

Emberi beszéd mesterséges előállítás és gépi felismerése

Az emberi hangok mesterséges előállítása a világon először a magyar Kempelen Farkasnak sikerült. Az ő készüléke mechanikus úton imitálta az emberi hangképzést az 1700-as évek végén.

A számítógépek őskorában, az 1960-as években újra feléledt a gondolat: hogyan állíthatnánk elő emberi beszédet mesterségesen? A gépi hangok emberi beszédként való felismeréshez még az 1980-as években is „sok jóindulat” kellett. Annyira rossz volt a hang minősége, hogy a hallgatónak önkéntelenül az az érzése támadt, a számítógépnek sürgősen meg kellene csinálna a fogait. Ennek az oka többek között a kimenő csatorna kis frekvenciaszélessége volt. A gépek növekvő teljesítménye mára lehetővé tette élvezhető és érthető hang előállítását számítógép segítségével. Ugyanakkor egy hibátlanul hangsúlyozó, valódi emberi hangon beszélő, az érzelmeket és hangerőváltozásokat hűen visszaadó rendszerig, amelyeket néha filmekben látunk, még hosszú utat kell megtennünk.

Magyar fejlesztések

Már a 80-as évek óta létezik egy magyar fejlesztésű hangszintetizátor, amit a <http://speechlab.ttt.bme.hu/> címen találhatunk meg. A www.vilaghallo.hu webhelyen kis is próbálhatjuk a rendszert, és tapasztalhatjuk, hogy a minősége kifejezetten jó. Először *multivox* néven egy tisztán szintetizált hangokból álló beszéd szintetizátort fejlesztettek ki, majd egy emberi hangra alapuló változatot, mely *profivox* néven ismert. Sajnos, mindkettő kereskedelmi termék, a *profivox* része az angol árjegyzék szerint körülbelül 850 dollárba kerülő *jaws* csomagnak, melynek demó változata a <http://www.vakalap.hu> oldalról tölthető le. Ugyanezt a <http://www.vilaghallo.hu> ingyenes felolvasóprogramként terjeszti, de ez a megoldás egyrészt állandó hálózatot követel, másrészt csak néhány mások által előre kiválasztott könyvet képes felolvasni, ami sokakat biztosan nem elégíti ki. 2001 óta létezik a szintén magyar fejlesztésű *speakboard* (<http://www.specht.com/speakboard/intro.php>), mely saját állítása szerint nyelvfüggetlen beszéd szintézis-rendszer. Magyar és angol változata vásárolható meg, ára körülbelül húszezer forint. Az előállított beszéd meg is hallgatható a fenti oldalon, véleményem szerint szintén jól érthető. **Ingyenes illetve nyílt forráskódú beszéd szintetizátorok** Mi kell tehát ma ahhoz, hogy számítógépünk beszéljen, vagy akár énekelni tudjon anélkül, hogy emiatt nagy pénzt

kelljen külön kiadnunk? Egy hangkártya, körülbelül 20 MB hely a merevlemezen és egy beszéd szintetizáló program. A jelenleg létező beszéd szintetizáló rendszerekről a <http://www.speech.cs.cmu.edu/comp.speech/Section5/Q5.5.html> címen találhatunk összefoglalót. A dolog gyakorlati oldalát tekintve az átlagos érdeklődő számára olyan rendszer jöhet számításba, amely ingyenes, amelyhez támogatja az új nyelvek felvételét, amelynek elfogadható a minősége, és amely lehetőleg Linuxon is és Windows alatt is működik. Ezeknek a feltételeknek jelenleg egyetlen rendszer felel meg, az *mbrola* beszéd szintetizátor. A nevét helyesen „*embérolának*” kell ejteni, amúgy az angol umbrella (esernyő) szóból ered. A rendszert a <http://tcts.fpms.ac.be/synthesis/mbrola.html> helyen találhatjuk meg.

