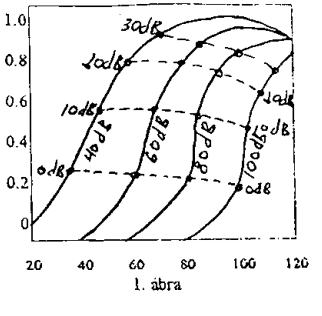
Beszédkezelési kérdések a számítógép hangprogramozásánál

A számítógépes hangkeltés még jobbára szórakozásnak számít, nincs komoly használati értéke. Ezért elsősorban a PC-s játékok kedvelői használnak hangkártyát. A házi használatra szánt PC-k elterjedésével viszont megnőtt a az igény a játékok iránt, és egy szép grafikával ellátott programhoz természetesen megfelelő zenei aláfestés is tartozik. Az Amiga és az Atari gépek MIDI interfésze gyorsan megjelent a piacon, de ez a hardverképességei alapján inkább zeneírásra mintsem zenehallgatásra való. Az első áttörést az AdLib és SoundBlaser kártyák megjelenése hozta. Ezeket a viszonylag olcsó áramköröket — egy átlagos erősítő közbeiktatásával — már többcsatornás, digitális hangzást lehet elérni. A kártyákat azóta is folyamatosan fejlesztik, a szoftvergyárakban egyre jobb zenét írnak a játékprogramokhoz. A felcsendülő hangok minden képzeletet felülmúlnak. Nem gondolnánk, de egy jól megírt hangzás teljesen meg tud változtatni egy játékot, „feldobja" még a gyengébb szoftvereket is. Mikrofont is csatolhatunk a hangkártyához, amellyel, hangfájlokat vehetünk fel. A felvétel több percig is tarthat, csupán a Winchesterünk szabad kapacitása szabhat korlátot. A mintavételi frekvenciák módosításával csökkenthetjük a fájlok méretét, de ez a hangminőség rovására megy. 60 mp felvétel 8 kHz-en 480 000 bájtot, 22 kHz-en 1 320 000 bájtot igényel. Legérdekesebb a Sound Editor. A hangszerkesztőben valós időben tekinthetjük meg a felvett hangzás burkológörbéjét, visszhangot keverhetünk a hang alá, akár fordítva is lejátszhatjuk, vagy más hangzást keverhetünk a régibe. A kibővített funkciók kiválasztásával szintetizátort is használhatunk. Ezenkívül megváltoztathatjuk a lejátszási sebességet, más mintavételi frekvenciákat állíthatunk be, illetve hangot keverhetünk. Az ilyesfajta lehetőségeket csak egy komoly technikai háttérrel működő stúdió tudja jól kihasználni!

A számítógép és az ember közti kapcsolat hang által — például számítógép vezérlése, irányítása — nem valósult meg. Számtalan probléma adódik: Nem egyforma mindenkinek a hangszíne, hangmagassága, hangerőssége, hangsúlyozása, kiejtése, hanghordozása, modorossága, és a sebessége. Ezért nem biztos, hogy a számítógép mindent és mindenkit megért.



1. ábra

A **beszédérthetőség** függ a fenti tényektől. Az általános beszéd-

érthetőségi összefüggésgörbéi a beszéd hangerősség és a környezeti zajszint függvényében változik. (1. ábra).

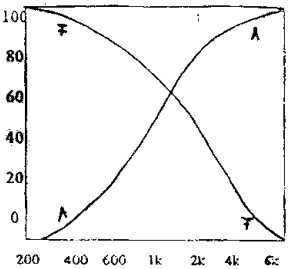
A Fletcher-Galf-féle görbesereg egy részlete, kiemelve az előadótermi, szabadtéri, az utcai közlekedési és az ipari zavaró zaj hatását szemlélteti. A hasznos hang és a zajszint különbségeinek hatása is leolvasható. Az érthetőség a hangerősséggel egy ideig növekszik, bizonyos hangosság fölött azonban csökkenni kezd. Az emberi hangban, de főleg az átvivő rendszerekben torzítások keletkeznek. A hangerősség meghatározásánál nem a beszélő saját teljesítményét, hanem a hallgató fülénél jelentkező hangerősséget kell figyelembe vennünk. Ebbe a hangterjedés a csillapítás, az átviteli torzítások és a lehallgatások körülményei is beszámítanak. Nem lehet jól mérni a beszéd átlagos hangerősséget. Erre sem a stúdiótechnikában alkalmazott VU-méterek, sem más átlagoló mérések nem alkalmasak. Figyelembe kell vennünk, hogy a beszédhangok akusztikai teljesítménye a leghangosabb magánhangzóktól (ó, á, é) a legmagasabb mássalhangzókig (f, h, angol th) 30-dB-t fog át. Valamilyen középértékhez csak statisztikai módszerrel lehet eljutni. Ehhez azonban ismét egyes nyelvek sajátos tulajdonságait kell figyelembe vennünk. Az 1. ábrán kiválasztott három érthetőségi görbe közül a 40-es jelzésű előadóteremben, a 60-as társalgási körülmények között, a 80-nal jelölt utcai zajban, végül a 100-as jelzésű zajos ipari üzemekben lebonyolított párbeszéd érthetőségére vonatkozik. A nehezebb esetekben mind nagyobb hangerő kell a jól érthető beszédhez, bár a nagy hangerő miatt az érthetőség csökken. Mivel az esetenkénti háttérzaj színképi eloszlása maga is változó paraméter, az ábra csak általános tájékozódásra alkalmas, pontosabb érthetőségi számításokhoz további színképi vizsgálatok szükségesek.

