

Mesterséges intelligencia és beszédtechnológia

Sukorói Értelmiségi Kör
2020.02.15.

Németh Géza

BME TMIT SmartLabs

Beszédkommunikáció és Intelligens Interakciók
Laboratóriumok

1980-2019

SmartLab
Intelligent Interactions

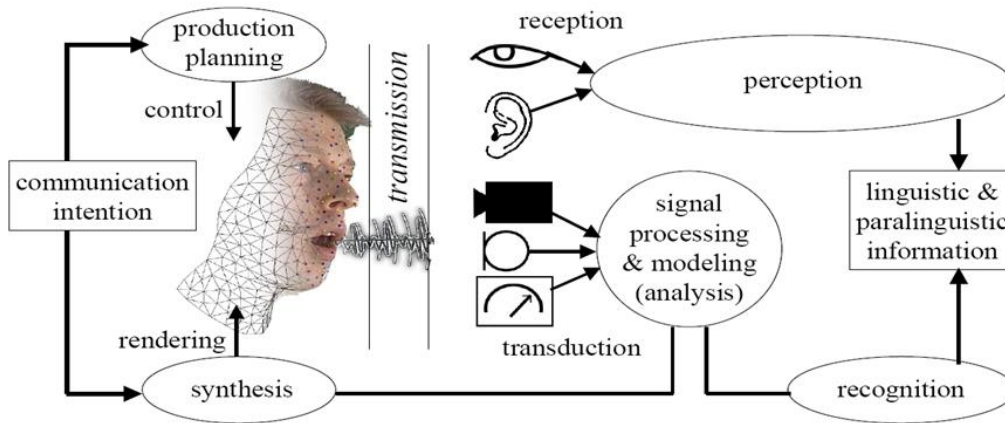
<http://smartlab.tmit.bme.hu>

SmartLab
Intelligent Interactions



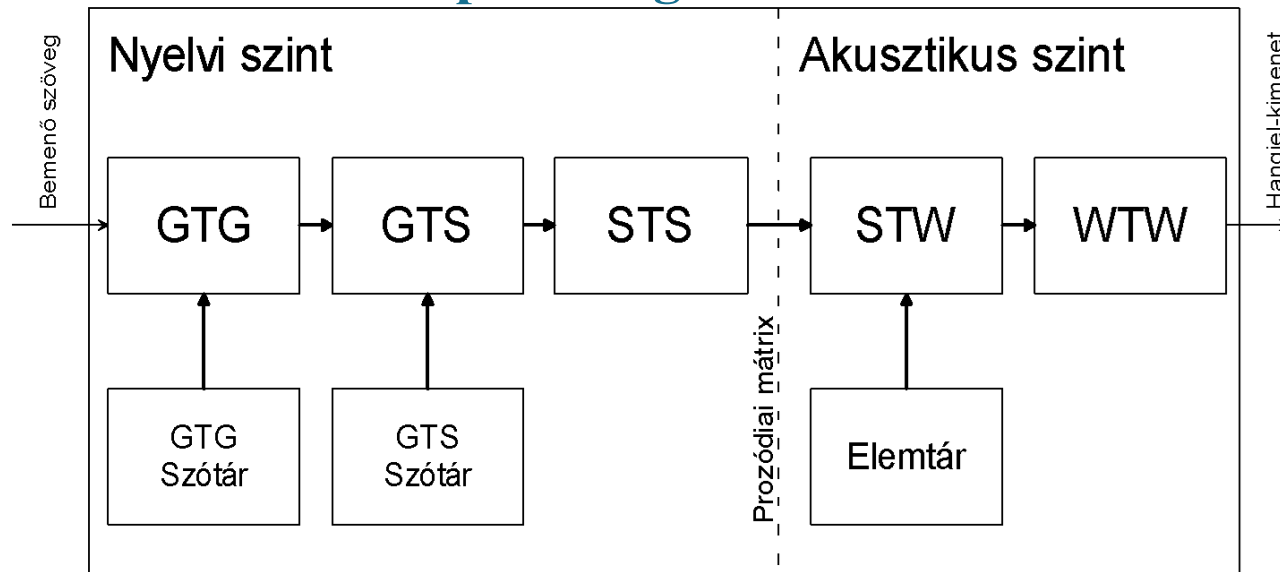
GPU
EDUCATION
CENTER

Mi is a beszédtechnológia?



