

A beszéd számítógépes feldolgozása Magyarországon


A Neumann János Számítógép-tudományi Társaság Informatikatörténeti Fóruma (NJSZT iTF) és az Óbudai Egyetem (ÓE) a magyarországi informatika fejlődésében meghatározó szerepet játszó Nagy Számítástechnikai Műhelyek sorozatában bemutatta a beszéd számítógépes feldolgozásának hazai történetét. A szervezők egy kiállításon megmutatták és megszólaltatták hajdan fejlesztett készülékeiket és alkalmazásaikat, valamint Kempelen Farkas beszélő gépének egy működő másolatát.


A rendezvény időpontja: 2018. szeptember 28. (péntek) 14:00-18:00

Helyszín: Óbudai Egyetem (Budapest III. Bécsi út 96/B) F06 terem.

A program:

Tick József: *Megnyitó* ▶

Takács György: *Kempelentől Gordos Gézáig, Posta Kísérleti Intézet, A Digitális beszédfeldolgozás c. könyv megszületése* ▶ 


Olaszy Gábor, Kiss Gábor: *A HUNGAROVOX szöveg-beszéd átalakító PDP-11/34 számítógépen (MTA Nyelvtudományi Intézet 1980)* ▶ 


Arató András, Vaspöri Teréz: *A Brailab beszélő számítógép család fejlesztése vakok segítésére (MTA KFKI 1982-1990)* ▶ 

Király József: *A PC-talker beszéd szintetizátor és digitális hangrögzítő-visszajátszó rendszer (1989)* ▶




Vicsi Klára: *A gépi beszéd felismerés kezdeti kutatásai (MTA Békéssy György laboratórium 1985)* ▶ 

Takács György, Olaszy Gábor: *A MEA 8000 formánsszintetizátor chip alkalmazása (BME, Scriptovox, Multivox, és a PCrobot PCF8200-zal 1985)* ▶ 

Olaszi Péter: *A hullámforma összefűzés bevezetése a beszéd szintézisben, Profivox TTS (BME TTT 1995-től)* ▶ 

Mihajlik Péter: *A folyamatos beszéd gépi felismerésének kísérleti fejlesztései (BME TTT 1990-től)* ▶ 

Németh Géza, Eisler Péter, Szabó András, Fegyó

Tibor: *HUNGAROVOX és RUSSZON szintetizátor (BEAG 1985); számváltás bemondó automata (Hungarocom 1991-től); gépi beszéd felismerés (BME TMIT, AITIA, SPEECHTEX)* ▶   



Nagy Számítástechnikai Műhelyek



A beszéd számítógépes feldolgozása Magyarországon

2018. szeptember 28.

Takács György

**Kempelentől Gordos Gézáig, Posta Kísérleti Intézet
A Digitális Beszédfeldolgozás c. könyv megszületése**

Már a régi görögök is....

1791

Kempelen beszédkeltő gépe



REKONSTRUÁLT, MŰKÖDŐ VÁLTOZAT

(2001 BUDAPEST)

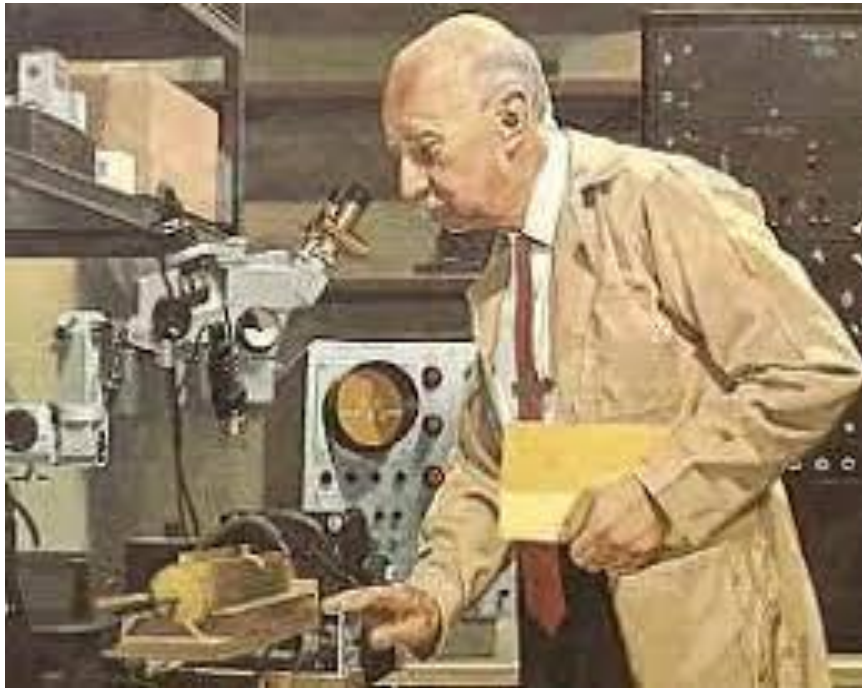
1910

Gáti Béla **mérnök** elektronikusan láthatóvá teszi a magyar magánhangzók rezgésképét

Gombocz Zoltán **nyelvész** megméri a magyar beszédhangok időtartamait

1940

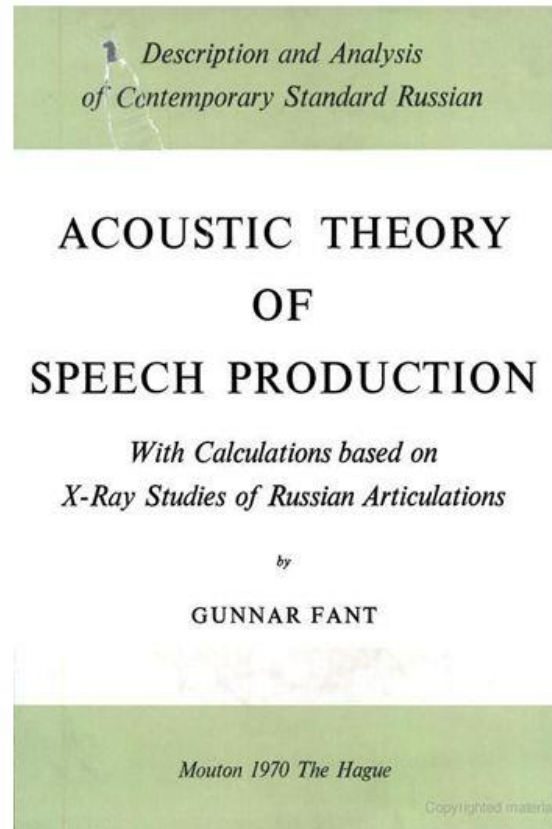
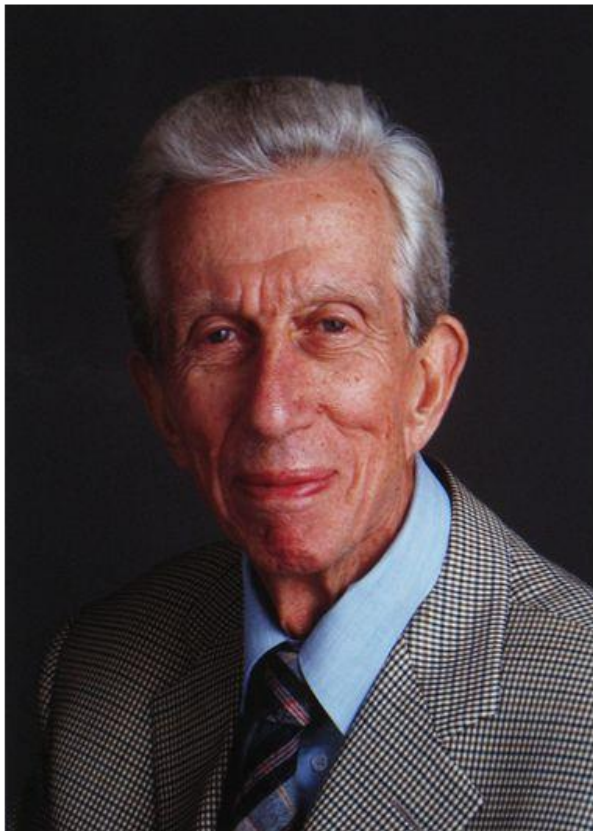
Tarnóczy Tamás **fizikus** kiszámítja a magyar magánhangzók formáns frekvenciáit



ITF, 2018. szeptember 28, A beszéd
számítógépes feldolgozása Magyarországon

Gunnar Fant -- KTH Stockholm

Source-Filter Theory of Voice Production



The first edition of Gunnar Fant's book was published in 1960.

OVE

- One of the pioneers of the development of speech synthesis in Sweden was Gunnar Fant.
- During the 1950s he was responsible for the development of the first Swedish speech synthesis OVE (Orator Verbis Electricus.)
- By that time it was only Walter Lawrence's Parametric Artificial Talker (PAT) that could compete with OVE in speech quality.
- OVE and PAT were text-to-speech systems using Formant (parametric) synthesis.

Quantization of formant coded synthetic speech

Fant, G. and Mártony, J. STL QPSR 1961. pp. 16-18

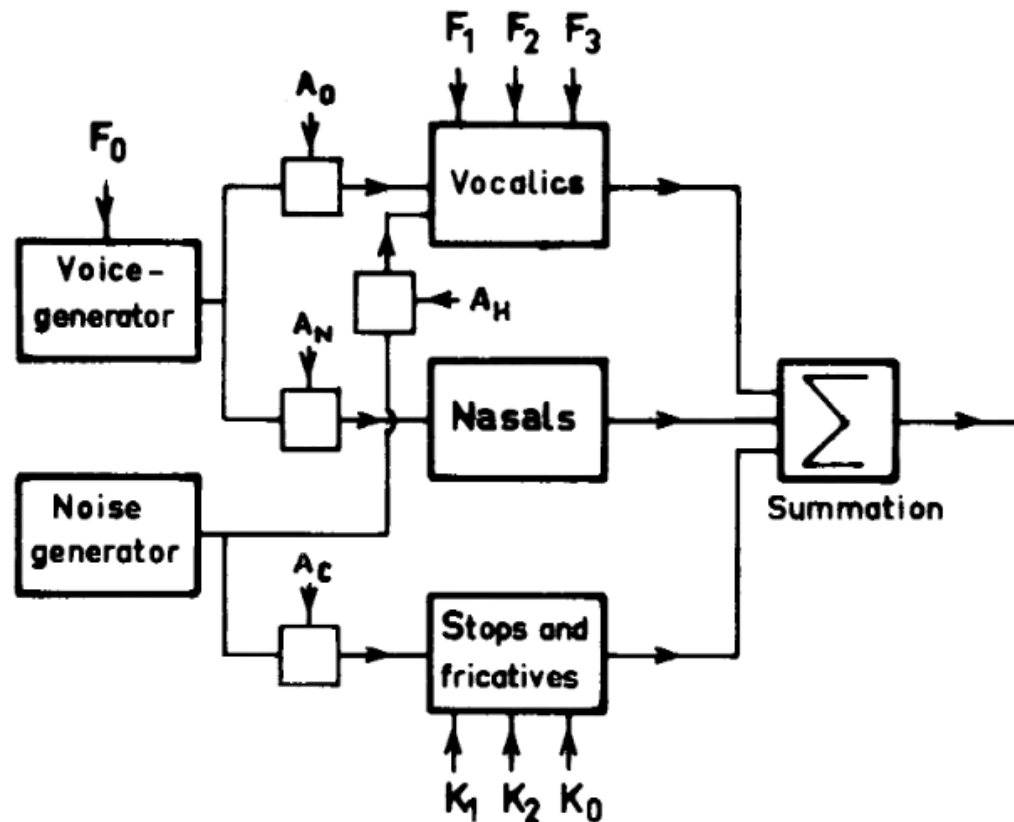
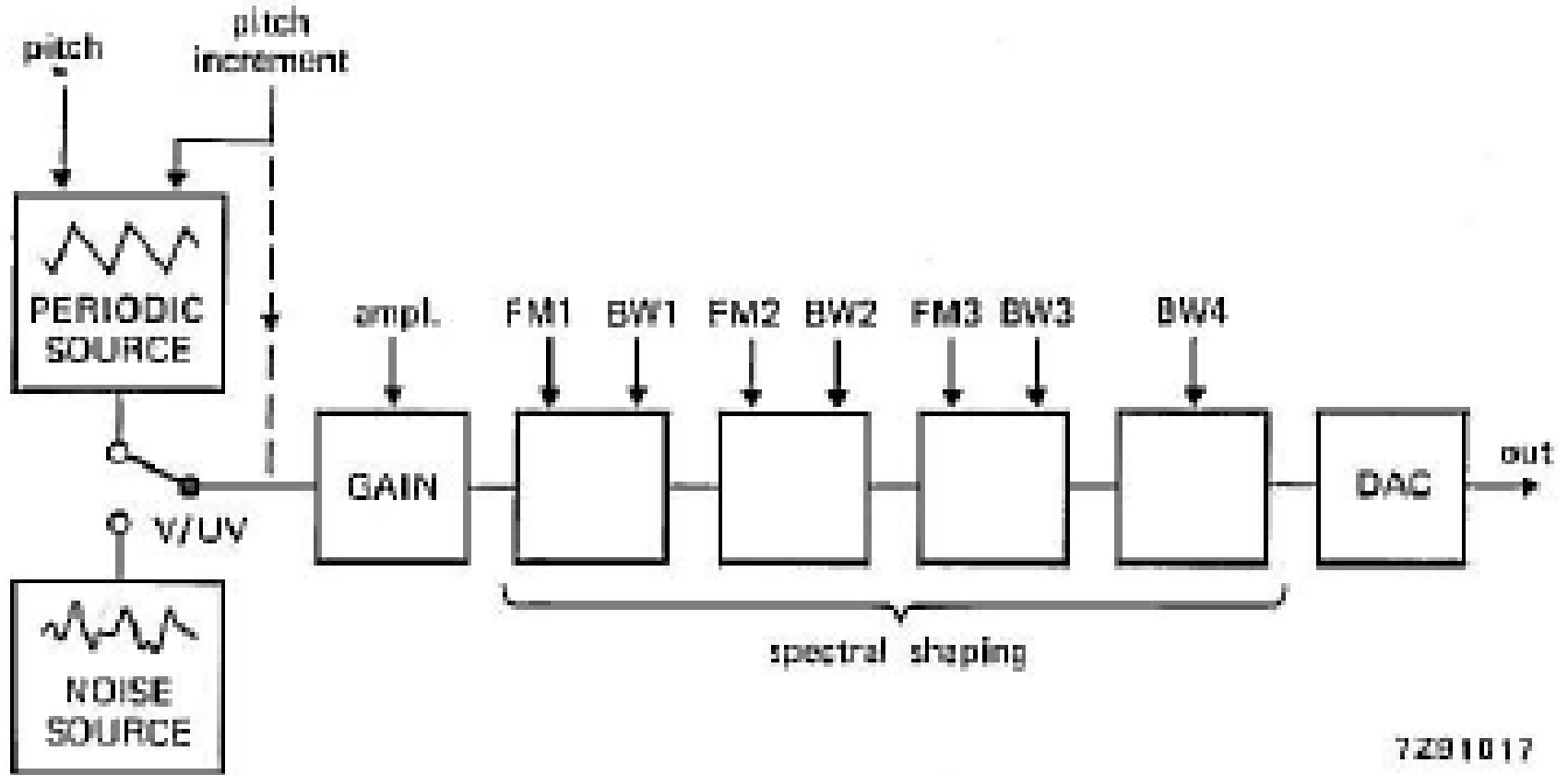


Fig. III-1 Block diagram of the speech synthesizer OVE II.

Philips MEA 8000 voice synthesizer



7291017

Posta Kísérleti Intézet

- PDP 8e mérésvezérlő automata megérkezett 1973-ban (számítógép) 8 kszó operatív memória DOS operációs rendszer, mágneslemez háttértár
- Eredeti analóg bemenettel, saját fejlesztésű analóg kimenettel rendelkezett.
- Alkalmas volt beszédelemzésre és beszédjel előállítására.....

Beszédérthetőség új mérési módszere

- A hallott logatomok kiválasztása egy készletből
- A válaszeredmények rögzítése lyukszalagon
- Statisztikák, konfúziós mátrixok készültek számítógépes feldolgozással (1972-74)



Sok panasz volt a telefon ellátottsággal, minőséggel kapcsolatban 1970-1990 között!

- Elavult rendszerek
- Szűkös beruházási források
- Tűzoltás jellegű és erőltetett fejlesztések
- Politikai eredetű kényszerpályák,
beavatkozások
- Várólisták, panaszok, kabarétémák....

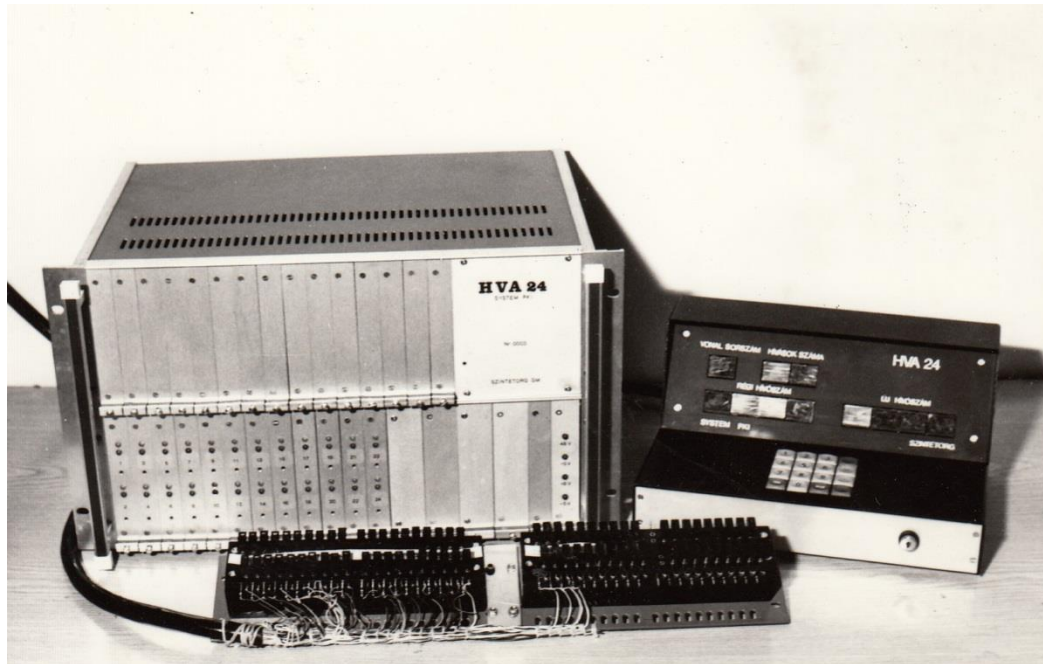


Tájékoztató hangbemondások

- Új központok üzembehelyezésekor
- Rendkívüli helyzetek kezelésére
- Meddő forgalom csökkentésére
- Tömeges téves hívások elkerülésére
- A meglévő hangbemondó eszközök kiváltására

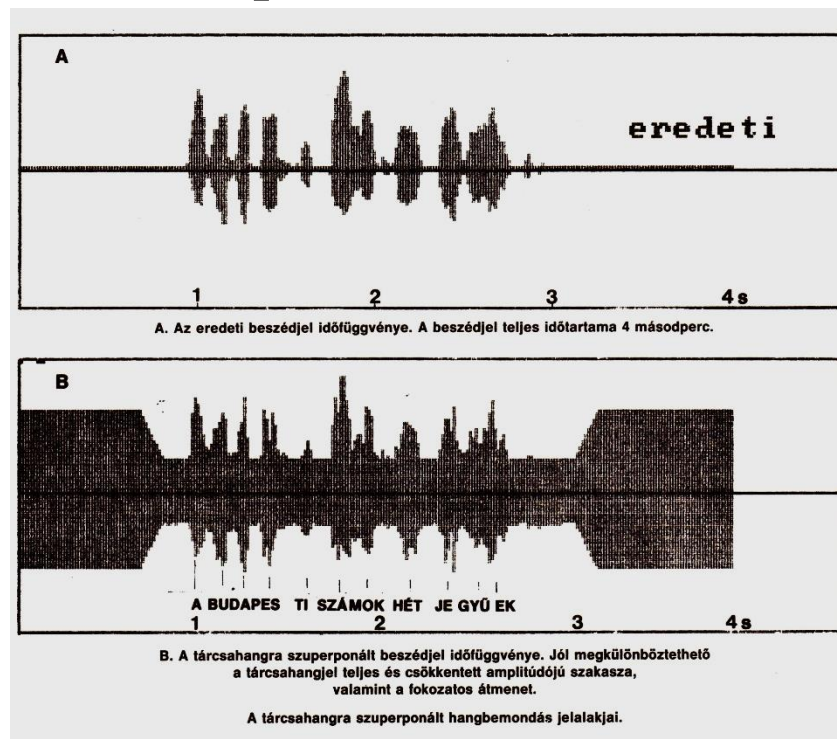
Hívószámváltások bemondására fejlesztett berendezés

- Először számítógépes szimuláció (1982)
- Üzemi próbára alkalmas mintaberendezés Z80 alapú mikroprocesszoros egységekkel, saját fejlesztésű szintetizátorral és vonali illesztőkkel, üzemi próbák (1983-1985)
- Saját fejlesztésű mikrogép egységek, gyártás (1985-90)



Budapesti hétszámjegyes rendszer átállás támogatása

- Tárcsahang generátor a bűgő hangra szuperponált figyelmeztető bemondással
- Élő rendszeren kipróbálva és üzembe helyezve mindenféle korú és működésű központnál



További beszédes szolgáltatások, tájékoztatások

- SIT hang nem adott tájékoztatást a felhasználóknak, hogy mit érdemes tenniük
- Automata ébresztő hangbemondása
- Digitális központokban szokásos bemondások átültetése analóg rendszerekre
- Totó-Lottó automata

A saját beszédinformatikai megoldások tanulságai

- Sikerült megoldásokat találni és alkalmazni hagyományos (7A1, 7A2, AR) központrendszereknél. Olyan többletszolgáltatásokkal bővültek, amelyek a következő központgenerációk elemei.
- Szabadalmakon alapuló, saját fejlesztésű műszaki megoldások és berendezések születtek.
- Hosszas üzemi próbák igazolása kellett.
- A digitális központok természetes részei ezek a beszédalapú intelligens szolgáltatás-elemek.

Gordos Géza szerepe

- Sokunknak tanára és példaképe volt.
- Sokunkat bevont oktató és kutató közösségébe.
- Béketeremtő és összefogásra ösztönző személy volt
- **HALLGASSUK ŐT A KEZDETEKRŐL (VIDEO 1,5 PERC)**

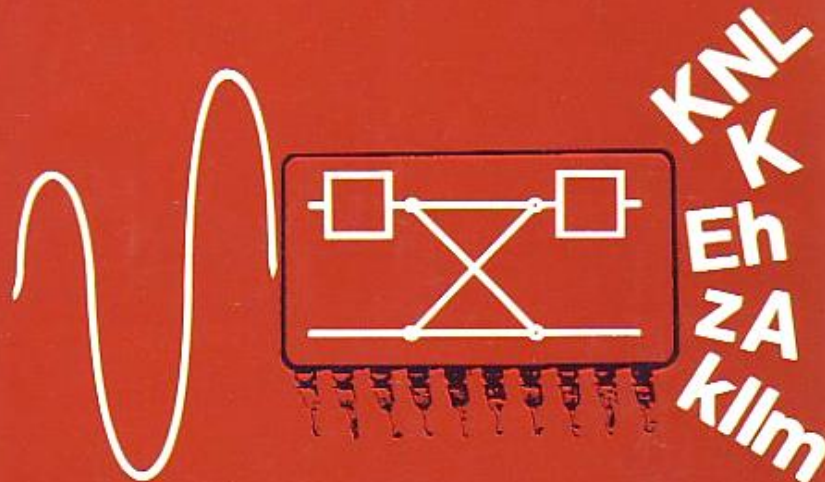
Korai publikációk

- Távbeszélő szolgáltatások fejlesztése szintetikus beszéd és beszédfelismerés alkalmazásával -- Posta Kísérleti Intézet Közleményei 1980. (Flanagan, Rabiner, Fant nyomán)
- Recognition experiments of numbers -- Speech Symposium Budapest 1980.
- Előfizetői szolgáltatások szintetikus beszéddel -- PKI Tudományos Napok 1982

- Korai publikációm és eredményeim alapján Dr. Lajtha György kezdeményezte egy összefoglaló könyv írását a Műszaki Könyvkiadó gondozásában.
- A tematika tervezetet megmutattam Gordos Gézának.
- Ő nem javítgatta, hanem kijelentette, hogy mi ketten együtt sokkal jobbat tudnánk megírni – igaza lett!
- Egy évig írtuk, minden héten egy egész napot és egy egész éjszakát együtt dolgoztunk.

GORDOS G.
TAKÁCS GY.

DIGITÁLIS BESZÉD FELDOLGOZÁS



cript:

ITF, 2018. szeptember 28, A beszéd
számítógépes feldolgozása Magyarországon

Köszönöm a figyelmet!



Nagy Számítástechnikai Műhelyek
A beszéd számítógépes feldolgozása
Magyarországon
2018. szeptember 28.

Olaszy Gábor - Kiss Gábor - Nikléczy Péter

**A „HUNGAROVOX” szöveg - beszéd átalakító PDP–11/34
számítógépen
(MTA Nyelvtudományi Intézet, 1980)**

LÉPJÜNK VISSZA AZ IDŐBEN 41 ÉVET

1977

A nemzetközi helyzet a gépi beszédelőállítás terén:

- USA MIT angol TTS fejlesztése (formáns szintetizátor)
- Svédország Royal Institute of Technology (KTH) OVE nevű formáns kódolású beszéd szintetizátor Svéd és angol TTS beszélő program.

-Magyarország: MTA Nyelvtudományi Intézet Fonetikai Laboratórium.

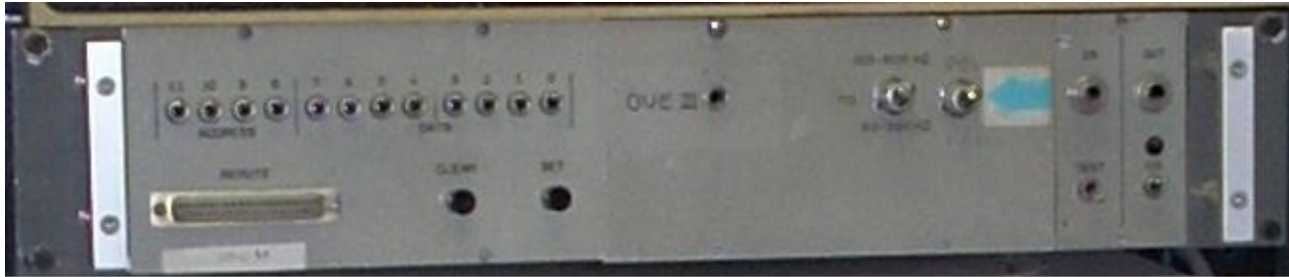
Bolla Kálmán osztályvezető kezdeményezésére felkészülünk egy új kutatási formára:

fonetikai alapú analízis gépi szintézissel

Hardver beszerzés

1977

OVE III svéd gyártmányú programozható formáns szintetizátor beszerzése gépi beszédes kísérletekhez
(Bolla Kálmán kezdeményezésére)



12 db kétállású, manuális kapcsoló= 12 bit

Szintetizált magyar magánhangzókat generálunk
kézi vezérléssel

Publikációk születnek a varázslatos kutatásról

A hardver beszerzés folytatása 1979

**MEGJÖTT AZ ÉTVÁGYUNK.
GÉPPEL AKARTUK VEZÉRELNİ .**

DEC PDP 11/34 számítógépet
vásároltunk (nem volt könnyű).

-Memória 64 kbyte

-2 db nagy floppy meghajtó

-Képernyős konzol

-OVE III összekötése a géppel

-A stáb: Olaszy G (36). Villamosmérnök
Nikléczy Péter (32) elektrotechnikus
Kiss Gábor (24) programozó matema-
tikus



A gépi beszédelőállítást céloztuk meg 1979

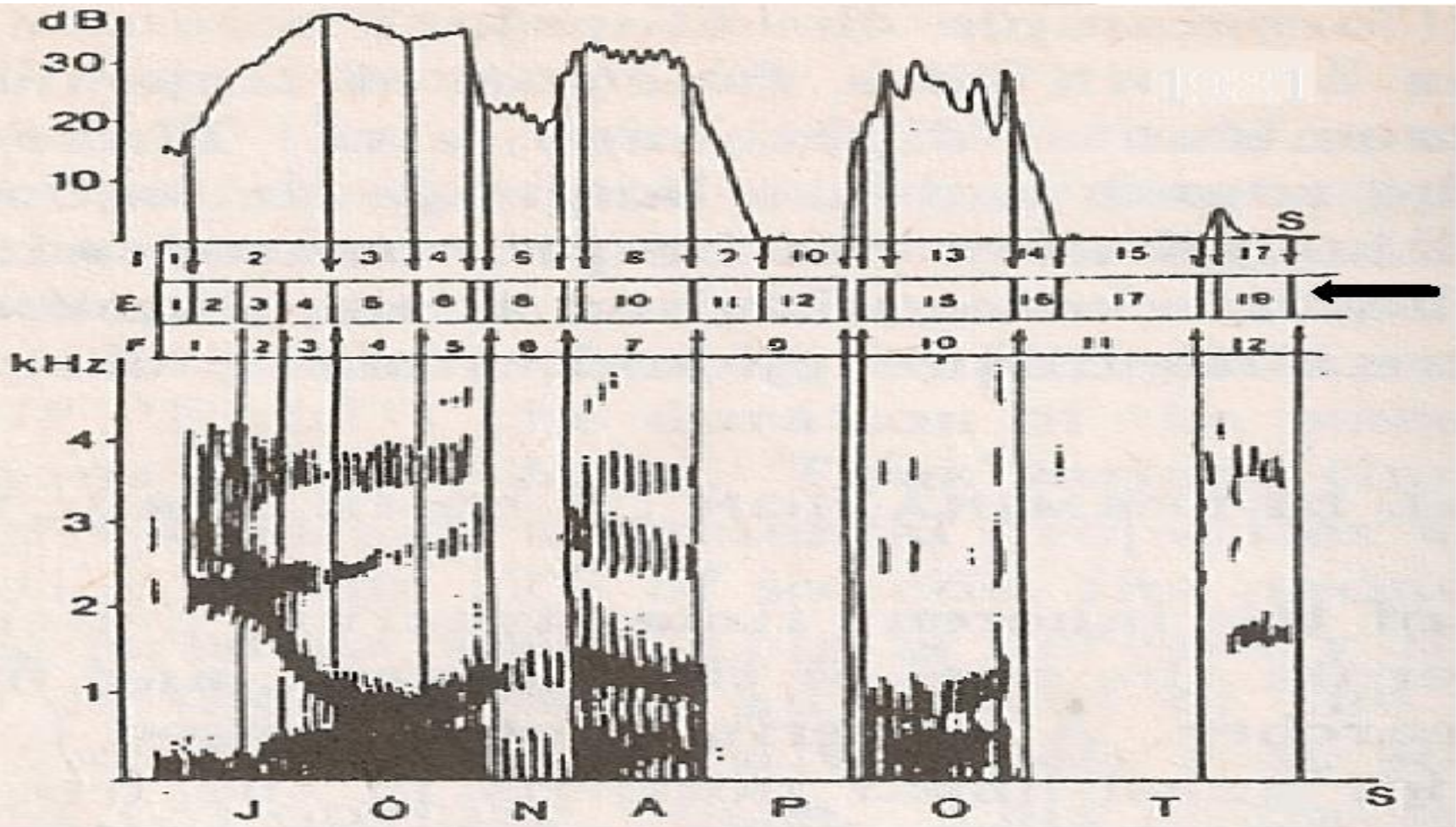
**FEJEST UGROTTUNK A TENGER KÖZEPÉBE
ÚSZNI TUDÁS NÉLKÜL.**

CSAK SAJÁT KREATIVITÁSUNKRA VOLTUNK UTALVA.

**FEJEST UGROTTUNK A TENGER KÖZEPÉBE
ÚSZNI TUDÁS NÉLKÜL.**

CSAK SAJÁT KREATIVITÁSUNKRA VOLTUNK UTALVA.

AZ ALAPVETŐ KÉRDÉSEK A KÖVETKEZŐK VOLTAK: MILYEN LEGYEN A VEZÉRLÉSI FOLYAMAT IDŐ STRUKTÚRÁJA?



HONNAN SZEDJÜNK ADATOKAT? HÁNYAT? HÁNY FÉLÉT?

