# 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

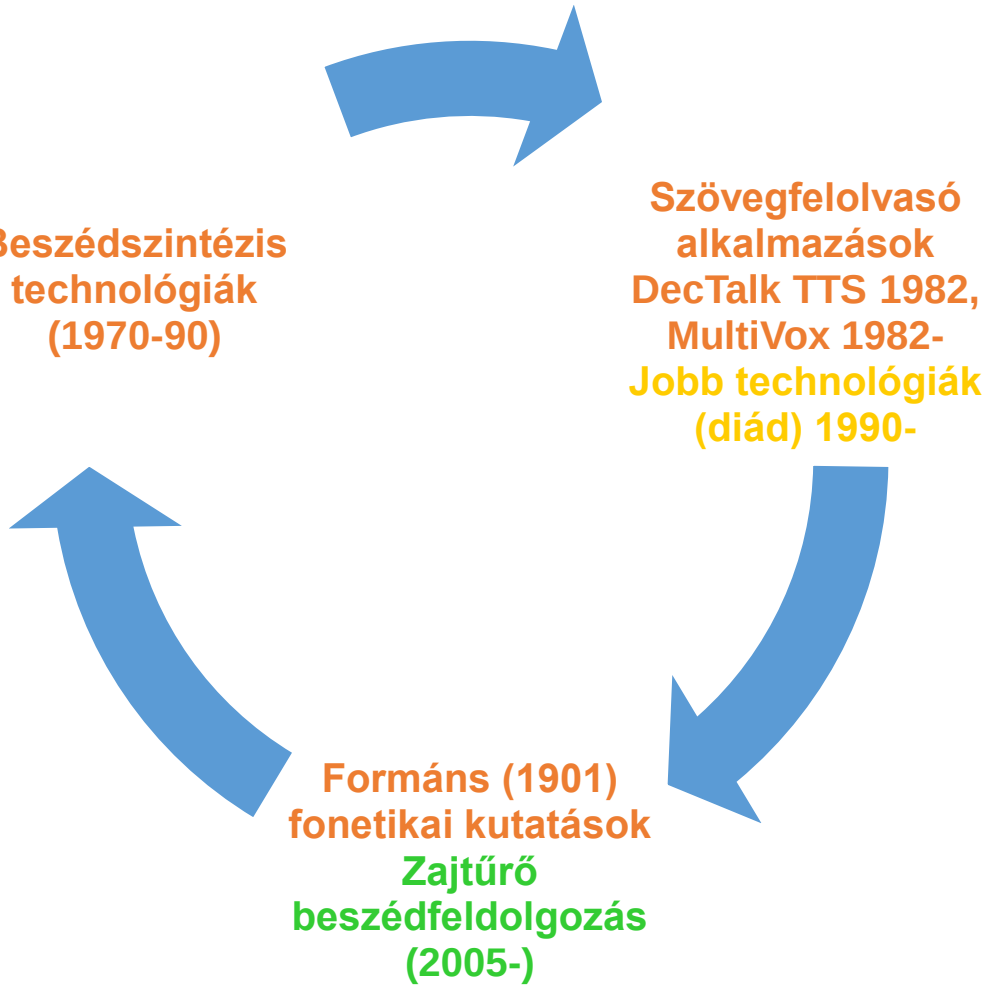
# Út a beszédszintézishez


- 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 órázás 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???

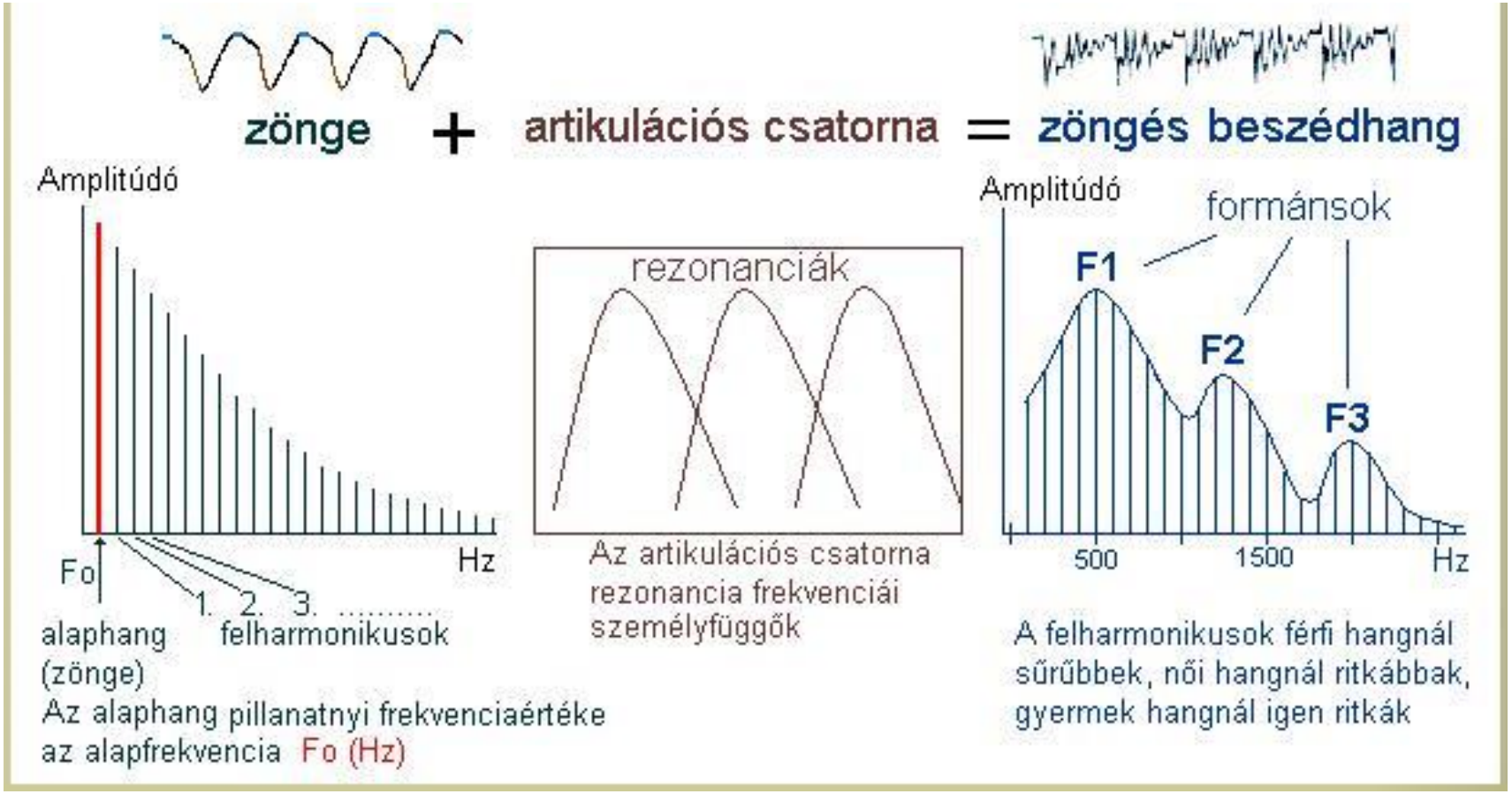
# Út a BEAG-ba

- 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)