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

Gépi szövegfelolvasás



Speech Groups within DTT TUB (1999, Eurospeech - 1100 résztvevő)

- Telecommunications and Signal Processing (TSP) Lab.
Head: Péter Tatai
- Speech Acoustics (SA) Lab.
Head: Klára Vicsi
- Speech Technology (MULTIVOX) Lab.
Head: Géza Németh and Gábor Olaszy



TSP Lab (4)

On-going projects

- Qualiphone/A Objective speech quality measurement software for NMT 450 (Westel 450)
- Multilingual (Japanese - Hungarian) speech recognition (NTT, Japan)
- Educational ASR and speech processing demonstration package

TSP Lab (5)

Project ideas

- Voice dialing system for mobile phones
- Large vocabulary speech recognition for agglutinative languages
- Speech quality assessment system

SA Lab (1)

Topics

- Basic acoustics-phonetics research
- Data-base construction
- Isolated word recognition
- Continuous speech recognition
- Speech enhancement in noisy speech
- Development of aids for handicapped people
- The group has constructed instruments, tools in the field of speech technology for different industries.

SA Lab (3)

Project ideas

- Speech processing for foreign language learning
- Multilingual dictation machine
- Multilingual mobile speech database collection
- Dialogue systems over the telephone

MULTIVOX Lab (4)

Projects

- Hungarian and German TTS on C64 (1984)
- Integrated OCR and TTS (1986)
- Intelligent announcement system (1992)
- 1st speaking bill service in Hungary (1995)
- 1st residential voice mail in Hungary (1996)
- High quality Number-to -Speech (1996)
- 1st Hungarian e-mail reader (1997)



MULTIVOX Lab (9)

Project ideas summary

- Multilingual number-to-speech
- Multilingual e-mail reader
- Multilingual text-reading environment (e-mail/fax access, speaking newspaper, reverse directory)
- Multilingual mixed recorded/TTS for public information systems (traffic, railways, insurance, process control)



Távközlési és Telematikai Tanszék

2001

■ Gordos Géza (tanszékvezető)



■ Beszéd laborok:

– Németh Géza és
Olaszy Gábor



– Tatai Péter
TSP



– Vicsi Klára
AKL



Beszédtechnológiai laboratórium

- Németh Géza, Olaszy Gábor
- 4-6 doktorandusz
- graduális hallgatók

Interspeech 2001



SmartLabs ma: 3 labor

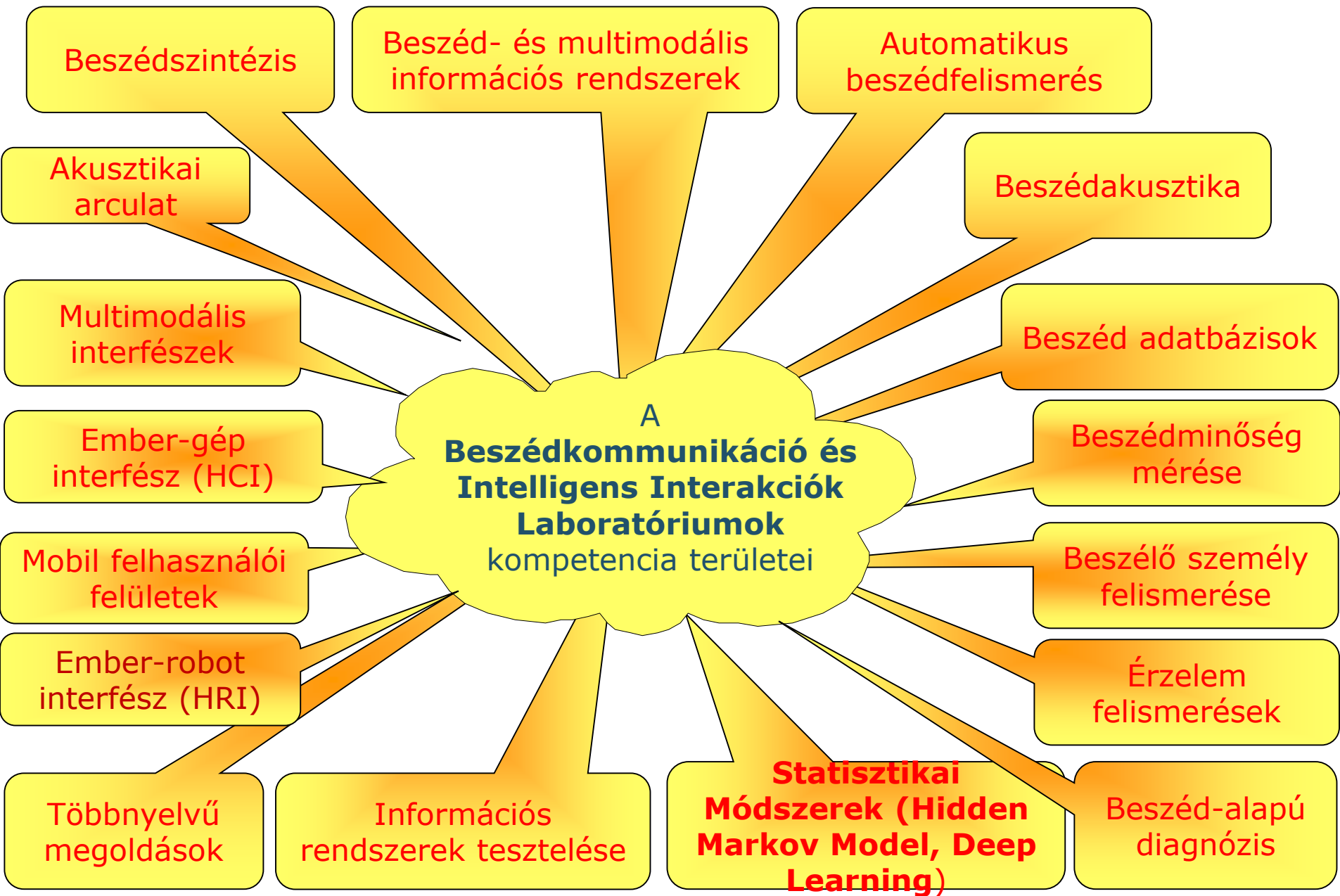
Beszédtechnológia és Intelligens Interakciók Labor (Németh Géza)