Az *mbrola* beszéd szintetizátor

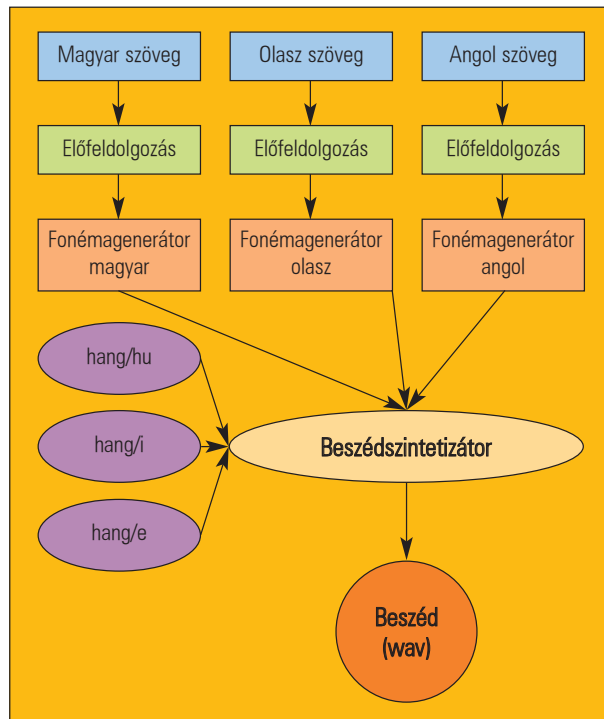
Az *mbrola* ugyan nem szabad forráskódú program, de a katonai és a kereskedelmi célú alkalmazásokat kivéve bárki ingyen használhatja. A programot a *Belgiumban* levő *Monsi Műszaki Egyetem* hangtechnika tanszéke írta. Körülbelül 10 éves, így ma már érettnak mondható, amit az is bizonyít, hogy 33 nyelvhez, köztük a japánhoz és koreaihoz is van már hangja. Egy adott nyelv támogatásához valakinek gépbe kell mondania az összes előforduló kettős hangzót, szaknyelven *difonémát*. Ha megvan ez az adatbázis, a szintetizátor képes az adott nyelven beszélni.

A difonéma adatbank

Egy difonéma adatbank elkészítése nem triviális feladat, erről részletes beszámoló olvasható a <http://tkltrans.sourceforge.net/magyar/bmrolamod.htm> címen. Ezt természetesen csak egyetlen egyszer kell megtenni, hiszen később a hangadatbank minden felhasználó rendelkezésére áll. A magyar nyelvhez 2005. januárja óta van adatbank, és további magyar adatbankok vannak készülében.

Szóveg-fonéma átalakító

Ha kész a difonéma-állomány, még szükséges egy nyelvfüggő fonémaelőállító program is, mely az elmondandó szöveget difonémákká alakítja. Ez a magyar nyelvhez



1. ábra

a <http://tkltrans.sf.net/magyar/hunpho.tar.gz> címen található. A „Helló világ!” szöveg például így néz ki fonémákká alakítva:

```

_ 50 0 160
h 70 0 170 30 200
E 90
l 65
l 65
o: 148
_ 10 0 160
_ 150 0 160
_ 10 0 160
v 70 0 210 30 200
i 90
l 65
a: 148
g 75
_ 10 0 160 30 190
_ 350 0 160

```

A fonémásorozatban egy sor értelmezése: Az első jel a kimondandó betű úgynevezett *sampa jelölésben*. Erről bővebben a <http://phon.ucl.ac.uk/home/sampa> címen olvashatunk. A második jel a kimondás hossza ms-ban, majd a hanglejtés (pitch) megadása következik, amely számpárokból áll: Az első szám a fonémán belüli hanglejtés kezdetét adja meg, a második a frekvenciát hertzben (Hz).

A példa első sora (_ 50 0 160) azt jelenti, hogy 50 ms-nyi szünetet tartunk úgy, hogy ennek 0%-ánál a frekvencia 160 Hz kell legyen. A hanglejtési pontok egy lineáris hanglejtési görbét adnak meg.

A hangerősség változtatására a *.pho* fájlban belül nincs lehetőség. Ugyanakkor ez is megoldott például a *de6* és *de7* hangadatbankokban. Az ötlet annyi, hogy ezek a hangadatbankok terjedelmüket megháromszorozva, minden difonémát hangos és halk módban is bemondanak.

A fonémásorozatot az *mbrola* szövegszintetizáló program közvetlenül *.wav* állománnyá képes alakítani, melyet a számítógép közvetlenül a *play* segédprogrammal vagy például az *mplayerrel* (*Linux*) vagy *Windows* esetén bármelyik lejátszó programmal (például *Media Player*) le tud játszani.