PROGRAM TERVEZÉS ÉS PROGRAMOZÁS

Kiss Gábor programozó matematikus készítette
FORTRAN nyelven
az INBERE és HUNGAROVOK szoftvert

```
.RUN DX1:FOPR25

~IMI <90>

~BEA E'1(TM:<30>,AD:<3-9>,FO:<96-98>)

~BEA E'1(F1:<546>,F2:<2100>,F3:<2435>)

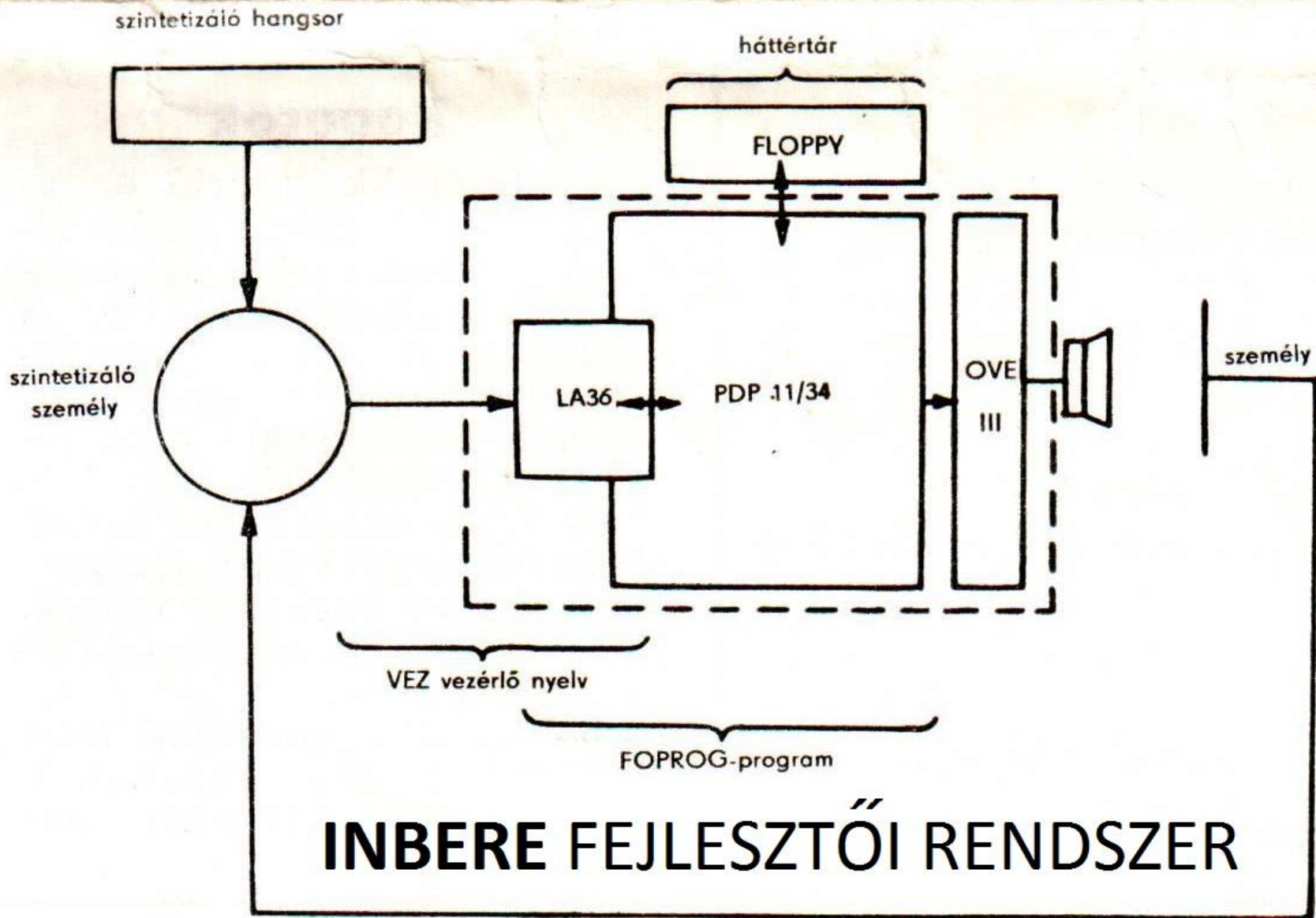
~BEA E'2(TM:<120-130>,F1:<546-564>)
**HIBA!** AZ IDO MEGADASA HELYTELEN.

~BEA E'2(TM:<230>,AD:<9-13>,FO:<98-154>)

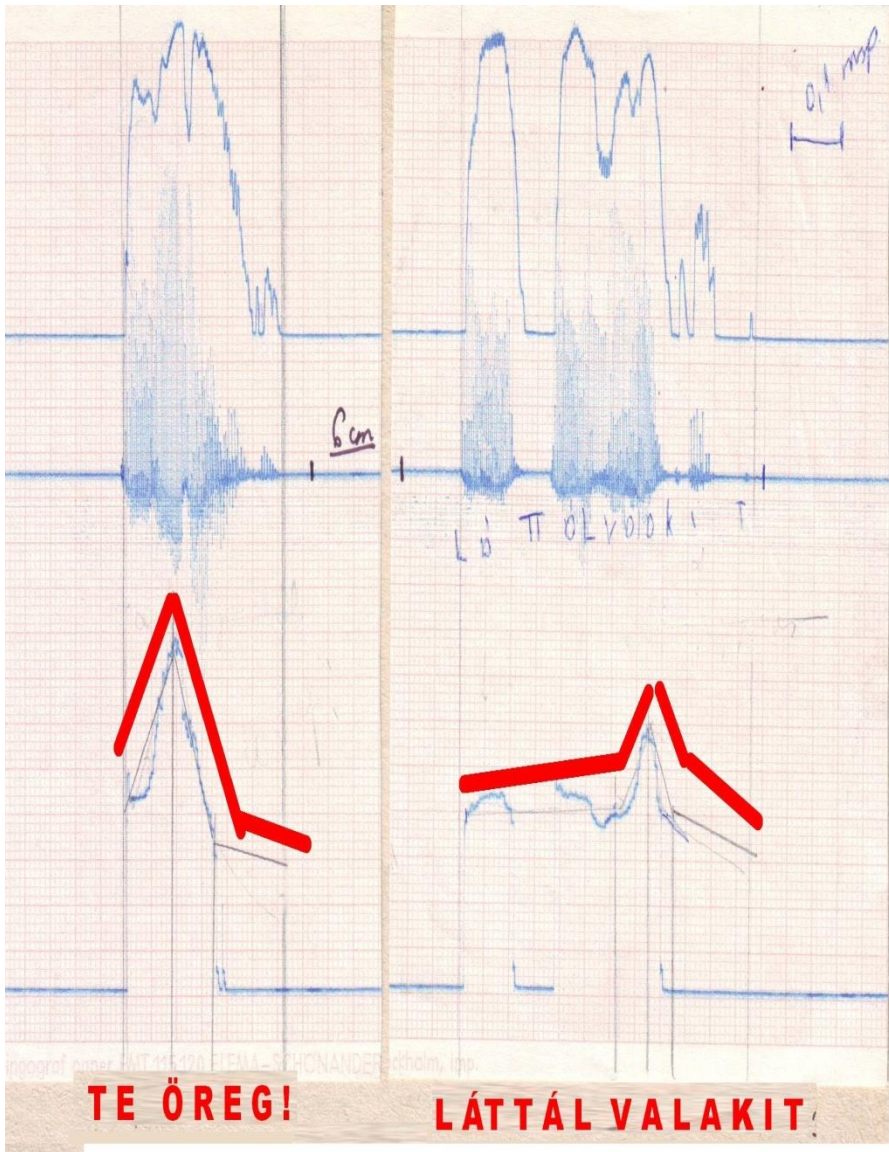
~BEA E'2(F1:<546-576>,F2:<2150>,F3:<2435>)

~JON (E'S,S1,S2)

~ETI (E'1,E'2,E'S,S1,S2)#ES
```

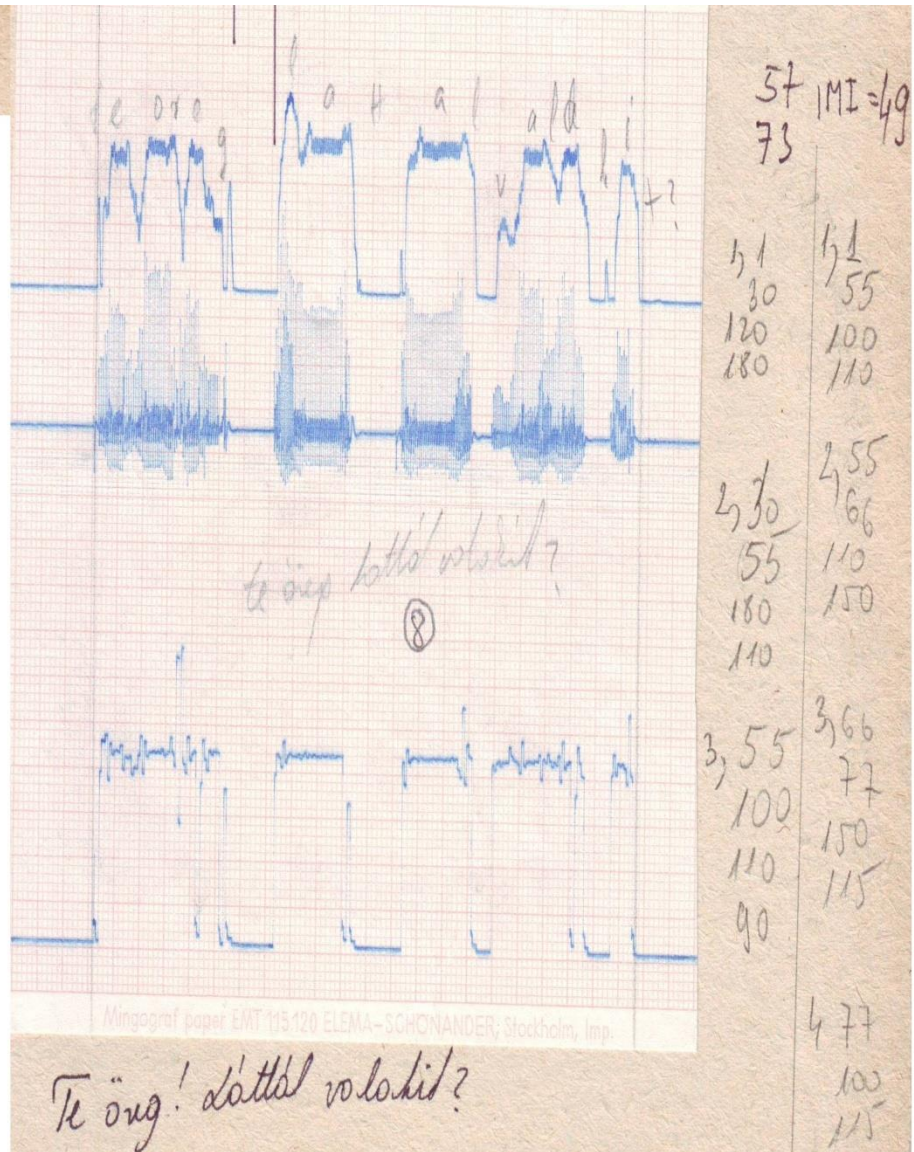


ANALIZÁLTUNK, SZINTETIZÁLTUNK, ANALIZÁLTUNK...



TE ÖREG!

LÁTTÁL VALAKIT!



MEGOLDOTTUK!

1980-ra megszólalt a magyarul beszélő gép.

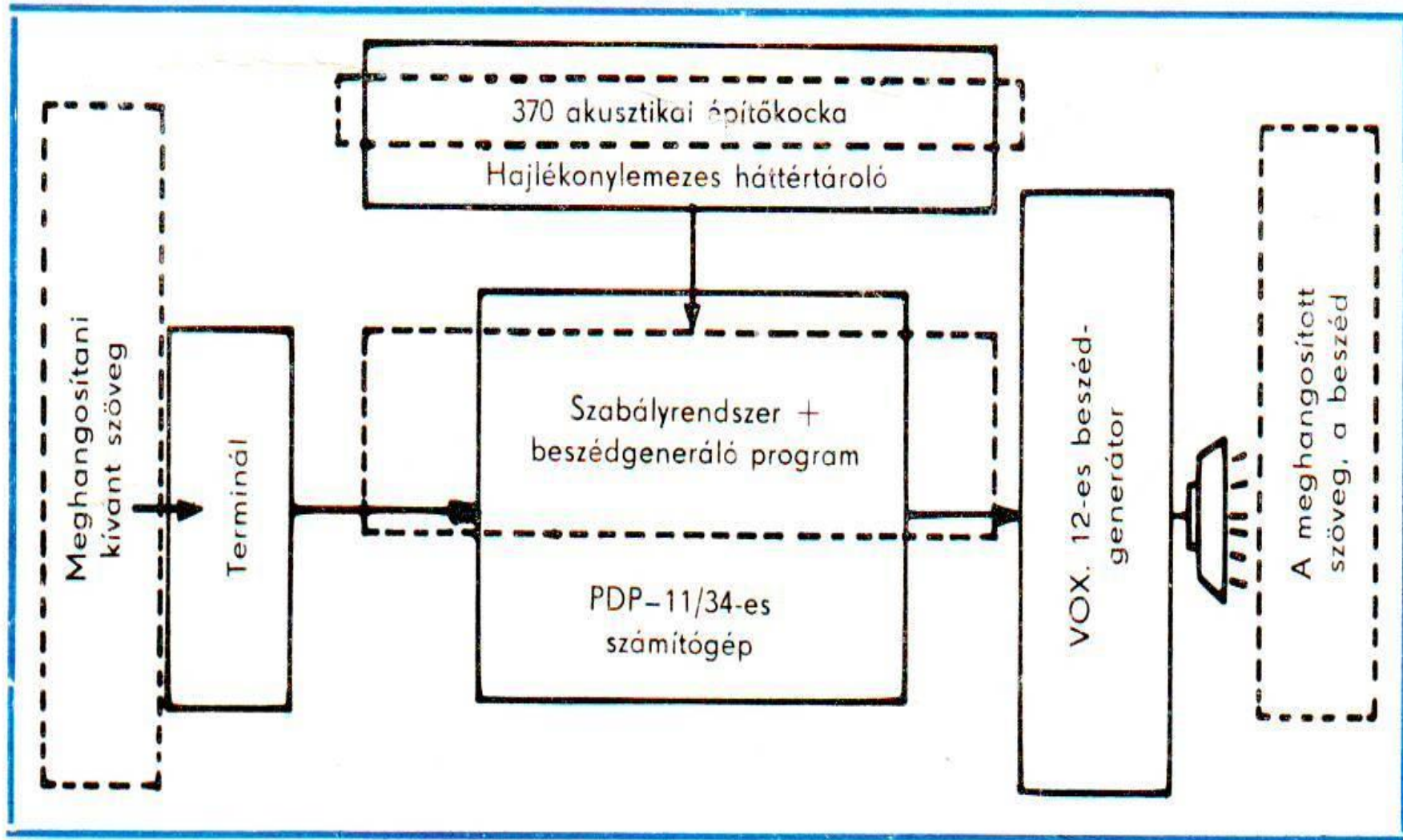
370 db hangszeletből bármilyen magyar tartalmú beszédet tudtunk szintetizálni ASCII szövegbemenet alapján.

Állítható volt a szeletek időtartama és a hangszetelek között automatikus interpoláció működött.

A folyamatos magyar beszéd akusztikai szerkezetét feltártuk a hanglejtéssel együtt.

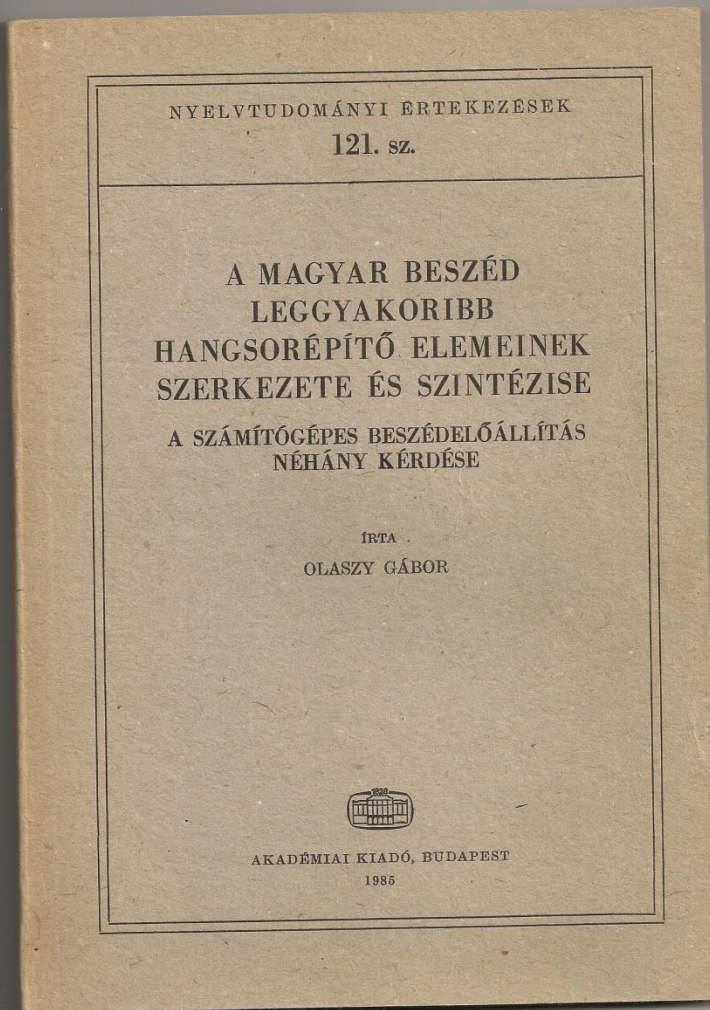
MEGÉPÍTETTÜK A VOX-12 BESZÉDSZINTTETIZÁTORT



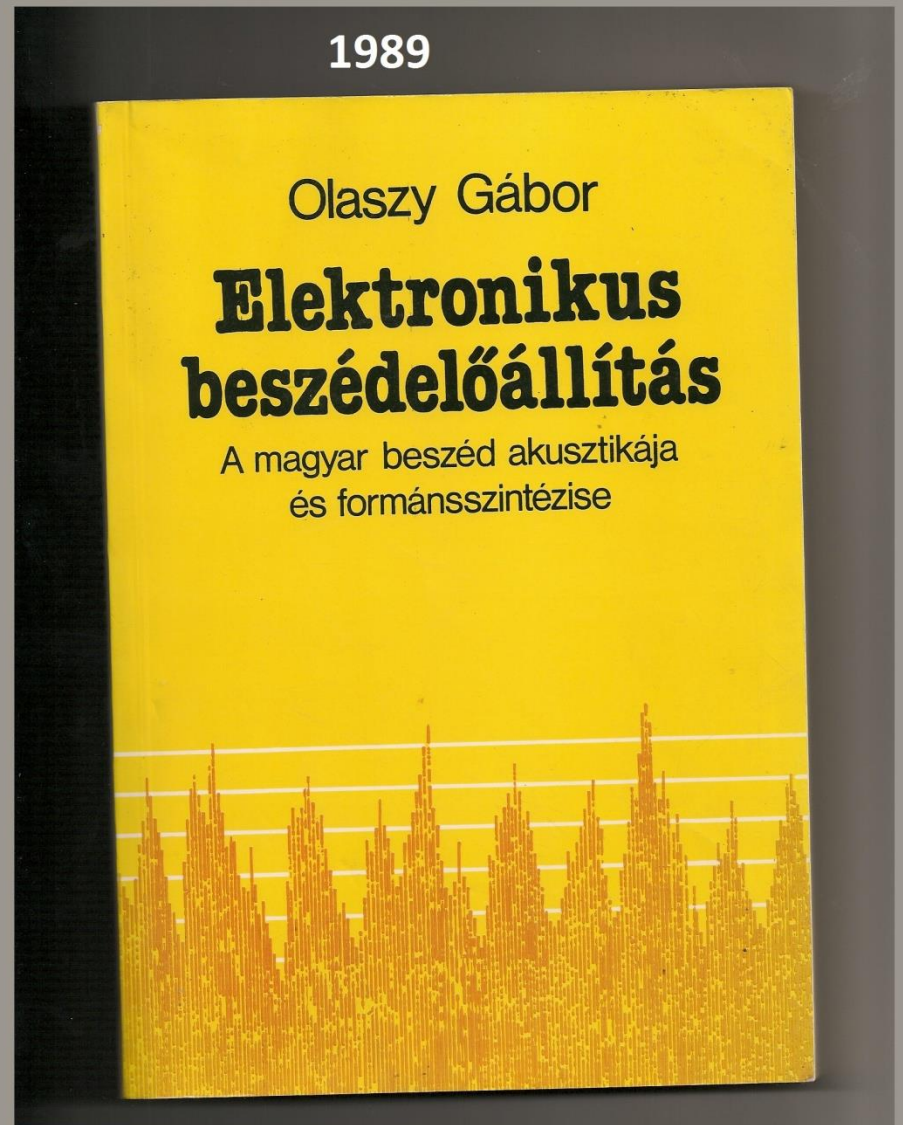


A HUNGAROVOX felépítése

1985



1989



SZOLGÁLATI SZABADALOM!

MTA Nyelvtudományi Intézet

Olaszy Gábor, Kiss Gábor, Nikléczy Péter

Eljárás és berendezés szintetizátor(ok) vezérlésére, szótár nélküli szintetizált beszéd a vezérléssel quasi azonos időben történő előállítására

Lajstromszám: 185527

Benyújtás éve: 1982.

Közzététel éve: 1982

Benyújtás helye: Magyarország

**A Budapesti Elektroakusztikai Gyár
1984-ben megvásárolta a szabadalmat**

1983 Budapesti Nemzetközi Vásár MTA PAVILLON

A HUNGAROVOX nyilvános bemutatója

NAGY SIKER



NAGY FELTŰNÉS

1 héten át beszélt.....beszélt.....beszélt.....

ITF, 2018. szeptember 28, A beszéd
számítógépes feldolgozása Magyarországon

NAGY SAJTÓ VISSZHANG

NÉPSZABADSÁG

1983. június 22., szerda

Beszélni kezd a gép

A hangszórókból még furcsán hangzanak a magyar szavak, visszhangoznak a laboratórium három kis szobájában, amelyeket zsúfolásig megtöltöttek a magnetofonok, a képernyők, s végül a számítógép. Ha szűk is a hely, az adatok tárolásához nem kellenek paloták, és sokaság sem: mindössze négy ember több éves munkájának eredményeként készült el a közelmúltban az első hazai beszélő számítógép.

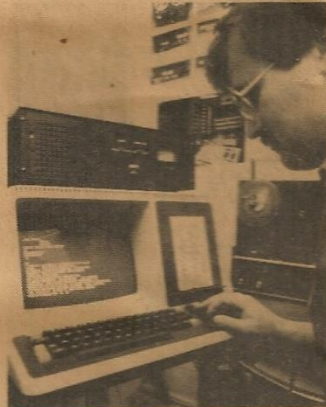
A hang rajza

A laboratórium a Magyar Tudományos Akadémia Nyelvtudományi Intézete fonetikai osztályának részlege. 1970-től folytak itt beszédvizsgálati kutatások. Dr. Bolla Kálmán osztályvezető irányításával Kiss Gábor programozó matematikus, Nikléczy Péter technikus és Olaszky Gábor villamosmérnök a hangok és a beszéd akusztikai szerkezetét kutatja.

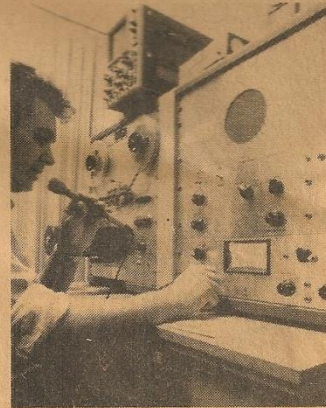
Fő feladatunk a hangok intenzitásának, frekvenciájának, idejének elemzése. E tulajdonságok jellemzőnek leginkább egy hangot — mondja Kiss Gábor. — Méréseinkkel segítjük a nyelvészek munkáját.

S máris mutatja a gépeket. Mindegyiket a hangspektográfot, amely egyike a Magyarországon található két ilyen berendezésnek, s amely vizuális képet ad a beszédről vagy akár csak egyetlen hangjáról is bármely szónak. A laikus számára talán nem is érdekes maga a működés elve, inkább csak a végeredmény meglepő: a keskeny speciális papírszalagon hogyan jelenik meg rajzban a kimondott szó, vagy akár egyetlen hangja, s a nyelvész számára hogyan segítheti ez a kép a kutatásokat.

Hiszen a kutató — a hallható beszéd alapján — hipotéziseket készít, ám ezek összevetése a műszeres mé-



A képernyőre írt szöveg a „fekete doboz”, a beszédgenerátor segítségével szólal meg.



zetközi Vá-
ázási javak
eterterm a
adémia pa-
a, jól ért-
llönös — a
röteljeseb-
hang foga-
A beszé-
edje meg,
1982-ben
lttudomá-
i Labora-
NGARO-
ilyen ma-
c. Megta-
sírását is.
alakítok.
bb haso-
írásjele-
tta, hogy
őtagolva,
is meg-
keresett
játékra.
sztérium
lyi, mű-
oratóri-
ág első,
dolgozó,
n folya-
szinteti-
ír szá-
áln-
rojgép-
gra-
ite dul-
n zte-
juk,
—
trét
ént, elő-
lis; zen
so; ról
sul,
rájuk,



Nevem: HUNGAROVOX

Munkánk során végül 370 ilyen pici, elemi hangot határoztunk meg, az így nyert „készletből” tetszőleges magyar beszédet...
Ismeri a magyar helyesírást és játszani is tud...

gépnek a már viszonylag jó beszédminőségét tovább finomítuk.
— Bár a mamár hangja is magyar

A HUNGAROVOX feltalálói:
Olaszky Gábor és Kiss Gábor
FOTO: BOROS JENŐ



NAGY SAJTÓ VISSZHANG

MAGYAR SIKER

A beszélő gép I.

Az emberiség egyik régi álma teljesül napjainkban: megvalósulóban vannak a beszélő gépek. Az emberek érintkezésének, gondolatcseréjének, kommunikációjának legtermészetesebb eszköze az élő nyelv. A kívánság tehát jogos: a gépekkel is úgy érintkezhessünk, mint embertársainkkal.

A vezető tőkés államok kutató laboratóriumaiban előrehaladott állapotban tartanak a kísérletek a szóból értő komputerekkel és az élőbeszéd leegyszerűsített szavaival programozható, irányítható berendezések-

„Egyszerre öt másodperces, a magyar helyesírás szerint betáplált szöveg szólaltatható meg”



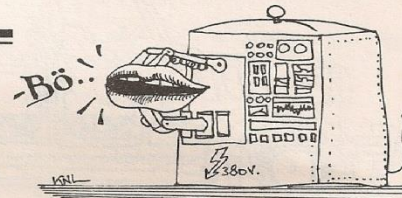
kel. A beszélő információs rendszerek, játékok, sakkozógépek, autók, chippek pedig már elkezdtek bevonulni a hétköznapi gyakorlatba.

A szóból értő és emberi hangon válaszoló rendszerek közül hazánkban az utóbbi, a beszélő számítógép kifejlesztéséért folytatnak eredményes kísérleteket többek között az MTA Nyelvtudományi Intézetének Fonetikai Laboratóriumában. A szövegszintetizáló rendszer az első valós idejű szöveg-beszéd átalakító eljárás, amelyet a magyar nyelvre kidolgoztak. Erre építve alkották meg a beszélő játékos tanítóeszköz prototípusát is. A munkát Kiss Gábor (28 éves) programozó matematikus és nyelvész, valamint Olaszy Gábor (40 éves) villamosmérnök végezte.

Üdvözljük Önt...

Rendszerüket legutóbb az Optikai, Akusztikai és Filmtechnikai Egyesületben mutatták be.

A laboratórium két kis szobában



kapott helyet a Művelődési Minisztérium épületében. Az asztalokon, polcokon körös-körül műszerek, integrált áramkörös modulok, oszcilloszkópok, stúdió magnetofonok, forrasztópákák, vezetékek, elektronikus alkatrészek, számokkal, vázlatokkal, diagramokkal teleírt papírlapok. Mindenből a legkorszerűbb és legszükségesebb. Olaszy Gábor ujjai könnyedén futnak a klaviatúrán. A képernyőn zöld betűk izzanak fel: Szeretettel üdvözljük önt laboratóriumunkban! A felirat keletkezésével szinte egyidőben kissé fémes csengetésű hang kíséri az írásos köszöntést, a gép kimondja a szöveget. Elismerően bólintok, és megkérem a szakembert, ismeresse a HUNGAROVOX szövegszintetizáló rendszer főbb technikai adatait.

– A beszédelőállítás digitális vezérlésű beszédgenerátorral (közismerten – szintetizátorral) történik. Ezt bármilyen 12 bittel rendelkező számítógép vezérelheti. Egy másodperces beszédhez 450 byte információ tárolása szükséges. A rendszer működtethető írógéppel, konzolterminállal, vagy automatikus beszédgenerálás esetén előre beprogramozott ASCII kód-sorozattal. Működése valós idejű, azaz a beszéd a beírás vagy az ASCII kód-sorozat megadása után gyakorlatilag azonnal hallható. Egyszerre öt másodperces, a magyar helyesírás szerint betáplált szöveg szólaltatható meg. Az előírt szöveg az ismétlő gomb megnyomásával tetszőleges számban újrhallgatható. A berendezés automatikusan szerkeszti a magyar beszédnek megfelelő időszerkezetet, a tempót, a dallamformálást pedig a megadott mondatvégi írásjelek sze-

NAGY SAJTÓ VISSZHANG

alábokat nevezik
inthyog a zöngés
ő hangzását a kü-
szerkezet okozza,
zatorokban olyan
áramköröket kell
eknek a szűrőkés-
s hangzások meg-
felelően változ-

kon kívül zörejes
éldául az s, az sz,
) is vannak beszé-
előállítására zö-
nálnak. Persze itt
angformáló szűrő-
rangok is más-más
rejgenerátor jelé-
is használhatók az
nkör rezonanciakö-
(mivel ugyanazt a
áramkört használ-
ben) kevert (zörej
mekből együttesen
elvileg nem állít-
latilag azonban —
izmus tehetetlensé-
— ezeket a problé-
thidalhatják.
ormánsszintetizátor-
tsanak elő, minde-
— és a zajgenerátor
áló szűrőáramkörök
ors egymásutánban,
sséggel változtatva
ogy hatásukra a be-
nek megfelelő rez-
ki. Formánsszinte-
sére általában 10—
megadni, amelyeket
t 50—70-szer kell
a másodpercenkénti

BESZÉLŐ GÉPEK

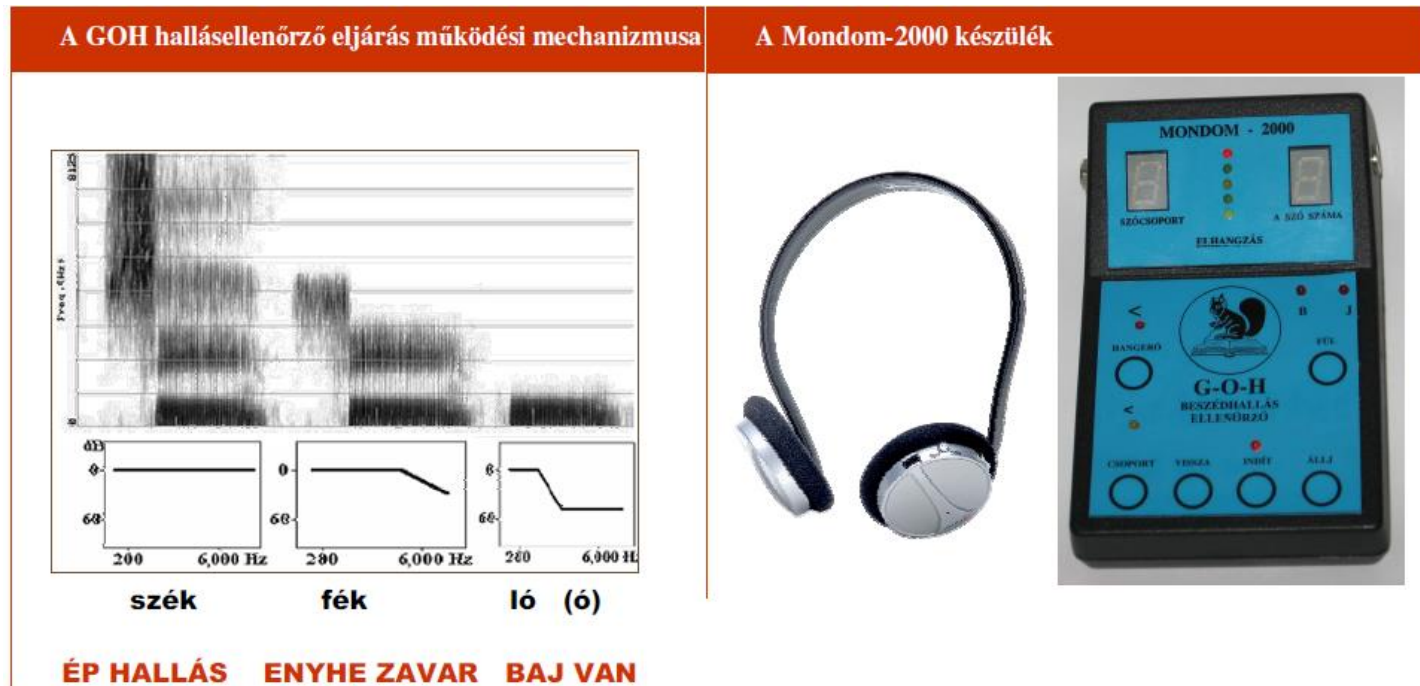
Régi vágya az embernek, hogy gépei értsék a szavát, s ugyancsak beszéddel közöljék legfontosabb információikat. Ez utóbbi óhaj napjainkra már teljesült, s ennek sikeres megvalósulásában magyar kutatóknak is számottevő szerepük van.

AZ MTA NYELVTUDOMÁNYI INTÉZETÉNEK hangspektográfja képet ad a beszédéről. A kimondott szó keskeny papírszalagon jelenik meg, természetesen rajzban



ITF, 2018. szeptember 28, A beszéd
számítógépes feldolgozása Magyarországon

A HUNGAROVOX (1982) technológia más jellegű leszármazott eredménye a **GOH beszédhallást vizsgáló eljárás**, amit a mai napig (2018) használnak. **Szabadalmaztatott.** *Lényege, hogy szintetikus beszéddel vizsgálják a hallást, főleg kisgyerekeknél.*



Köszönjük a figyelmet!

BraiLab beszélő számítógépcsalád vakoknak

„Only for me”

MTA KFKI MSzKI 1982-1991

Vaspöri Teréz, Arató András, Lukács testvérek

Sulyán János, Holtzer Lóránt, Vajda Ferenc

Zimányi Magda, Kiss Gábor

**(akik nélkül nem jött volna létre a BraiLab
számítógép család)**

Braille vagy beszéd

- Egykarakteres Braille kijelző 1978
- Paul Fortier: használjunk beszéd szintézist, mert gyors
- Ekkor még mikroprocesszoros magyar TTS nem volt
- Fejlesztő rendszer: Digitalker + 8085 SDK + 1char Braille + egykezes Braille keyboard + TPA1140 cross assembler + Text editor + RS232 kapcsolat + MEA-8000 (4 formánssal)

HomeLab -> BraiLab

- 2x4Kb-át szabad hely:
- 4Kb-át beszéd frame-eknek átlapolva
- 4Kb-át a magyar Text-to-Speech-nek egyszakaszos diádós struktúra (I. kandidátusi disszertáció)
- Beépítés az ernyőszerkesztős alap és CP/M-es HomeLab-ba
- 1985-től 400 db „kis” BraiLab készült
- 1987-től 70 db BraiLab plusz (CP/M) készült
- Német TTS kísérlet Olasz Gábor segítségével

BraiLab PC

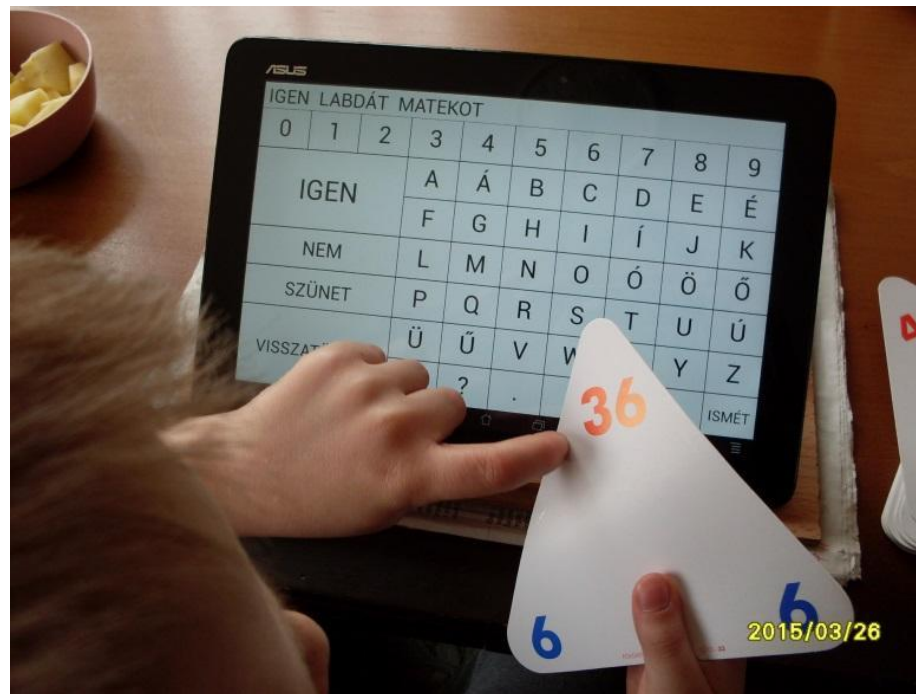
- 1990-re OTKA támogatással elkészült a BraiLab PC Philips gyári formáns szintetizátor fejlesztő rendszer segítségével PCF-8200 (5 formáns)
- 3200 különböző, összesen 3241 beszéd frame az új TTS-ben diádonként két szakaszra osztva
- Beépítés a DOS operációs rendszerbe
- 2300 magyar vak BraiLab PC felhasználó
- A BraiLab-okat eljárásra és berendezésre szóló szabadalmi oltalom védte

További fejlesztések: Mobil SegítőTárs (MOST) vakoknak

- MOST Windows Mobile-ra majd Android-ra 2004-től
- Közel 200 felhasználó a napjainkig használja
- GVOP támogatással Veszprémi Pannon Egyetem Juhász Zoltán
- Kezdetben BraiLab beszéddel
- Később más TTS-ekkel (eSpeak, Mariska, Profivox stb.)

Beszélő kommunikációs segédeszköz Autisták számára

- 2006-ban TalkPad (támogatott kommunikáció)
- Közel száz beszédképtelen felhasználónk van
- <http://rehab.wigner.mta.hu/?q=hu/node/6>



Rezgő Morse és Braille MOST változat egy siket-vak embernek

- Több mint 5 éve használja önállóan a MOST funkcióit (kivéve beszéd)
- A Morse nyelvet néhány hónap alatt tanítottuk meg neki
 - <http://rehab.wigner.mta.hu/node/32>
 - <https://youtu.be/YQsPncM3GVQ>

A nyelvek kiterjesztett értelmezése

- Identitással rendelkező emberek közössége használja kommunikációra
- Hierarchikus mentális szótár kell az agyban
- A percepció valamely érzékszervvel történik
- A reprodukció motorikus programmal történik
- A percepció és a reprodukció a tanulásban egymást segítik, és mindkettő része a mentális szótárnak
- Beszédnyelvek, siketek jelnyelvei, siket-vakok jelnyelvei, morze nyelv, Braille nyelv (írás-olvasás)

A morze nyelv tanításának kísérlete

- 20 WPM sebességgel tanultunk csak meg venni 9 emberrel. Feljegyeztük mindenkinek a hibaarányát.
- 6 WPM-el kíséreltünk meg venni alanyainkkal.
- Csak 1 ember volt képes megőrizni a hibaarányát!
Ő lejátszotta magában gyorsan!
- → A mentális szótárat fel kell tölteni a lényegesen eltérő sebességekkel
- Ezt az eredményt a siket-vak emberek tanításában felhasználtuk

Minden lényegesen eltérő sebességet külön kell tanulni!