Nehéz a beszéd sebességétől függő érthetőség vizsgálata. A sebesség növekedhet a szóközi szünetek csökkentésével, a hangok arányos rövidítésével (például mesterséges úton), vagy akár sajátságos egyéni hanghordozással (a magánhangzók rövid ejtése, szótag gyorsítás, hangkihagyások). Ezek meghatározásában a „hadarás" ítéletén túl kevés objektív lehetőségünk van.

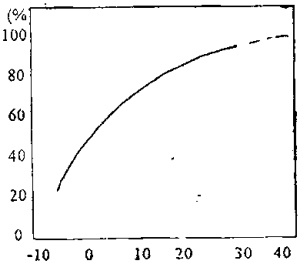
Az átviteli rendszer hatásának vizsgálata a leggyakoribb kísérleti eljárás. Ilyenkor feltételezzük, hogy a beszélő jó szövegkiejtésű személy, a megfigyelők pedig kitűnően hallanak, s mindkét oldal megfelelően gyakorlott a kísérleti technikában.

Változhatnak az átvivő közeg vagy a műszakiláne paraméterei, például frekvenciaátvitel, torzítás, alapzaj, vagy a terem visszhangja, utózengés, stb. A frekvenciaátvitel érthetőségcsökkentő hatására korai ismereteink vannak.

A 2. ábrán azt mutatjuk be, hogy az átvitel alsó és felső határfrekvenciájának korlátozása hogyan rontja az átlagos szóérthetőséget. Ha 3000 Hz felett minden összetevőt levágunk, elsősorban a zár- és réshangok



2. ábra



3. ábra

érthetősége vész el, így az átlagos szóérthetőség mintegy 78-82%-ra esik vissza. Persze, ugyanaz a baj akkor is bekövetkezik, ha az említett felső összetevőket maga a hallás vágja le (öregkori hallásveszteség!) Az átviteli sávkorlátozásoknál is fontosabb a zavaró zajok érthetőségcsökkentő hatásának ismerete. Súlyosabb esetekre felkészülve nem az érthetőség szokványos vizsgálatával kell törődnünk, hanem eleve olyan emberi megoldásokat kell választanunk, amelyekkel jobb érthetőséget érhetünk el. Legegyszerűbb módszer a hangosabb és tagoltabb beszéd, a szavak megismétlése, betűzése. Közlekedési zajban, rossz légköri körülmények, elektromos zavarok közepette az információ pontossága helyett a redundancia fokozására kell törekednünk. Ebből a szempontból például a többszótagú szavak érthetőbbek, mint a rövidek. A repülőgépek zajában a rádiós üzenetek „yes" és „no" egytagú szavai helyett az „affirmative" és „negative" kifejezéseket tették kötelezővé. A felsorolt okok ma is a legfontosabbak érthetőségvizsgálati feladatok. Ezekre fejlesztették ki a szokásos eljárásokat, méréseket és számításokat. Gyakorlati cél eleinte a telefonfejlesztés, később az elektroakusztikai átvivő láncok tökéletesítése, ma pedig ezek mellett a teremakusztikai ellenőrző mérések egyszerűsítése.

Egy másik tapasztalat, hogy zajos környezetben három küszöböt kell átlépnünk.