# Alapkutatás (formáns) 2

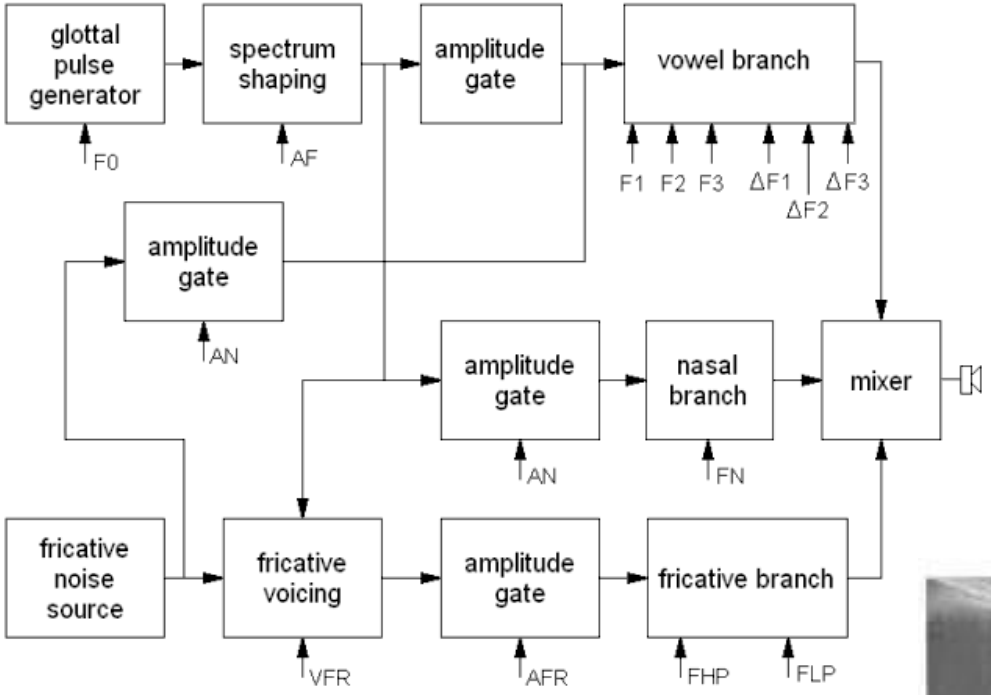


# Alapkutatás (formáns 1791-) 1



# Előzmények

## Kempelen Farkas 1791




## HungaroVox 1982

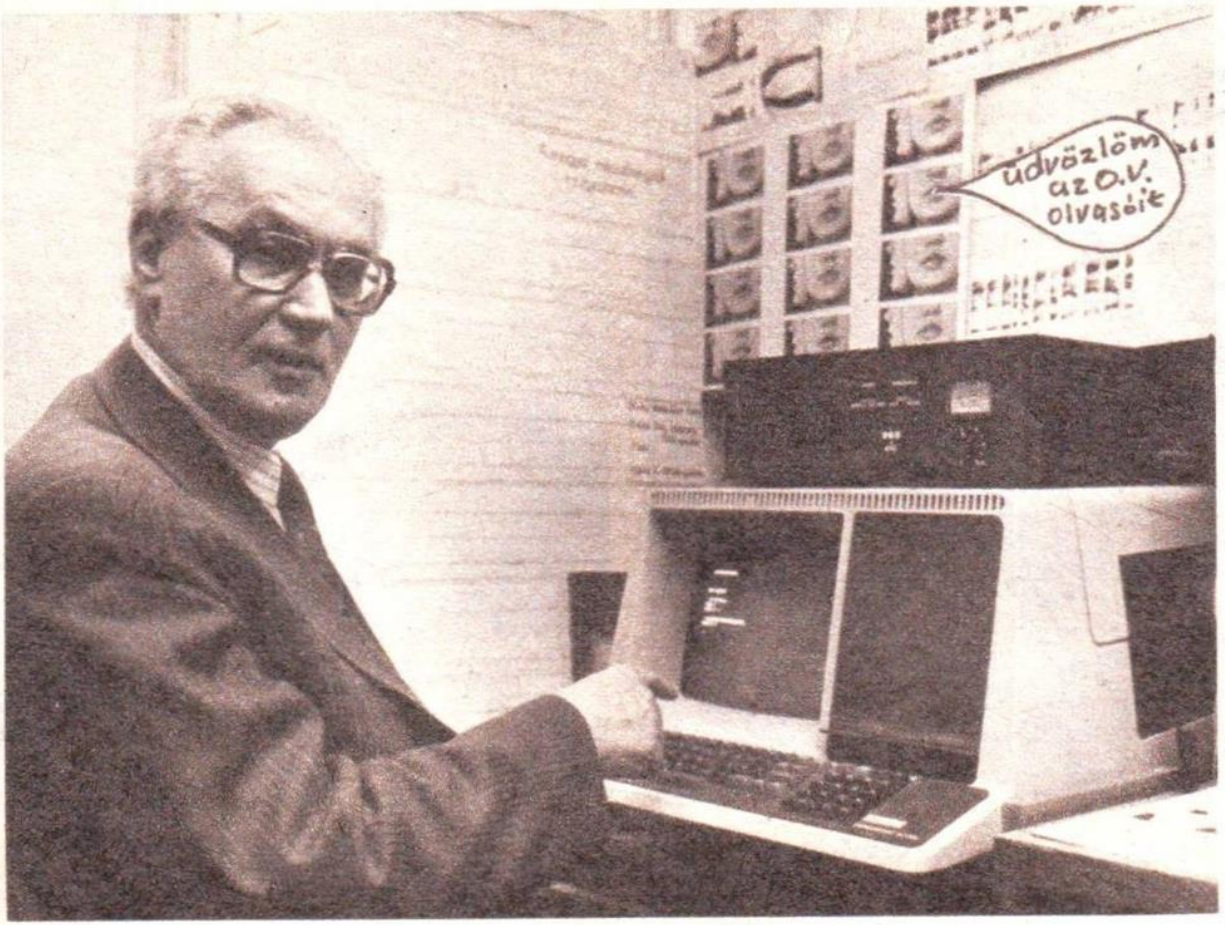
# Magyarul és oroszul szólal meg a Voxton – a beszélőgép

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

Nem sajtóhiba: a harmadik ugyanis nem személy, hanem *gép, amelyik beszélni 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 szonuszból kapta a nevét. A Voxonnal — az első magyar nyelvű személytelen**



# HungaroVox és Russon

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

# Még Moszkvába is eljutott

ПРОЛЕТАРИИ ВСЕХ СТРАН, СОЕДИНЯЙТЕСЬ!

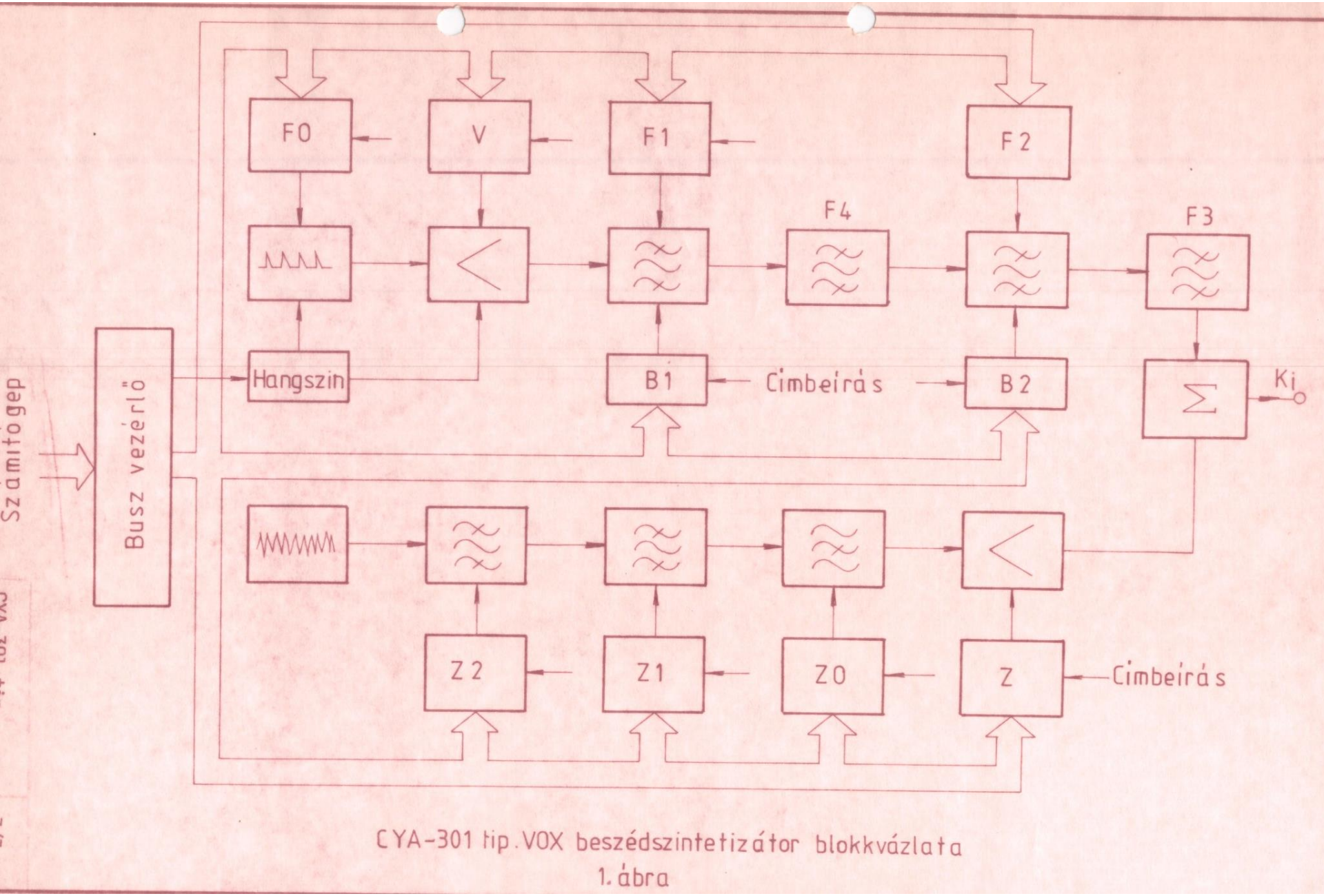
Орган Центрального Комитета ВЛКСМ

# КОМСОМОЛЬСКАЯ ПРАВДА

Основана в мае 1925 г. ● № 78 (18282) ● Пятница, 5 апреля 1985 г. ● Выходит ежедневно, кроме понедельника ● Цена 3 коп.