Feltételezhetően ez igaz más nyelvi tanulásra is.

A gépi beszédfelismerés kezdeti kutatásai

Vicsi Klára

BME TMIT Beszédakusztikai Laboratórium

Első beszédfelismerő Magyarországon - 1971

Készítő: Prof Tarnóczy Tamás - MTA Akusztikai Kutatólaboratórium

Analóg erősítőkből és szűrősorból álló berendezés

Folyamatos lassú szöveg bemondásakor az **egyes magánhangzókat mutató** lámpasor elemeiből az a lámpa gyulladt ki, amelyik magánhangzó éppen elhangzott.

A berendezést a Budapesten megrendezésre került 7th ICA kongresszuson mutatták be.

MTA TTK BÉKÉSY GYÖRGY AKUSZTIKAI KUTATÓLABORATÓRIUM

(Budapest, XI., Budaörsi út 45.)

Vezető: Dr. Illényi András, PhD (1981-1995)

Munkatársak:

*Fehér köpenyben Dr. Illényi András,
Horváth Attila műszerész,
Körmöczy László technikus,
Dániel István tud.s. munk.,
Berényi Péter tud.s. munk.,
Török Kálmán műszerész,
Vicsi Klára tud.munk.,
Kosik László elektromérnök,
Illényi András mellett
egy fizikus kolléga*



Az 1980-as évek közepétől kezdve a beszéd kutatás végig kiemelt és előremutató témája volt a Kutatólaboratóriumnak, Vicsi Klára vezetésével. Központi téma a kötött szótáras gépi beszéd felismerés volt.

1985-ben elkészült az első magyar kötött szótáras beszéd felismerő rendszer.



*Vicsi Klára
tudományos munkatárs*



*Tihanyi László
matematikus, szoftverfejlesztő*

A kész termék

Személyfüggő, kötött szótáras berendezés.

Szótárkészlet:

100 szabadon választható szó vagy tőmondat.

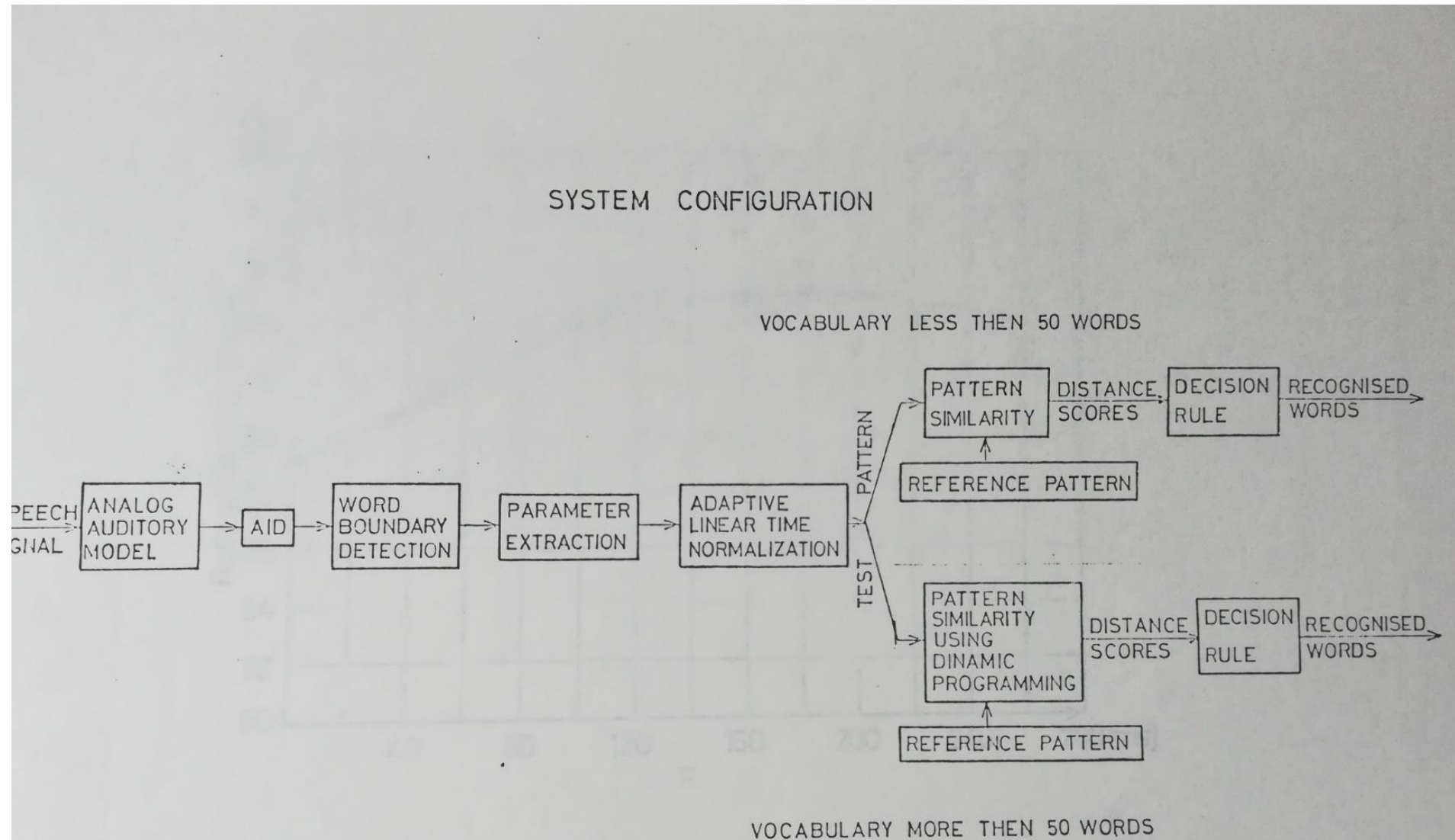
Többfelhasználós rendszer:

egyidőben a felhasználók száma 10.



A műszer a Budapesti Elektroakusztikai Gyár megrendelésére készült

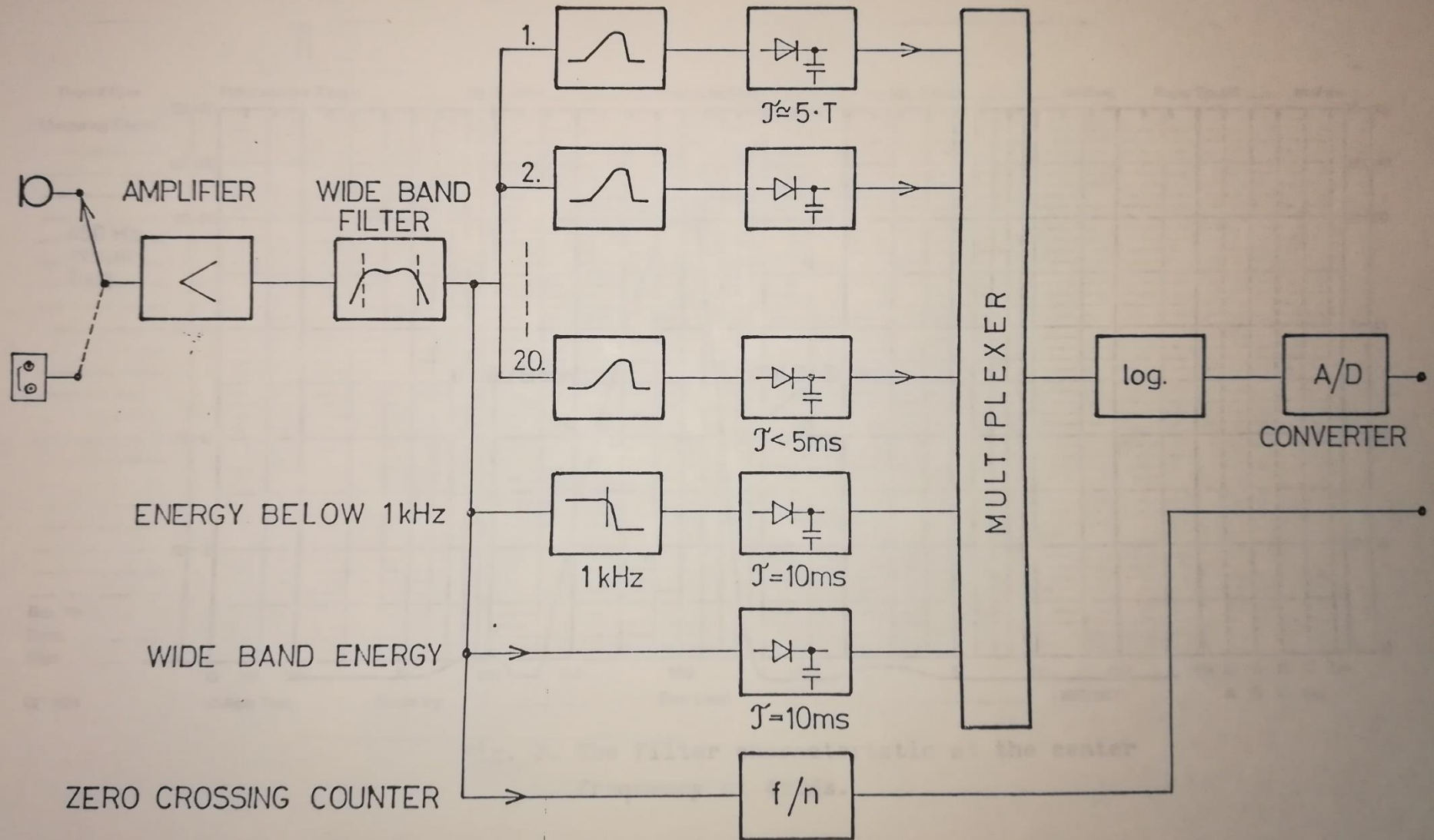
A rendszer felépítése



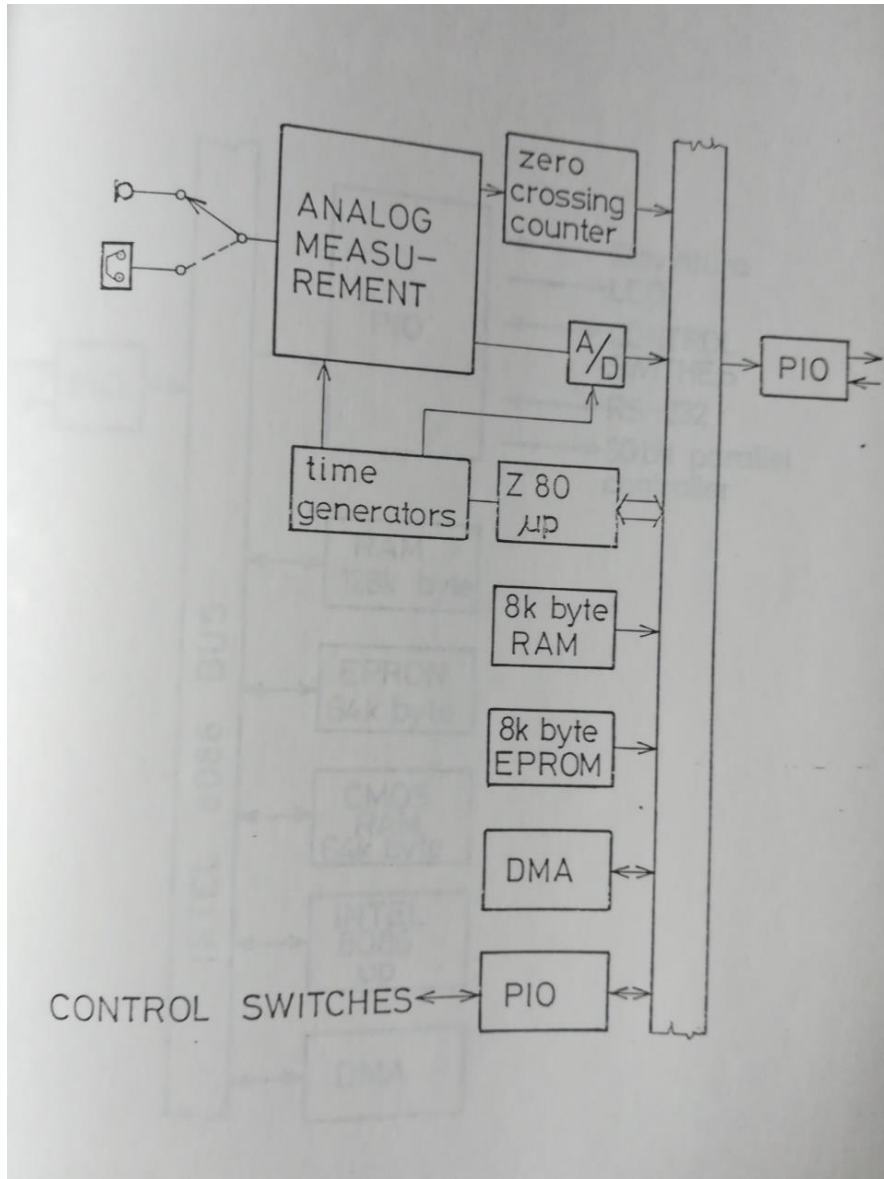
ANALOG PART OF THE PREPRO- CESSOR

FREQUENCIES ANALYSIS
IN 1 BARK

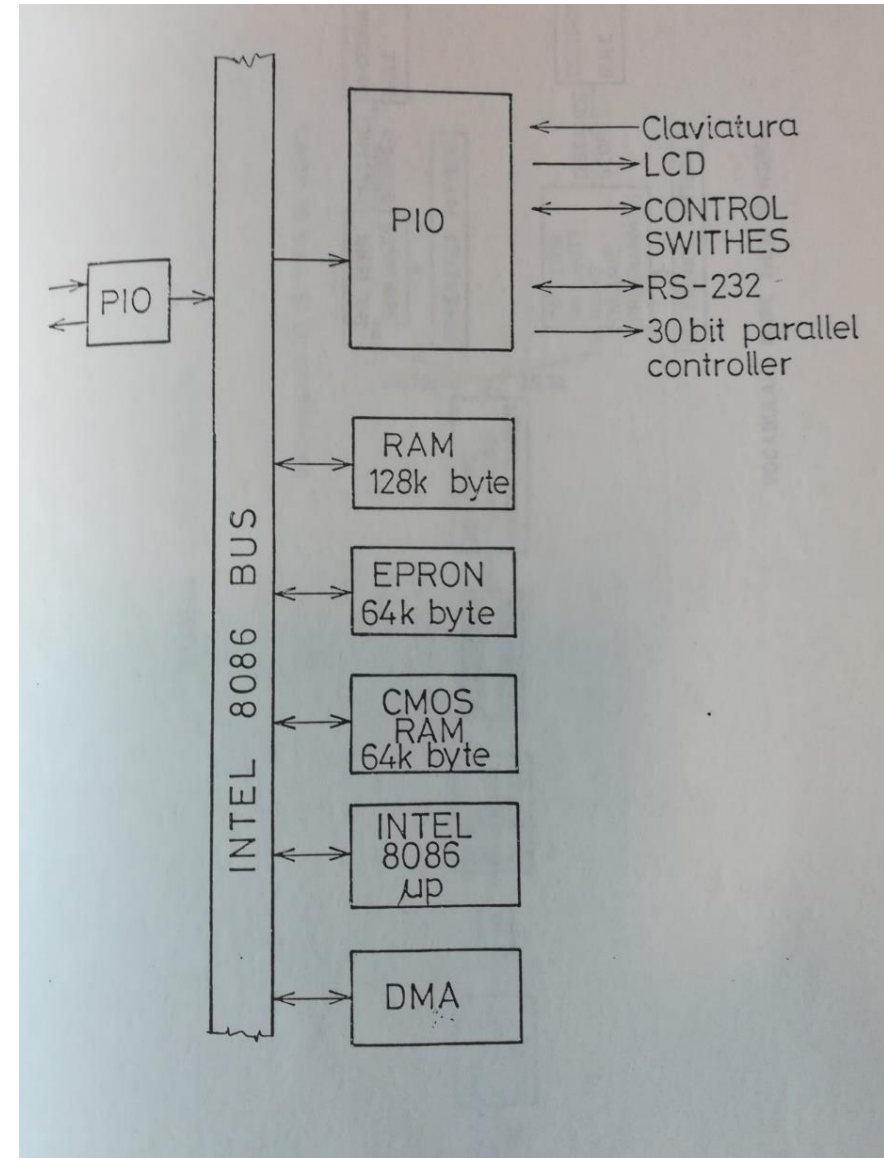
RECTIFF



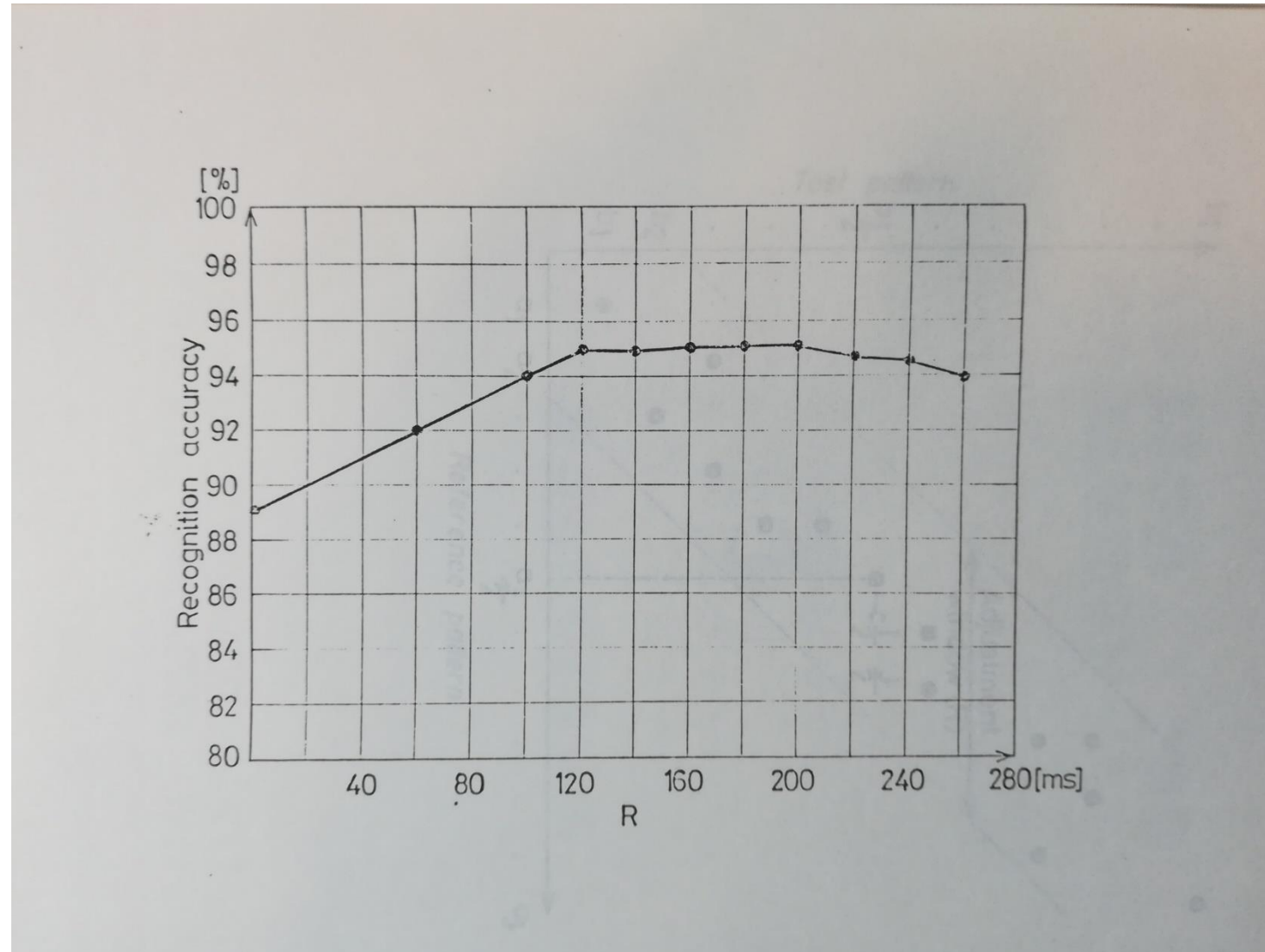
A preprocessor digitális része



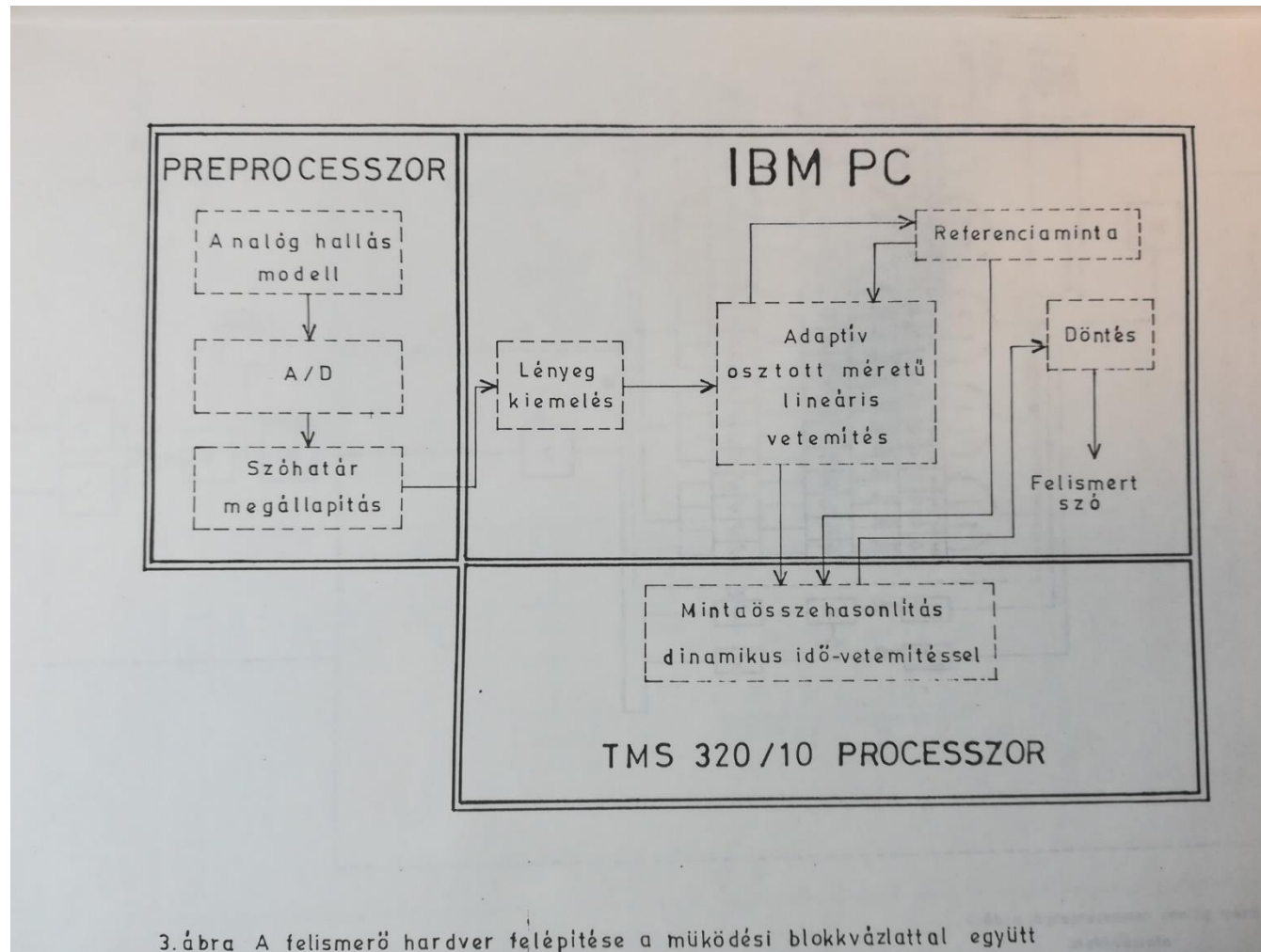
A beépített számítógép



A készülék szófelismerési pontossága a szavak hosszának a függvényében



1986- ra elkészült a beszédfelismerő IBM XT/AT kompatibilis számítógépeken futó változata



3. ábra A felismerő hardver felépítése a működési blokkvázlattal együtt



MTA MŰSZERŰGYI ÉS
MÉRÉSTECHNIKAI SZOLGÁLAT

Békésy György Akusztikai Kutatólaboratórium

BESZÉDFELISMERŐ RENDSZEREK IBM XT/AT KOMPATIBILIS SZÁMÍTÓGÉPEKEN

Az MTA MMSZ Akusztikai Kutatólaboratóriumában kifejlesztett beszédfelismerő rendszerek lehetővé teszik, hogy személyfüggően tetszőlegesen választott szavakkal és egyszerű mondatokkal számítógépeket, robotokat vagy egyéb berendezéseket lehessen vezérelni. Minden olyan területen igen hatékonyan alkalmazhatóak, ahol a berendezések szóval történő vezérése a kezek felszabadítását teszi szükségessé más tevékenység elvégzésére. Testi fogyatékosok, betegek, akik kezükkel nem tudják a megszokott módon a környezetükben lévő berendezéseket irányítani, szóval történő vezérlés alkalmazásával több hasznos tevékenységek elvégzésére is képessé válnak.

Típusos felhasználási területek:

GYÁRAK, ÜZEMEK:

- automaták, robotok, egyéb berendezések szóban történő irányítása
- Funkciók: adatgyűjtés, folyamatszabályozás, vezérlés, minőségellenőrzés

HIVATALOK:

- Billentyűzet részben vagy egészben történő kiváltása.
- CAD/CAM rendszereknél, térképkészítésnél

POSTA:

- Levelek és csomagok szortírozása

KUTATÓ-FEJLESZTŐ INTÉZETEK:

- magasszintű ember-gép kapcsolat

TESTI FOGYATÉKOSOK:

- környezetben lévő berendezések vezérése, pl. telefon, rádió

EGYÉB:

- leltárkészítés, megrendelésfelvétel, cím és üzenettrójtás, stb.

Kisszótár kezelő rendszer (ST-03) /specifikáció/

Szótárkészlet: szabadon választható, változtatható szavak és tömodulok, nyelvfüggetlen

Megengedett bemondási hossz: 2 s

Bemondások közötti minimális szünet: 0,5 s

Felismerési pontosság: 95-99% (szótármérettől és jel-zaj viszonytól függ; pontosság > 95% ha a szótárméret < 80 és a jel-zaj viszony > 20 dB)

Felismerési idő: < 500 ms

Többfelhasználós rendszer: felhasználók száma egyidőben max. 10

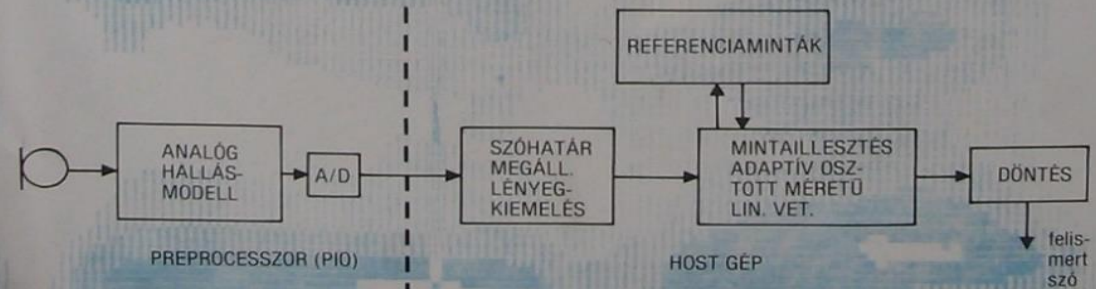
Előfeldolgozó egység: mikrofon bemenettel egy IBM kártyán

Felismerő rendszer: működik minden IBM XT/AT kompatibilis számítógépen

VEDELEM TEVESZTÉSEK ESETÉN!

A preprocesszor a mikrofonnal felvett jel akusztikus elemzését 19 párhuzamos szűrőcsatornára alapuló hallásmodell segítségével végzi. A 10 ms időintervallumban mintavett paraméterek A/D átalakítás után a „párhuzamos porton” keresztül a HOST gépbe kerülnek, ahol a további feldolgozás és a felismerés történik. A mintaillesztés adaptív időcsoportokban lineáris idővetítéssel történik.

A RENSZER FELÉPÍTÉSE:



A kiállításokra készült
„korabeli” prospektus

Beszéd felismerő bemutatása a Budapesti Akusztikai Konferencián

1986 –ban a Központi Fizikai Kutatóintézet által készített robot vezérlésére



Azóta, sok szép eredmény született Magyarországon

A BME TTT-n izolált szavas felismerő készült,

-
-

Takács György stokholmi KTH béli tartózkodása során neurális háló alapú folyamatos szöveget lejegyző rendszert készített.

-
-

BME TMIT-n Fegyó Tíbor-Mihajlik Péter HMM alapú folyamatos szöveget lejegyző rendszereket állított és állítanak elő
a mai napig is,

-
-
-

stb. (a lista nem teljes).

32 év után ma

Google által fejlesztett kereső-, diktálórendszerek
okostelefonokon magyar nyelven is jól működnek.

Igaz, hogy még a mondathatárokkal bajok vannak,

de

Szaszák György a Beszédaesztikai laboratórium munkatársa magyar nyelvű prozódia felismerő rendszerével a mondathatárok felismerése már lehetséges, és beilleszthetővé vált az általános tartalomfelismerő rendszerekbe.

Köszönöm a figyelmet

viczi@tmit.bme.hu

<http://alpha.tmit.bme.hu/speech/>



Nagy Számítástechnikai Műhelyek

A beszéd számítógépes feldolgozása Magyarországon

2018. szeptember 28.

Takács György, Olaszgy Gábor, Podoletz György, Tihanyi Attila

A MEA 8000 formánsszintetizátor chip alkalmazása a BME-n: Scriptovox, Multivox, Pcrobot 1983-tól)

Philips 1981

A Budapesti Nemzetközi Vásáron
kiállította a MEA8000 beszéd-processzorát.



Mea VOX

Embargós termék a NATO-nak fejlesztve

PHILIPS



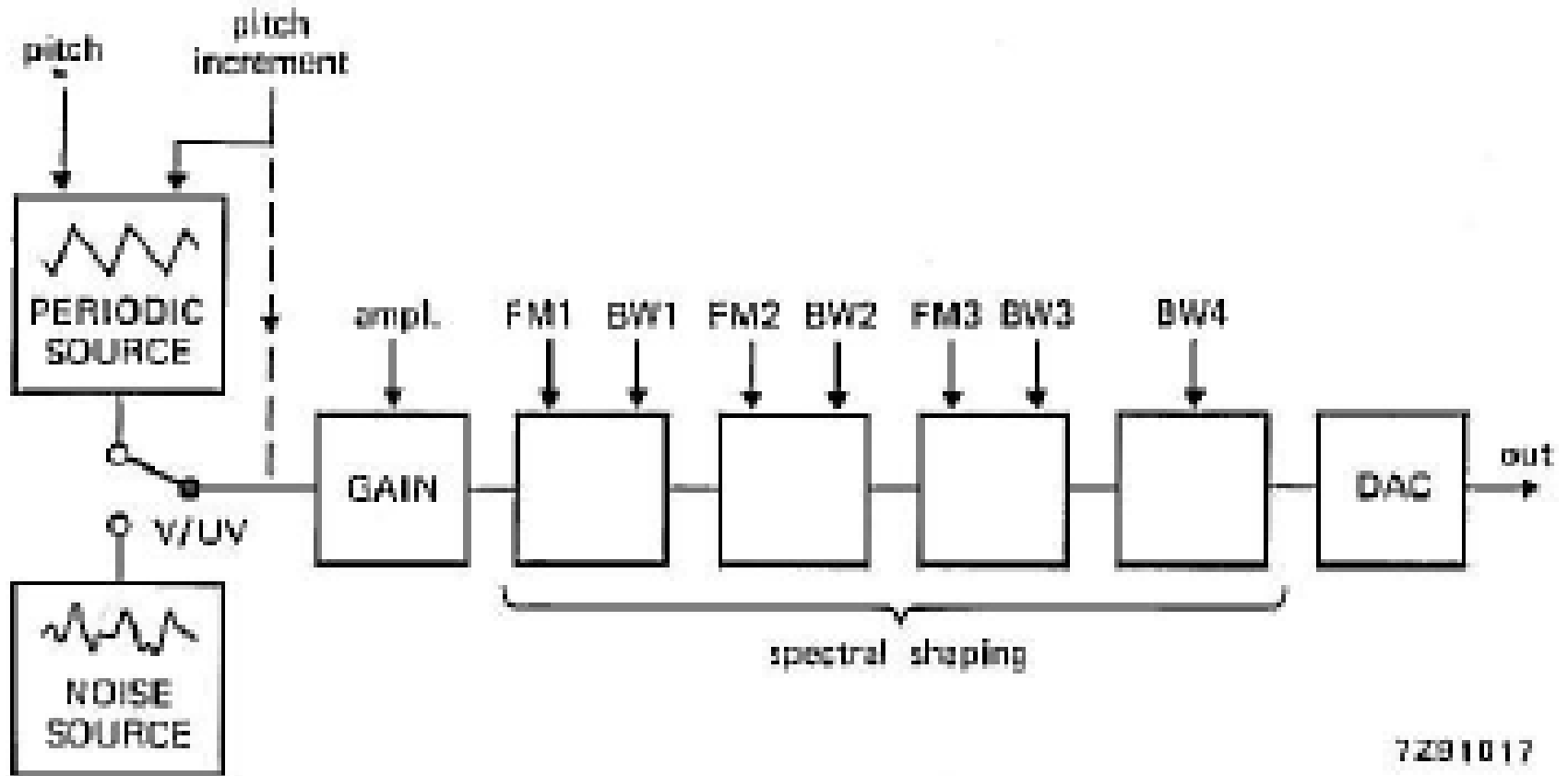
Electronic
components
and materials

Technical
publication

101

MEA8000 voice synthesizer: principles and interfacing

Philips MEA 8000 voice synthesizer



7291017

1983-ban BME diplomaterv született a MEA 8000 első alkalmazására

Németh Géza: VOXALARM beszélő riasztó



Kisméretű (egy kártyás) beszéd szintetizátorok fejlesztése a BME-n

MEA 8000 chipeket már lehetett vásárolni

Kutatási partner az MTA Nyelvtudományi Int.

A Scriptovox rendszer írott szöveget alakított át beszéddé MEA 8000 vezérelhető chip-el

Gép: SGS (olasz) **A PC-k EMBARGÓSAK !**

Beszélő egység: egyedi hardver (Tihanyi Attila)

A PHILIPS MEGJELENTETI A MEA 8000 ÚJ VÁLTOZATÁ A PCF 8200 CHIPET

1985

Beszélő Commodore 64

Fejlesztő: BME

Gyártó: Megamicro szövetkezet
(Podoletz György igazgató)

Hardver: Microvox-64 doboz PCF 8200-al

Vezérlés: Commodore Basic program

Hangkimenet: 3 pólusú dugó a TV-hez



European Conference on Speech Technology, Edinburgh

1987.

- Magyar előadás (Olaszy-Gordos) és hangbemutató.
TTS konverzió real time.

Automatic text-to-speech system applied in a reading machine

- Feltűnést keltett.
- Nem hitték el, hogy TTS konverziót csináltunk
Commodore 64 számítógéppel.

SZABADALOM 1986-89

Olaszy Gábor, Gordos Géza, Takács György,
Podoletz György, Czap László, Magyar Gábor
feltalálók

***Eljárás beszédet előállító jelrészletek halmazának
nyelvfüggetlen meghatározására szintetizált beszéd
létrehozásához***

Lajstromszám: 195022

Benyújtás éve: 1986.

Közzététel éve: 1989.

EZT A SZABADALMAT EURÓPAI CÉGEK MEGVÁSÁROLTÁK A BME-TŐL

- **Seibersdorf Versuchszentrum** (Ausztria, 1990)
Amiga géphez illesztettük, német TTS szoftverrel (1993)
- **Vertical Technology** (Németország, 1992)
Macintosh géphez illesztettük
német TTS szoftverrel
(TMS 320 mikroprocesszorral
szimuláltuk a PCF 8200 –at)



A hardvereket Tihanyi Attila és Osváth László tervezték.

