni. Még ha e rendszerek összetevőit folyamatosan javítják is, ennek a területnek is megvannak a maradandó alapismeretei, amikről a következőkben ejtünk szót.

6.1. Hangszóró

Az első személyi számítógépekbe piezoberregőket építettek be, majd később apró hangszórókat, amelyek hangjelek kiadására voltak képesek. Ebben az időben a számítógépek esetében a hanghatásoknak nem tulajdonítottak különösebb jelentőséget, egyedül a hiba felismerését segítették pittyegéssel (beep-code).

6.1.1. Felépítés és működés

A hangszórók olyan elektroakusztikai átalakítók, amelyek az alacsony frekvenciájú elektromos rezgéseket hallható mechanikus rezgésekké alakítják. A legtöbb hangszóró a hangot egy hártya (membrán) rezegtetésével idézi elő, majd a levegő segítségével juttatja el az emberi fülhöz.

6.1.2. A hangszórók jellemzői

A frekvenciagörbe a frekvenciagrafikonon (frequency response diagram) megmutatja, hogy ugyanolyan teljesítménnyel különböző frekvenciák esetén milyen hangerősséget érhetünk el. A frekvenciagörbe a következőképpen ábrázolható: egy hanggenerátor táplálja a megfelelő dobozt, és egy szabványos mikrofon veszi fel az adott frekvenciához tartozó hangerőt. Az ideális hangszórónak lineáris a frekvenciagörbéje, a lehető legmélyebb alsó és a lehető legmagasabb felső határfrekvenciával.

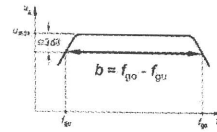
Meg kell jegyeznünk, hogy az emberi fül csak a 3 dB-nél magasabb rezgéseket érzékeli, így az ennél kisebb eltéréseket a gyakorlatban elhanyagolhatjuk.

Torzítási tényező alatt a felhullám és az alaprezgés közti különbséget értjük. Ha a hangszóróba hangjel érkezik, annak kizárólag ezt a hangjelet szabad visszaadnia (ebben az esetben a torzítási tényező 0 %). A gyakorlatban azonban nemkívánatos felhullámok sugárzására is sor kerül. A legerősebb felhullám általában kétszerese, háromszorosa az alaprezgésnek. A régi DIN 45 500-as hifi-norma (a német szabvány) szerint a torzítási tényező legfeljebb 1 % lehet.

A hangszóró hatásfoka megadja, hogy a hangszóró adott teljesítmény mellett milyen hangerőre képes. Mértékegysége a dB-W/m. A szabványos hifi hangszórók 86 ... 90 dB-W/m hatásfokkal rendelkeznek, ami viszonylag kicsinek számít. Az érzékenyebb hifi berendezések 90 ... 95 dB-W/m hatásfokúak.

A hangteljesítmény alapvetően semmit nem árul el arról, hogy a hangszóró hogyan képes visszaadni a hangjeleket. Szerepe akkor válik jelentőssé, ha kicsi az erősítő kimeneti teljesítménye, vagy amikor egy óriási termet kell behangszórunk.

A hangszóró váltóáramú ellenállását impedanciának nevezzük, mértékegysége az Ohm, amit azonban nem szabad összetéveszteni az egyenáramú ellenállással, ami multiméterrel mérhető és 20 %-kal kisebb, mint a váltóáramú ellenállás. Vagyis az a hangszóró, amelynek váltóáramú ellenállása 4 Ω , annak ellenállása egyenáram esetén 3,2 Ω .



3.42 ábra

Frekvenciagrafikon. A b sávszélesség mutatja meg – a felső (f_{gu}) és alsó (f_{gd}) határfrekvencia különbségeként – az átviteli terület nagyságát.

Ohm = az elektromos ellenállás mértékegysége, melyet Georg Simon Ohm-ról neveztek el;
1 Ohm ellenállás az az áramkör, amelyben 1 voltnyi feszültségkülönbség 1 amper áramerősséget eredményez

A gyakorlatban arra kell ügyelnünk, hogy a hangszóró váltóáramú ellenállása növekvő frekvencia hatására jelentősen megváltozik. A hangszóró váltóáramú ellenállásának sosem szabad kisebbnek lennie, mint az adott esetben használt erősítő váltóáramú ellenállása, vagyis ne használjunk 4 Ω ellenállású hangszórót 8 Ω-os erősítővel. Amennyiben a hangszóró váltóáramú ellenállása meghaladja az erősítő váltóáramú ellenállását, az csökkenésként csekély mértékben csökkenti az erősítő kimeneti teljesítményét.