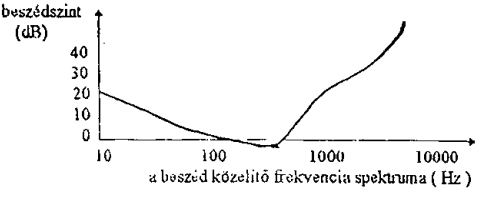
A fokozatosanjavuló jel/zaj viszony mellett a szöveg megértéséhez jutunk el. Ha a zajszint 10 dB-lel magasabb az átlagos beszédszintnél, meg tudjuk állapítani, hogy beszédet hallunk. További 6 dB beszédszint emelés esetén már sok szóalakot fölismerünk, végül 0 dB jelszint-zajszint helyzetben 40–60%-os szótagérthetőség tapasztalható, ami nyelv és téma ismeretében elérheti a 85-90%-os beszédérthetőséget is (3. ábra)

A telefonbeszélgetés ma is 300–3400 Hz sávszélességű vonalon folyik, amit alapzaj és mikrofon torzítás is terhel. Egy elképzelt beszélgetés során a hívó fél elkapkodott bemutatkozása nem csak azért nem érthető, mert felületes az artikuláció, hanem azért sem, mert a figyelem fölkeltéséhez

bizonyos időre, semleges bevezető szavakra — például köszönésre — van szükség. Gyorsan mondott szöveg azért marad kevéssé érthető, mert a beszélőt sem azonosítottuk és a témát sem ismerjük előre. Ha ezeket a körülményeket egy kulcsszó megvilágosítja, a tartalom hirtelen érthetővé válik. Agyi feldolgozás szempontjából teljesen más az egyes hangok vagy szavak felismerése és a mondandó szöveg megértése. Az 50-60%-os szótagérthetőség akár 90-95%-os beszédérthetőséget jelenthet. Ennek oka, hogy a téma ismeretében a megfigyelő agyműködése az elvesző részinformációk nagy százalékát pótolja vagy korrigálja.

Beszédkezelés azokat a műveleteket jelenti, amelyeket a beszéd átvitele, felismerése és mesterséges előállítása során végeznek. Ezeknek a feladatoknak minél pontosabb megoldására törekednek.

Az elektromos jellé átalakított beszéd jellemezhető frekvencia-karakterisztikájával:



A beszédkezelésben előnyös volna szétválasztani a beszéd információhordozó részét, amely az írott szöveggel egyenértékű, a beszélő egyénre jellemző részét, a hangulati hangsúly és dalamrészeit, amelyek a szöveg és a beszélő kölcsönhatásának jellemzői.

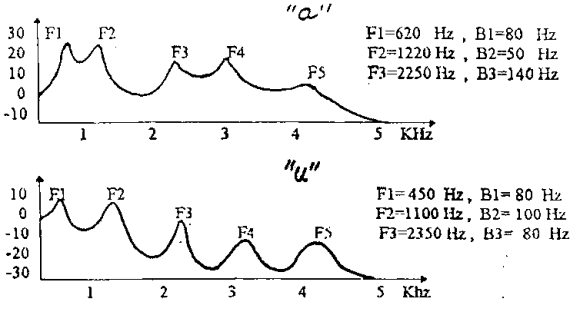
A felsoroltak érdekében az időtartományban lévő nullátmenet, a jelek néhány 10–100 Hz frekvenciájú burkológörbéje, csúcstényezője és aktivitási tényezője hasznos adatok lehetnek. A beszéd információ tartalmát nagyrészben a jel nullátmenetei hordozzák. Ezért ha a beszéd csúcsait levágjuk, közelebb kerülünk a hasznos információkhoz. Ezt nevezik *csúcslevágásnak*. A csúcslevágás különböző méretű lehet: az idő ε hányadában (ε=0,01 vagy 0,001) előforduló, minőség szempontjából lényegtelen amplitudokat vágják le. A beszéd 40 dB-es dinamikáját 10–20 dB-re komprimálják a zajérzékenység csökkentése érdekében. Így az egyéni és hangulati jellemzők egy része elvész. Csak a nullátmenetet kezelik, mialatt az elemek rendkívül leegyszerűsödnek, digitalizálhatók és számítógéppel feldolgozhatók lesznek. Így monoton, de szolgálati célokra érthető beszédet kapunk. Ezt az eljárást *végtelen csúcslevágásnak* nevezzük. A végtelen csúcslevágás után kapott jel mellett a 0–10 Hz periódusú burkolóingadozásokat kezelve javul az érthetőség. A beszédkezelés nemcsak az amplitudók csökkentésével egyszerűsíthető, hanem az időbeli jellemzők kihasználásával is. A beszédfelismerés célja az