FEJLESZTÉSEK PC-RE 1990-től






BME



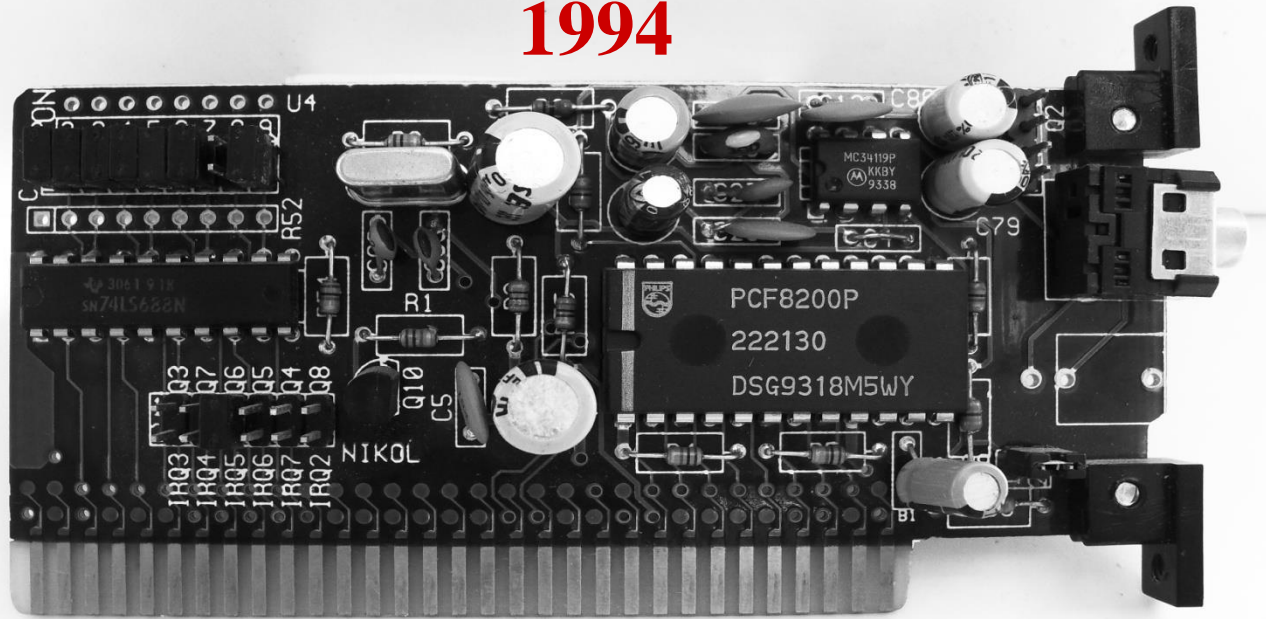
A MULTIVOX HANNOVERBEN A CEBIT-EN 1994



FEJLESZTÉSEK PC-RE 1990-től

- **Hardver:** hangszóró+nyák PCF 8200-al (Tihanyi Attila)
- **Vezérlés:** PC párhuzamos portról
- **Szoftver:** Multivox TTS (kis floppy lemezen)
- **Nyelvek:** 8 nyelv (magyar, német, olasz, finn, holland, spanyol, eszperantó, arab).
1 floppy/nyelv 
- Multivox mutatványok a hanggal    
- **Gyártás:** Nikol KKT
- **Forgalmazás:** Nikol KKT, Európában

PC kártya magyar beszédszintetizátor (PC-ROBOT) 1994



Tihanyi Attila tervezése

A szoftverhez oktató programokat is adtunk (mai néven app-okat)

Éneket is lehetett készíteni.



Köszönjük a figyelmet!



NJSZT Informatikatörténeti Fórum –
A beszéd számítógépes feldolgozása Magyarországon

A hullámforma összefűzés bevezetése a beszédszintézisben – Profivox TTS

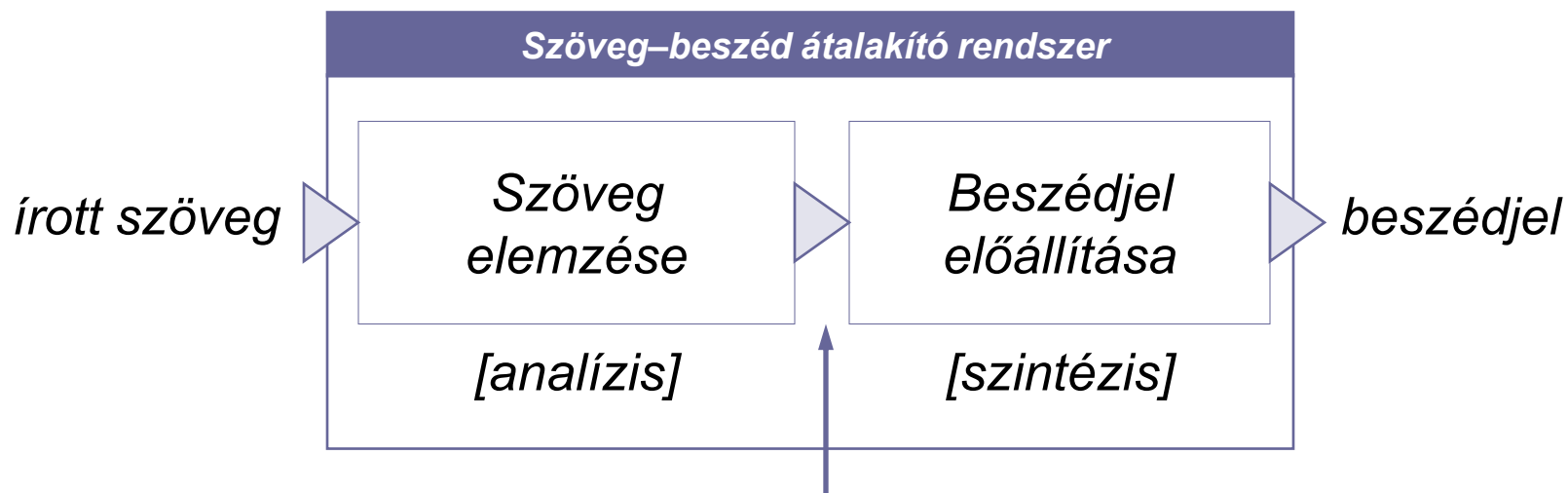
Olaszi Péter, Kiss Géza, Olaszky Gábor, Németh Géza

- **Parametrikus szintézis:** A hangképző szervek működését modellezzük. A beszédjel előállítása mesterséges gerjesztéssel és szűrőssorral – formánszintézis (HUNGAROVOX, MEA8000 chip)
Jellemzői: gépies jellegű hangzás; kis számítási és tárigény; tág határok között változtatható paraméterek
- **Hullámforma-kódolás:** Természetes emberi beszédből kivágott részleteket tárolunk, és ezek összefűzésével építjük fel a beszédjelet
Jellemzői: természetesebb hangzás; nagyobb számítási és tárigény; szűkebb határok között változtatható paraméterek
Példa: a *hullámforma-összefűzéses* technológia
- Megvalósítható a fenti két módszer kombinációja is

A technológiánk a 2000-es évek előtt

- **A 80-as évek közepéig:** lassú processzorok, kevés memória
Cél: érthető beszédjel előállítása – parametrikus szintézissel
Univoice és Hungarovox: 1982, MTA Nyelvtudományi Intézet
„1982 májusára elkészült az első automatikus, magyarul beszélő szövegszintetizáló rendszer, az Univoice alapváltozata.” 
- **A 90-es évek közepéig:** gyorsabb processzorok, több memória
Cél: többnyelvű szintézis, valamint a *prozódia*, azaz a beszéddallam, az idő- és intenzitás-szerkezet, a hangsúlyozás javítása
Multivox 3: 1986–94, Olasz Gábor – Németh Géza, BME TTT
„Das Multivox Sprachausgabe spricht jetzt mit einer Sound Blaster Karte.” 
Multivox 4: *„Most a Multivox beszéd-szintetizátor négyes változata hallható.”* 
- **90-es évek közepétől:** gyors processzorok, rengeteg memória
Cél: az emberi ejtést megközelítő hangminőség és prozódia elérése; mélyreható nyelvi elemzés; hullámforma-összefűzéses technológia
Profivox: 1999, Olasz Gábor – Németh Géza, BME TTT
„Most a Profivox professzionális szövegfelolvasó beszél.” 

A szöveg–beszéd átalakító rendszer szerkezete

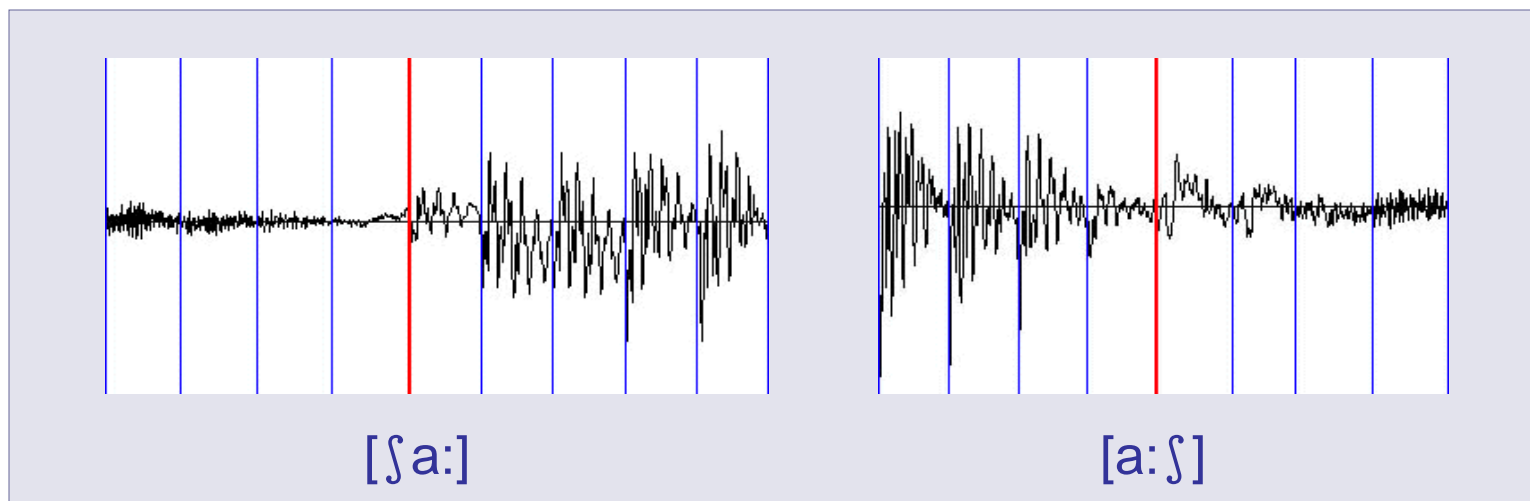


Beszédhangsor: fonetikus szöveg

Prozódiai vezérlés:

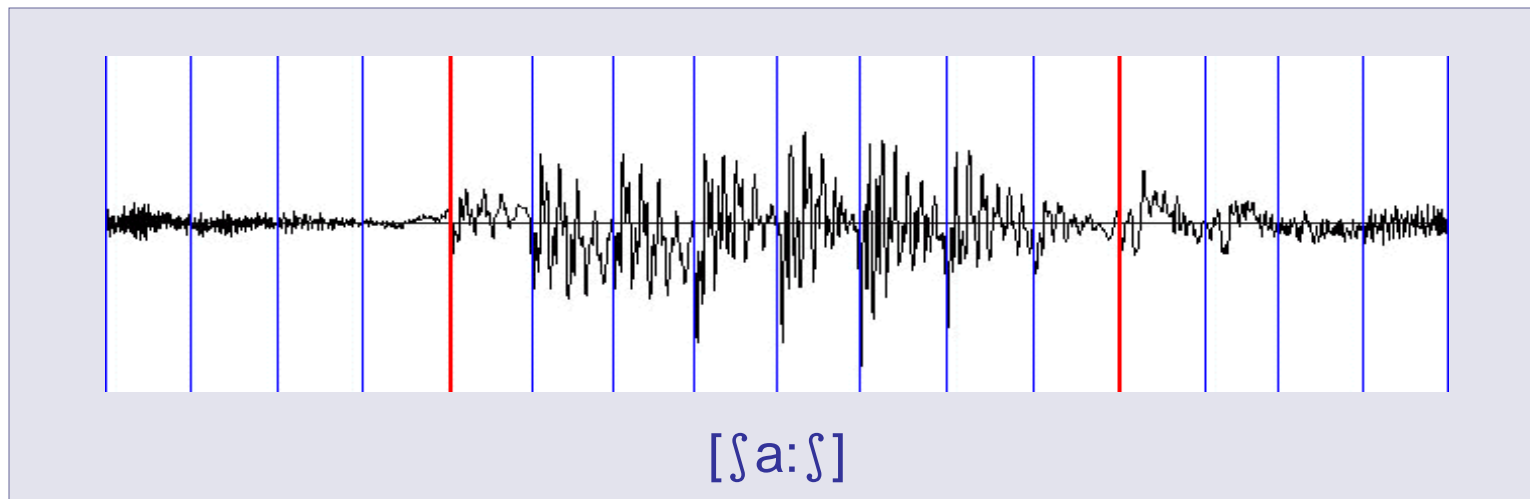
hangidőtartamok, beszéddallam,
hangsúlyozás, intenzitásmenet

A beszédhang-hangadatbázis felépítése diád-elemekkel



- Diádelem: két beszédhang kapcsolata;
a vágási pont a beszédhangok középső, stabil fázisában van
- A beszédjelet a diádelemek összefűzésével állítjuk elő
Példa: *ötven* → [_ø] [øt] [tv] [vε] [εn] [n_] (6 elem)
- Az elemek illesztésénél torzítás léphet fel

A beszédhang-hangadatbázis felépítése triád-elemekkel



- Hanghármasok: kevesebb illesztési pont
- Többnyire msh-mgh-msh kapcsolatok
Példa: *ötven* → [_øt] [tv] [vɛn] [n_] (4 elem)
- Az adatbázisban lehetnek diádok, triádok és hosszabb elemek is
Példa: [øtvɛn]

- Minden beszédhangnak minden beszédhanggal való kapcsolatát rögzíteni kell
- Diádelemek száma: ~1000
Hanghármasok száma: ~4000
- Szisztematikus módszer: hangkapcsolatok felvétele vivőszótagokba ágyazva



1. báb	abába
2. báp	abápa
3. bád	abáda
4. bát,c,cs	abáta
5. bág	abága
6. bák	abáka
7. bágy	abágya
8. báty	abátya
9. bám	abáma
10. bán	abánat
11. bány	abánya
11. báj	abája
13. báh	abáha
14. báv	abáva
15. báf	abáfa
16. báz	abáza
17. bász	abásza
18. bázs	abázsa
19. bás	abása
20. bál	abála
21. bár	abára

1. bab	ababa
2. bap	abapa
3. bad	abada
4. bat,c,cs	abata
5. bag	abaga
6. bak	abaka












- Prozódia: időtartamok, hangmagasság, intenzitás változásai
- Egymás mellé fűzött adatbázis elemek, prozódiai vezérlés nélkül



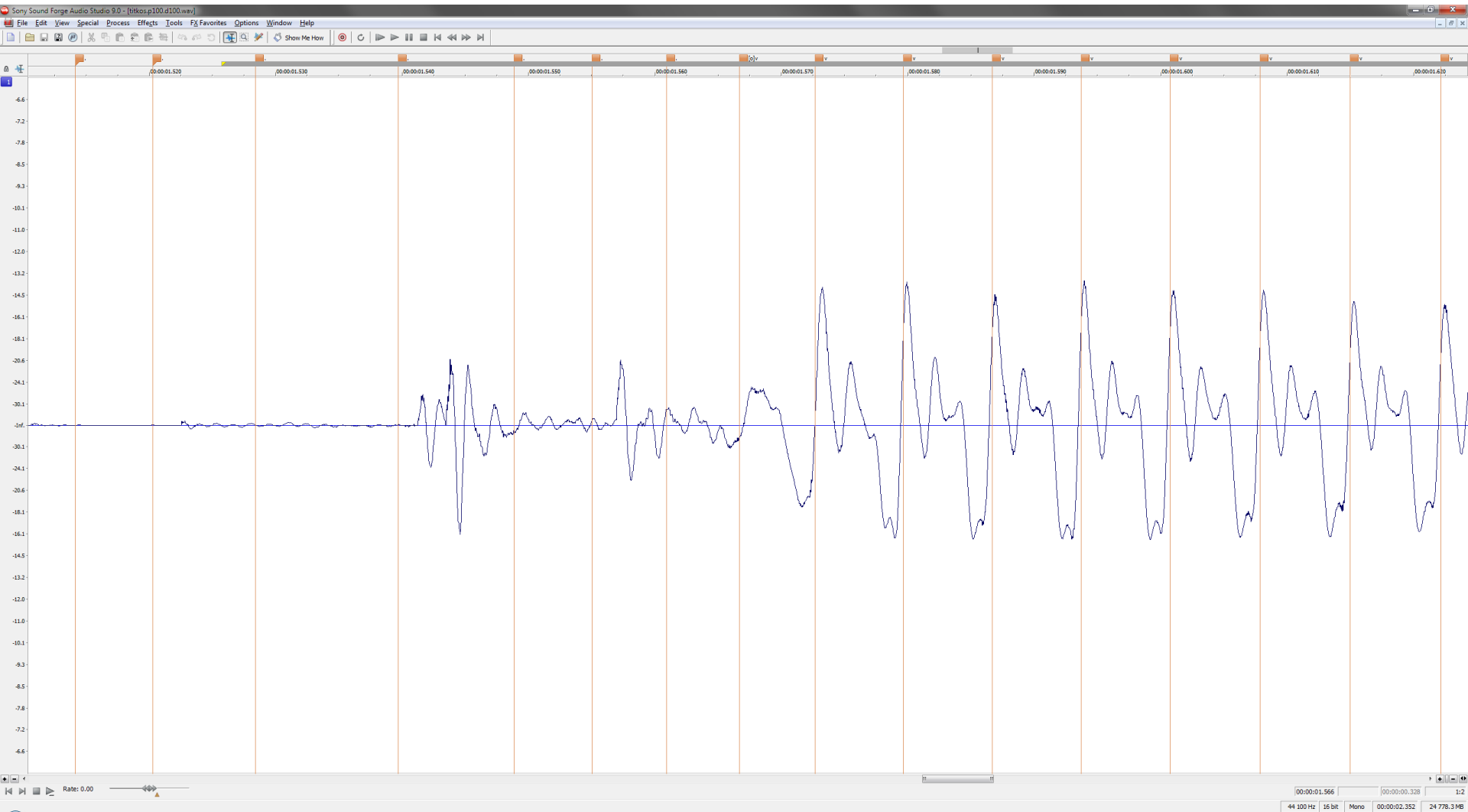
- Ugyanaz a mondat, ráültetett prozódiával



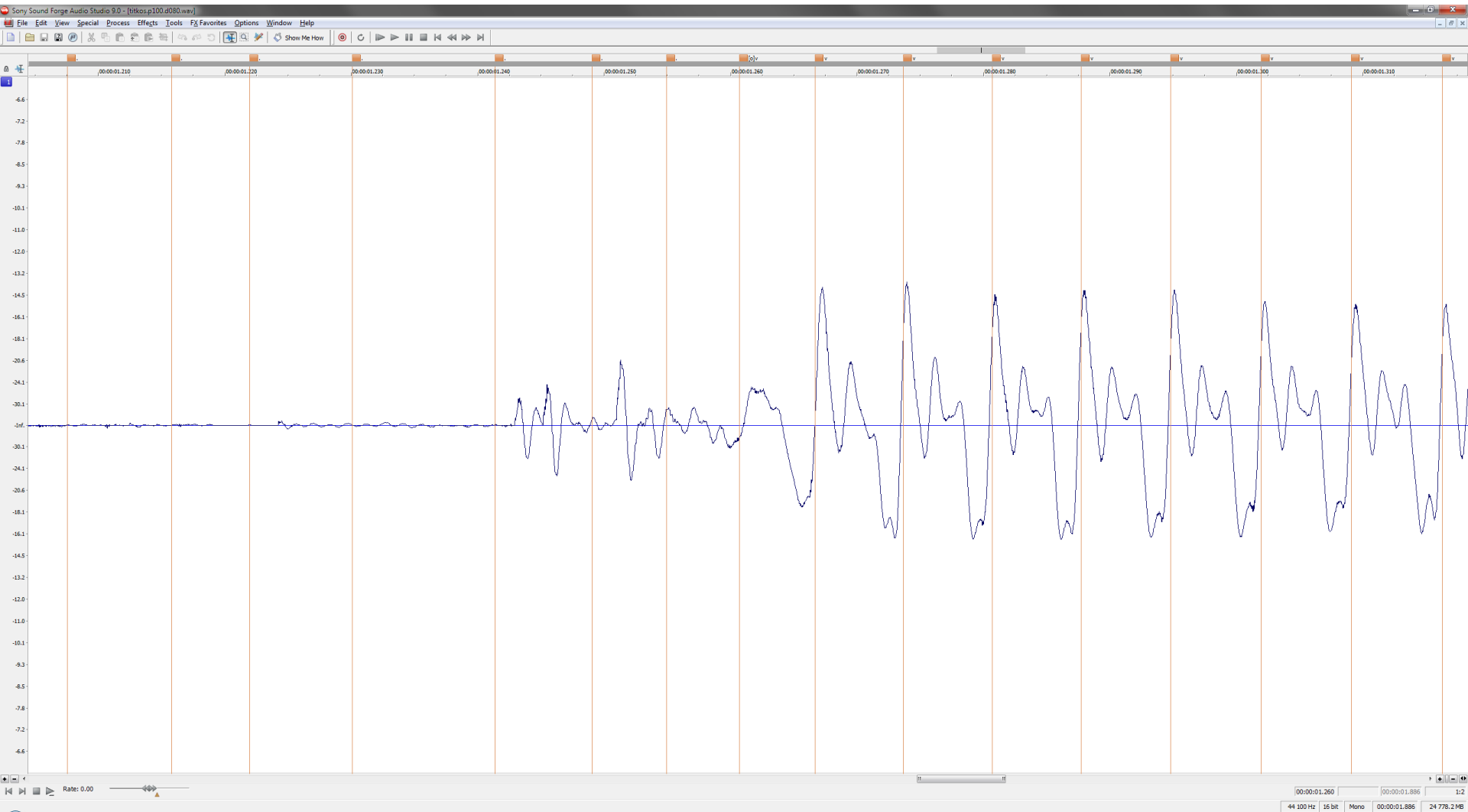
Prozódiai módosítások – összehasonlító táblázat

		Hangmagasság [%]							
		40	60	80	100	120	140	160	180
Időtartam [%]	80								
	100								
	140								
	200								

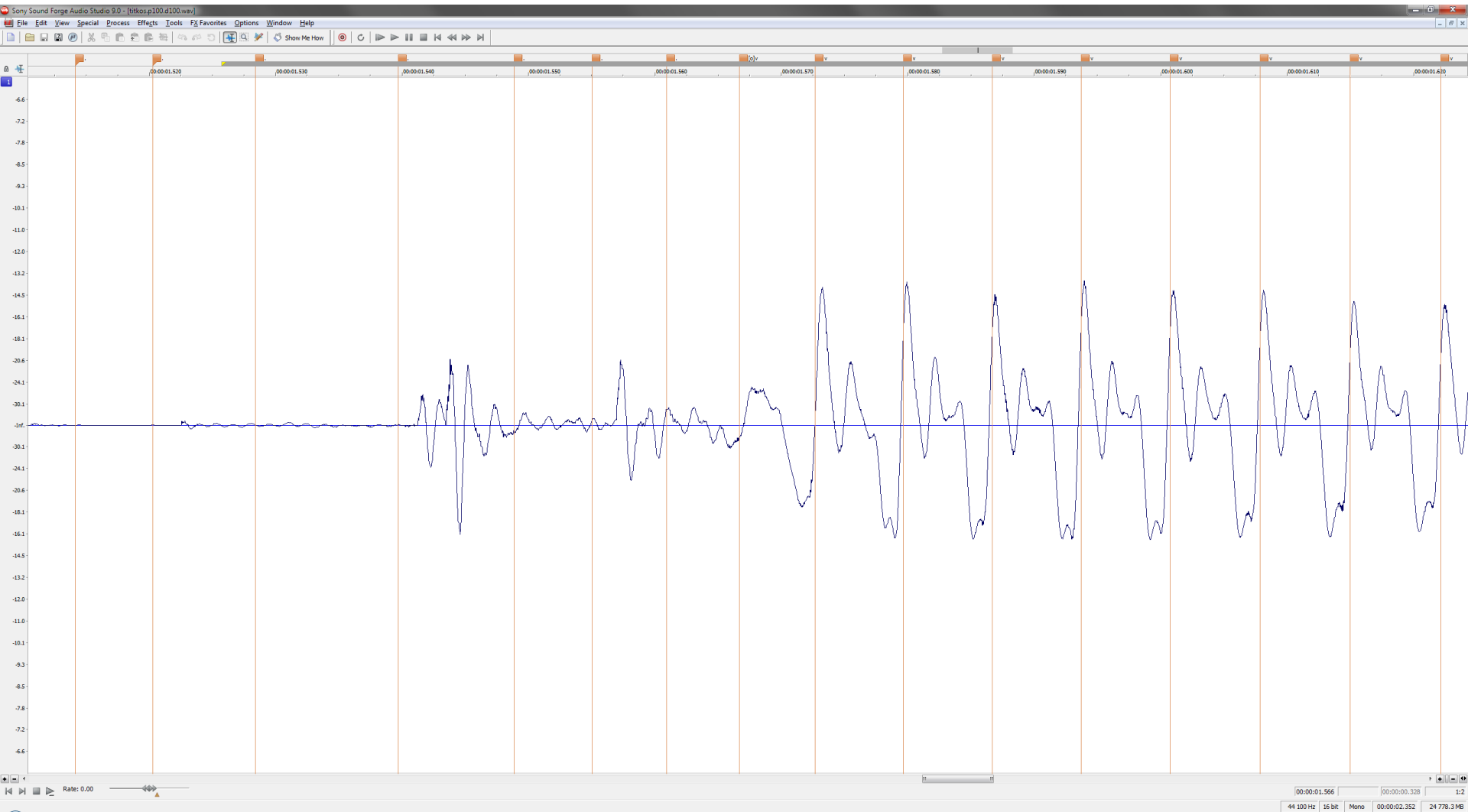
A hangidőtartam módosítása – eredeti beszédjel



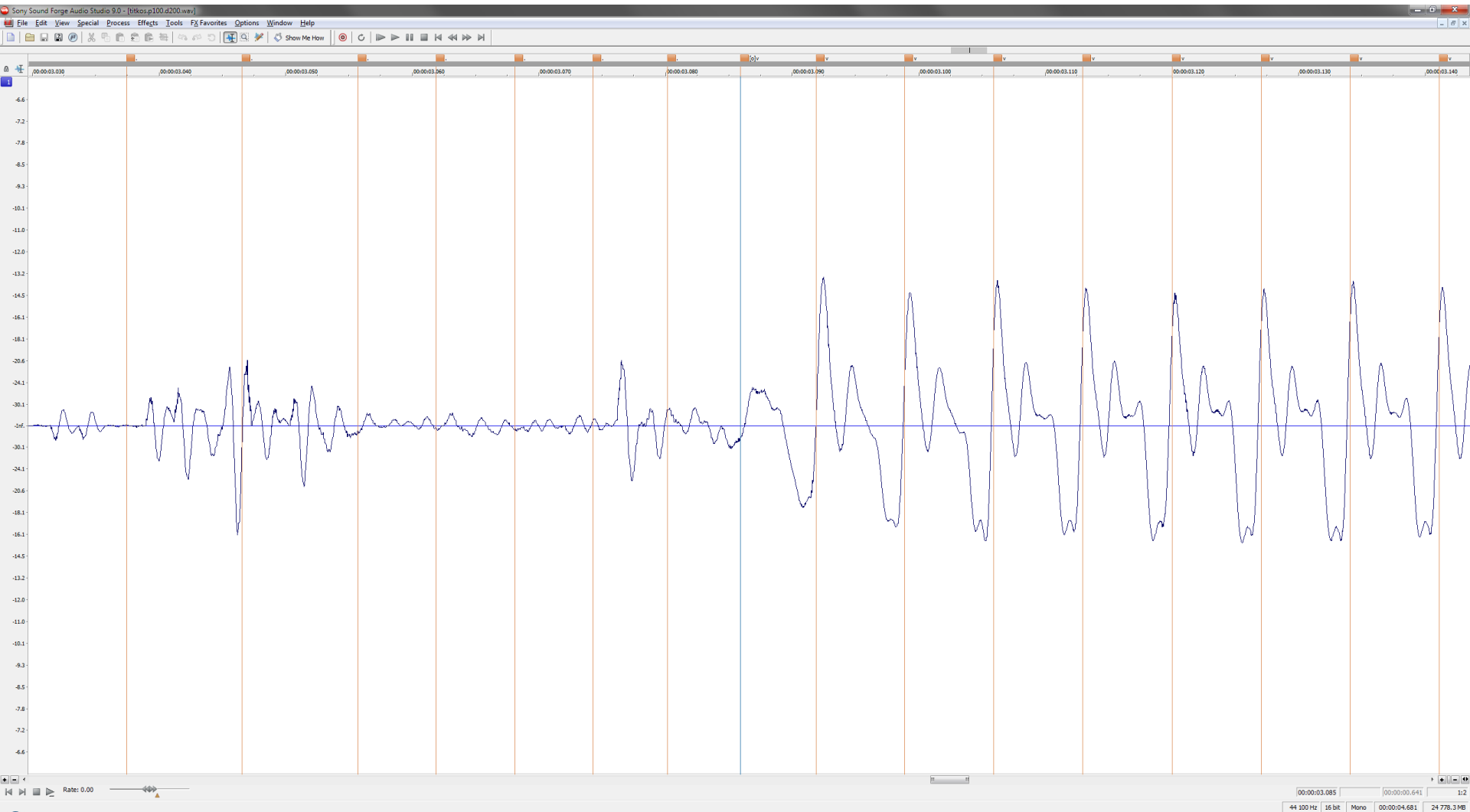
A hangidőtartam módosítása – csökkentett hangidőtartam



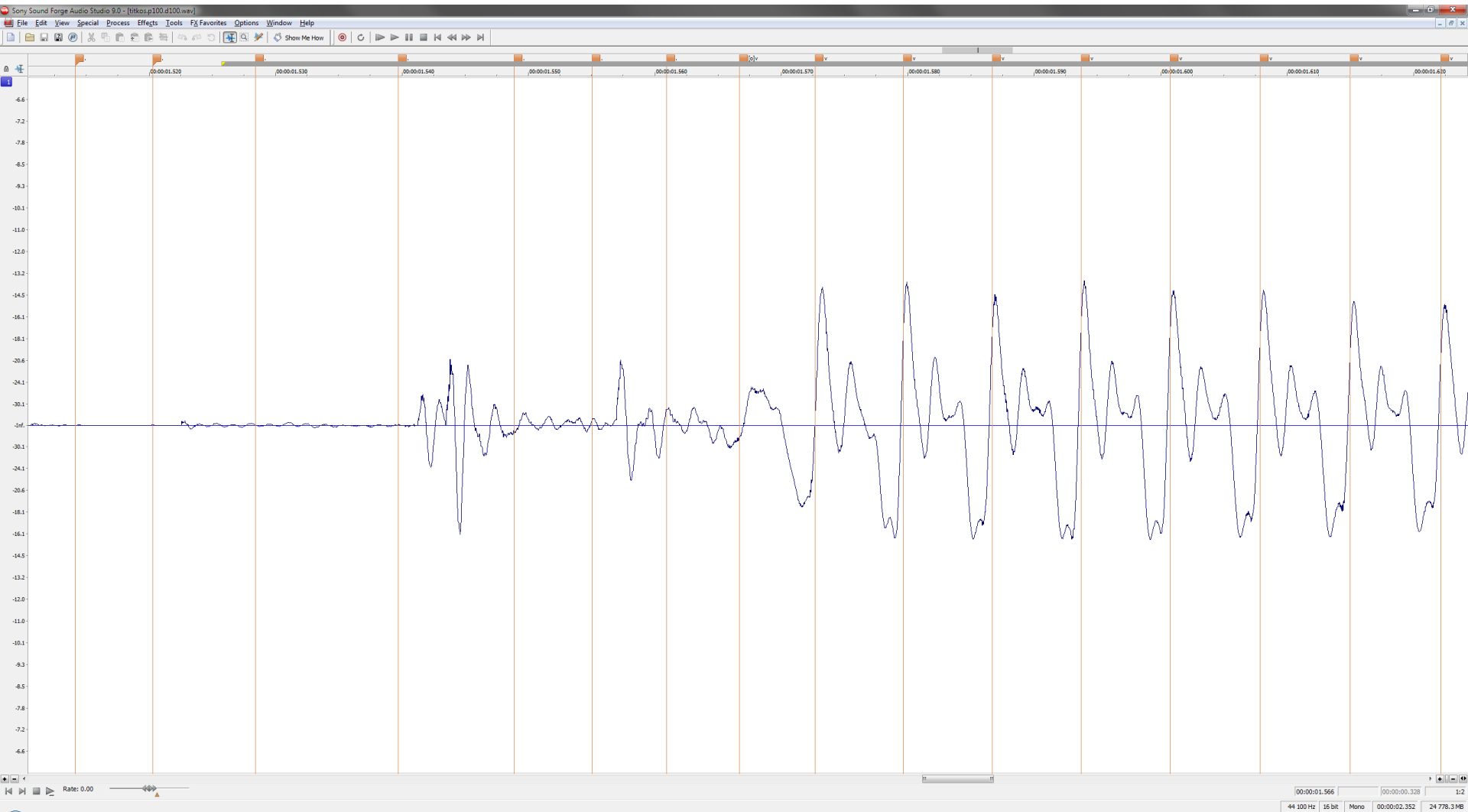
A hangidőtartam módosítása – eredeti beszédjel



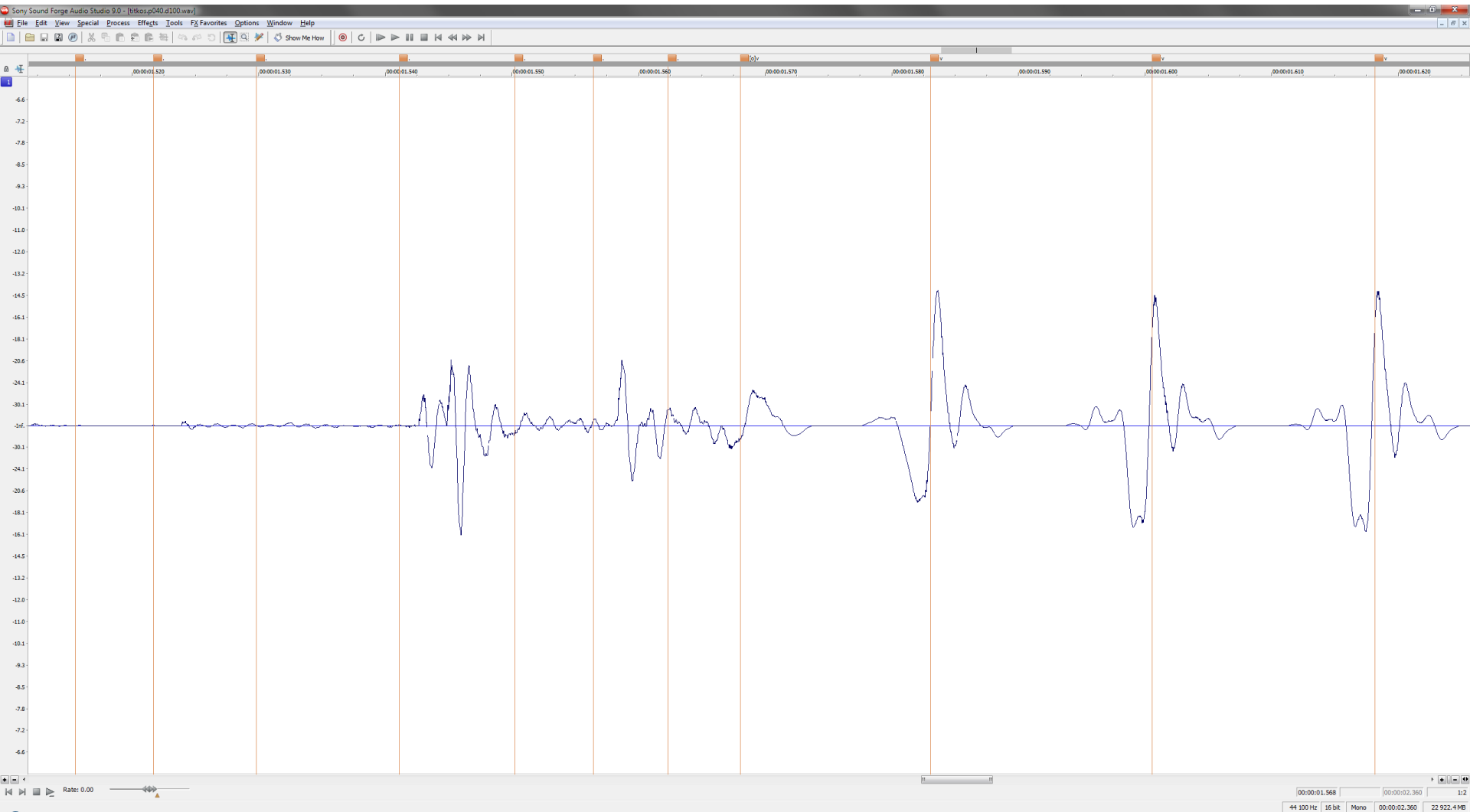
A hangidőtartam módosítása – növelt hangidőtartam