ГОВОРЯЩАЯ МАШИНА „РУССОН"

МУЖЕСТВО




CYA-301 tip. VOX beszédszintetizátor blokkvázlata
1. ábra

# VOX beszédszintetizáló

## Műszaki leirás

Az ipari felhasználásra kifejlesztett valósidejü magyar nyelvű beszédszintetizáló rendszer a MTA Nyelvtudományi Intézetben folyó beszédakusztikai és szintetizálási kutatások egyik eredménye. Ez az első olyan beszédszintetizáló rendszer, amelyben a beszéddé átalakitani kivánt szöveget a magyar helyesirás szerint kell megadni és a mikroszámitógép ebből a betüsorozatból állitja elő a folyamatos köznyelvi beszédet. Hangja, beszéde nem emberi alapú, tehát hangszinében sem hasonlit valamely személy hangjához. A beszédet előre megtervezett akusztikai épitő elemek sorozatából állitja elő. A szótárnélküli, automatikus beszédelőállitáshoz ilyen, a nyelv akusztikai szerkezetét leiró elemtárat, elembázist kell képezni. Ez az elemtár 370 akusztikai épitőkockát /hangszeletet/ tartalmaz, A beszédelőállitás során ebből az elembázisból válogatja ki a számitógépprogram az éppen aktuális hangsor felépitéséhez szükséges épitőelemeket. Ezeket sorrendbe állitja, majd az épitőelemekben 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 intezitásszerkezetű elemi beszédhangrész, vagy hangkapcsolódási szakasz. Egy-egy hangszeletnek nyelvi szempontból sem önálló jelentése, sem értelmes hanghatása nincs. A magyar beszédre az egyes hangszeletek paramétereit és paraméter értékeit az MTA Nyelvtudományi Intézetben végzett kutatások alapján állapitották meg és kódolták be a számítógép adattára számára. A magánhangzókat és az egyszerü szerkezetü beszédhangokat általában három hangszelettel 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,

A bonyolult szerkezetű [illegible]

Szer

Kábelezési rajz:

Elvi, kapcsolás:

...egy csőhöz/ való csatlakozás része,

A bonyolult szerkezetű mássalhangzóknál a hangszeletkiosztás jóval összetettebb. A hangsorépitéshez egy megfelelő számítógépes program /a HUNGAROVOX/ az elembázis elemeiből a megadott betüsor szerint meghatározza a kiválasztandó hangszeleteket és ennek kódjaival vezérli a CYA-301 tipusú beszédgenerátort. Ez egy digitális vezérlésű analóg, soros beszédszintetizátor, amely áramköreivel az ember hangképző mechanizmusát utánozza. Zenés hangok képzéséhez periodikus jelre, zörej hangok képzéséhez zörej forrásra van szükség. Ezt a szintetizátor hangalkotó oszcillátorral, zajgenerátorral és változtatható karakterisztikájú szűrőkkel, erősitő és összegző elemekkel valósitja meg. /lásd. 1.ábra/

A beszédszintetizátor /CYA-301/ működtetését a SYSTER mikroszámitógép végzi. Az ezen futó számitógép programrendszer /HUNGAROVOX/ az alábbi feladatokat látja el:

- az akusztikai elembázis kezelése,
- az irás - hangszelet konvenció elvégzése,
- helyesirás értelmezése,
- leggyakoribb hasonulások és egyéb hangsorépitési szabályok értelmezése
- intonáció generálása,
- beszéd tömöritett tárolása,
- tárolt mintákból a beszéd előállitása,
- a beszéd manuális szerkesztése (pl. időváltoztatás, ritmus meghatározás, intonációs szerkezet generálás, dinamikai változtatás, beszédtempó meghatározás, hangnyújtás, hangrövidités stb.)

| | | | | |
|---|---|---|---|---|
| **EAG Elektroakusztikai Gyár BUDAPEST** | Rajzoló | | | CYA-301 tip. |
| | Szerkesztő | | | |
| | Témafelelős | | | Műszaki leirás |
| | Laborvezető | | | |

Műszaki adatok:

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

El. anyaglista: Kábellista: Műszaki leírás:

# 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

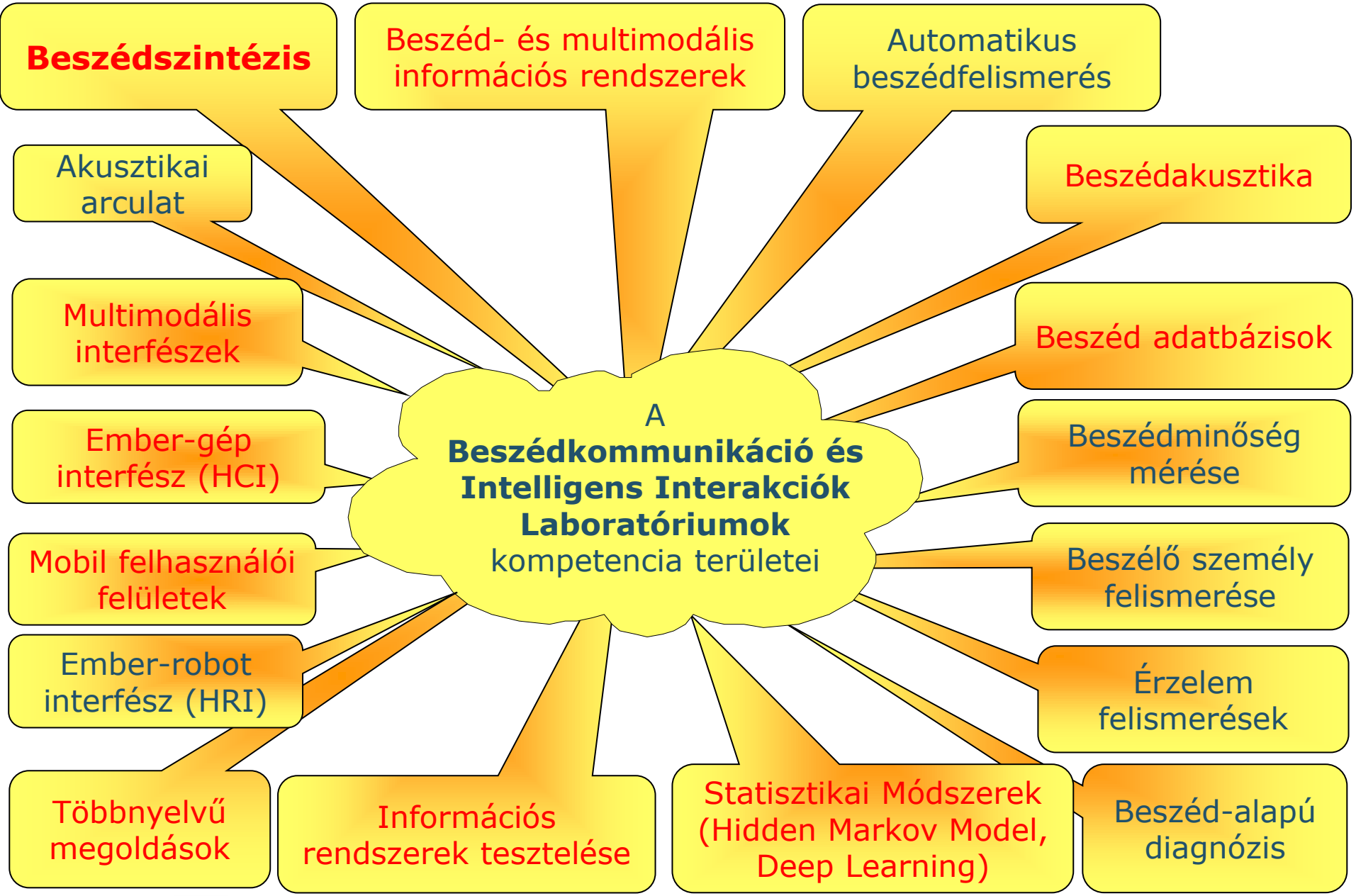
Beszédtechnológia és Intelligens Interakciók Labor (Németh Géza)