Az *mbrolával* való hangelőállítás menetét talán legvilágosabban ez az ábra szemlélteti:

Mint látható, a hangelőkészítést és a szöveg fonémákká átalakításának nyelvspecifikus részét nem az *mbrola* szintetizátor végzi, hanem külső modulok. Ez a modularizált felépítés nagy szabadságot jelent, és lehetővé teszi, hogy a fejlesztő mindig csak az adott nyelvre koncentráljon, hiszen a hallható beszédalakítás végső soron független a célnyelv szabályaitól, sajátosságaitól.

A szövegszintetizáló rész

A felhasználó számára az *mbrola* program és a fonémaképzéshez szükséges eszközök telepítése talán a legnehezebb a megoldandó feladatok közül. A programot és a hangadatbázist a <http://tcts.fpms.ac.be/synthesis/mbrola/mbrcopybin.html> kell letölteni.

Egy linuxos gépen az *mbr301.zip* tartalmát kicsomagolás után a */usr/local/mbrola* könyvtárba kell írni, és ez alatt létre kell hozni egy *hu1* alkönyvtárat. Ide kerül a *hu1* adatbank, vagyis ide kell kitéríteni a *hu1.zip* nevű fájlt.

Az egész folyamat a *Linux* alatt így néz ki (*/en/konyvtaram* az a hely, ahová a letöltött csomagokat helyezzzük).

```

bash# cd /usr/local
bash# mkdir mbrola
bash# cd mbrola
bash# unzip /en/konyvtaram/mbr301h.zip
bash# unzip /en/konyvtaram/hu1.zip

```

A */usr/local/bin* könyvtárban hozzunk létre egy a végrehajtható *mbrola* állományra mutató szimbolikus linket:

```

bash# cd /usr/local/bin
bash# ln -s /usr/local/mbrola/mbrola-linux-i386
mbrola

```

A szöveg fonémákká alakítása

Ezek után a szöveg-fonéma átalakítót (<http://tkltrans.sf.net/magyar/hunpho.tar.gz>) kell beüzemelnünk. Ennek a részegységnek szüksége van a Perl nyelvre, valamint az *awk*-ra.

A feldolgozandó szövegnek egyszerű szöveggé kell rendelkezésre állnia, vagyis a *.doc*, *.pdf*, és egyéb ehhez hasonló fájlokat előbb tiszta szöveggé kell átalakítanunk. A szöveg két szűrőn halad át, mielőtt hanggá (vagyis fonémák sorozatává) válik.

A *sam.awk* kiszűri a zárójeleket és idézőjeleket, az önálló betűket, egyes rövidítéseket kiejthetővé tesz, problémás szókapcsolatokat szétválaszt a számjegyeket pedig szavakká alakítja.

Az *xttp.pl* aztán az így megszürt szöveget fonémákká alakítja. Az keletkezett fonéma állományt kell átadni az *mbrola* programnak úgy, hogy az a *hu1* adatbázist használja.

```
awk -f szam.awk <$1.txt >$11.txt
perl xttp.pl $2 < $11.txt >$1.pho
mbrola /usr/local/mbrola/hu1/hu1 $1.pho $1.wav
play $1.wav
rm -f $11.txt
```

xttp.pl paramétere, melyet nem kötelező megadni, lehet *m*, *f1*, *f2* vagy *f3*. A paraméter hiánya a legmélyebb női hangot jelenti, azaz megfelel az *f1* paraméternek, *f2* és *f3* egyre magasabb hangokat jelent, *m* pedig a férfihang. Ha egy szövegállományt sikeresen átalakítottunk szöveg.pho formátumba, akkor az *mbrola* azt *wav* formátummá alakítja, amit aztán egy tetszőleges hanglejátszó program le tud játszani. A legegyszerűbben a *play* szöveg.wav

paranccsal tehetjük hallhatóvá művünket. (*Alsa* meghajtó esetén az *aplay* parancsot kell használni).

Íme két példa a *do.sh* (vagy *do2.sh*) használatára:

```
sh do.sh szoveg f1
sh do.sh szoveg
```

Eljutottunk tehát odáig, hogy bármely szöveget fel tudunk olvasatni a géppel. Most koncentráljunk a finomabb igényekre. A *mbrola* fel tud olvasni hosszabb szövegeket, tud énekelni, a szöveg hanglejtését pedig grafikusán is megjeleníthetjük vele. Ez utóbbi szolgáltatás a fonémaátalakító program javításánál jól használható.