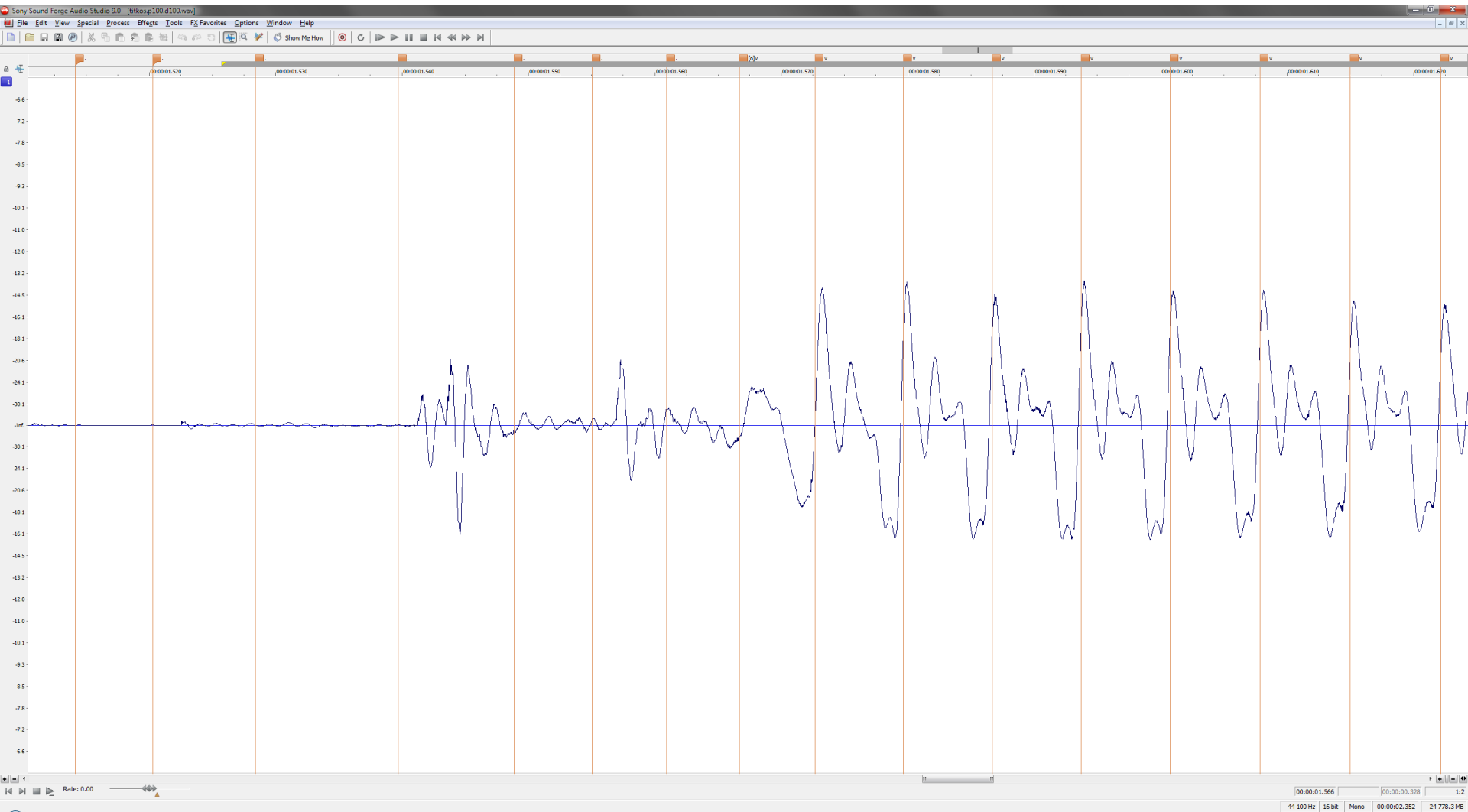
A hangidőtartam módosítása – eredeti beszédjel



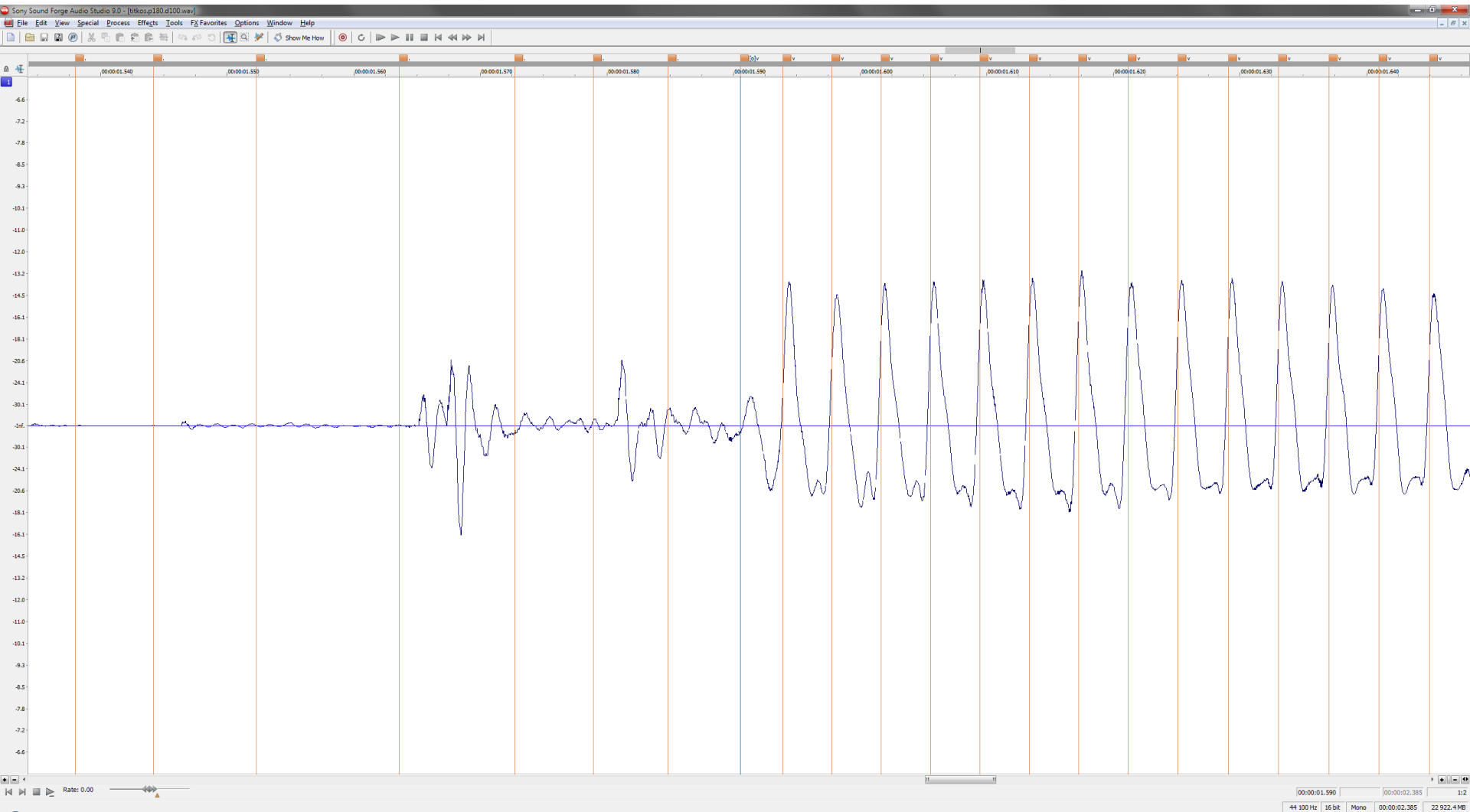
A hangidőtartam módosítása – csökkentett hangmagasság









A hangidőtartam módosítása – eredeti beszédjel



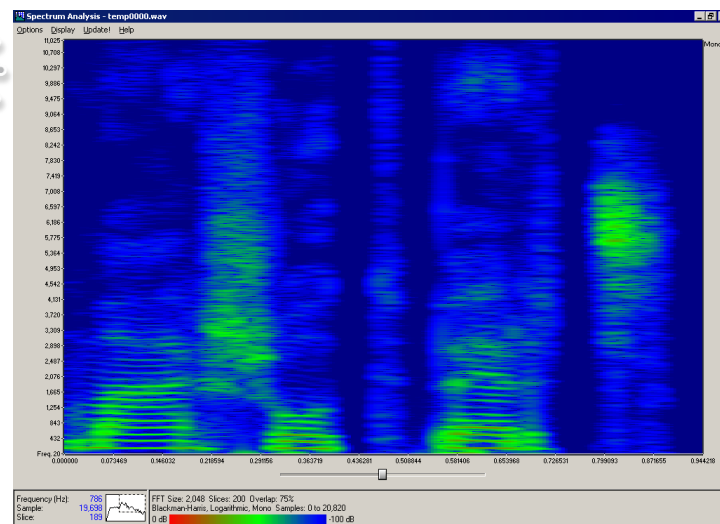
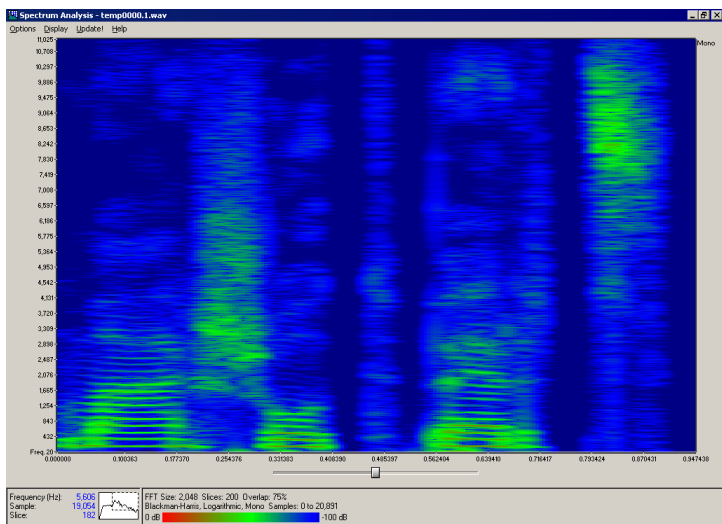
A hangidőtartam módosítása – növelt hangmagasság



- A hangidőtartamok, intonáció és intenzitás-változások átültetése egyik beszédjelről a másikra
 - csak ugyanarra a beszédhangsorra működik
- Természetes beszédjel, férfi beszélő: 
- Az eredeti mondat prozódiaja átültetve a szintetizált beszédjelre:
 - Azonos beszélő: 
 - Másik beszélő, szintén férfi: 
 - Női beszélő, időtartamok átültetve, és
 - hangmagasság változtatás nélkül: 
 - hangmagasság eltolva, és
 - eredeti hangmagasság-dinamikával 
 - női beszélőhöz igazított hangmagasság-dinamikával 

Spektrum-kompresszió telefonos beszédátvitel céljára





- A G.722 ITU-T standard szélessávú beszédkódoló felső sávkorlátja 7 kHz.
- A beszédjelben a réshangok és zár-réshangok tartalmaznak akár 10 kHz-es összetevőt is. Ezek az összetevők elvesznek a telefonos beszédátvitel során
- A spektrum-kompressziós módszer az érintett beszédhangok spektrumát összenyomja – az így módosított hang minden lényeges komponense visszaállítható a vevőoldalon.



- Jaws for Windows – képernyőolvasó szoftver vakok és gyengénlátók számára
 - Nagy mértékben gyorsítható a beszéd
 - 2018-ban is ez működik
- Robobrace fájl konverter (2018)

RoboBraille.org



- WESTEL SMS mondó 
- *Dömdödöm* kötött szótáros szövegfelolvasó 
- Banki számlainformációs rendszer 
- Általános szövegfelolvasó 



„Szóltam, meghallgattatok, értitek, ítéljeteK.”

Lüsziasz
(i.e. 445–380)





M Ű E G Y E T E M 1 7 8 2

SpeechTex 

The Speech Technology Expert

A folyamatos beszéd gépi felismerése –
a kezdetektől (BME-TTT 90-es évek) napjainkig

DR. MIHAJLIK PÉTER

BME-TTT '90 „Beszédfelismerők”



Gordos Géza

Tanszékvezető



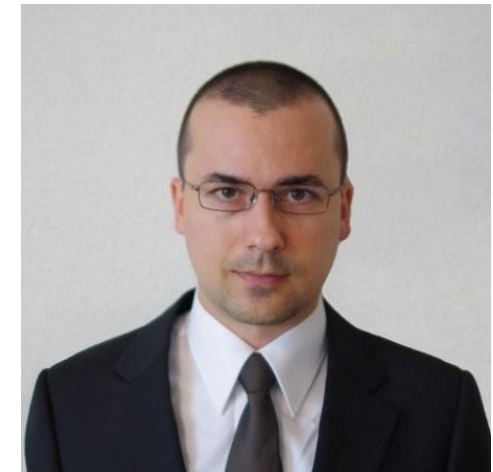
Tatai Péter

Telecom. Signal Proc.
Labor vezető



Lükő Bálint

Fegyő Tibor



Szarvas Máté

Doktoranduszok

Mihajlik Péter

Az első,
folyamatos
beszédfelismerési
alkalmazások
('90)

- ❑ 1993 IBM Personal Dictation System, the first dictation system for the personal computer
- ❑ 1995 PHILIPS developed SpeechNote, a dictation and transcription software.
- ❑ 1997 Dragon Systems released NaturallySpeaking 1.0 as their first continuous dictation product.

Jellemzőik:

- Angol nyelv
- 20-30 ezres szótárméret
- Szűkített témakör (pl. radiológia, jog)
- Nagy költségvetésű kutatóhelyi háttér, az 50-es 60-as évektől

A folyamatos beszédfelismerés kezdetei

~ 1975: IBM, Fred Jelinek: **HMM**

"Every time I fire a linguist, the performance of the speech recognizer goes up".

$$\hat{W} = \arg \max_W P(W) \cdot P(O | W)$$

'80-as évek vége: **NN**

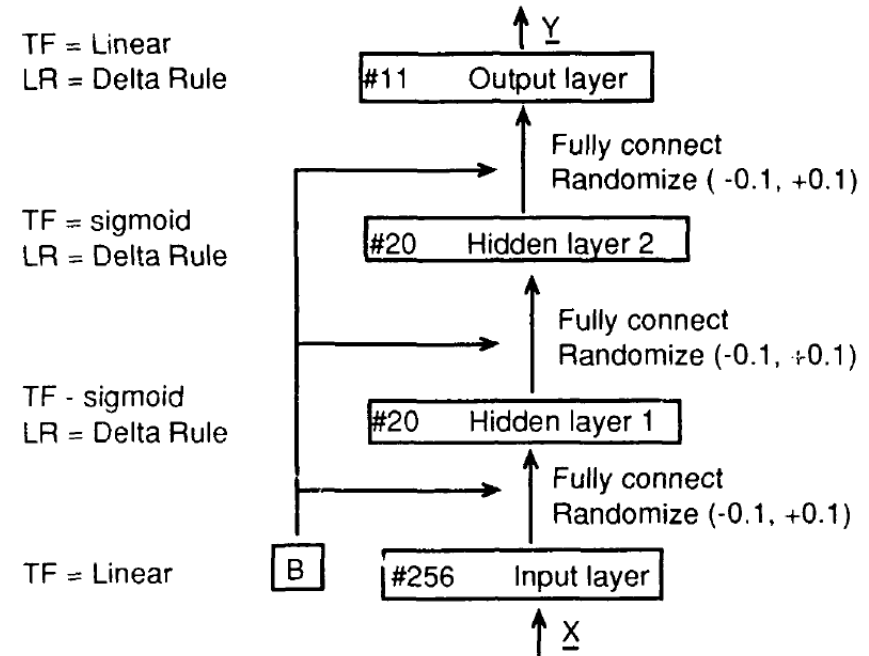


Figure 3. Back-propagation neural network topology

Mi kell(ett) a folyamatos magyar nyelvű beszéd felismeréséhez?

- Technológia...
- A hangjel megfelelő reprezentálása (időtartománybali jel nem alkalmas)
- A beszéd dinamikájának modellezése
- Kiejtési szótárak
- A koartikuláció (beszédhangok egymástól függése) modellezése
- Szóhatárokon fonológiai változások (egybeolvadás, hasonulás) modellezése
- Ragozás miatti hatalmas szóalakszám csökkentése
- Szavak visszaállítása morfémaszerű nyelvi egységekből
- Folyamatos beszédet (és leíratot) tartalmazó adatbázisok
- Sok-sok adat és erőforrás...

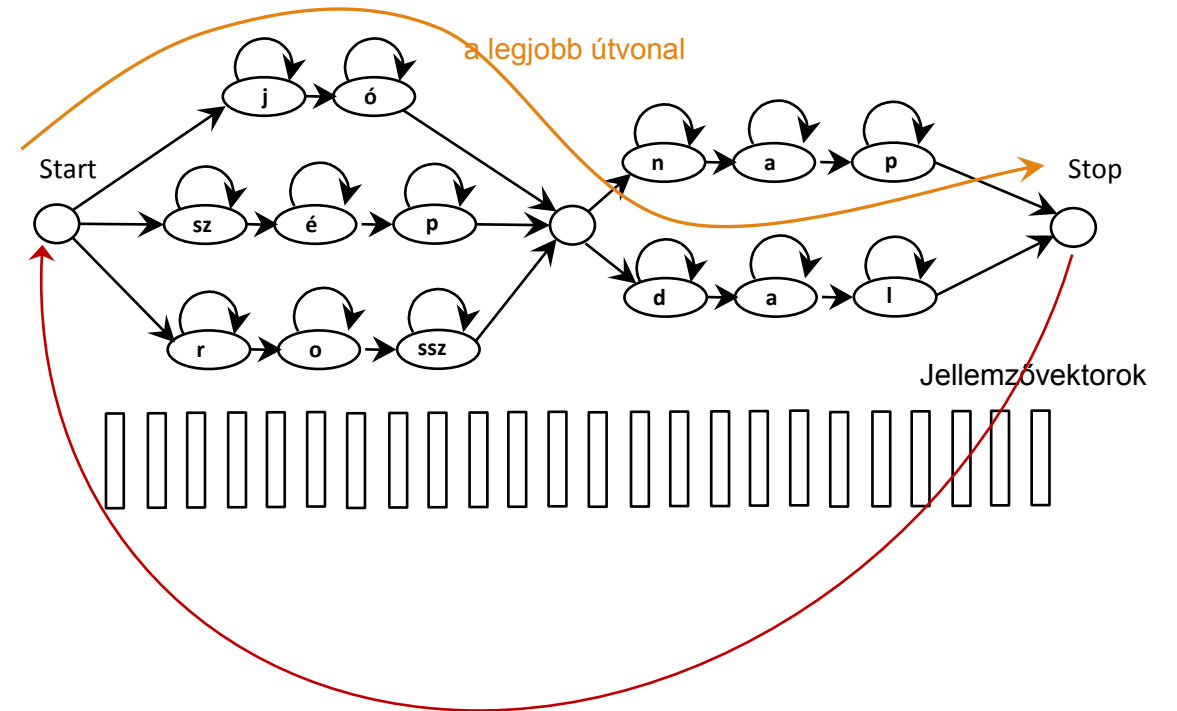
Folyamatos beszédfelismerési technológia

Rejtett-Markov modell

□ Implementáció

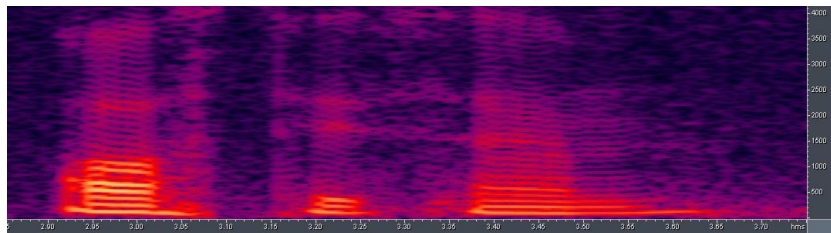
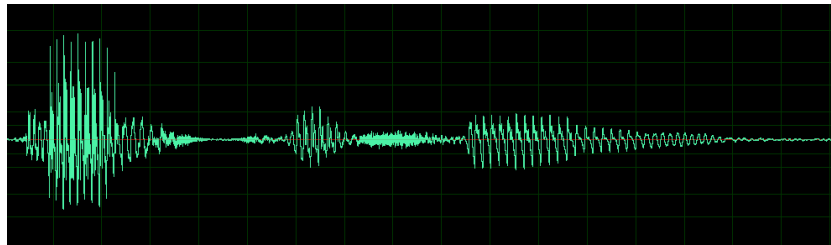
- Szarvas Máté: '96 diplomamunka
- Cambridge Entropic Research:

HTK – Hidden Markov-Model Toolkit

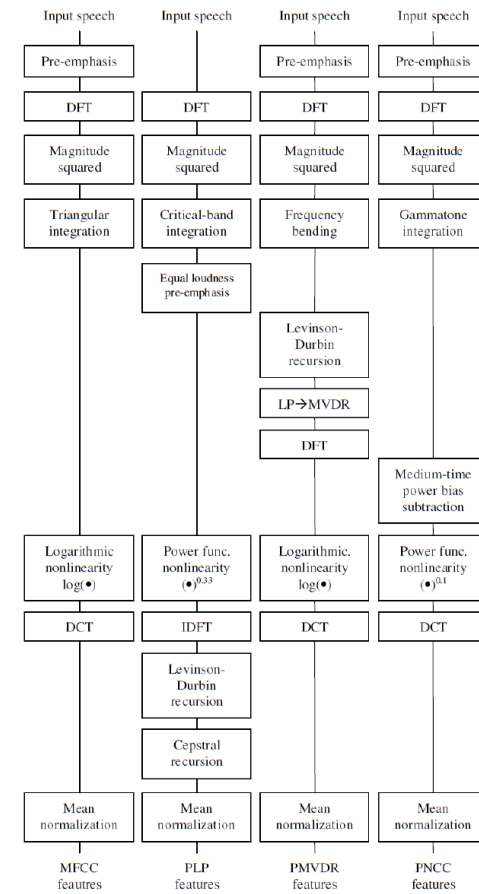


A hangjel megfelelő reprezentálása

„ a z t h i s z e m ”



MFCC / LPC / PLP ...



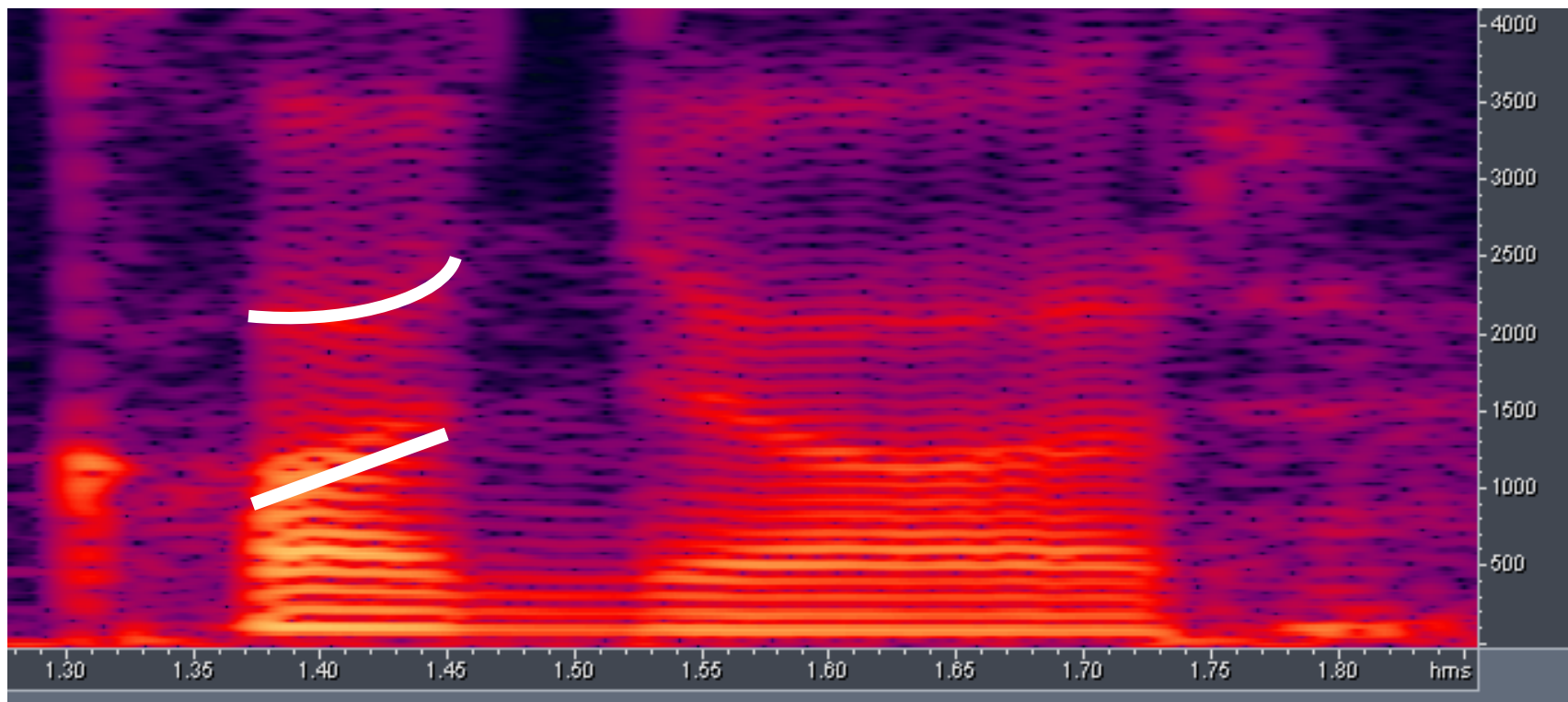
[k]

[a]

[ny]

[a]

[r]



+ Δ ?

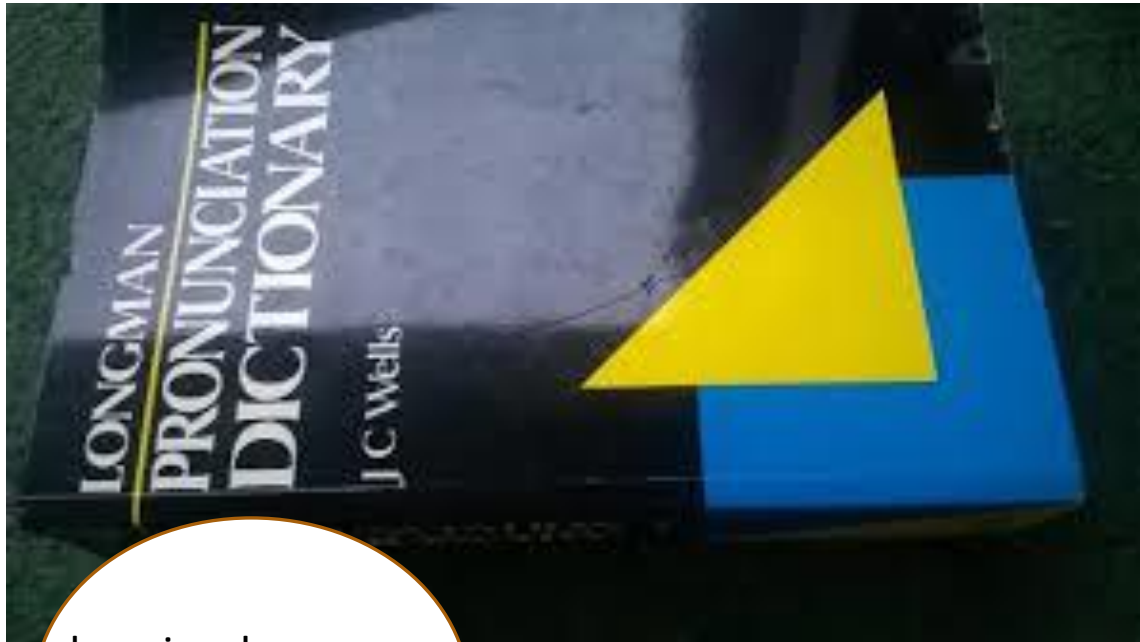
+ $\Delta\Delta$?

+ $\Delta\Delta\Delta$?

Beszéddinamika-modellezés

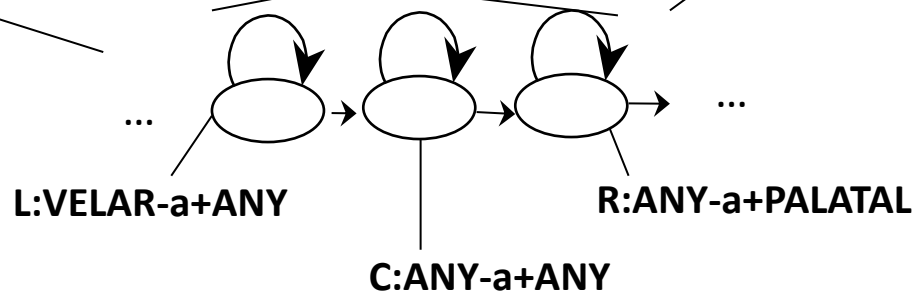
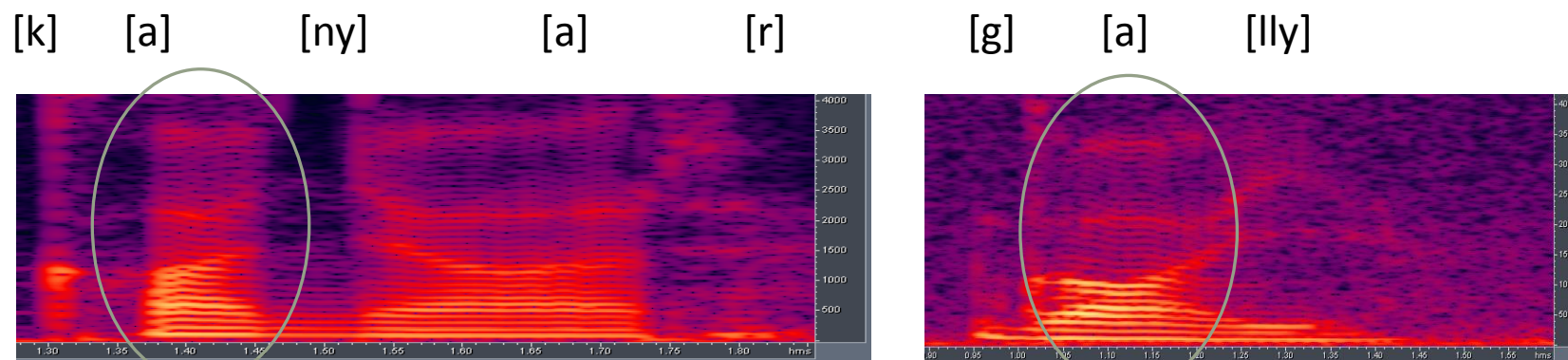
Kiejtési szótár

- ❑ Szabály alapon?
 - ❑ $t + sz = c$
- ❑ Kivételek?
 - ❑ Churchill = cs ö r cs i ll
- ❑ Kiejtési változatok? Gyakoriságok?
 - ❑ miért = m é r t, m i é r, stb...
- ❑ Ambiguitások?
 - ❑ Lacheგი vs. Lachema, malacsült, meggyógyít...



hagyja h a ggy a
hangya h a ny gy a

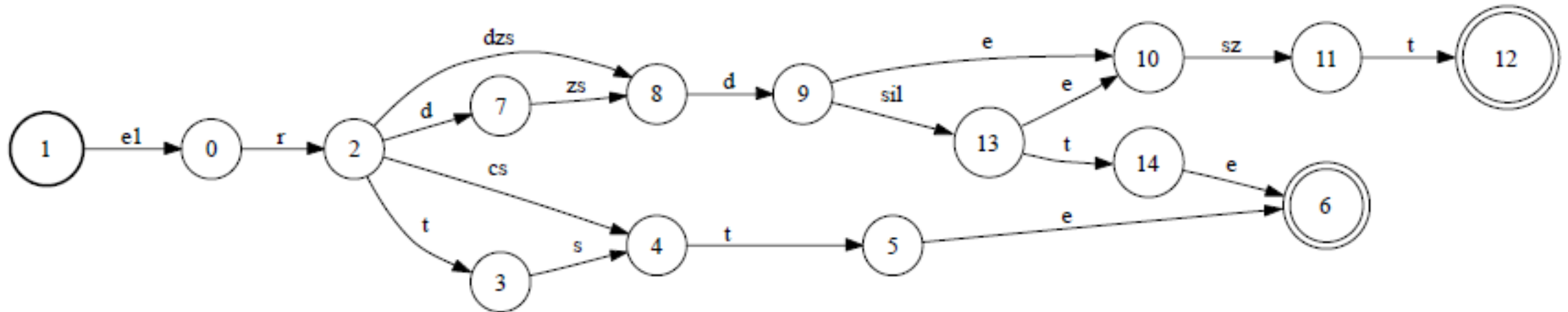
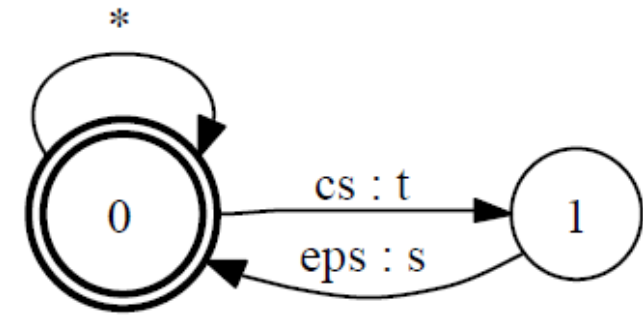
Koartikuláció modellezés



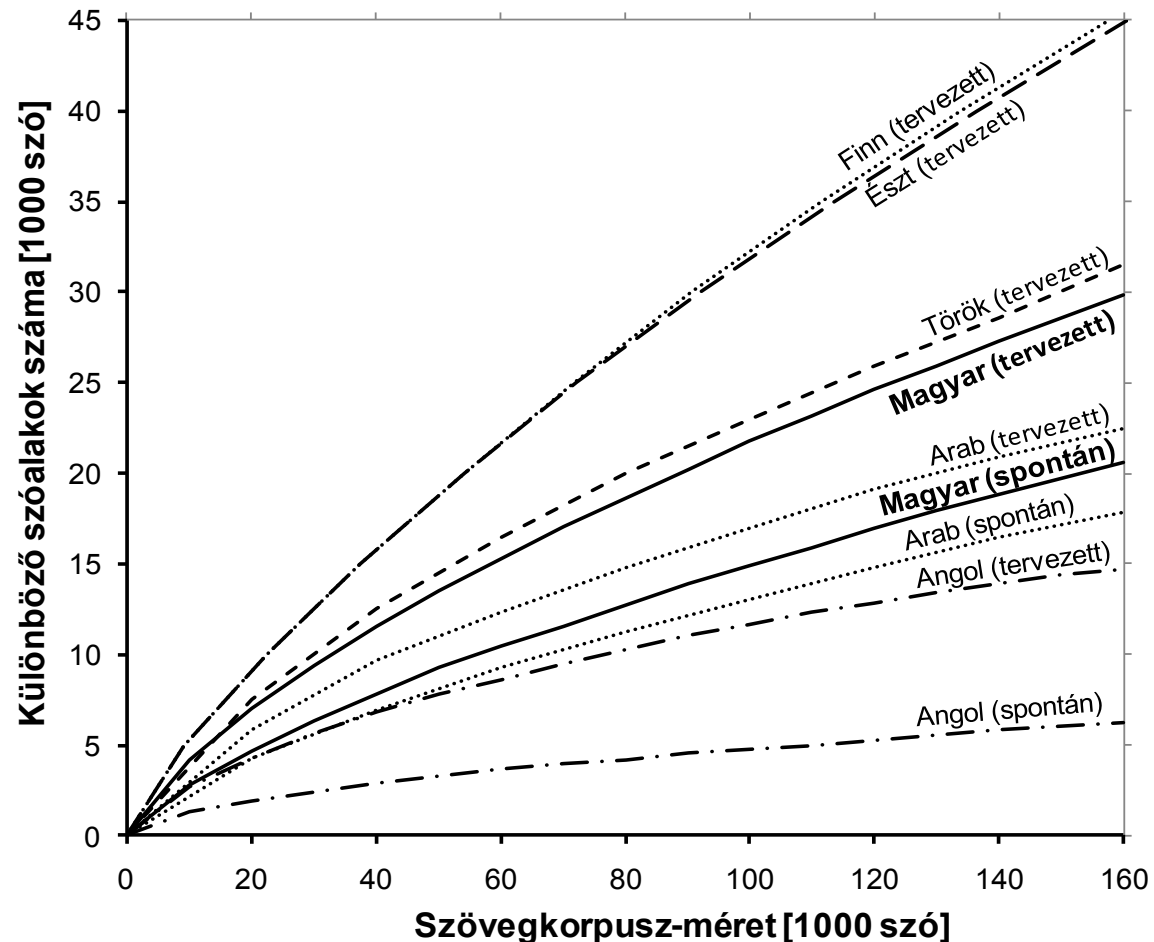
Megosztott állapotú
környezetfüggő
beszédhangmodellekkel

Egybeolvadás, hasonulás stb. modellezése

é r d z s d sil t e
ért sd [te|ezt] → é r cs t e
é r d z s d [sil] e sz t ...