Beszédfelismerés és Hangbányászat Labor
(Mihajlik Péter)

Beszédakusztikai Labor
(Sztahó Dávid)

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


www.ai4eu.org


Beszédszintézis
Beszéd- és multimodális információs rendszerek
Automatikus beszédfelismerés
Akusztikai arculat
Beszédakusztika
Multimodális interfészek
Beszéd adatbázisok
Ember-gép interfész (HCI)
Beszédminőség mérése
A
Beszédkommunikáció és Intelligens Interakciók Laboratóriumok
kompetencia területei
Mobil felhasználói felületek
Beszélő személy felismerése
Ember-robot interfész (HRI)
Érzelem felismerések
Többnyelvű megoldások
Információs rendszerek tesztelése
Statisztikai Módszerek (Hidden Markov Model, Deep Learning)
Beszéd-alapú diagnózis

# SmartLab munkatársak

Németh Géza
PhD 1997, Habil 2013
(Laborvezető)

Olaszy 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

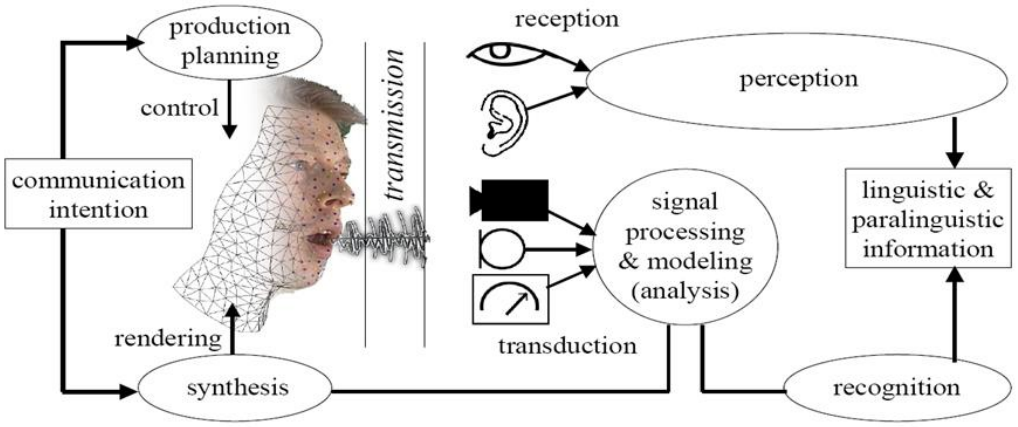
Hajgató Gergely
PhD hallgató

Hamdi Abed
PhD hallgató

# SmartLab kutatási területek

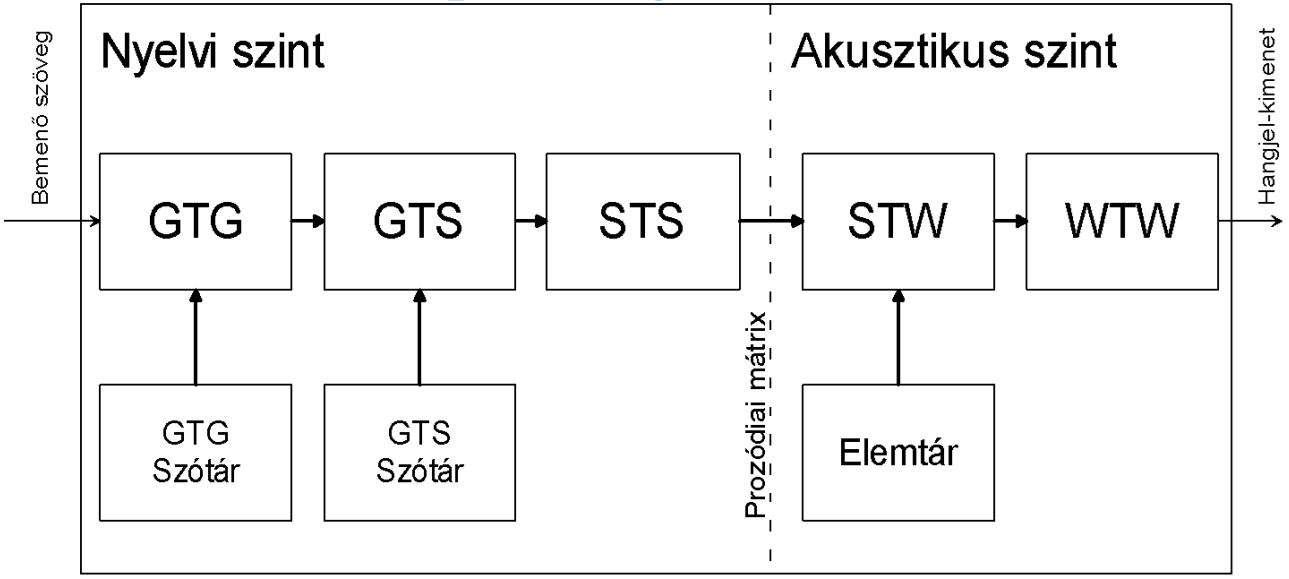
- Gépi szövegfelolvasás (text-to-speech, TTS)
  - Elemösszefűzéses és korpusz-alapú
  - Gépi tanulás alapú (Deep Learning, Hidden Markov-model)
- Beszédszintézis részproblémái
  - Parametrikus kódolás, gerjesztési modellek
  - Intonációs modellek
  - Rövid- és kérdő mondatok prozódiája
  - Kommunikációs 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



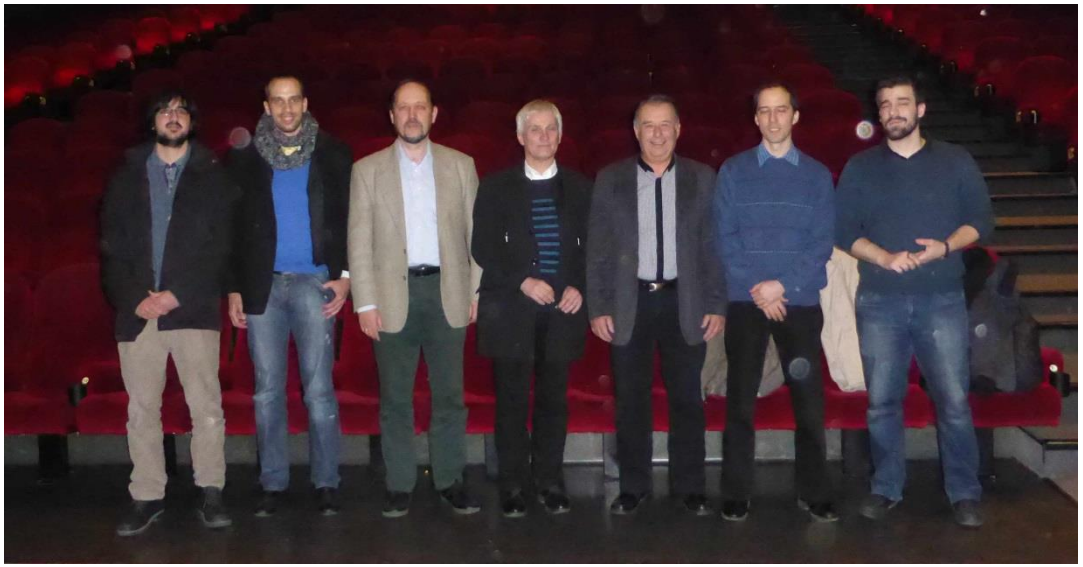
# Történelem
## közlekedés és beszédtechnológia

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)