elhangzott beszéd írásban való rögzítése. Erre különböző kísérleteket végeztek, tökéletes eredményt azonban még nem értek el. Az eddigi megoldások lehetővé teszik, hogy a beérkező beszédet véges szókészlettel összehasonlítva meg lehessen állapítani, hogy a vett szó melyik szóval azonos. 70–100 szavas készletekből nagy biztonsággal tud a berendezés kiválasztani. A rendszer alkalmas a számítógépek beszédvezérlésére. A beszédfelismerésre kidolgozott módszerek a beszédet vagy időben vagy frekvenciában darabolják és az így előállított rácsokat illetve szegmenseket vetik össze a rendelkezésre álló szókészlet megfelelő darabjaival. Időbeni darabolásnál jelöljük ΔT-vel a periódust és x(t)-vel az elektomos jellé átalakított beszéd frekvenciáját a *t* pillanatban. A j*Δt pillanatban az x(j*Δt) értékeket kodifikáljuk és a memóriában rögzítjük. A visszajátszásnál az értékeket dekodifikáljuk és egy D/A átalakítón keresztül egy hangszóróra küldjük. Az impulzusok kódja hordozza az információ átalakítást:ezt PCM (Pulse Code Modulation) módszernek nevezik. Hogy jó mínőségű beszédet kapjunk, melynek dinamikája eléri a 60 dB-t, az

A(dB)=20*lg (Umax/Umin) képlet alapján kiszámítható,
hogy 60=20*lg(Umaz/Umin)=0, Umax/Umin=±1000

Tehát a beszédérthetőség érdekében kell biztosítani ±1000 szintet. Ezen értékek kódolására minimum 11 bit memória szükséges. A 80–8000 Hz frekvenciasávok átfogása érdekében a Shannon-féle digitalizálási tétel alapján 1 másodpercnyi időt legalább a frekvenciasávok számának a duplájára kell osztani, azaz minimum 16 Khz-re. Ilyen feltételek mellett a memóriaszükséglet 16 KHz*11=176 Kbit/sec. Ha a frekvencia sávokat 300–3000 Hz-re szűkítjük és a diamikát úgy állítjuk be, hogy egy jel hossza 8 bit legyen akkor a szükséges memória 64 Kbit/sec, de a beszéd érthetősége erősen lecsökken.

A frekvenciában való daraboláskor a beszédet úgynevezett hangképekre daraboljuk. Minden hangkép F1, F2, F3, F4, F5 csúcsokból és hozzátartozó B1, B2, B3, B4, B5 sávokból tevődik össze. A következő ábrák az „a", illetve az „u" hangképét mutatják.



A beszéd karakterisztikáját az F0 mindenkire jellemző alaphang és az F4, F5 csúcsok határozzák meg. Ezeknek a váltakozása a beszéd ideje alatt nagyon kicsi. A beszéd információtartalmát nagymértékben az F1, F2, F3 csúcsok hordozzák. A mesterséges beszéd célja, hogy írásban vagy a memóriában rögzített jelek érthető beszéddé formálódjanak. Ez egy bizonyos mértékig megvalósítható a frekvenciában való darabolás módszerével.

Első lépésként a számítógépbe bevitt szövegből kiszűrjük a nem kiejthető jeleket (pont, vessző, köz, zárójelek stb.), majd a megmaradt karaktereket átalakítjuk a hozzájuk tartozó hangképekké. Második lépésként az így keletkezett hangtömböt a számítógépes memóriában levő szótárhoz hasonlítva elvégezzük a megfelelő kiigazításokat. Harmadik lépésként a már elkészített nyelvi szabályokat tartalmazó könyvtárból kikeressük az illető szóhoz tartozó legmegfelelőbb kiejtést. Ezt a lépést csak azokra a megmaradt hangképrészekre kell elvégeznünk, amelyeknek a megfelelőjét nem kaptuk meg a második lépésben. Ilyenek például a számok. A módszer nehézségét fokozza, hogy a karakterekhez tartozó hangképek megszerkesztését csak megfelelő technikai körülmények között oldhatjuk meg.

Könyvészet:

1. Fizikai szemle, 1995/3
2. Computer Panoráma, 1992. május
3. Introducere în microprocesoare, Ed Ştiinţifică şi Enciclopedică, Bucureşti, 1986.

Varga Elemér – tanuló

Gábor Áron Szakközépiskola Kézdivásárhely

Tudománytörténet

Imre Lajos

A XX. század hajanlán, 1900. március 21-én született a magyarországi Litke községben szegény, földműves családban. Kiváló szorgalmával és tehetségével korán felkeltette a helység lelkészének figyelmét, aki támogatta elemi és középiskolai tanulmányainak elvégzésében. Az érettségi vizsga után a budapesti egyetemen szezett matematika, majd kémiatanári oklevelet, mint kormányzógyűrűs - ami azt jelentette, hogy minden vizsgáját jeles eredménnyel tette le.