Hosszabb szöveg felolvasása

Mivel hosszabb szövegek feldolgozása nagyon nagy *wav* állományt eredményezne, célszerű ezeket kisebb darabokra szabdálni és úgy fölolvastatni. Erre való a *mesel.pl* program, amely a *segédesszkozok* könyvtárból futtatható. Az olvasandó állomány nevét a program paramétereként adhatjuk meg (a *.txt* végződés nélkül), az eredménynek pedig hozzuk létre a *test* nevű alkönyvtárat.

A program kizárólag olvassa a neki átadott állományt, tehát annak tartalmát semmilyen módon nem változtatja meg. Az éppen fölolvastott részt a *kwrite* segítségével a képernyőn is mutatja. A megjelenítéshez használt program a *\$szerkeszto* változóval ízlés szerint változtatható. A *\$sor* belső változó az egyszerre fölolvastott sorok száma, szintén ízlés szerint beállítható. A *mesel.pl* program hívása, ha a fölolvastandó szöveg a *szoveg/olvasnivalo.txt* állomány:

```
perl mesel.pl szoveg/olvasnivalo
A mellékelt mesel.sh állományt használva:
```

```
sh mesel.sh szoveg/olvasnivalo
```

A *mesel.pl* a *mesel_sorok.txt* fájlban feljegyzi, hogy éppen hol tart az olvasásban. Ha indításakor a második paraméter a *last* szó, akkor a *mesel_sorok.txt* állományban levő szám által kijelölt sornál kezd el a szöveg olvasását. Ha a harma-

1. táblázat *A zenei alaphangok frekvenciái*

c3	d3	e3	f3	g3	a3	h3	c3
262	294.5	327.5	349	393	440	491	524

2. táblázat *„Boci, boci, tarka...”*

{Singing: 8,c2, 8,e2, 8, c2, 8, e2, 4,g2 4,g2 }.
bo ci bo ci tar ka

dik paraméter szám, akkor az annyiadik sorban kezd meg a fölolvastást. Íme néhány példa:

```
perl mesel.pl szoveg/olvasnivalo last
```

Ez esetben *mesel_sorok.txt* tartalma szabja meg, hogy hol kezd el az olvasást.

```
perl mesel.pl szoveg/olvasnivalo 256
```

Ekkor a 256-odik sornál kezd el a fölolvastást.

Éneklés

Mivel a számítógép szemszögéből az éneklés sem más, mint a hangképzés egy különleges esete, a *mbrola* énekelni is tud. A *hunpo.tar.gz*-ben mellékelt *boci.txt* állomány például a boci-boci tarkát tartalmazza. A további mellékelt dalok a *damdam* (*damdam.txt*), a király és a bolond (*bolond.txt*), szeretnék szántani (*szantani.txt*). Ha valakinek netán komponálni támadt volna kedve, annak ajánlom figyelmébe az 1. táblázatot, amely az alaphangok frekvenciáit tartalmazza (hertzben). Gépünket a következő egyszerű paranccsal fakaszthatjuk dalra:

```
sh do.sh szoveg/boci.txt
```

A rendszer hangzókészlete négy oktávnyi távolságot fog át, ami egy géptől egészen szép teljesítmény. A hangok *cn*, *cszn*, *dn*, *diszn*, *en*, *fn*, *fiszn*, *gn*, *giszn*, *an*, *aiszn*, *hn* alakban adhatók meg. Az *n* egy szám, ami a hangmagasságot jelzi. 1 a legalacsonyabb, 4 a legmagasabb oktáv. A normál zenei á hang (440 Hz) jele tehát *a3*. Minden hang előtt meg kell adni a hosszúságát is. 1 a leghosszabb, 16 a legrövidebb hang. Ezzel a jelölésrendszerrel a „Boci boci tarka” a következőképpen fest:

```
{Singing: 8, c2, 8, e2, 8, c2, 8, e2, 4, g2, 4, g2}.
```

Az egyes szakaszok kiénekelte formáját a 2. táblázat mutatja. Látható, hogy az első helyen a {Singing: betűsorozatnak kell állnia, majd ezt követik a hosszúság/hang párok, vesszővel elválasztva. A sorozatot egy „.” jelpár zárja. (A pont fontos a sor végén!) A {Singing: kezdetű sort követő sorban van az énekelendő szöveg. Ennek ugyanannyi szótagból kell állnia, mint ahány hangot megadtunk, és ezt is pont zárja a végén. Szintén lényeges, hogy a {Singing:-et megelőző szövegrészt ponttal kell lezárni, különben ezt lenyeli a program.