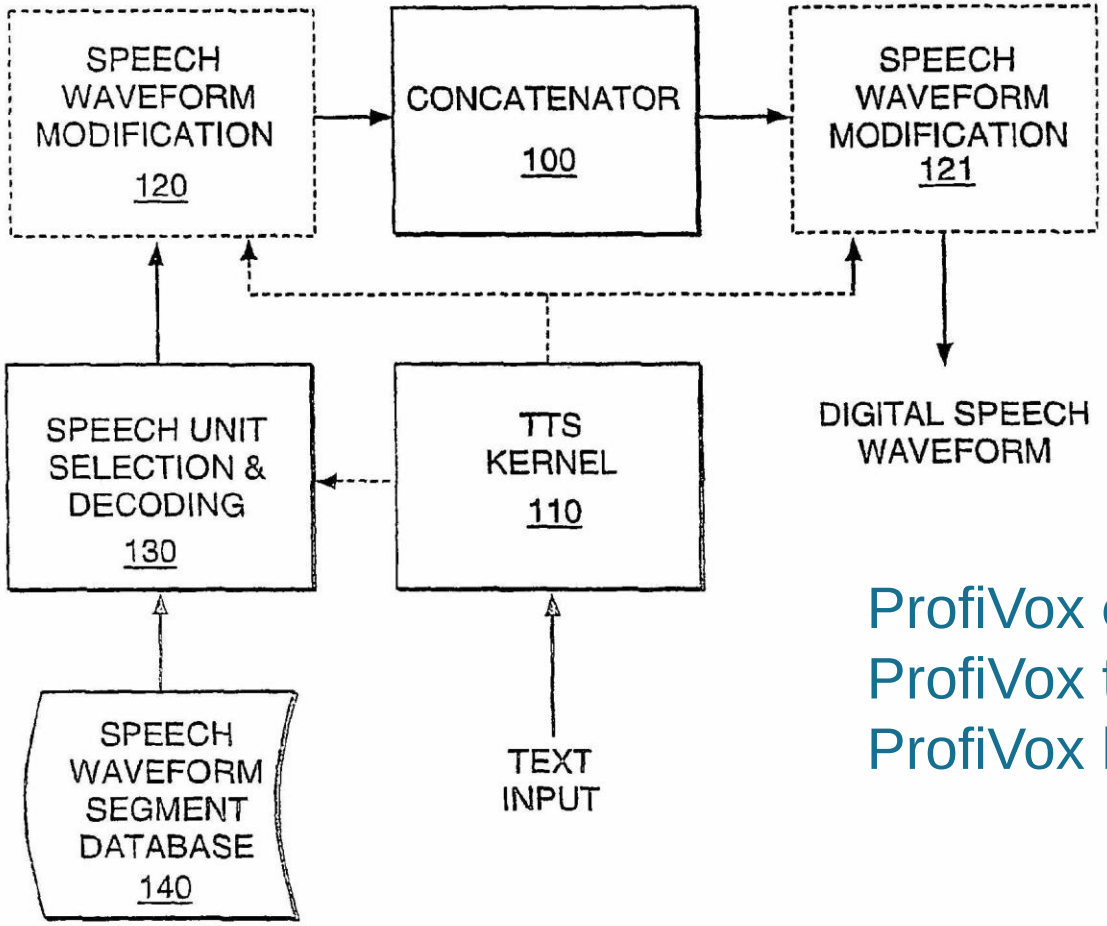
- mellett és helyett

Természetes elemek

- egyre nagyobb halmaza
- statisztikai modellépítés
- minimális jelfeldolgozás

Egységes(re törekvő) kiértékelés

# Hullámforma összefűzés (termé




Hírfelolvasás (2004)