A hangszóró hosszútávon megengedett terhelését szinusz-terhelhetőségnek nevezzük. Ennek ellenőrzése úgy történik, hogy a hangszórónak 48 órán keresztül egy meghatározott ritmusú, állandó jelet kell sugároznia, és a DIN 45573 szabvány szerint e próbát károsodás nélkül kell kiállnia.

A hangszóró zenei terhelhetősége alatt azt a rövidtávú (például 2 másodpercet meghaladó) terhelhetőséget értjük, amit hallható torzítás vagy egyéb károsodás nélkül kibír.

Minden zeneműben vannak olyan részek, amelyek rövidtávú magas terhelhetőséget jelentenek (például ütősök). Általában egy tized másodperc időtartamú behatást tekintünk a hangszóró impulzusteljesítményének.

A hangszórók elhelyezésekor figyelemmel kell lennünk arra, hogy a hangszóró hangtartomány lehetőleg fülmagasságban legyen, és hogy a hangszórók a hallgatók füle felé fordítsuk.

6.2. Mikrofon

A dinamikus mikrofonok a dinamóelv szerint működnek. A mikrofon belsejében lévő vékony lemez hanghullámok hatására rezgésbe jön. Ettől egy parafas tekercs is rezegni kezd. A tekercs egy mágnes pólusai között helyezkedik el. A tekercsben az elmozdulások hatására váltóáram keletkezik.

A váltóáram a hanghullámok leképezése. Megfelelő szerkezettel olyan dinamikus mikrofonokat lehet készíteni, amelyek a teljes hallható hangtartományban viszonylag egyenletes érzékenységűek. A kiváló minőségű mikrofonok zenei hangfelvételre is tökéletesen alkalmasak.

Az elektro-mikrofonokban levő vékony lemez az ellenpólussal egy mindig töltött kondenzátort képez. Hanghullám hatására a vékony lemez elhajlik, és a mikrofon elektromos tárolóképessége megnő. Mivel a mikrofon kimenetileg nem terhelhető túl, a mikrofonba beépítenek egy váltóáramú ellenállású működő erősítőt. Mivel ezek a mikrofonok igen kis méretűek, ezeket általában a mikroportokban, és ezeket építik be különféle készülékekbe.

A kondenzátor-mikrofonok az elektro-mikrofonokhoz hasonló elven működnek, de sokkal költségesebbek. Ezek a mikrofonok kiválóan alkalmasak stúdió-felvételek készítésére.

Mikrofonokat gyakran építenek be közvetlenül képernyőkbe is.

6.2.1. A mikrofonok jellemzői

A mikrofon a hanghullámokat elektromos rezgésekké alakító készülék. Fő jellemzője az **érzékenység**, ami megmutatja, mekkora a mikrofon azon hasznos váltakozó feszültsége, ami 1 Pascal váltakozó hangnyomás hatására keletkezik. Az átlagos mikrofon érzékenysége 1 mV/Pa ... 10 mV/Pa között mozog. A mikrofonok érzékenységét gyakran az 1 V/Pa-hoz viszonyítva adják meg, decibelben.

Ennek megfelelően a -60 dB érzékenység $1/1000 \cdot 1 \text{ Pa/V} = 1 \text{ mV/Pa}$.

Egy átlagos hangú ember beszéde 1 méter távolságból 0,1 Pa hangnyomást fejt ki. Egy szimfonikus zenekartól 5 méter távolságban kb. 1 Pa hangnyomás mérhető.

A mikrofon kimeneti feszültsége igen csekély. Ha fel akarjuk dolgozni, ezerszeresére kell erősíteni, hogy digitalizálható legyen.

A mikrofon további fontos jellemzője a **hatómező**, vagyis a mikrofon alakja. Az érzékenység a mikrofon fajtájának és felépítésének megfelelően a hangnyomástól is függ. A leggyakoribb hatómezők a gömb, buzogány vagy vese.