Egy mondat hanglejtésének grafikus megjelenítése

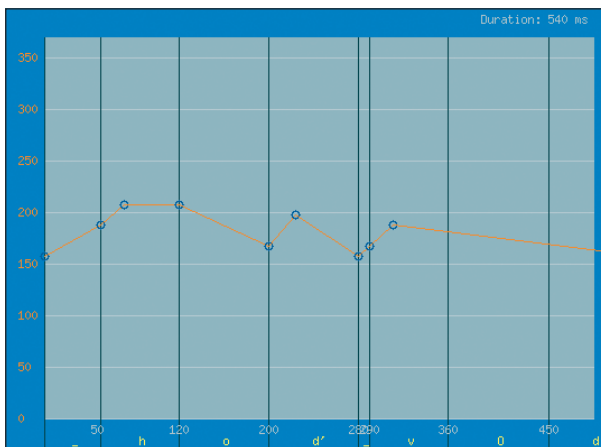
Néha nagy segítséget jelenthetnek a *show.pl* illetve a *graph5.pl* nevű Perl programok, amelyeket szintén

3. táblázat *Az mbrola program kapcsolói:*

Kapcsoló	Leírás	Példa
h	segítség a kapcsolókhoz	mbrola -h
i	fonéma-adatbank információ	mbrola -i /usr/local/mbrola/hu1/hu1
e	nem létező difonéma ne okozzon megállást	mbrola -e /usr/local/mbrola/hu1/hu1 test.pho test.wav
v	hangerő 1: alapállás, <1: halkabb, >1: hangosabb	mbrola -v 1.1 /usr/local/mbrola/hu1/hu1 test.pho test.wav
f	hangmagasság 1: alapállás, <1: magasabb, >1: mélyebb	mbrola -f 0.6 /usr/local/mbrola/hu1/hu1 test.pho test.wav
t	beszédsebesség 1: alapállás, <1: gyorsabb, >1: lassabb	mbrola -t 1.2 /usr/local/mbrola/hu1/hu1 test.pho test.wav

A beszéd során néha suttogunk és néha kiabálunk, Ezt, mivel a suttogás és kiabálás nem képezhető le ugyanarra a difonéma adatbankra csak úgy lehet megoldani, ha egy helyett háromszor készletet hozunk létre a difonémákból: egyet a suttogáshoz, egyet a normálshoz és egyet kiabálva. Ez egyetlen szöveg keretén belül lehetővé teszi a háromféle szöveg kiadását, amit a német de6 és de7 difonéma adatbankok esetébe már meg is oldottak.

© Kiskapu Kft. Minden jog fenntartva



2. ábra

a hunpho.tar.gz csomagban találunk. Ezek a megadott mondat hanglejtését egy képfájlba írják ki, amit aztán bármelyik böngészővel meg lehet tekinteni: show.pl:teszt.png graph5.pl:file3.png

Ezek az eszközök a jobb fonetika kidolgozásában segítenek, vagyis az xttp.pl program javításában. A háló kedvelői számára a szolgáltatás online is elérhető a következő helyen: <http://pascal.kgw.tu-berlin.de/expressive-speech/online/synthesis/hungarian/en/ia-en.php>. (Köszönet Astrid Paeschke-nek a Berlini Egyetemről a magyar adatbank integrálásáért!).

A fonéma-átalakító és a nyelvi adatbázis továbbfejlesztése

Első lépésben a hangsúlyozást kell tökéletesíteni, hogy a kérdő, felszólító, óhajtó, stb... mondatok intonációja helyes legyen. A hely és időbeli viszonyt kifejező szavakat nem hangsúlyozzuk, (például „asztal alatt”). így a fonéma-átalakító már ma is helyesen kezeli ezt.

A német fonetika-átalakító (txt2pho) képes egyes szavak különféle hangsúlyozására. A „felá**st**am a kertben a zö**ld**ség-á**gy**ást” mondat esetében például a hangsúlyozandó szót ki lehet emelni a szórenddel, de a hangsúlyozással is.

A következő lépés a különféle érzelmeket (kétely, félelem, letörtség, jókedv, buzdítás, stb) kifejező szöveg helyes hangsúlyozása lenne.