Futball hírek

ProfiVox diád 1995-

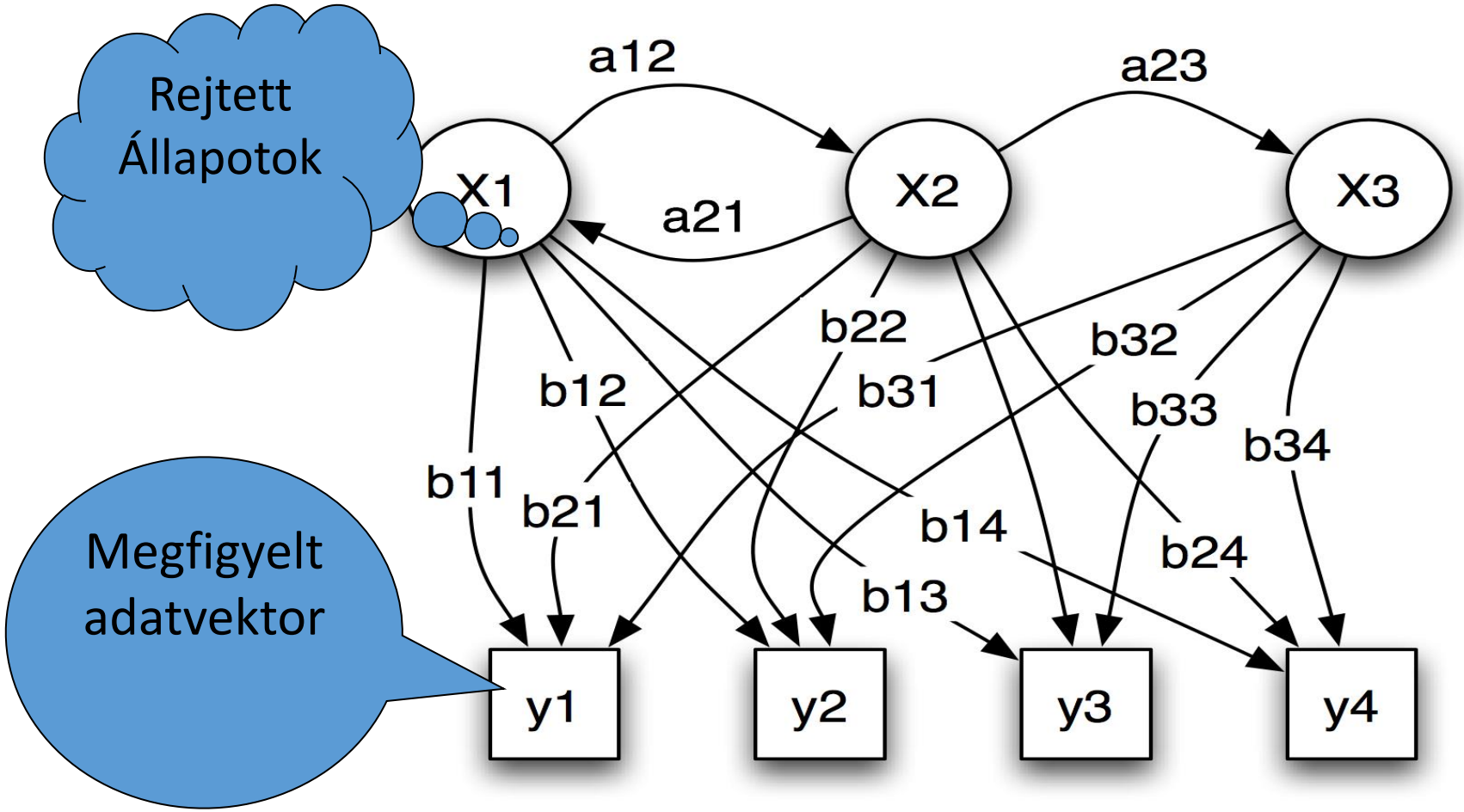
ProfiVox triád 2000-

ProfiVox korpusz 2002-

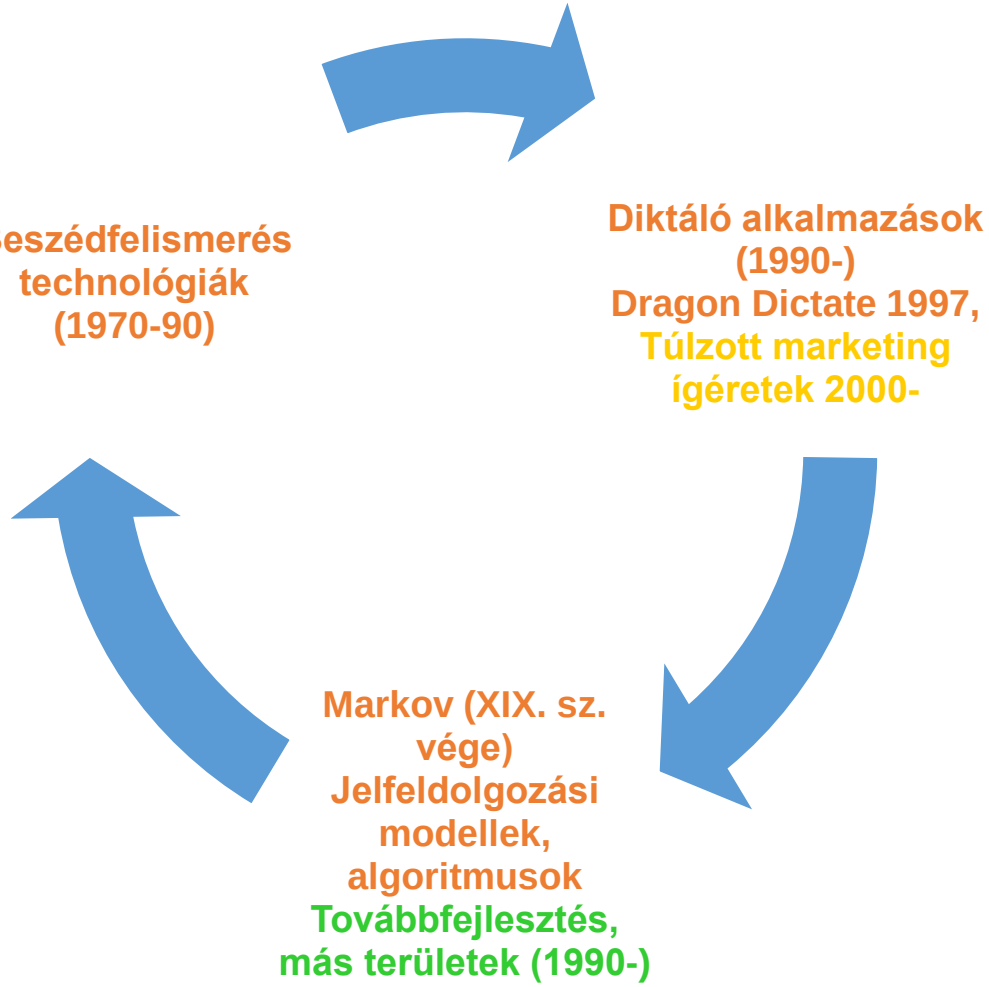
Utas információ (többnyelvű)

# Új hang a magyar vasútállomásokon

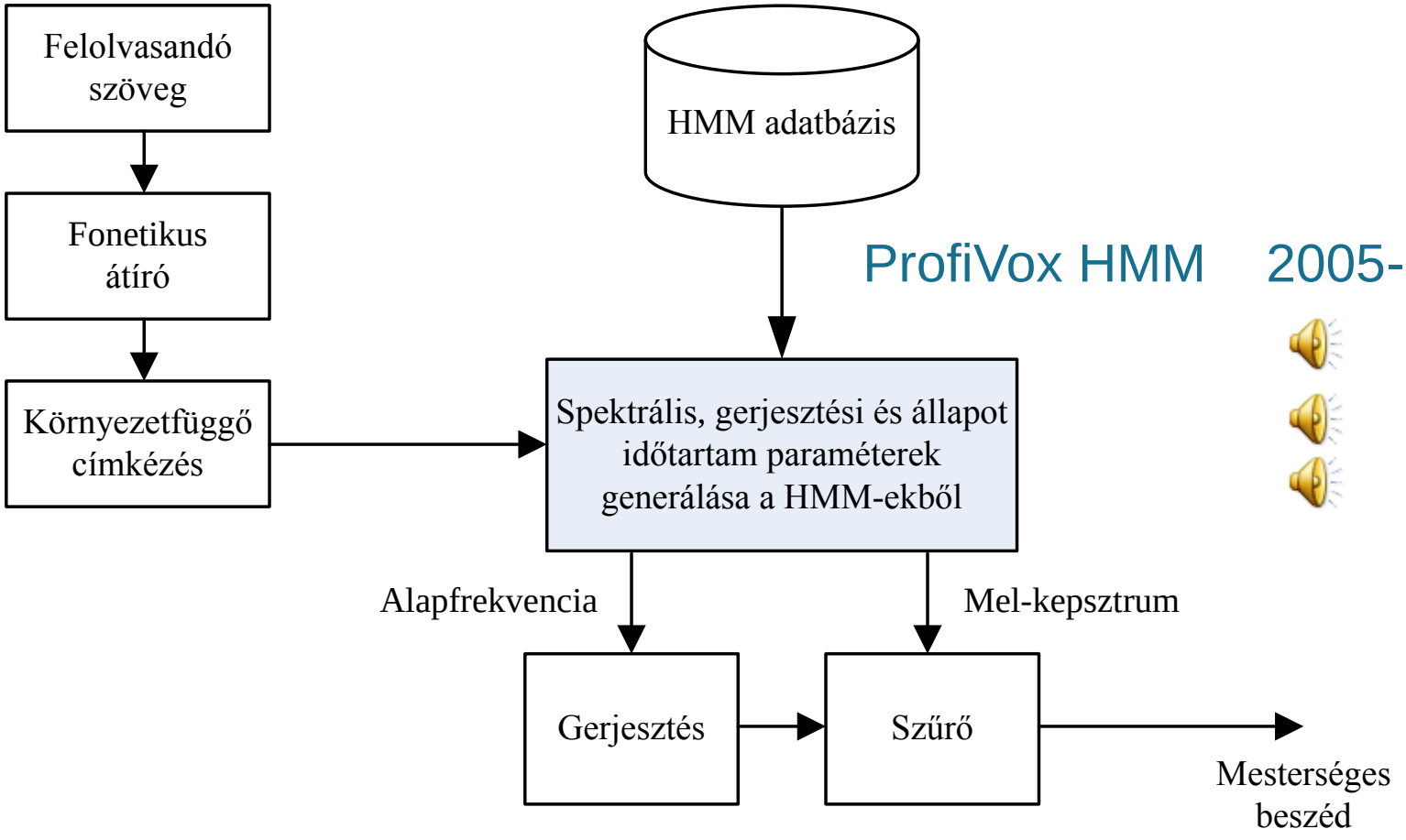
# Alapkutatás (Hidden Markov Model, HMM 1970-) 1



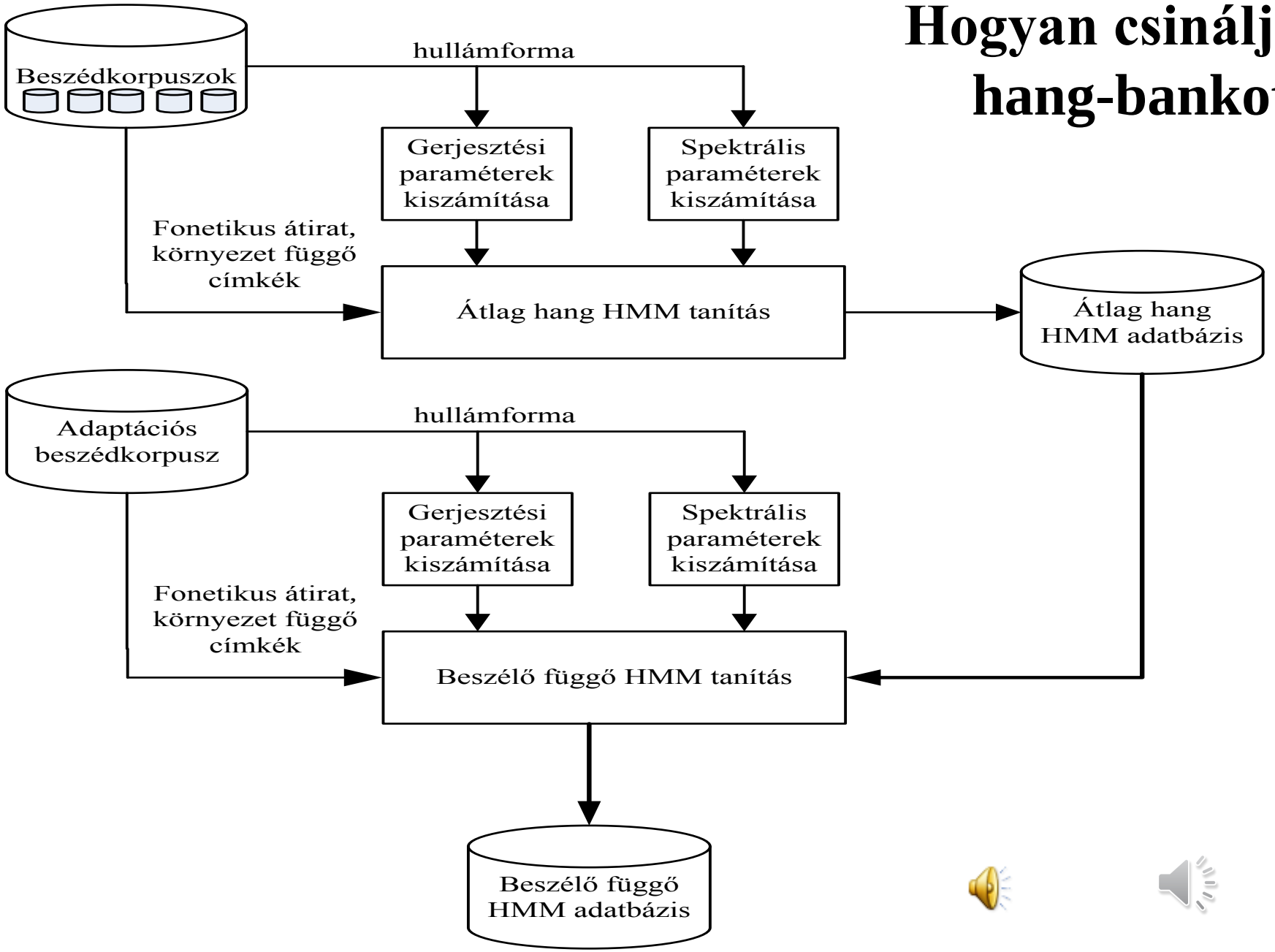
# Alapkutatás (Rejtett markov modell, HMM) 2