Mivel a buzogány, illetve vese alak esetében a fő irányvonal egyértelmű, ezeket irányított mikrofonnak hívjuk. A fő irányvonal mentén teljes érzékenység figyelhető meg. Az irányultság frekvenciafüggő is.

6.3. Hangkártya

A hangkártyák a személyi számítógépbe helyezhető különböző csatolófelületű kártyák, amelyek lehetővé teszik hangjelek bevitelét és kezelését. Különböző kivitelűek lehetnek; a következőkben áttekintjük az alapvető technikai jellemzőiket.

6.3.1. A hangkártyák jellemzői

A kimeneti feszültség a vonalkimenetnél (line out). Teljes kihasználtság mellett az analóg szintetikus, digitális jel (1 kHz-es tesztjel) visszaadásánál használható kimeneti feszültség 1000 mV körüli.



3.43 ábra
Hangkártya.

Linearitás. Ez alatt azt értjük, mennyire valóságghűen alakítja át az ADC (analóg–digitális átalakító) és a DAC (digitális–analóg átalakító) a különböző erejű (hangos–halk) jeleket. Legjobb esetben az eltérés 0 dB lenne, de az érték általában 0,1 dB körüli.

Jel–zaj arány. A hasznos és zavaró jelek (zúgás, zörejek stb.) közti különbség. Minél több a zavaró jel, annál kisebb ez az arányszám. Az aránynak legjobb esetben a végtelenhez kellene közelíteni, de az érték általában 80 dB(A) körüli.

Torzítási tényező. Akárcsak a hifi készülékek esetében, a nem lineáris torzítás (a jelgörbétől való eltérés) nagyságának jelzésére szolgál. Ideális esetben a torzítási tényező nulla körüli, a DIN 45 500 (1970-es hifi-szabvány) 1 %-os is megenged, a ma jellemző értékek pedig 0,02 % körül vannak.

Dinamika. A rendszer dinamikáját a nyugalmi állapothoz képest vesszük figyelembe. Mérése általában úgy történik, hogy az ellenőrző jeleket egyre halkabbra vesszük, egészen addig, amíg már csak a zavaró jelek (zúgás, zajok) maradnak. Minél tovább tart ez a folyamat, annál jobb, a leggyakoribb érték 80 dB körül van.

Frekvenciagörbe vagy frekvenciaválasz. A frekvenciatartományt adja meg és egyben a lineáris torzítás (a frekvenciagörbétől való eltérés) mérésére is szolgál. Külön görbe írja le a felvételt és a lejátszást (lásd a 3.42-es ábrát). Minél lineárisabb a frekvenciatartomány a kívánt sáv szélességen belül, annál jobb. Az eltéréseket dB-ben is meg lehet adni; legjobb esetben 0 körül van, de a leggyakrabban 3 dB körüli.

Bemeneti feszültség a mikrofonbemenetnél. Ez határozza meg azt a bemeneti feszültséget, amely 1000 Hz esetén a (szoftver-) kihasználást mérő eszköz teljes kihasználásához vezet.

6.3.2. A hangkártyák felépítése

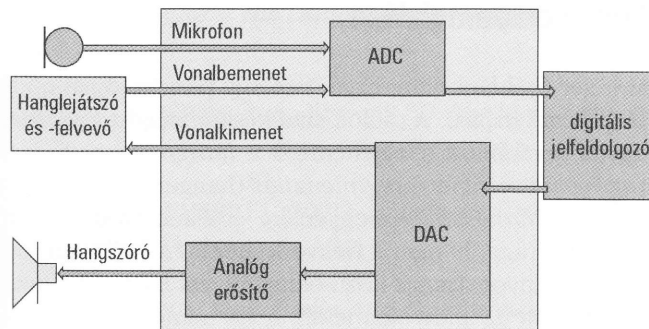
A SoundBlaster (SB) 1989 óta létezik; az 1987-es AdLib MSC-1-nek megfelelő

A Creative Labs (Creative Technology Ltd.) cég SoundBlaster hangkártya családja széleskörű elterjedtsége miatt akár szabványnak is tekinthető: a legtöbb hangkártyagyártó termékeit úgy készíti el, hogy azok együtt tudjanak működni a SoundBlaster termékekkel.