A szegény ember beszéd szintetizátora

A mbrola mögött megbúvó zseniális ötlet tulajdonképpen a difonémák használata. Ha van valaki, aki összeállítja egy adott nyelv difonémakészletét, azt kellő monotonitással a gépbe mondja (azaz wav állományokat hoz létre) és ezekből adatbankot készít, akkor ezzel a rendszer tetszőleges szöveg felolvasására képessé válik. Ráadásul ehhez elegendő például egy egyszerű Perl program, mely a felolvasandó szöveg difonémáit leképezi az adatbank difonémáira.

Tegyük fel például, hogy az adatbank tartalmazza a következő difonémákat:

_b, _a, ab, ba, aa, bb, _a, _b,

ahol a _ jel szünetet jelent.

Ezzel az adatbankkal a gép ki tudja mondani a „baba”, „abba”, „a”, „bab”, „ba” szakavak és szóelemeket, illetve minden más, csak „a”-t és „b”-t tartalmazó szót. A megfelelő Perl program mondjuk a „baba” szóhoz kikeresi a

b ba ab ba a

difonémákhoz tartozó wav jeleket, majd ezeket egyetlen wav állománnyá alakítja. És ezzel meg is oldotta a baba szó kimondását.

Problémát jelent, hogy az egyes difonémák összekötésekor a találkozási pontban egyes esetekben keletkezhet egy kattanás, ha a hangerő vagy a hangmagasság nem ugyanaz. Nyilván célszerű úgy kialakítani a szoftvert, hogy állítható legyen a hangmagasság (pitch), a beszéd sebessége, valamint a hangerő. Ezeknek a paramétereknek a helyes megválasztásával kelthetjük a beszéd valódiságának érzetét. Mindezt az *mbrola* program már most is lehetővé teszi.

Más, nyílt forráskódú beszéd szintetizátorok

Létezik nyílt forráskódú beszéd szintetizátor is. Ilyen például a <http://www.cstr.ed.ac.uk/projects/festival/> helyen elérhető *Festival*. Ehhez jelenleg csak angol, wales és spanyol difonémakészlet létezik, valamint egy olyan kapcsolat, melynek segítségével az *mbrola* hangadatbankjait és szintetizáló képességét is felhasználja a *Festival*.

Ehhez nagyon hasonló a CMU egyetem *Festvox* nevű terméke (<http://festvox.org>), melyhez szintén csak a fenti három hangadatbank létezik, de a dokumentációja leírja, hogyan lehet japán adatbankot készíteni hozzá (<http://festvox.org/bsv/bsv-jpdiphone-ch.html>). A Festival szintetizátor Mandrake Linuxon 10.0 alatt nekem rögtön működött, de a hang kétszeres sebességű, vagyis csipogó volt. Elolvasva aztán a gyakran ismételt kérdések oldalt rögtön a második kérdésére adott válaszban megtaláltam ennek az okát (http://www.cstr.ed.ac.uk/cgi-bin/cstr/lists.cgi?config=festival_faq&entry=arunning_festival/speed.html). A trükk annyi, hogy létre kell hozni a lib könyvtárban egy siteinit.scm állományt a következő tartalommal:

```
(Parameter.set 'Audio_Method 'Audio_Command)
(Parameter.set 'Audio_Command "sox -t raw -sw -r
$SR $FILE -c2 -t ossdsp /dev/dsp")
```

Alsa hangmeghajtás esetén a második sor:

```
(Parameter.set 'Audio_Command "aplay -fs16_LE -r
$SR $FILE ")
```

A *Festival* beszédszintetizáló és a *Sphinx* beszédfelismerő programcsomagok azoknak az elszántabb embereknek jók, akik szeretnének a forráskódhoz hozzájutni és akiket érdekelnek azok a programtechnikai eszközök, melyeket ezek a programcsomagok használnak, illetve akik kíváncsiak mondjuk arra, hogy más hangadatbankok milyen egyszerű fonémákból, trifonémákból vagy fölvetett szövegekből állnak, és ezekkel milyen hangminőséget eredményeznek.

Mivel ezek a programcsomagok temérdek lehetőséget kínálnak, megértésük lényegesen időigényesebb, mint a most bemutatott *mbrola* használata.

sAz beszédszintetizáló alkalmazások száma manapság egyre nagyobb, képességeik pedig egyre jobbak. Könyvek, cikkek, emailek fölolvása mellett lehetővé teszik, hogy a számítógépet vakok is kezelhessék. Ez utóbbira több linuxos projekt is létezik:

<http://developer.gnome.org/projects/gap/AT/Gnopernicus/>
<http://www.kde-apps.org/content/show.php?content=18806>
<http://emacspeak.sourceforge.net/>

Ezek közül talán a *Gnopernicus* a legelőrehaladottabb, de még ez sem kész.