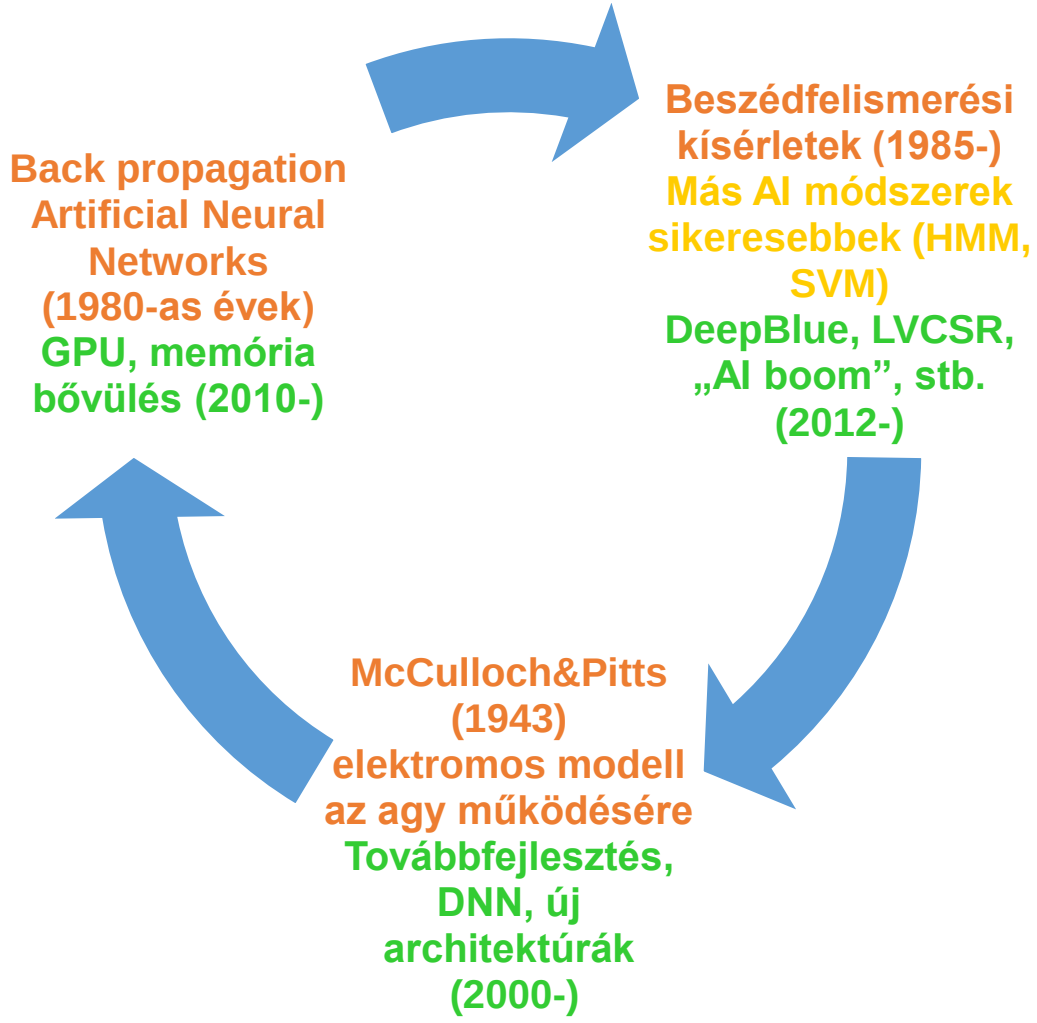
# Technológia fejlesztés HMM-alapokon (rugalmasság, 200x-)



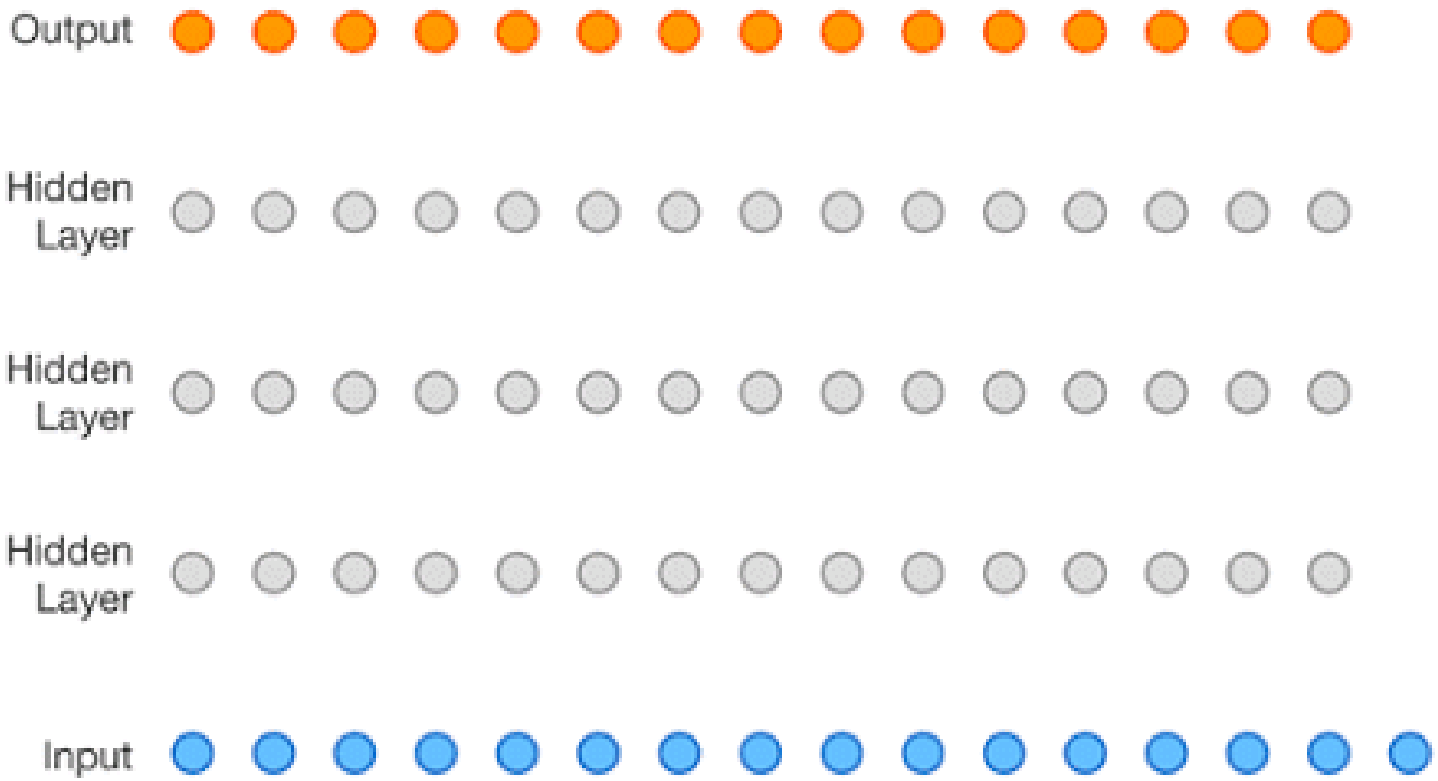
# Hogyan csináljunk hang-bankot?