Szóalakszám csökkentése – morfológiai változatosság kezelése



láthattuk kóstolhattuk milyen jól tudnak főzni
ha akarnak

lát hat tuk kóstol hat tuk mily en jól tud nak
főz ni ha akar nak

/we could see and taste how well they can
cook if they want/

Szóadarabolás: szabály és/vagy statisztikai
alapon

Szóalakok visszaállítása

□ Szóvég jelekkel

lát hat tuk# kóstol hat tuk# mily en# jól# tud nak# főz ni# ha# akar nak#

□ Non-initial jelekkel

lát -hat -tuk kóstol -hat -tuk mily -en jól tud -nak főz- ni ha akar -nak

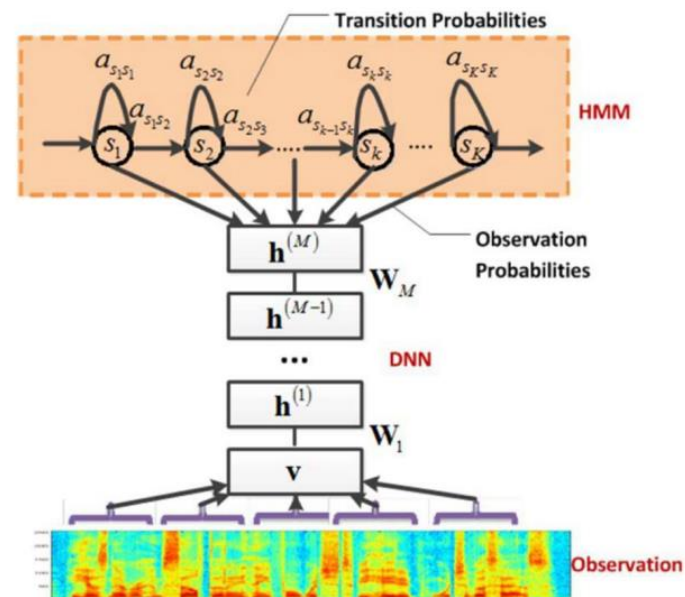
□ Szóhatár jelekkel

lát hat tuk # kóstol hat tuk # mily en # jól # tud nak # főz ni # ha # akar nak

Erőforrások – anno és ma...

- ❑ Operatív memória: n Mbyte -----> n Gbyte
- ❑ CPU: 80486, 50MHz, 50 MIPS/TMS320 DSP -----> core i7 3+ GHz, RTX 2080: 13.2 TFLOPS
- ❑ Beszédatbázisok (magyarra): 30perc-3 óra ----- > 300-1000 óra
- ❑ Szövegadatbázisok (adott témakörből, magyarra): 200 ezer szó -----> 200 millió szó
- ❑ Szótárméret: 200 szó -----> több millió szó
- ❑ (Projekt méretek: n x 100MFt -----> n x 10 MFt...)

Deep Learning



Szóalakok visszaállítása

□ Szóvég jelekkel

lát hat tuk# kóstol hat tuk# mily en# jól# tud nak# főz ni# ha# akar nak#

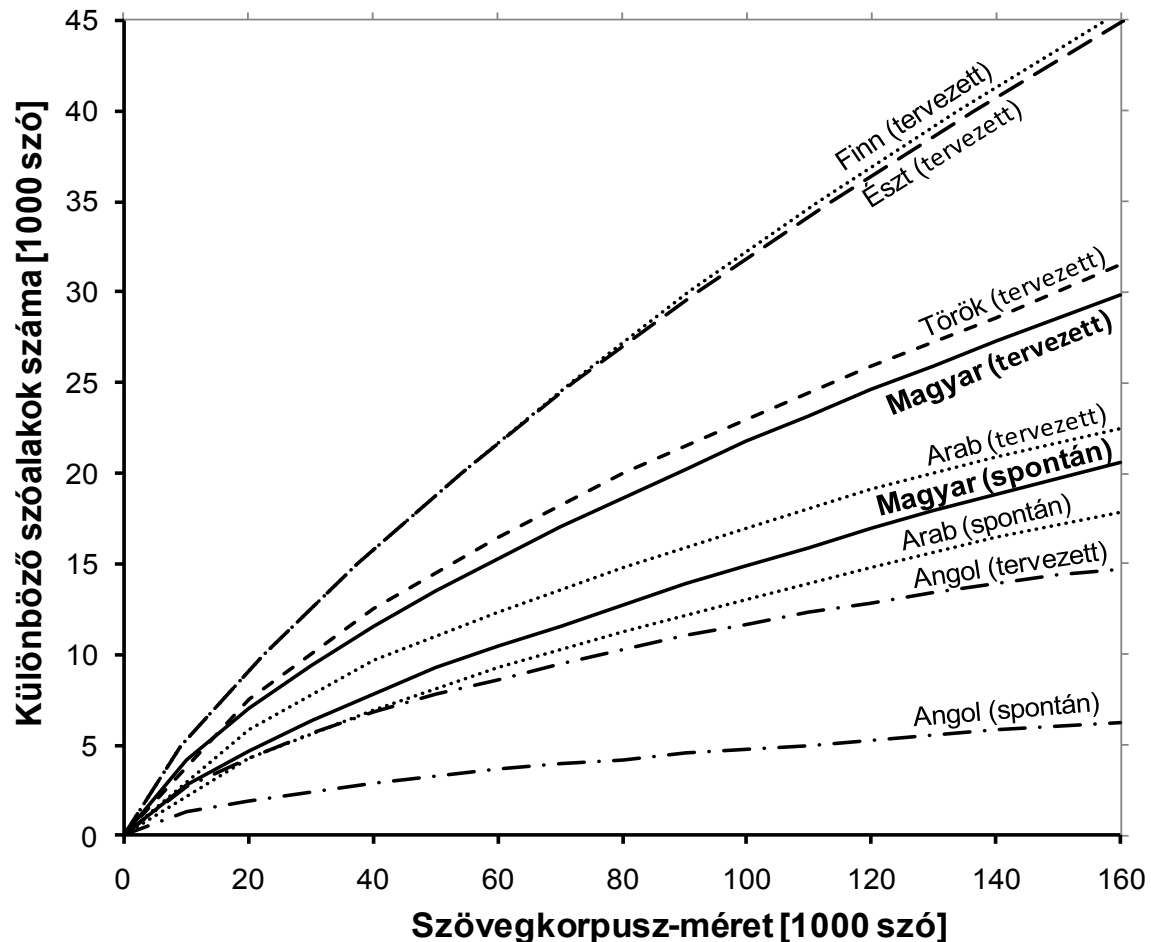
□ Non-initial jelekkel

lát -hat -tuk kóstol -hat -tuk mily -en jól tud -nak főz- ni ha akar -nak

□ Szóhatár jelekkel

lát hat tuk # kóstol hat tuk # mily en # jól # tud nak # főz ni # ha # akar nak

Szóalakszám csökkentése – morfológiai változatosság kezelése



láthattuk kóstolhattuk milyen jól tudnak főzni ha akarnak

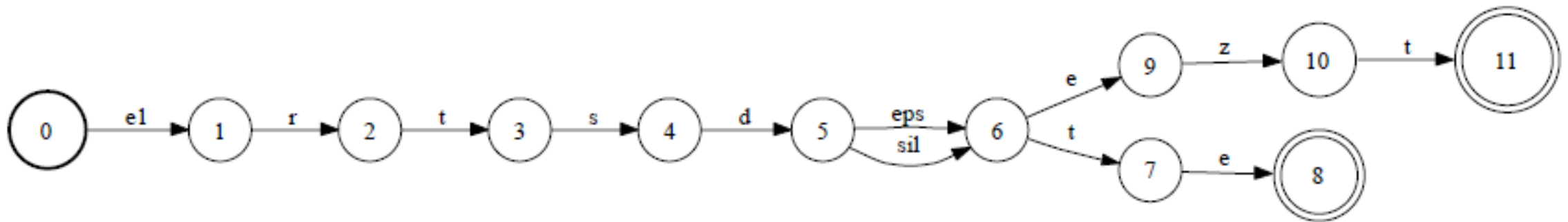
lát hat tuk kóstol hat tuk mily en jól tud nak főz ni ha akar nak

/we could see and taste how well they can cook if they want/

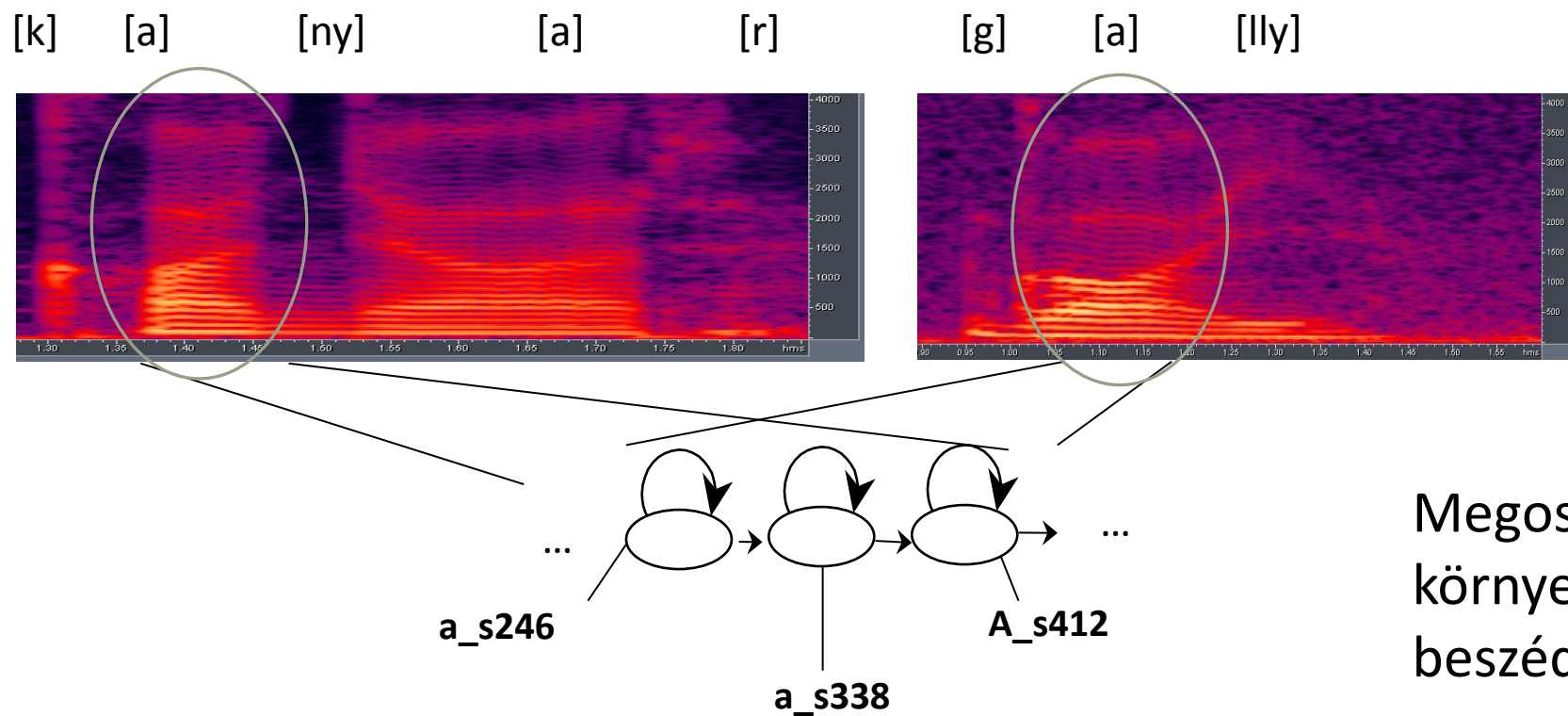
Szóadarabolás: szabály és/vagy statisztikai alapon

Egybeolvadás, hasonulás stb. modellezése

*é*rtsd [te|ezt] \rightarrow *é*rtsd [te|ezt]



Koartikuláció modellezés ?



Megosztott állapotú
környezetfüggő
beszédhangmodellekkel

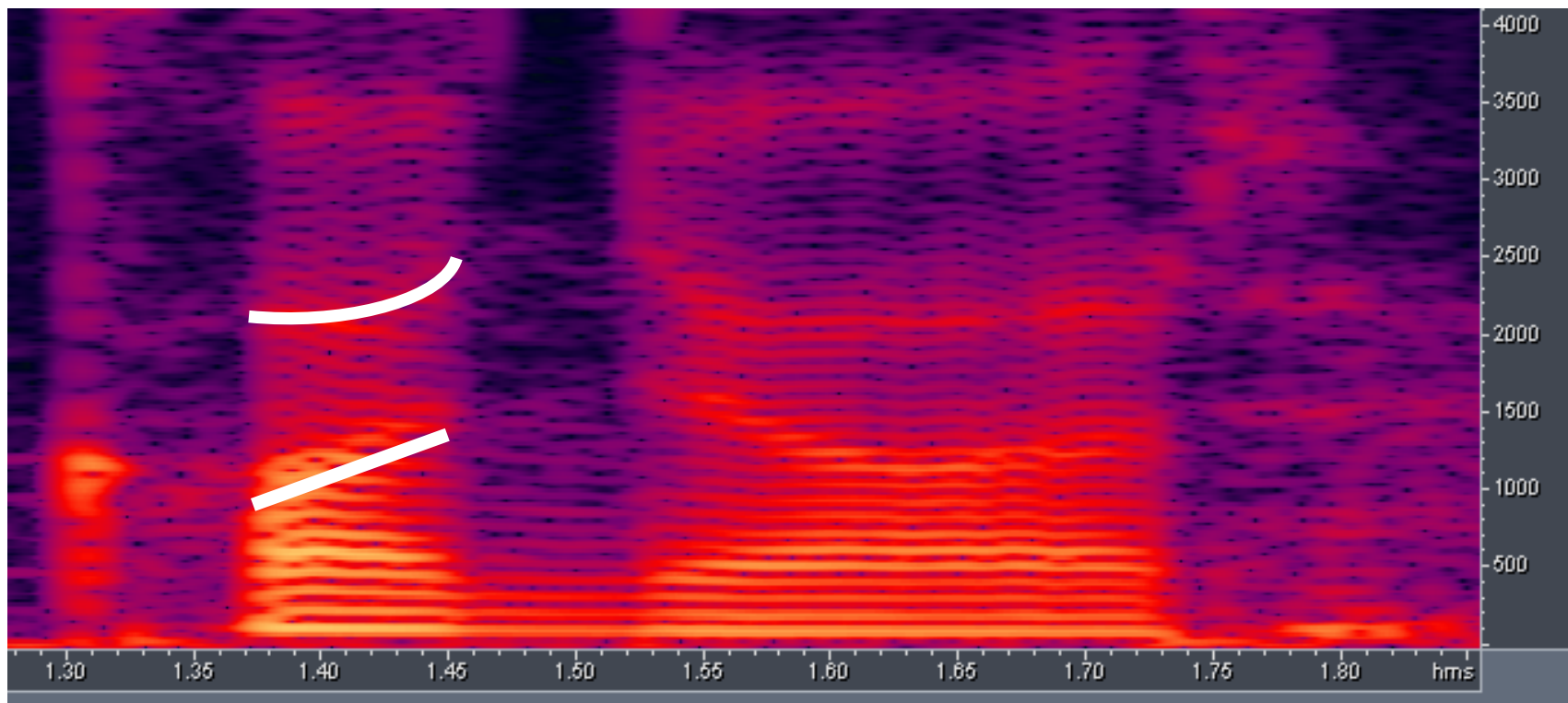
Kiejtési szótár



hagyja h a ggy a
hangya h a ny gy a

- ❑ Szabály alapon?
 - ❑ t + sz = c
- ❑ Kivételek?
 - ❑ Churchill = cs ö r cs i ll
- ❑ Kiejtési változatok? Gyakoriságok?
 - ❑ miért = m é r t, m i é r, stb...
- ❑ Ambiguitások?
 - ❑ Lachegeyi vs. Lachema, malacsült, meggyógyít...

[k] [a] [ny] [a] [r]

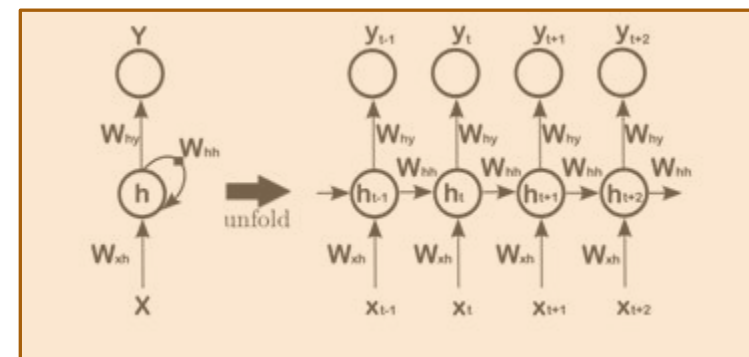


~~+ Δ ?~~

+ $\Delta\Delta$?

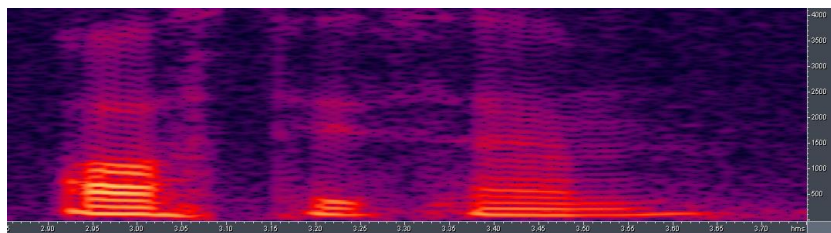
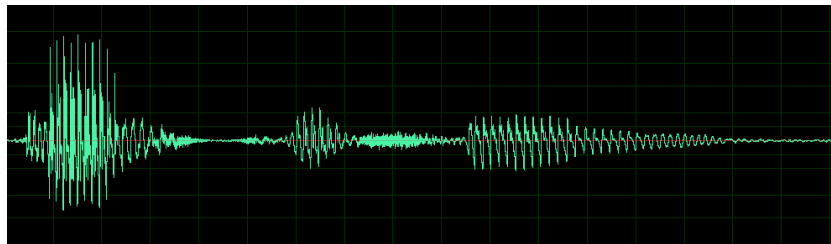
+ $\Delta\Delta\Delta$?

Beszéddinamika

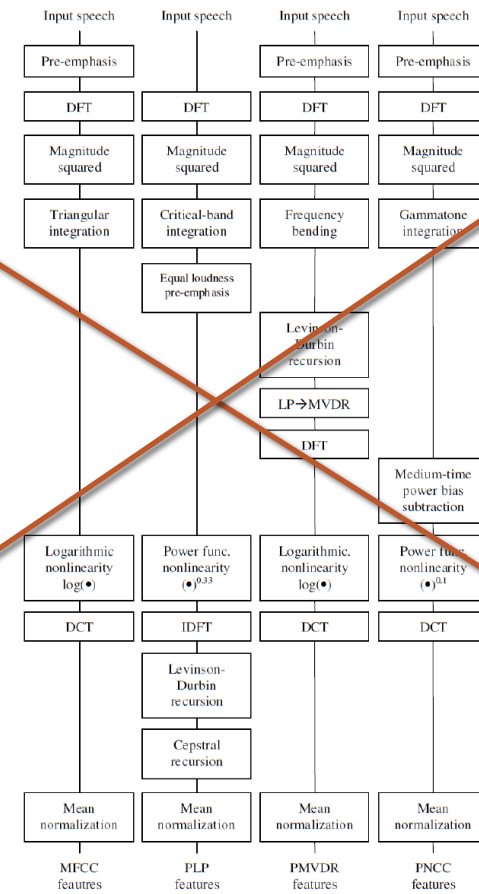


A hangjel megfelelő reprezentálása

„ a z t h i s z e m ”



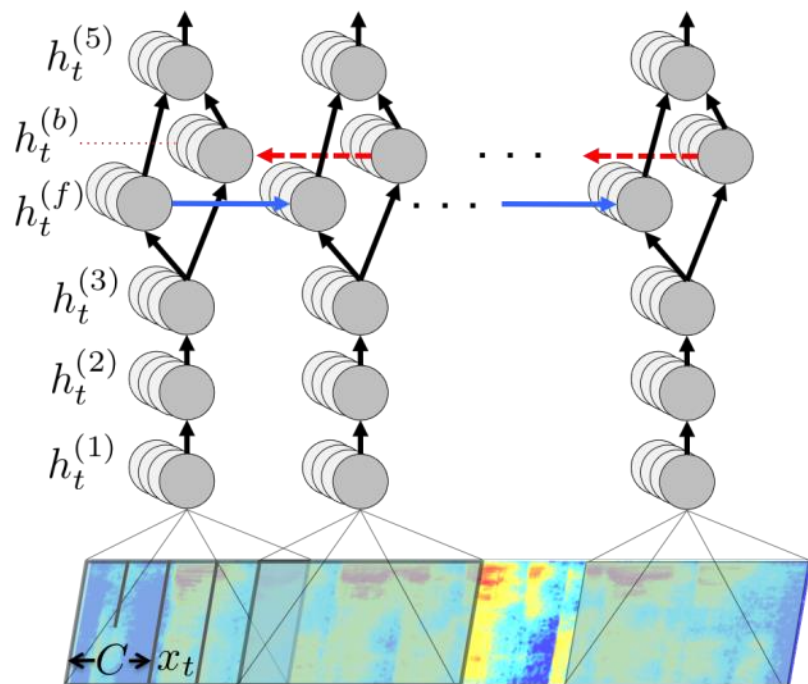
MFCC / LPC / PLP ...



A folyamatos beszédfelismerés jövője

$$\hat{W} = \arg \max_W P(W) \cdot P(O | W) ?$$

End-to-end (tisztán neuronhálós) megközelítés...



'80-as évek vége: NN

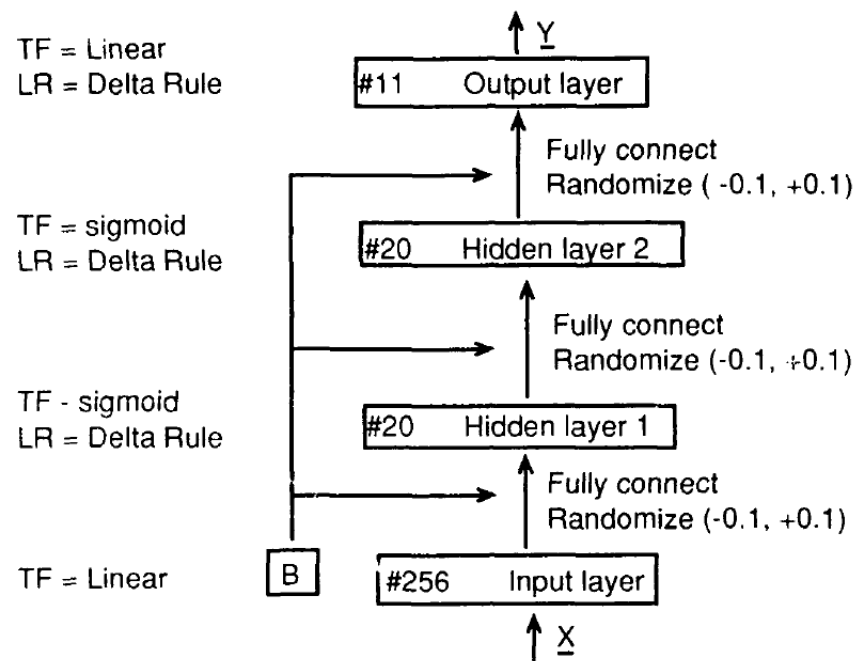
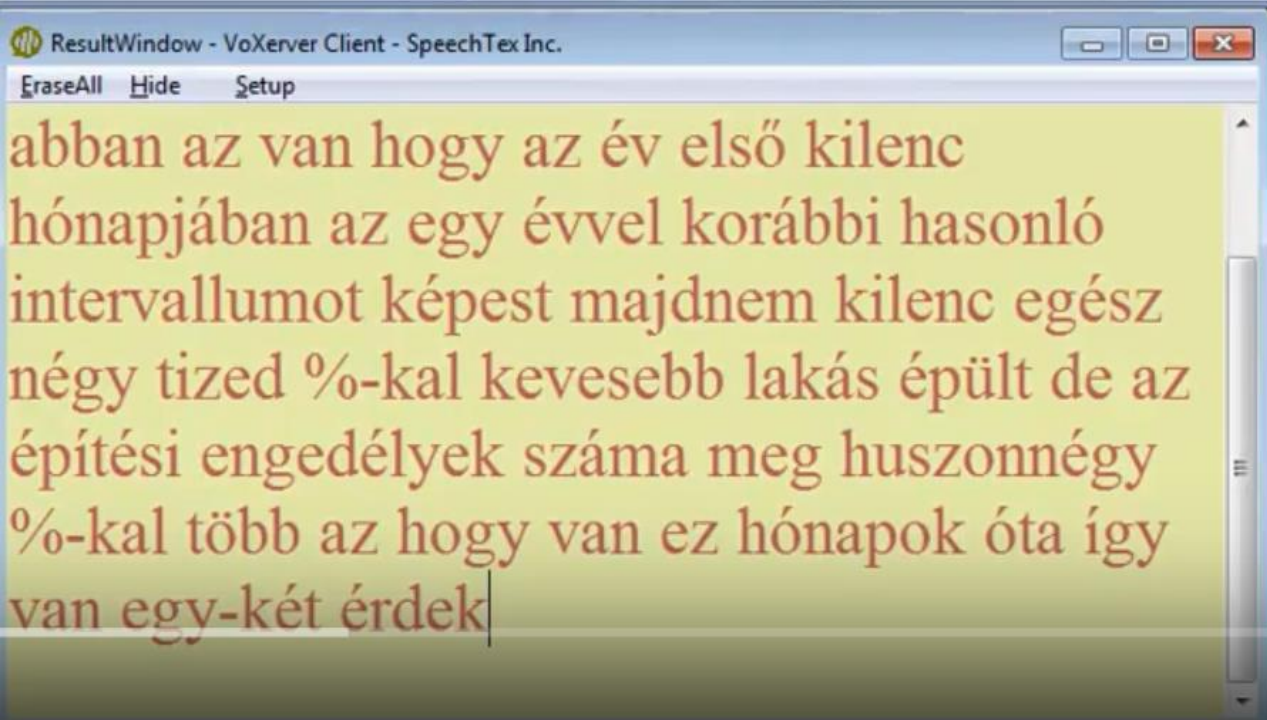


Figure 3. Back-propagation neural network topology



Jobb-e már a gép mint az ember?



https://www.youtube.com/watch?time_continue=1&v=p_oKK4xzZg8

Köszönöm a figyelmet!

SZÁMVÁLTOZÁS BEMONDÓ AUTOMATA

**A beszéd számítógépes
feldolgozása Magyarországon**
szakmai megemlékezés

Óbudai Egyetem 2018.szeptember 28.

dr. Eisler Péter HUNGAROCOM
Szabó András HUNGAROCOM
Németh Géza BME-HEI
Olaszy Gábor BME-HEI
Tihanyi Attila BME-HEI

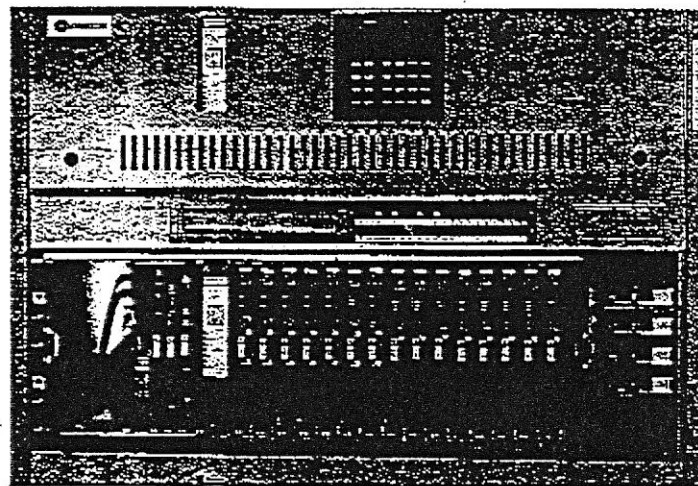
1993. június 1.

A hívott szám megváltozott...

Az új digitális központok üzembe helyezésével elkerülhetetlen a tömeges számváltozás. Ám hiába közlik az előfizetői körrel, hogy hívószáma megváltozott, azok, akik az ország távolabbi pontjairól, netán külföldről tárcsázták azt, nem tudják mire vélni a sikertelenséget. E meddő próbálkozások ráadásul olyan mennyiségű többletorgalmat generálhatnak, ami az amúgy is túlterhelt központok üzembiztonságát veszélyezteti. Gazdaságossági okokból olyan egyedi készülékek szöbeba sem kerülhetnek, amelyek bemondják a megváltozott telefonszámot.

E probléma megoldására dolgoztak ki intelligens szövegbe-

mondó rendszert a Hungarocom Híradástechnikai Kft. szakemberei. A központban elhelyezett IMS (Intelligent Message System) érzékeli, ha a hívott szám megváltozott, megkeresi az új számot, és három nyelven bemondja azt a hívó félnek. Az intelligens szöveg-bemondókat célszerű a központok azon egységeire kapcsolni, ahol a későbbiekben – amikor már nincs szükség a változás közlésére – szolgáltatásbővítésre használhatók. Ilyen értéknövelt szolgáltatás lehet például: Ne zavarj! – hívásátírányítás fix szövegre; hívásátírányítás kiválasztható szövegre; hívásátírányítás egyedi, összeszerkeszthető szövegre. IMS segítette a telefonforgalom zök-



Digitális vonalcsatlakozó a Hungarocomtól

kenőmentes lebonyolítását Szegeden, Újpesten és Szolnokon, a következő készüléket Kaposváron helyezik üzembe.

A Hungarocom mintegy 30 fős csapata a digitális központok telepítésével telmerülő műszaki

problémák megoldására fejleszt berendezéseket. Néhány termék a kínálatból: elektronikus kódadó-vevő, elektronikus forgalom-mérő, digitális vonalcsatlakozó, koncentrált előfizetői tarifaadó.

M. J.



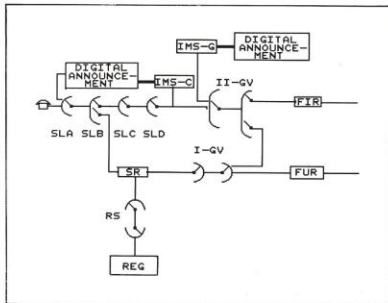
INTELLIGENS SZÖVEGBEMONDÓ RENDSZER (IMS) INTELLIGENT MESSAGE SYSTEM IMS

A hálózat dinamikus fejlesztése során, az új digitális központok üzembehelyezésekor elkerülhetetlenek a tömeges számváltások. A számváltásokkal kapcsolatos tájékoztatások egy adott helyiségen belül még kezelhetők, azonban az időben és térben szétvált üzembehelyezések kapcsán történő számváltásokról országosan szinte lehetetlen tájékoztatást adni. A tömegesen megváltozott számok hívása olyan mennyiségű többletforgalmat generálhat, amely az amúgyis túlterhelt központok üzembiztonságát veszélyezteti. Ha csak a számváltás tényéről tudjuk a hívó előfizetőket, úgy a tömeges tudakozás aligha megoldható, tömeges számváltásánál az egyedi jellegű, a megváltozott hívószámot tájékoztató bemondó eszközök gazdaságossági okokból szóba sem kerülhetnek. Problémaként merül fel továbbá, hogy a számváltásokat követő viszonylag rövid idő után ezek az eszközök feleslegessé válnak.

Ezért olyan megoldás kidolgozása vált szükségessé, továbbá a központok azon egységeire célszerű kapcsolódni, amely a későbbiekben lehetővé teszi, hogy a telepített berendezéseket a központok szolgáltatásbővítésére lehessen felhasználni. Az intelligens szövegbeszélő rendszer (IMS) megfelel a fenti célkitűzéseknek.

Tömeges számváltások bementésére két változat kerül kifejlesztésre. Gyakran előforduló eset, hogy a hagyományos központokban teljes ezres nyálábok száma meg. Erre az esetre alkalmazható az IMS-G rendszer, amelynek kapcsolódását az ábra mutatja. Amikor a központba bejövő hívás megazúnt nyáláb felé irányul a GV fokozat felkapcsolja a berendezést, amely MFC jelzések segítségével kikéri a hívott előfizető hívószámának utolsó három számjegyét. A rendszer az adatbázisból kikeresi a megváltozott hívószámot, amit a digitális szövegbeszélő közlő a hívó fél felé.

Amennyiben az ezres előfizetői fokozaton belül van jelentős számváltás, az IMS-C rendszer kerül alkalmazásra, amelynek kapcsolódását az ábra mutatja. A berendezés az előfizetői fokozat vezérlő egységeihez csatlakozva kikéri a hívott előfizető hívószámát, majd amennyiben úgy értékeli, hogy a szám megváltozott, az adatbázisból kikeresi a megváltozott hívószámot, amit a digitális szövegbeszélő közlő a hívó fél felé.

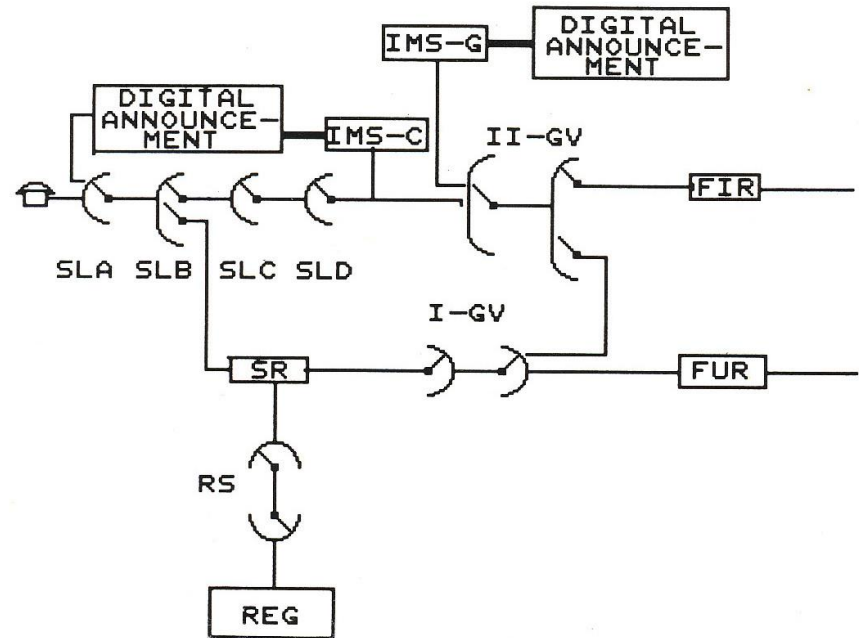


In the course of the dynamic expansion of the network the change of the directory number of the subscribers in a large scale is inevitable. The information regarding the changed numbers within a location is still manageable but in case of changes in connection with putting into service new exchanges all over the country the provision of information in a traditional way is almost impossible. Calls towards changed numbers generate excess traffic which can overload the traditional exchanges. If only the fact of the change of the directory number is announced it draws a mass demand for inquiry services which excess demand can not be easily handled. The application of individual announcers per line is uneconomical. That is also a problem what to do with these devices after the period when the new directory numbers are already evident.