Számos hangkártya az OPL-3 lapka (Yamaha gyártmány) segítségével hozza létre a hangokat, FM szintézis (frekvencia-moduláció szintézis) segítségével. Az OPL-4 lapka a hullámtáblás eljárást is támogatja. FM szintézisnél a hang több rezgés átfedéséből tevődik össze, így a legkülönbözőbb (szintetikus) hangokat halljuk. A hullámtáblás eljárás során a kártya memóriájában tárolt (valódi) hangszerek eredeti hangjai biztosítják a zene természetes hangzását. A 3.44-es ábrán egy jellegzetes hangkártya látható.

A hangbemenet segítségével az analóg hangjeleket 16 bit adatmélységgel és legfeljebb 22,05 kHz frekvenciával (44,1 kHz mintavételi frekvencia) digitalizálhatjuk. A MIDI csatlakozón keresztül a kártyát szintetizátorral és más MIDI készülékekkel kapcsolhatjuk össze. Számos kártya rendelkezik „3D-Audio” képességekkel: ekkor egy újabb processzor segítségével térbeli hanghatást érhetünk el. Az alapvető szolgáltatások a következő blokkdiagram segítségével ábrázolhatók:

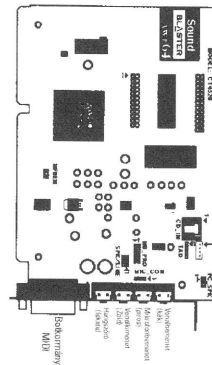
Az analóg jelforrások, a mikrofon, a bemenet és a CD (analóg kimenet) jeleit az ADC digitális jelekké alakítja, amelyeket aztán továbbít a digitális jelfeldolgozóhoz. A DSP azután a digitális bemeneti jeleket különféleképp változtatja meg: változtathatja a hangerőt, a hangszínt, keverheti és sűrítetheti a hangot. Az így feldolgozott hangjeleket aztán a PC-busz segítségével továbbítják, és WAV fájlként tárolják. A tárolt WAV fájlokat a digitális jelfeldolgozóhoz továbbítják, amely például beállítja a lejátszási hangerőt. A megváltoztatott lejátszási jeleket a DAC elemzi, hogy hangszóró segítségével lejátszhatóvá váljanak.



A MIDI jeleket másképp kell felvenni, lejátszásuk viszont WAV jelként történik. A billentyűzet például, mint MIDI-beviteli egység, MIDI-beviteli adatáramot hoz létre, ami már digitális jelekből áll és MIDI fájlként lehet menteni. A MIDI fájlok lejátszásához a MIDI nyelvét megfelelő digitális hangjelekké kell alakítani, amit a MIDI processzor végez. Ez egy táblázatból, mégpedig az úgynevezett hullámtáblából olvassa ki, hogyan kell az adott hangszert utánozni. A digitális hangjel hangerejét, hangzását stb. ezután a digitális jelfeldolgozóban lehet alakítani, majd a jelet a DAC-n keresztül lehet a hangszóróhoz továbbítani.

Az egyszerűbb hangkártyák nem hullámtáblával dolgoznak, hanem FM szintézissel.

A legtöbb hangkártya 16 bit felbontású és mintavételi mértéke 44,1 kHz-ig terjedhet. Ezzel elméletileg az SNR akár 90 dB, a frekvenciaválasz pedig akár 22 kHz is lehet.



3.44 ábra

A hangkártya vázlatos rajza.

3.45 ábra

A hangkártya felépítése.

6.4. Keverőpult

A keverőpulton több különböző forrású hangjel jelszintjét igazíthatjuk egymáshoz, és elegyíthetjük egy végső hangjellé.

Egyszerű esetben a keverőnek minden hangbemenethez egyetlen hangerő-, illetve hangszint-beállító gombja van, összegük (a jelek összeadódása) pedig egy hangkimeneten jelentkezik. A keverő **bemeneti csatornáinak** nevezzük a bemeneteket, és a hozzájuk tartozó beállító gombokat (amiket szabályozónak is hívnak).