# Alapkutatás (Neurális hálózatok)

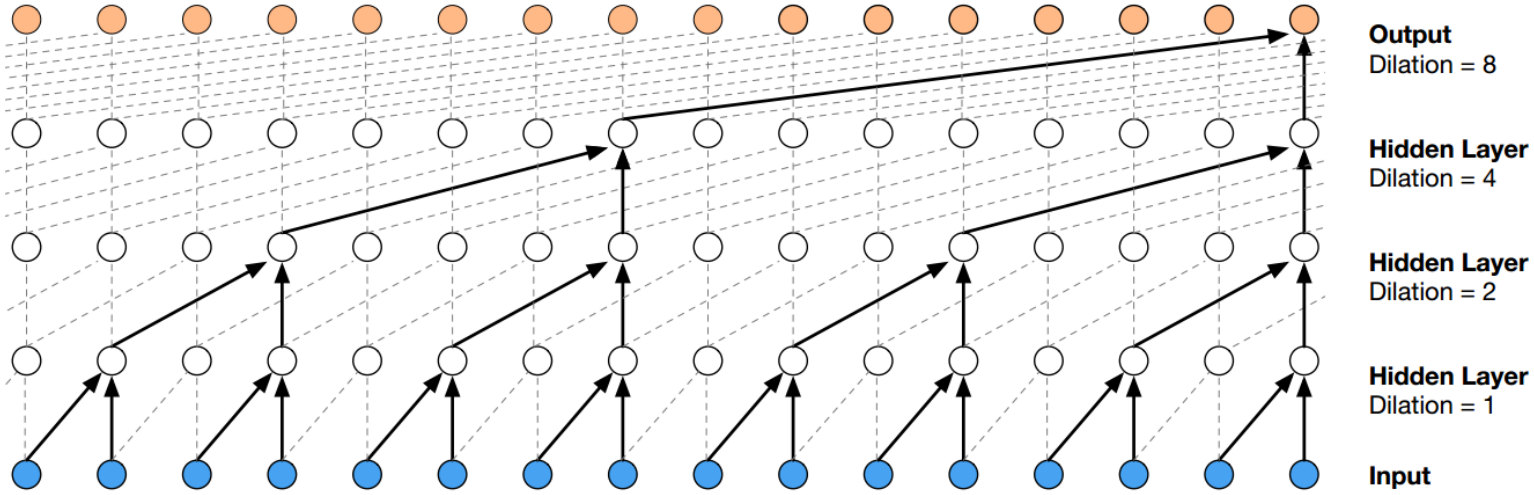


# Wavenet (2016. szept.- )




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

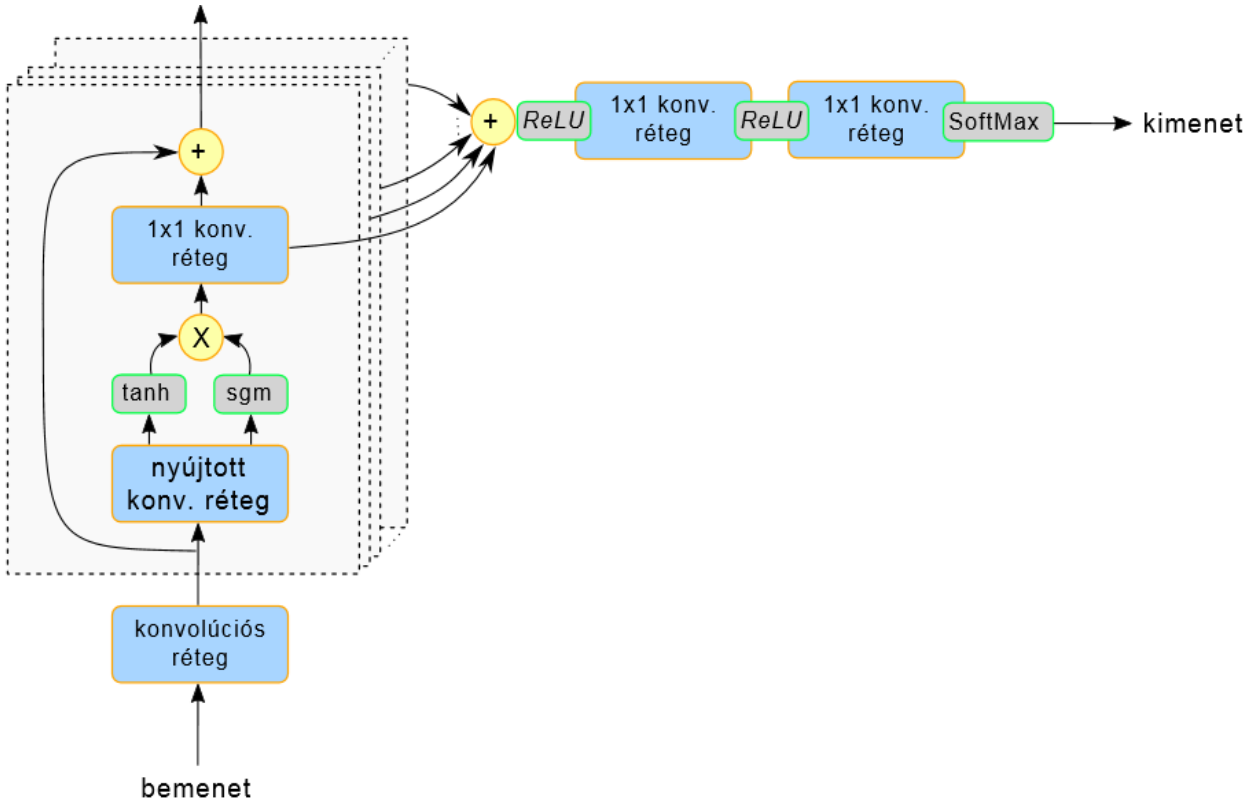
# Generálás




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) |
| 2008 | 4,8 | 4,1 | 2.0 | UK English (15 óra)<br>+ 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

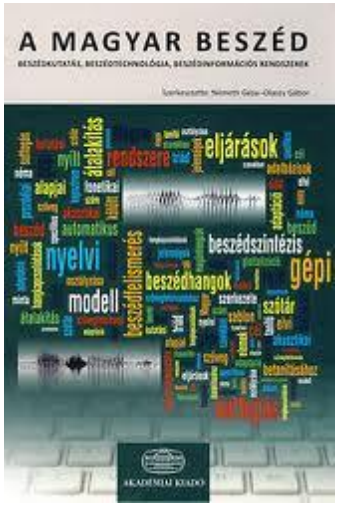
Szöveg és beszédfeldolgozás DNN integráció

Kognitív infokommunikáció/robotika

Életközeli alkalmazások

- Egészségügy
- Idős emberek támogatása
- Ipari, gyártási alkalmazások

# Hozzászólások




**Mélyebb érdeklődőknek:**

**http://smartlab.tmit.bme.hu/**
**http://magyarbeszed.tmit.bme.hu/**


**Köszönjük**

**az** NKTH Nemzeti Kutatási és Technológiai Hivatal; ÚJ SZÉCHENYI TERV; MAGYARORSZÁG MEGÚJUL; Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal; EUROPEAN COMMISSION; Magyar Tudományos Akadémia 1825 **támogatását.**

BelAmi; ETO COM; CESAR CEntral And South-East EuropeAn Resources; PaeLife; VUK

(Teleauto, BelAmi, EtoCom -TÁMOP-4.2.2-08/1/KMR-2008-0007- , BME Kutatóegyetemi -TÁMOP-4.2.1/B-09/1/KMR-2010-0002- , CIP CESAR, AAL PAELIFE, VUK, Eureka DANSPLAT projektek)