Beszédfelismerés

A gépi beszéddel rokon, de lényegesen bonyolultabb téma a gépi beszédfelismerés. Műszaki szempontból, illetve nehézségét tekintve a beszédfelismerés a kézírásos szöveg gépi felismerésével egyenértékű feladat. (A kézírásos szövegnek az írások különbözősége mellett az igazi nehézséget az jelenti, hogy az egyes szavak közti üres jel nincs mindenhol bejelölve.)

A *Carnegie-Mellon Egyetem* több projektje is foglalkozik a beszédtechnikával. (<http://www.speech.cs.cmu.edu/>, <http://cmusphinx.sourceforge.net/html/cmuspinx.php>).

A beszédfelismerés módszertana mára már annyira érett, hogy nemsokára bizton számíthatunk a gyakorlatban is használható beszédfelismerők megjelenésére. Ezek a gépbe mondott szöveget írott alakra hozzák, ami szintén sok esetben hasznos lehet. Egy értekezleten például a gép vezetheti a jegyzőkönyvet, de diktálhatunk a gépi titkárnök leveleket és más szövegeket is.

Összehasonlításként talán nem árt áttekinteni a kereskedelmi termékek piacát sem.

Magyarul összesen egyetlen olyan program létezik, amely beszédfelismerésre és gépi szövegbevitelre képes. Ez a *Philips SpeechMagic* nevű programja, mely azonban ármegjelölés nélkül szerepel a magyar weben. Az idegen nyelvű változat 850 dollárba kerül.

Angol és német nyelven az *IBM* illetve a *Scansoft Dragon* nevű programja versenyez egymással, az *IBM* terméke a *ViaVoice* körülbelül 160 dollárba kerül, a *Dragon* pedig kiépítéstől függően 90 és 850 dollár közötti áron kapható. A fölismerési pontosság cégek szerint ma még csak 95%-os, ami nyilvánvalóan komoly utólagos feldolgozást tesz szükségessé. A pontosság a program betanítása után általában javul, de ezt minden felhasználónak külön-külön kell elvégeznie.

Nyelvi megfontolások, algoritmusok

A magyar nyelv esetén a ragozott alakok fölismerése nyilvánvalóan nem triviális, de mint a magyar helyesírásellenőrzők minősége mutatja, egyáltalán nem lehetetlen feladat. Szerencsére az egyértelmű kiejtés a szükséges szótár méretét nagyon lecsökkenti (különleges kiejtést kívánó nevek mint Széchenyi, Desewffy, vagy idegen szavak mint pszichológia tartoznak a magyar szótárba.). Az angol nyelv esetén a ragozott szavak problematikája ugyan minimális, a szótárnak viszont a teljes angol szókészletet fel kell ölelnie a kiejtés szóspecifikus volta miatt.

Az összehasonlításra az elterjedt módszer a *HMM (Hidden Markov Model)*, mely alapjában statisztikai módszer. Előnye a hullámösszehasonlítási módszerekkel szemben a gyorsaság. A szavak határának megállapítására a szintén statisztikai *Viterbi* algoritmus terjedt el, de elképzelhetőek természetesen más módszerek is. A szintetizálásra az egyszerűbb *LPC (Linear predicting code)* vagy a komplikáltabb *PSOLA (Pitch-Synchronous-OverLap-Add)* módszerek a legelterjedtebbek. A szintetizálási módszerek iránt érdeklődők jó bevezetőt találhatnak a <http://tcts.fpms.ac.be/synthesis/introts.html> címen.

Korlátozott felismerés illetve szövegszintézis

Olyan területeken, ahol kevés szó felismerése elegendő, már próbaképpen bevezették a gépi szövegfelismerést. Ilyen például a vonat- vagy repülőjegy telefonos megrendelése, vagy a menetrendi tájékoztatók. Egyes bankok a korlátozott szövegfelolvasást használják például a folyószámla állásának lekérdezésére. Ilyenkor a felhasználó a telefon billentyűin keresztül adja be a folyószámlaszámot, a gép pedig a fölolvassa a folyószámla állását.

Klein Eleonóra