To avoid the above mentioned problems the Intelligent Message System (IMS) has been developed which besides solving the problem of mass inquiry assures the future use of the equipment when providing new value added services for the subscribers.

For the announcement in case of large scale of changes in the directory numbers two versions has been developed. It often occurs very often that complete exchange branches are changed. For this purpose the IMS-G system is used as shown in the figure. When the call is connected to such a direction the group selector stage connects the inlets of the system, which requests the last digits of the called subscriber's number by means of MFC signaling. The system selects the new directory number from the data base and communicates it towards the calling party by means of a digital speech announcer.

In case of large scale of changes within a subscriber stage the system IMS-C can be applied. The equipment which is connected to the control part of the subscriber stage requests the digits of the called subscriber. In case of change of the received number the new directory number of the subscriber is selected from the data base which is transferred to the calling party by means of the digital speech announcer.



Feladat: „futási időben” dinamikusan összeállított mondatok bemondása „sok” csatornán

Jól használható technológia: formáns szintézis

Előnyök az adott feladat szempontjából:

- Rendkívül kis memória igény (1-2 kbit/sec) (1993-ban 16Kbit-256Kbit) méretű EPROM-ok álltak rendelkezésre
- A fenti tulajdonság lehetővé teszi a sok csatornás stream-elést (wave table – ből történő sok csatornás stream-elés az akkori CPU és memória sebességek mellett még nehézkes volt)
- Az „összeolvasott” szöveg részek hangzása homogénebb, mint ha külön-külön felvett hangmintákból raknánk össze a mondatot. A mondat végi intonáció is megoldható volt
- A BME által fejlesztett PC-hez csatlakoztatott formáns szintetizátor eszközzel az egyes szöveg részek text to speech módon könnyen előállíthatók és editálhatók voltak, nem kellett időt tölteni a hangminták felvételével és azok utólagos editálásával/vágásával

Szövegbemondások elkészítésének menete:

- Szóelemek, illetve kifejezések előkészítése PC-n text to speech módon, „finomítás”
- Szóelem és kifejezés táblázatok előállítása, csatolás a működtető szoftverhez
- Szövegbemondás valós időben futási időben összeállított szóelem illetve kifejezés sorrend alapján

Megvalósított hardware: VSS (Voice Synthesis Subsystem)

VSS kártya specifikáció:

- 16 független bemondó csatorna (16 x PCF 8200 chip, dupla euro méret)
- CPU: 80C186 12MHz
- Memória: 256 Kbyte RAM, 64Kbyte EPROM
- Vezérlés: RS 422 (8250 UART) 9600 bit/sec
- Analóg kimenetek: 16 darab, 600 ohm, szimmetrikus

Lehetséges vezérlések:

- Magyar nyelvű számváltozás bemondás
- Szóelem bemondása egyszer
- Kifejezés bemondása kétszer ismételve
- Számváltozás bemondás, 1 magyar, két orosz
- Számváltozás bemondás, 1 magyar, 1 német, 1 angol
- Üres emelet hang, 8-szor ismételve
- Számváltozás bemondás 1 magyar, 2 német
- Számváltozás bemondása 1 magyar, 2 angol
- Szóelem bemondása kétszer ismételve



M Ű E G Y E T E M 1 7 8 2



Egy gimnazista útja a HUNGAROVOX – RUSSON –ig a BEAG-ban (és tovább)

Németh Géza

Beszédkommunikáció és Intelligens Interakciók Laboratóriumok

BME Távközlési és Médiainformatikai Tanszék

A BESZÉD SZÁMÍTÓGÉPES FELDOLGOZÁSA MAGYARORSZÁGON

NJSZT ITF

2018. SZEPTEMBER 28..

SmartLab
Intelligent Interactions

<http://smartlab.tmit.bme.hu>

 NVIDIA

GPU
EDUCATION
CENTER

Út a beszéd-szintézishez

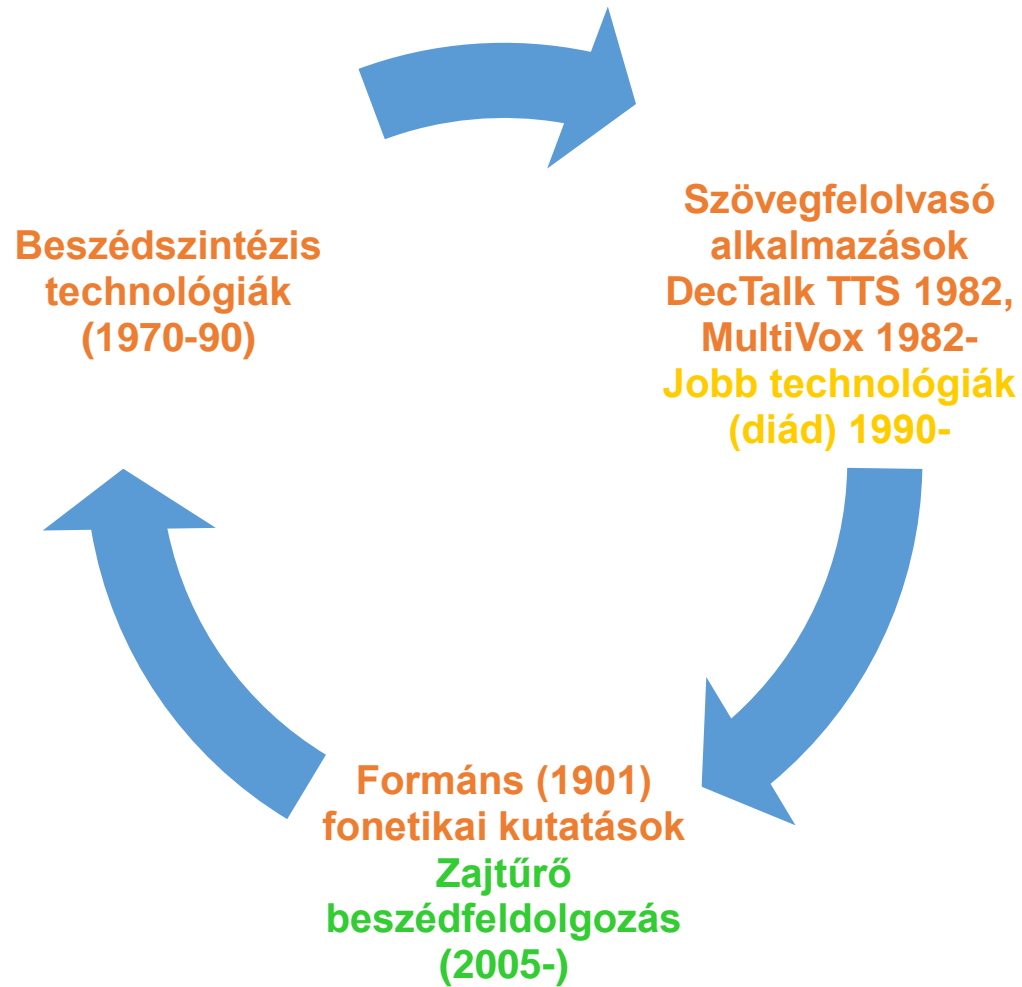
- 1977 érettségi + felvételi (de hova???)
- BME Vill. -> hangmérnökség, mint kompromisszum
- 1 év katonaság
- 2 év nagyon alap képzés
- TDK témakiírás -> Gordos Géza – Ferenczy Pál
- 18 órás az osztályon (Podoletz György, Kovács Pál, Ambrus Sándor, Bárányné Sülle Gabriella és még sokan mások)
- Beszéddetektor, beszéd-zaj adatbázis
- Diplomaterv – VoxAlarm (Takács György) - Hogyan tovább???



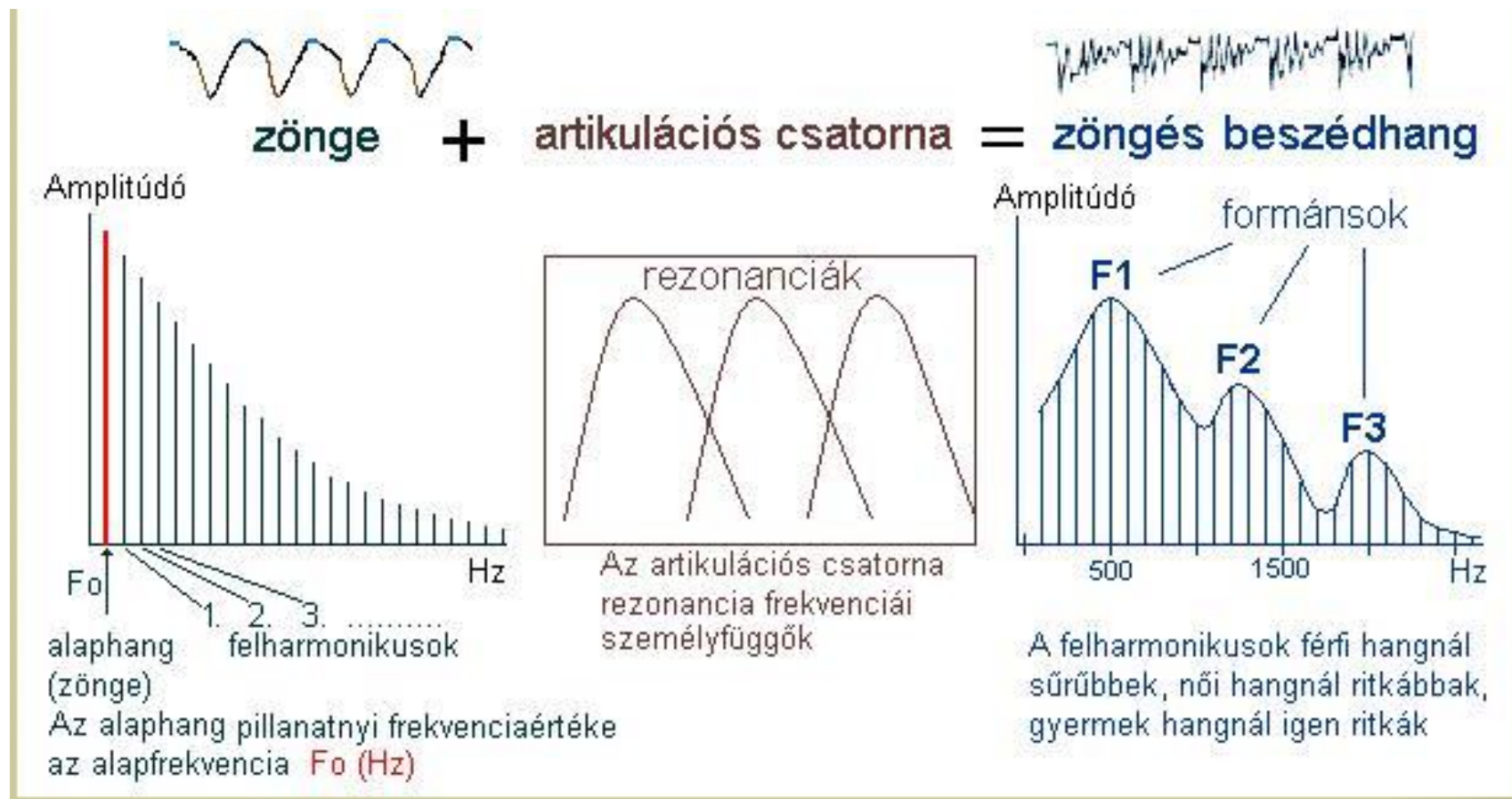
Út a BEAG-ba

- Szakmérnöki képzés, mint kutatói pálya lehetőség
- Kiinduló téma: digitális jelfeldolgozás -stúdiótechnika
- Szakdolgozat: MEA 8000 fejlesztői rendszer
- Végzés után:
 - Stúdiótech főosztályvezető: „Az egyetemről úgysem jön semmi használható”
 - AHFF főosztályvezető (Balogh Géza) felvett beszédtechnológiai fejlesztésekhez.
 - Laborvezető: Vinkovits Sándor, majd Bálint Zoltán
 - Beszédfelismerő és szintetizátor fejlesztés (CP/M Syster)

Alap kutatás (formáns) ²

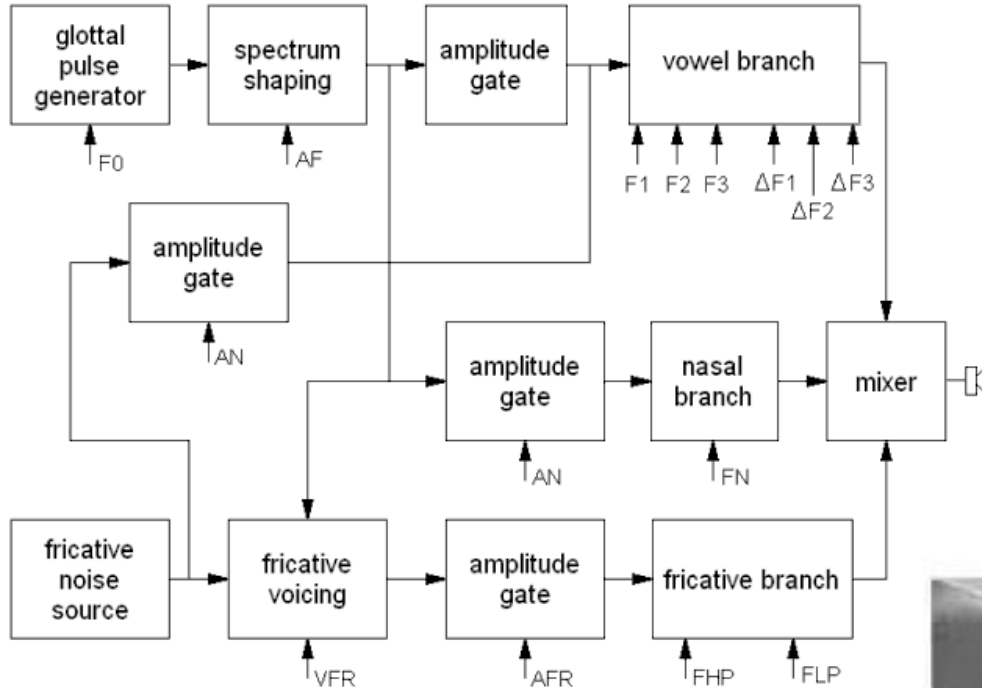


Alap kutatás (formáns 1791-) ¹



Előzmények

Kempelen Farkas 1791



HungaroVox 1982



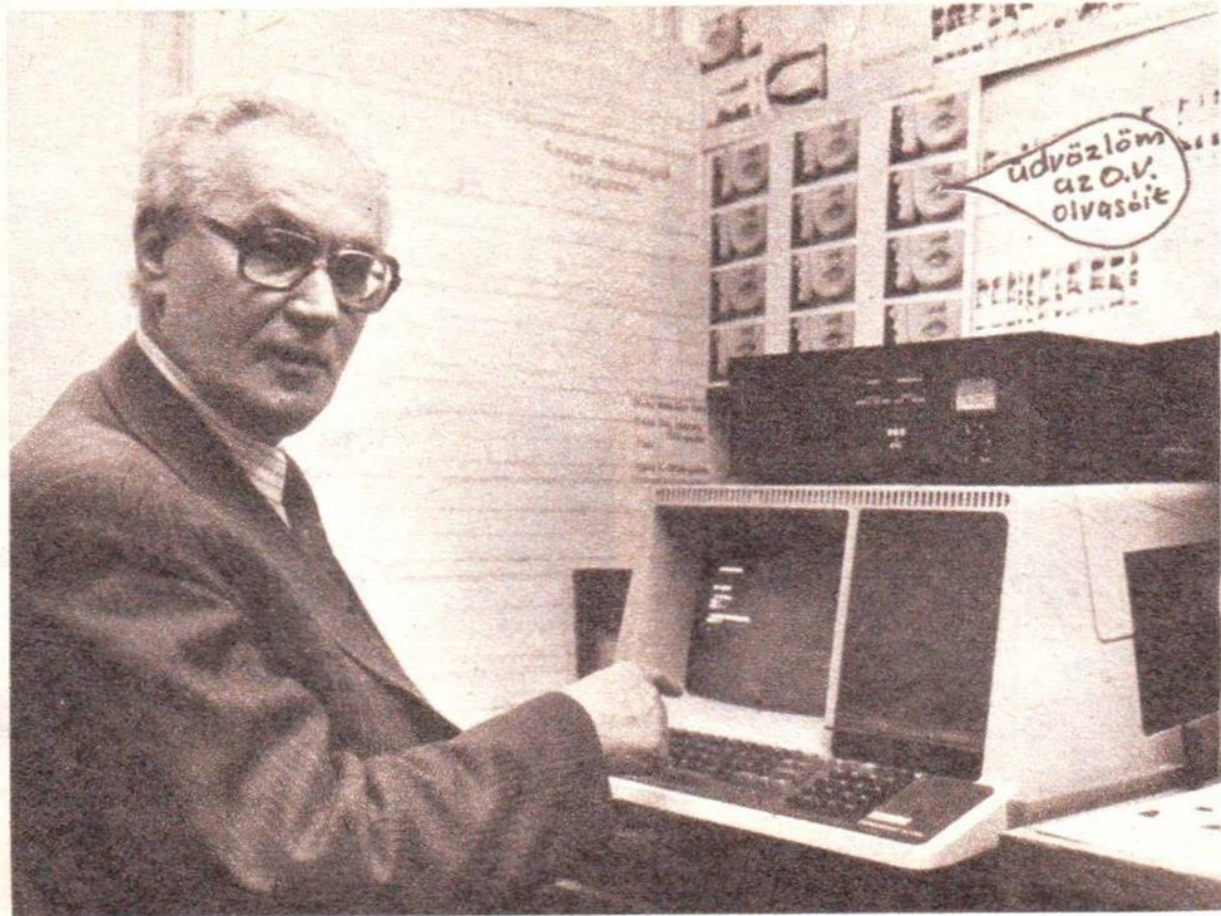
Magyarul és oroszul szólal meg a Voxton – a beszélőgép

A helyiségben csak *ketten* vagyunk. Mindketten hallgatunk. A *harmadik* beszél!?!

Nem sajtóhiba: a harmadik ugyanis nem személy, hanem *gép*, amelyik *beszél*ni tud. Neve: *Voxton*. A Magyar Tudományos Akadémia Nyelvtudományi Intézetének fonetikai osztályán *dr. Bolla Kálmán* osztályvezető mutatja be a magyarul és oroszul tudó beszélő rendszert.

A világon magyarul először megszólaló gép folyamatos szövege a következő volt:

„Figyelem, figyelem! Jó napot kívánok! A Voxton köszönti hallgatóit. Magyar nyelvű szintetikus beszédet hallanak. A Voxton elnevezés a vox és a ton szavak összevonásából származik. Elődje a Voxon, a vox- és szonusból kapta a nevét. A Voxonnal — az első magyar nyelvű személytelen

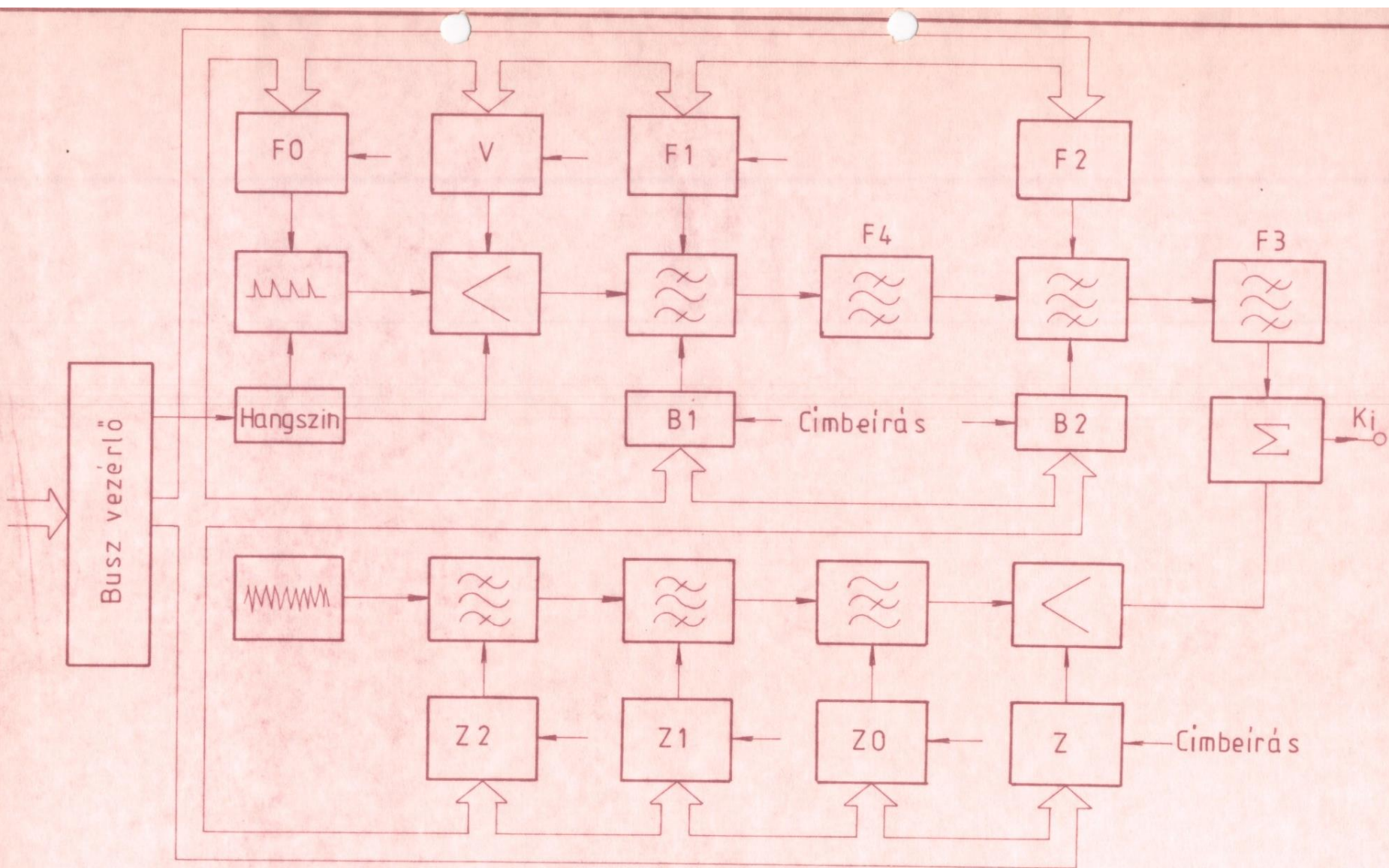


HungaroVox és Russon

- Azonos hardver (VOX / CYA 301) formánszintetizátor
 - MTA NYTud licenz
- Vezérlés számítógéppel Centronics porton
 - SZTAKI licenz
 - HungaroVox: Olaszy Gábor + Kiss Gábor
 - Russon: Bolla Kálmán + Kiss Gábor
- Komplet hardver újratevezés, számítógépes módszerek bevezetésével (pl. NYÁK tervezés)
- Érthetőségi tesztek

Még
Moszkvába
is eljutott





CYA-301 tip.VOX beszédszintetizátor blokkvázlata

1. ábra

VOX beszédszintetizáló

Műszaki leírás

Az ipari felhasználásra kifejlesztett valósidejű magyar nyelvű beszédszintetizáló rendszer a MTA Nyelvtudományi Intézetben folyó beszédakusztikai és szintetizálási kutatások egyik eredménye. Ez az első olyan beszédszintetizáló rendszer, amelyben a beszéddé átalakítani kívánt szöveget a magyar helyesírás szerint kell megadni és a mikroszámítógép ebből a betűsorozatból állítja elő a folyamatos köznyelvi beszédet. Hangja, beszéde nem emberi alapú, tehát hangszíneben sem hasonlít valamely személy hangjához. A beszédet előre megtervezett akusztikai építő elemek sorozatából állítja elő. A szótárnélküli, automatikus beszédelőállításához ilyen, a nyelv akusztikai szerkezetét leíró elemtárat, elembázist kell képezni. Ez az elemtár 370 akusztikai építőkövet /hangszeletet/ tartalmaz, a beszédelőállítás során ebből az elembázisból válogatja ki a számítógépprogram az éppen aktuális hangsor felépítéséhez szükséges építőelemeket. Ezeket sorrendbe állítja, majd az építőelemekben megadott bitkombinációk sorozatát a program közli a beszédgenerátorral és a beszéd hallható lesz a hangszóróban. Az elembázis egy-egy eleme a hangszelet, amely meghatározott időtartamú, frekvencia- és intenzitás szerkezetű elemi beszédhangrész, vagy hangkapcsolódási szakasz. Egy-egy hangszeletnek nyelvi szempontból sem önálló jelentése, sem értelmes hanghatása nincs. A magyar beszédre az egyes hangszemek paramétereit és paraméter értékeit az MTA Nyelvtudományi Intézetben végzett kutatások alapján állapították meg és kódolták be a számítógép adattára számára. A magánhangzókat és az egyszerű szerkezetű beszédhangokat általában három hangszellettel jellemezzük. Ez a három hangszelet a következő hangfázisokat reprezentálja:

- az előző hanghoz való csatlakozás szakasza,
- a tiszta fázis szakasza,
- a következő hanghoz /vagy csendhez/ való csatlakozás része,

Műszaki adatok:

- frekvenciatartomány: 70-7000 Hz
- interface Centronics
- üzemi hőmérséklettartomány: 0-50 °C
- táplálás 220 V váltakozó feszültségről elő-
állított stabilizált +5V, +12V, -12V
- teljesítményfelvétel 22W
- névleges súly 5 kp
- névleges méretek:
- magasság 80 mm
- szélesség 430 mm
- mélység 290 mm
- érintésvédelmi osztály
beszerelt állapotban I.
- relatív páratartalom 70 %
- külső hangszóró impedancia 8-16 ohm
- külső hangszóró terhelhetőség 2 W

Mi lett belőle?

- Néhány prototípus készült (pl. érintésvédelmi jóváhagyáshoz)
- A rendszerváltozás elsodorta a BEAG-ot
- A műszaki fejlődés elsodorta a diszkrét elemekből építkező megoldásokat
- Jöttek az egychipes megoldások

MultiVox

1986-2002

Beszélő óra (1988)



Ma: BME TMIT SmartLabs: 3 labor

[Beszédtechnológia és Intelligens Interakciók Labor](#) (Németh Géza)

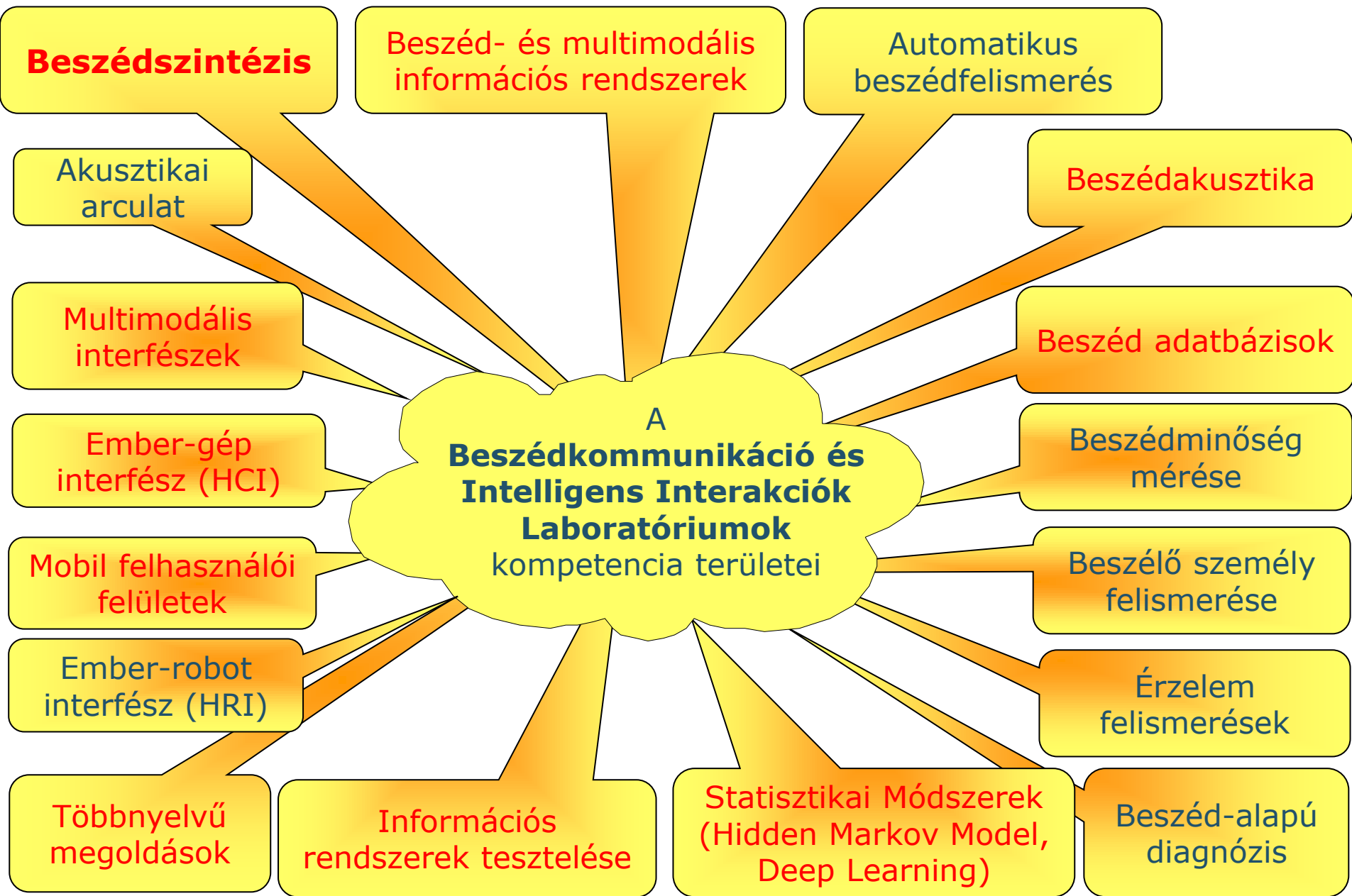
[Beszédfelismerés és Hangbányászat Labor](#)
(Mihajlik Péter)

[Beszédakusztikai Labor](#)
(Sztahó Dávid)

~20 munkatárs, 5 állami
finanszírozású (2 DSc, 9 PhD)



www.ai4eu.org



SmartLab munkatársak



Németh Géza
PhD 1997, Habil 2013
(Laborvezető)



Olaszgy Gábor
DSc 2003



Zainkó Csaba
PhD 2010



Gyires-Tóth Bálint Pál
PhD 2013



Mohammed Al-Radhi
PhD hallgató



Csapó Tamás Gábor
PhD 2014



Bartalis Mátyás
Msc



Nagy Péter
PhD jelölt



Laczkó Klára
asszisztens



Sevinj Yolchuyeva
PhD hallgató



Hajgató Gergely
PhD hallgató

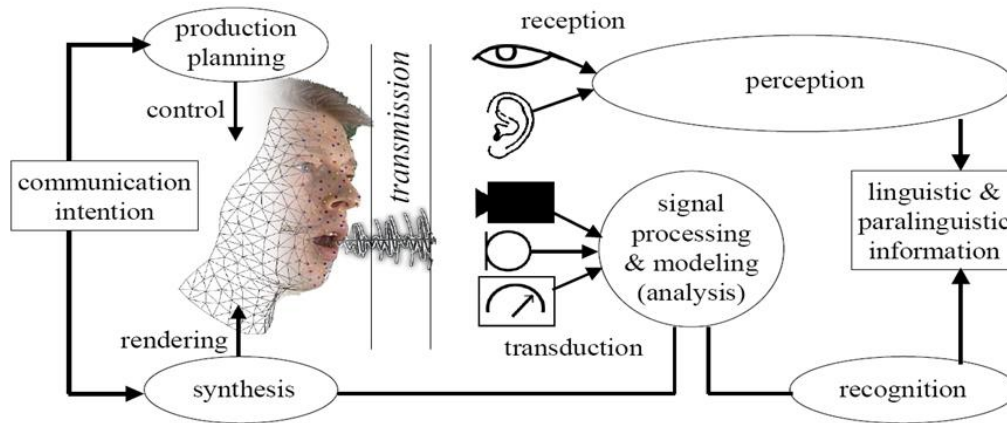


Hamdi Abed
PhD hallgató

SmartLab kutatási területek

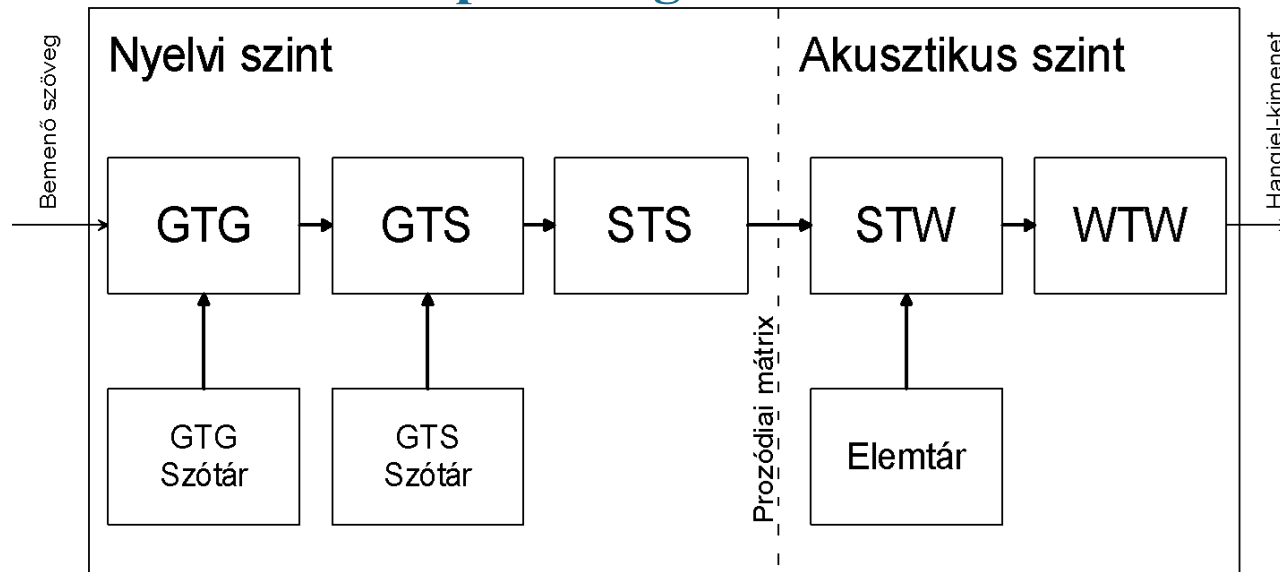
- Gépi szövegfelolvasás (text-to-speech, TTS)
 - Elemösszefűzéses és korpusz-alapú
 - Gépi tanulás alapú (Deep Learning, Hidden Markov-model)
- Beszédszintézis részproblémái
 - Parametrikus kódolás, gerjesztési modellek
 - Intonációs modellek
 - Rövid- és kérdő mondatok prozódiaja
 - Kommunikációs kontextus figyelembe vétele
 - 2D ultrahang-alapú artikuláció vizsgálat
- Ember-gép interakció
 - Humanoid robotok
 - Beszédkommunikációs segédeszköz
- Mély tanulás (Deep Learning)
- Alkalmazási lehetőségek

Mi is a beszédtechnológia?