Sztereó berendezéseknél legalább két csatornát találunk, saját kimenettel a jobb és a bal oldali jelnek, valamint minden bemeneti csatornához tartozik egy panoráma-szabályozó. A panoráma-szabályozóval a bemenő hangot hallás alapján tetszőlegesen lehet jobbra és balra „elhelyezni a térben”. Nagyobb keverőpultokon a bemeneti jelek hangszínét hangszabályozóval (equalizer-rel) külön-külön tetszőleges hanghatásúvá tehetjük.

7. A modul rövid összefoglalása

Minden fizikai inger (például hangrezgés) az agy pszichológiai észleléseinek egy fajtája (például hallás). A csillók által kódolt, majd a hallóidegeken keresztül az agyba továbbított hangjellemzők a *hangfrekvencia* (hangmagasság), a *hangnyomásszint*, a *hangintenzitás* (hangerő), a *hangirány* és *-távolság*, valamint a *hangszín*. Az egyénileg érzékelt hangerő nem csak a hangnyomásszinttől függ, hanem a frekvenciától is. A hallás tartományában a hangerő és hangnyomásszint közti összefüggést *hallási görbékkel* ábrázoljuk, ahol a dB csak az arányokat jelöli. A fon mértékegység az egyéni hangerőérzékelést próbálja leképezni.

A beszéd kiejtett hangokból álló, hangjeleken alapuló információs forma. Az ASR (gépi beszédfelismerés) az *ismertetőjel-vektorok* segítségével lehetséges.

A nyershang kódolása többek között PCM, GSM és μ -law eljárással történik.

A személyi számítógépek körében leggyakrabban használt hangformátumok a Wave és a MIDI. A WAV fájl a mikrofon (bemenet) hangnyomásának digitális másolata, és tetszőleges jellegű hang tárolására alkalmas. A MIDI (Music Instrument Digital Interface) a zenei adatok legsűrítettebb ábrázolása. Hangszerhez kötött ábrázolásmód, amelynek kódolása tartalmazza a hangerőt, az alapfrekvenciát és a hangszer megnevezését.

A továbbiakban gyakran használjuk az AU, MPEG és *MPEG Audio Layer III* (röviden MP3) kifejezéseket is.

A hangkártya–hangszóró rendszer az ember és a számítógép közti közvetítő-funkciót látja el. A hangszórók legfontosabb jellemzői a *torzítási tényező*, a *hatásfok*, a *hangteljesítmény* és a *terhelhetőség*. A mikrofonok legfontosabb jellemzője az érzékenység, ami megadja, mekkora a tényleges feszültség, ha 1 Pa *hangnyomásnak* teszik ki. A -60 dB érzékenység 1 mV/Pa-nak felel meg. Egy szimfonikus zenekartól 5 m távolságban körülbelül 1 Pa hangnyomás mérhető. A hangkártya általában *FM szintézis* segítségével hoz létre hangokat. Keverőpult segítségével több hangforrás jeleit igazíthatjuk egyaránt.

EA modul függeléke

III. Irodalomjegyzék

III.1. Könyvek

BURFORD-KOEGEL, JOHN F. (1994): *Multimedia Systems*. New York: Addison-Wesley.

COOKE, DERYCK (1962): *The Language of Music*. London: Oxford University Press.

GRITCHLEY M.; HENSON, R. A., EDS (1977): *Music and the Brain*. London: Heinemann.

ISSER, LEONARD EVERETT (2001): *Alexander Graham Bell*. North Lebanon, TN: Greenleaf Press.

SMITH, STEVE (1999): *Multimedia and Communications Technology*. 2nd Edition. Oxford: Focal Press, Butterworth, Heinemann. 7-43.

MCLAUGHLIN, TERENCE (1970): *Music and Communication*. London: Faber and Faber.

ISSER, STEVEN (1994): *The Language Instinct*. New York: William Morrow.

HEIDMAN, ROBERT D. (1999): *Computer Speech Technology*. Artech House.

WINSTONSON, JOHN (1994): *The art of digital audio*. Second edition. Woburn (MA): Focal Press.

WHITE, GLENN D. (1987): *The Audio Dictionary*. Second Edition. Seattle (WA): University of Washington Press.

WU, CHWAN-HWA; IRWIN, DAVID J. (1999): *Emerging Multimedia Computer Communication Technologies*. Upper Saddle River (NJ): Prentice Hall. 27.