**A természetes beszédlánc
bármely elemének gépi
megvalósítása
(interdiszciplináris
tudomány, AI???)**

Gépi szövegfelolvasás



Történelem

közlekedés és beszédtechnológia



1791



2015



Hawking gépi hangja angolul és magyarul

Dectalk 1982



ProfiVox 2000 – 2014



Mindenség elmélete film magyar szinkronhangja

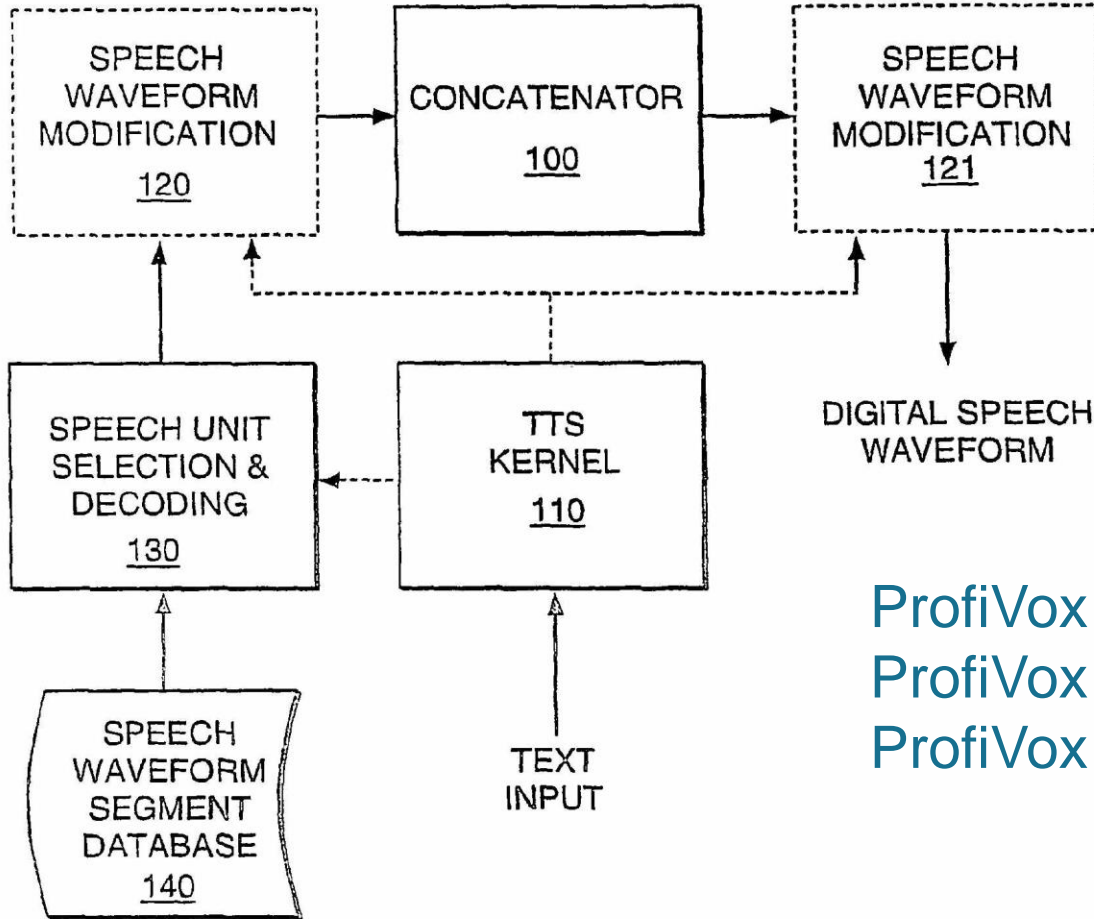
A fejlődés útja

A szabály-alapú modellek
(artikulációs csatorna, prozódia)

mellett és helyett

Természetes elemek
egyre nagyobb halmaza
statisztikai modellépítés
minimális jelfeldolgozás
Egységes(re törekvő) kiértékelés

Hullámforma összefűzés (termés)



Hírfelolvasás (2004)



Futball hírek

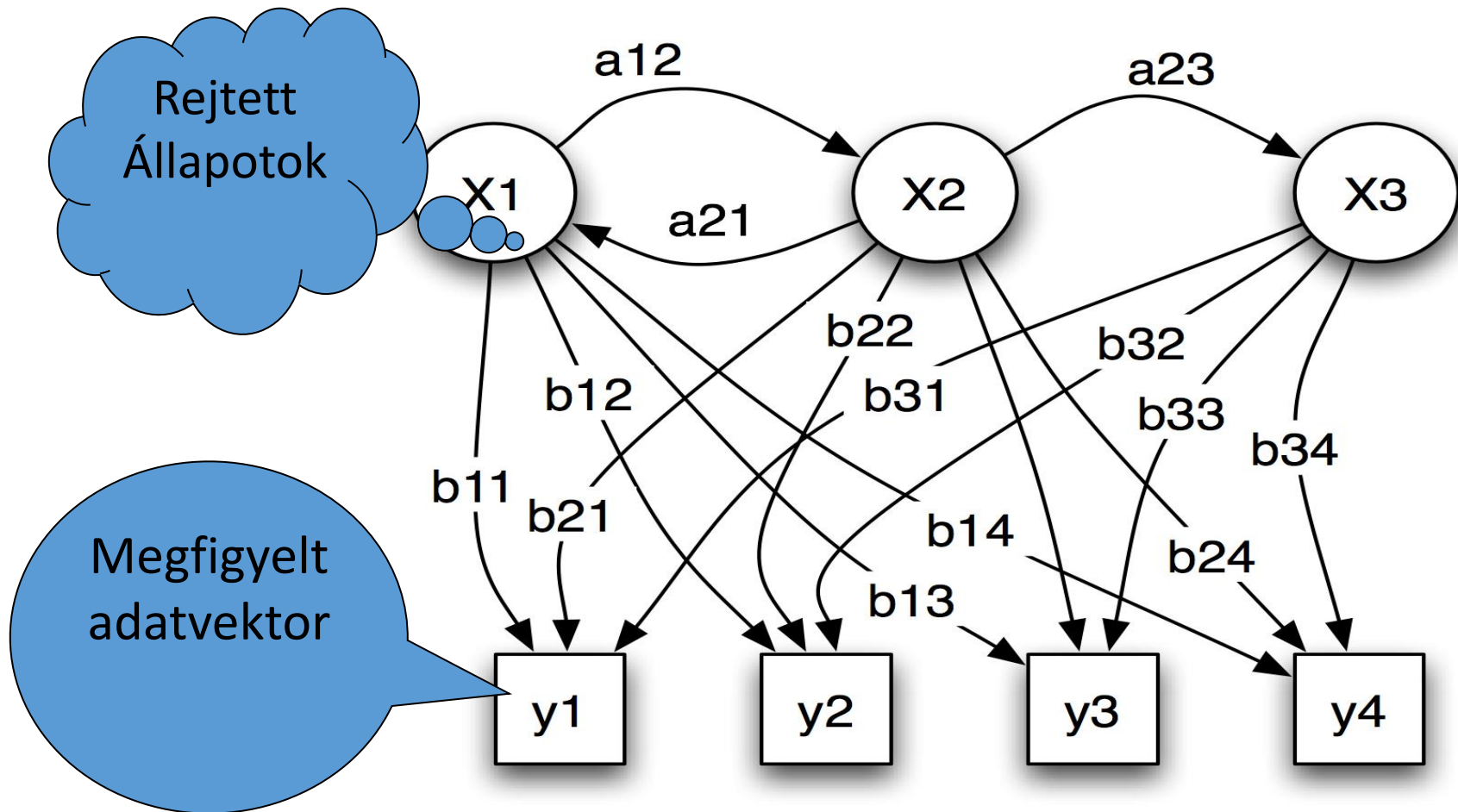
- ProfiVox diád 1995- 
- ProfiVox triád 2000- 
- ProfiVox korpusz 2002- 

Utas információ
(többnyelvű)

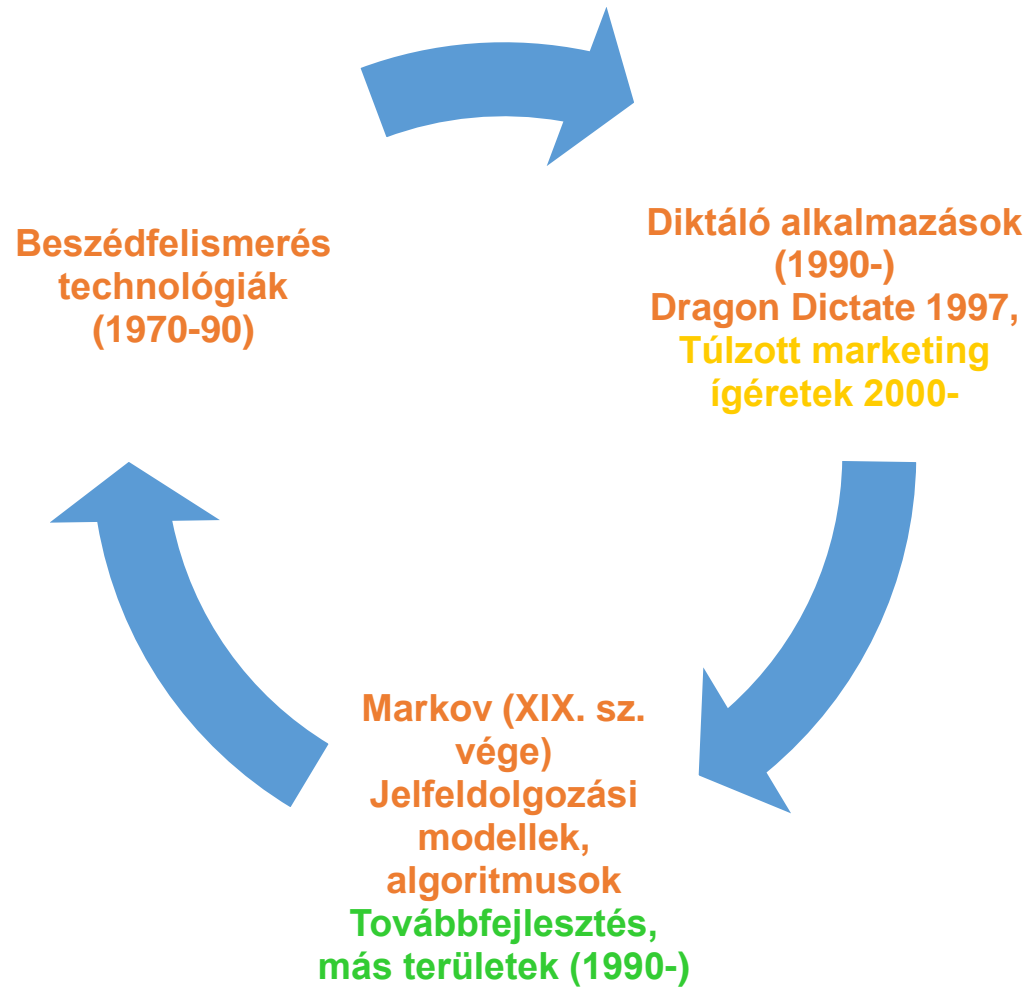


BLOCK 120 AND 121 ARE OPTIONAL IN CORPUS-BASED SYTHESIS

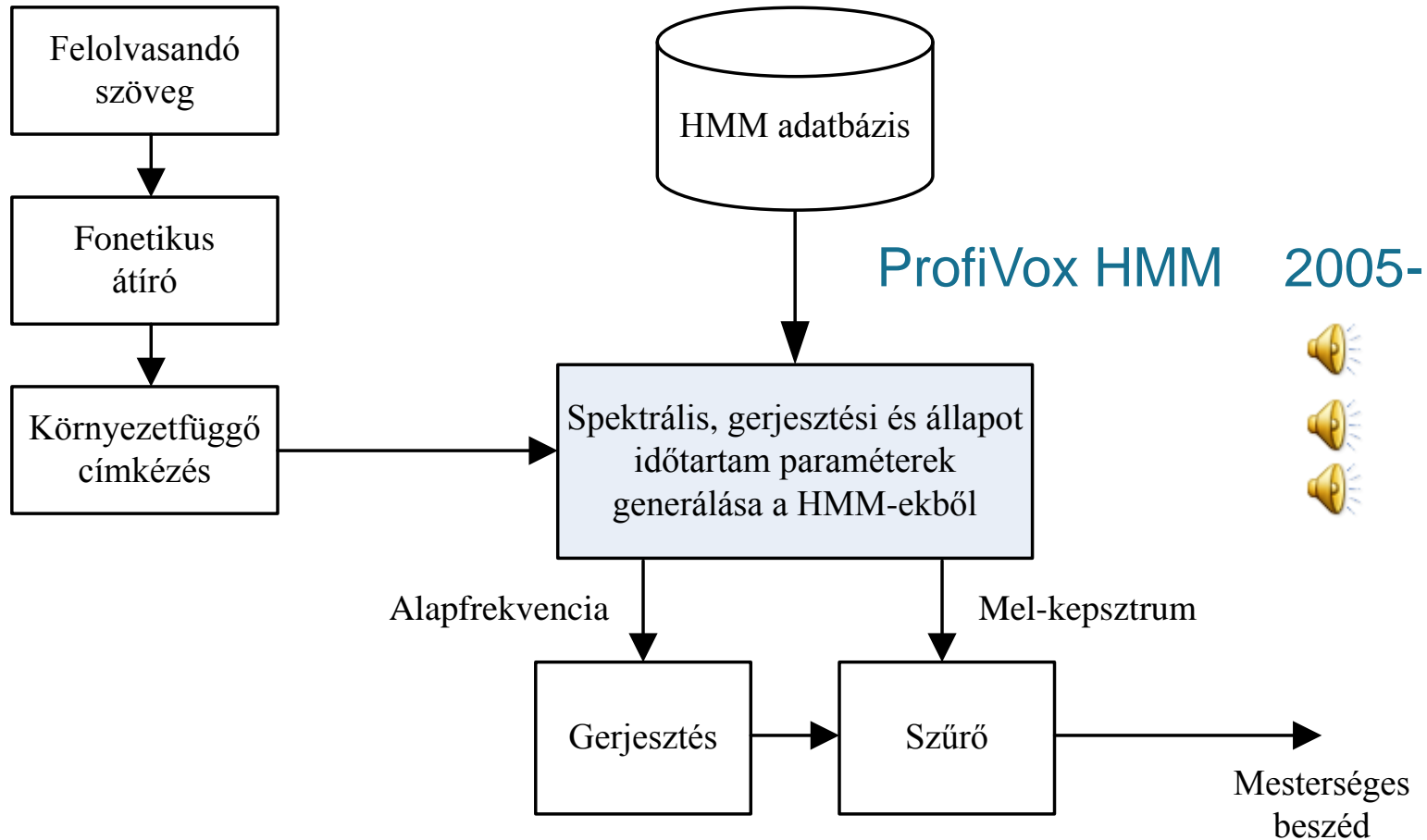
Alap kutatás (Hidden Markov Model, HMM 1970-) ¹



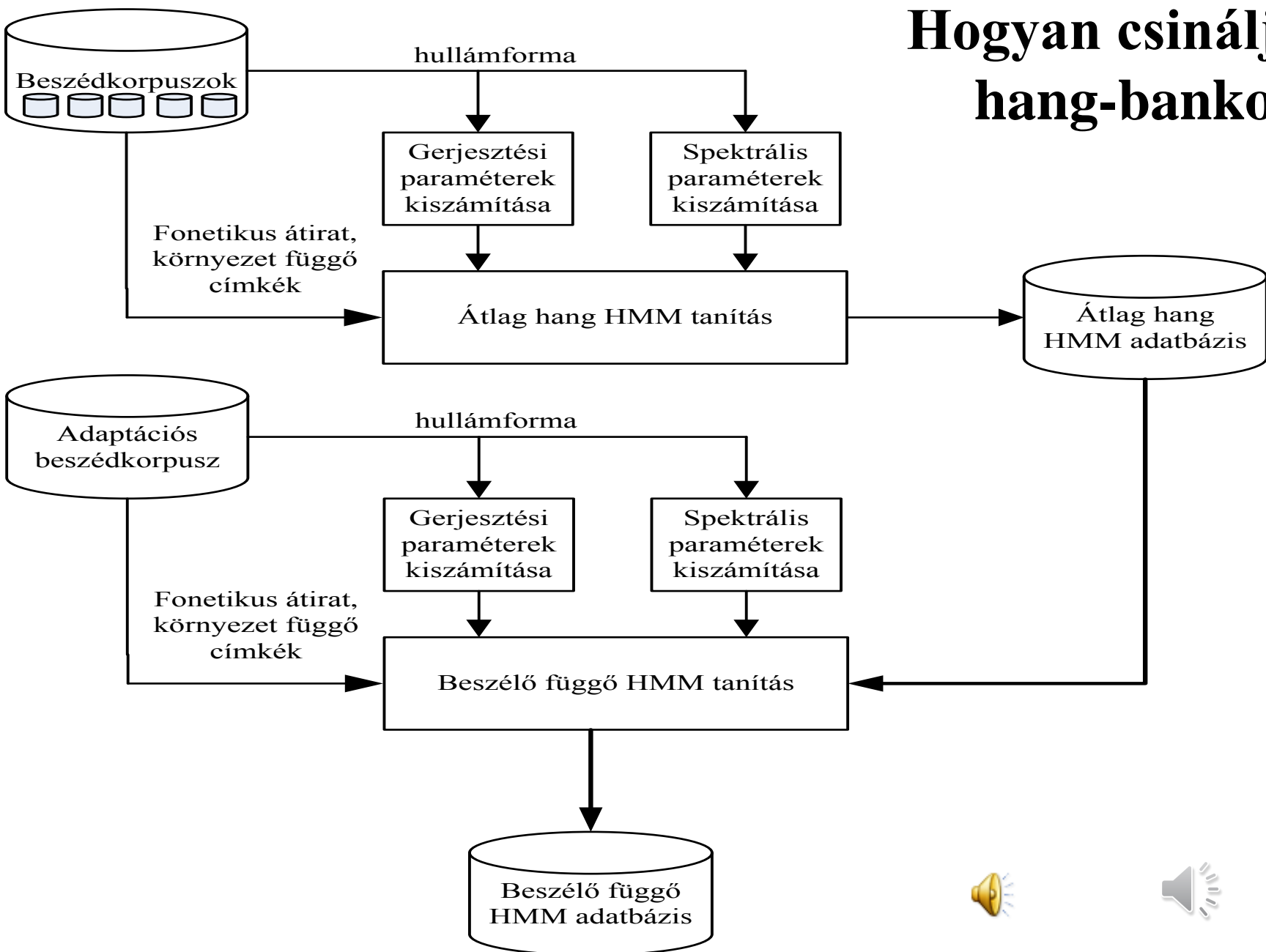
Alap kutatás (Rejtett markov modell, HMM) ²



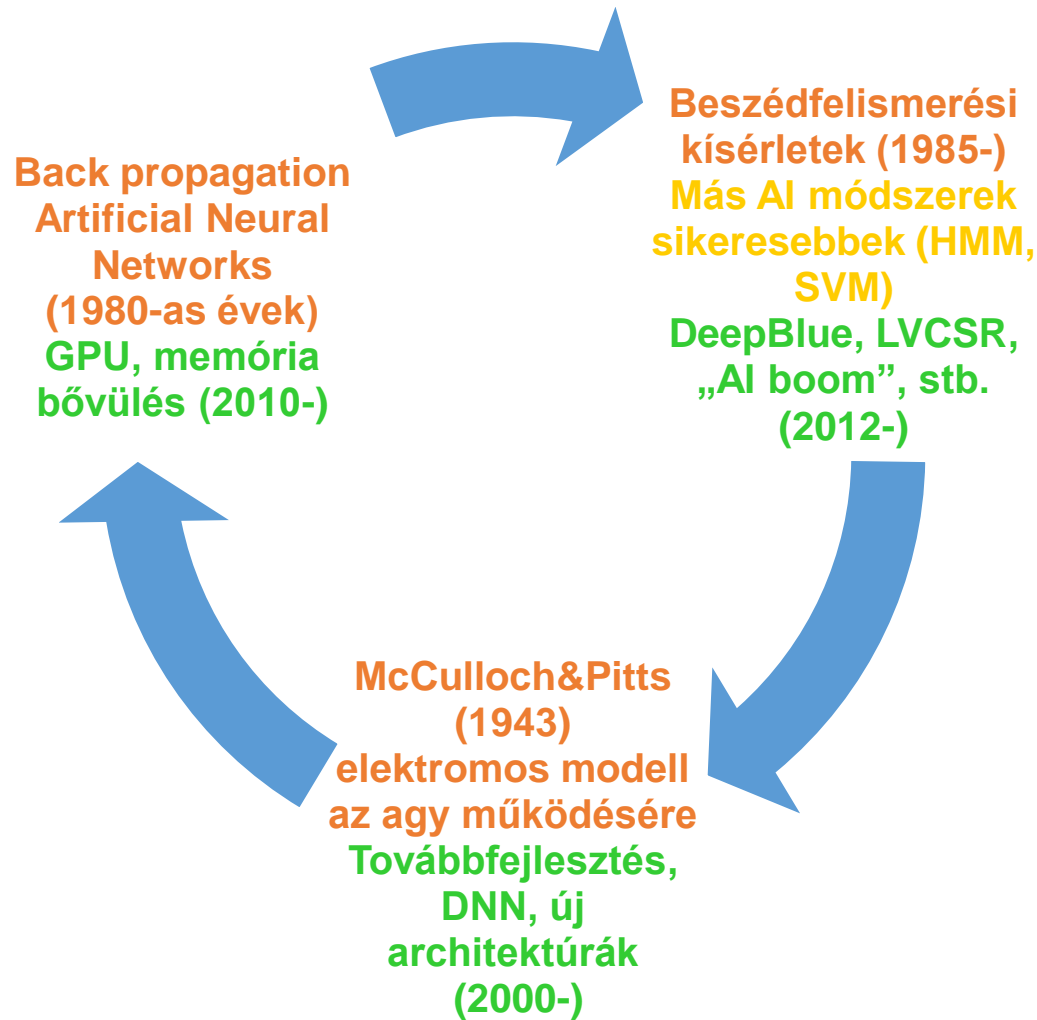
Technológia fejlesztés HMM-alapokon (rugalmasság, 200x-)



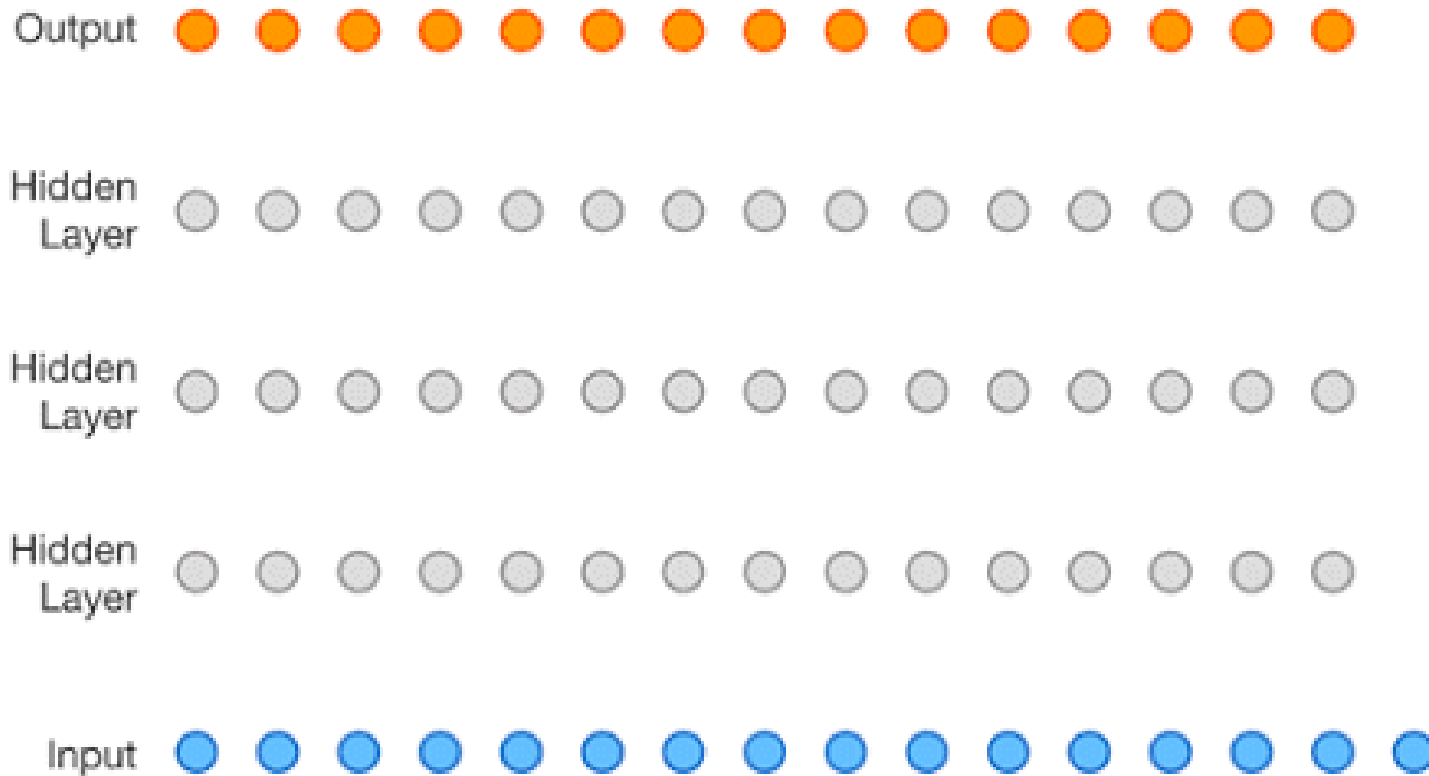
Hogyan csináljunk hang-bankot?



Alap kutatás (Neurális hálózatok)

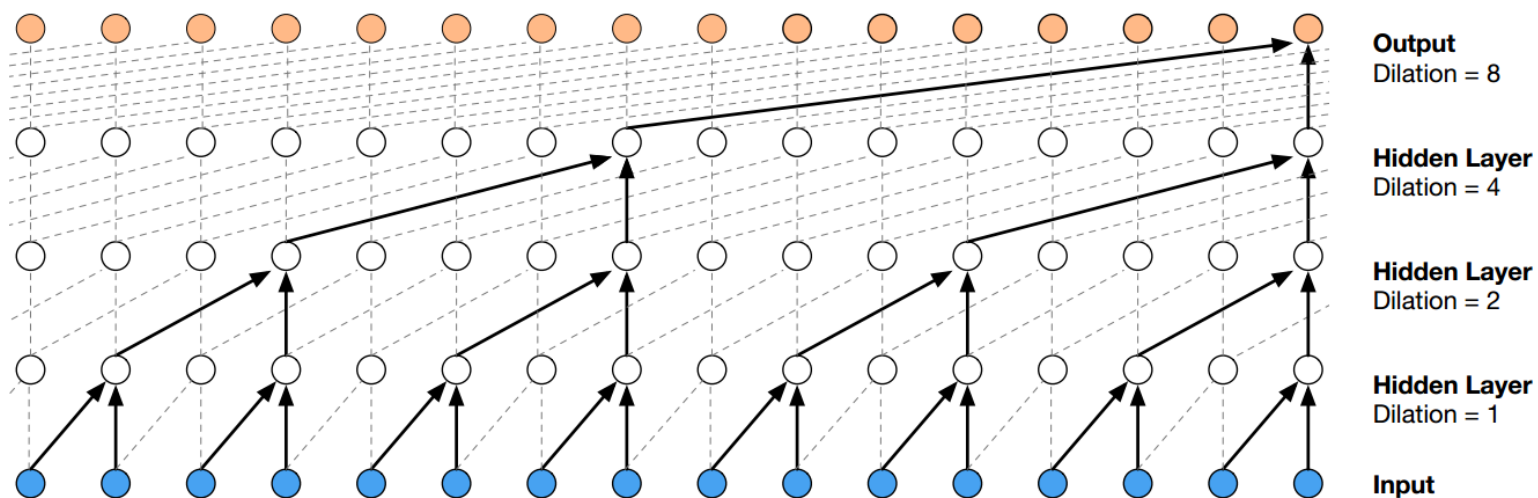


Wavenet (2016. szept.-)



Ábra forrása: <https://deepmind.com/blog/wavenet-generative-model-raw-audio/>

Generálás

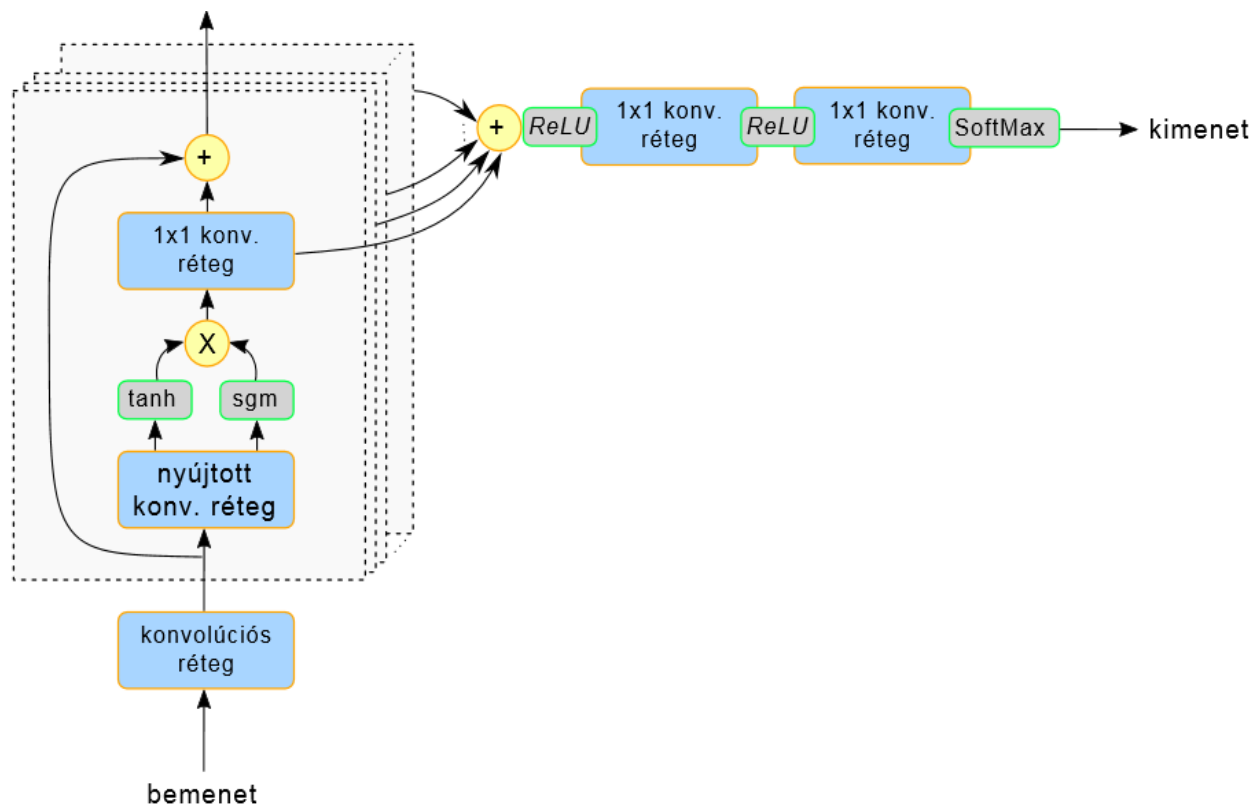


Google 2017. okt
(US angol és japán Google Assistant „élesben”)



Wavenet-alapú magyar TTS

- Női hang:
Mátyus Kati
- Állomási bemondás
 - 3225 mondat
 - 44.1kHz, 16 bit
 - 27826s= 7 h 44
- Szövegből generálva:



Amit nem tudsz egyszerűen elmagyarázni,
azt nem is érted egészen.

2017. január 

november 

Albert Einstein

A fejlődés egy mértéke

Blizzard Challenge (<http://festvox.org/blizzard>)

Év	Legjobb ember	Legjobb TTS	Legrosszabb TTS	Megjegyzés
2005	4,76	3,19	1,98	
2006	4,66	3,74	1,34	nagyobb adatbázis (5000 mondat)
2007	4,7	3,9	1,3	nagyobb adatbázis (8 óra)
				UK English (15 óra)
2008	4,8	4,1	2.0	+ Mandarin (6.5 óra)
2009	4,9	4,2	1,9	
2010	4,8	4,2	1,6	zaj, kisebb adatbázisok
2013	4,8	3,9	1,2	300 órányi angol hangoskönyv címkézés nélkül
2017	4*	3,3*	0,7*	6,5 órányi angol hangoskönyv (56db) gyermekeknek (változatos stílus)*

Kutatási kihívások

Pontos referencia beszédfeldolgozási infrastruktúra
(platform)

Spontán interakciók feldolgozása, kontextus függő
beszédstílusok (színészet)

Elégséges (?) adat gyűjtése és annotálása

Hibrid (szabály-adatvezérelt) kombináció

Szöveg és beszédfeldolgozás DNN integráció

Kognitív infokommunikáció/robotika

Életközeli alkalmazások

- Egészségügy
- Idős emberek támogatása
- Ipari, gyártási alkalmazások



Gépi beszédfelismerés alapú alkalmazások

szubjektív történeti áttekintés

Tibor Fegyó
SpeechTex Kft.



A kezdetek, első lépések

- Honnan indultunk? (1995)



- Saját készítésű mini adatbázisok
- BABEL adatbázis (Vicsi Klára)
- Hidden Markov Model Toolkit (HTK)
- Saját fejlesztésű motor (Szarvas Máté)

1. mérőkö

- Izolált szavas felismerő - néhány ezer szó/mondat
 - Név szerinti kapcsolás telefonközpontokban (2004-)
 - EU csatlakozásról információkat nyújtó interaktív robot (2004)
 - Helyfüggő szolgáltatások (2005)
 - Gyógyszernév felismerő (2005)
 - Név szerinti tudakozó (2009)
 - TV távirányító beszédvezérléssel (2010)



2. mérőkö

- Az első nagyszótáros beszédfelismerő (2005-2008)
 - Holokauszt túlélők visszaemlékezéseinek feldolgozása
(Malach projekt)
 - Johns Hopkins University
 - University of West Bohemia

The screenshot displays the USC Shoah Foundation Visual History Archive Online interface. At the top, it reads "USC Shoah Foundation Visual History Archive Online" and "Viewing Screen". The main content area features a video player showing an elderly woman, Zsuzsanna Akos, with the name "Zsuzsanna Akos" above her. Below the video player, there is a list of segments. The current segment is "Segment#: 9 +", titled "Ilona Grosz - mothers, biological". Below it, "Segment#: 10 +" is visible, titled "Endröd (Hungary)", with a sub-entry: "Hungary 1918 (November 11) - 1939 (August 31) Hungary 1939 (September 1) - 1944 (March 18) schools". To the right of the video player, there is a map of Central Europe with red location pins and a "Tape: 1 of 8" dropdown menu. At the bottom right, there are navigation controls like "Maximize/Minimize Data", "Next Result", "Previous Result", "Back to Search Results", "New Search", "Save to Projects", and "Print Testimony".

3. mérőkö

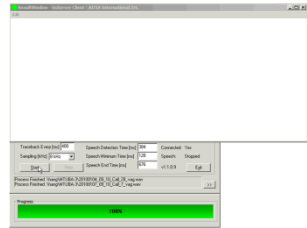
- Folyamatos beszéd felismerésére alkalmas saját fejlesztésű technológia
 - Voxearch - hírműsorokban keresés (2004-2007)
 - Kormányzóvivői tájékoztatók leiratozása (2008-10)

4. mérőkö

- Szélesebb körű alkalmazások (2009-)
 - Technológiai fejlődés
 - Megfelelő számítási kapacitás
 - Mély neuronháló alkalmazása (2015-)

4/1 Telefonos ügyfélszolgálati rendszerek

- Telefonbeszélgetések elemzése üzletfejlesztési és minőségbiztosítási célból.



- Vodafone (2009-12)
- Invitel (2009)
- Aegon (2011-16)
- OTP (2012-)



4/2 Hírműsor felismerés (2009-)

- ~100m szavas tanító korpusz
- 1.5+ millió szavas szótár
- Integrált eszközök (pl. teletext)
- Alapvetően hírek + sport kiegészítés
- Kitekintés a többnyelvű megoldások felé



- Mindroom (2009-2012)
- MTVA (2014 -)
- NMHH (2018)

digital natives

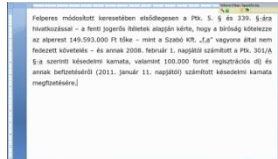


mindroom®

NMHH
Nemzeti Média- és Hírközlési Hatóság

4/3 Szaknyelvi diktálási megoldások

- A szaknyelvek szókincséhez és dokumentum formázási elvárásaihoz is igazodó megoldások



- EU jogi szakfordítók (2015-)



- Bíróságok (2015-)



- Műszaki szakfordítók (2016-)



- egyéb szakterületek, többnyelvű megoldások (2018-)

Köszönöm a figyelmet!