Beszéd felismerés és Hangbányászat Labor (Mihajlik Péter)

Beszédakusztikai Labor (Sztahó Dávid)

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





K+F+I példák, alapkutató

- Gépi szövegfelolvasás technológiái
- Automatikus beszéd felismerés technológiái
- Hangsúly modellezés deep neural network –DNN- alapon
- Rövid és kérdő mondatok intonációja modellezése
([SP2 SCOPES](#) project, IDIAP, Switzerland, 2014-2016)
- Beszédkódolás és rendhagyó beszéd modellezése
- Nyelvmozgás elemzése ultrahanggal
(Fulbright 2014, Momentum Lendület 2016, ELTE – MTA Nyl - BME)
- Érzelmi és egészségi állapot meghatározása beszédből
([COALA](#) Psychological Status Monitoring by Computerised Analysis of Language Phenomena (ESA, 2013-2018))

K+F+I példák, alkalmazott kutatás

- Personal Assistant to Enhance the Social Life of Seniors using speech recognition and synthesis (www.paelife.eu, EU AAL, 2011-2014)
- Navigáció vak emberek számára
 - (Visionless Supporting Framework, EU AAL, 2016-2019)
- Speech Technology Platform for the Danube Region
 - (DANSPLAT, Eureka, 2016-19)
- Szám szerinti tudakozó (T-Mobile, Vodafone)
(Név- és címfelolvasó TTS, 2008-)
- Beszédfelismerés telefonos ügyfélszolgálatokon
- Automatikus TV feliratozás (MTVA, 2015-)
- Audio-vizuális kiejtés oktató rendszerek
- SmartPhone / Tablet demo alkalmazások
(Weather for All – Windows 8 >40.000 letöltés; NAO robot magyarul)
- *Szolgáltatás: termék- és rendszer támogatás*
(Automatikus szám-szerinti tudakozó, Jaws for Windows képernyőolvasó)

Oktatás és hallgatók támogatása

- Beszédinformációs rendszerek (BSc)
- Ember-gép interfész, Human-Computer Interaction (MSc)
- Infocommunication (BSc)
- Beszéd- és hallásdiagnosztika (MSc)
- Multimédia tartalmak intelligens feldolgozása (MSc)
- Kiváltott agyi jelek informatikai feldolgozása (SZV)
- Deep Learning a gyakorlatban Python és LUA alapokon (SZV)
- Ember-robot interakció (SZV)
- Téma-, projekt laboratórium, BSc és MSc szakdolgozat
- PhD képzés, iMSc mentorálás
- Nemzetközi együttműködés (Fulbright, DAAD, EU)
- Kutatók Éjszakája, BME Nyílt Nap, ...

SmartLab Beszédtechnológia és Intelligens Interakciók Labor 1981-



Németh Géza
PhD 1997
Laborvezető



Olasz Gábor
DSc 2003



Zainkó Csaba
PhD 2010



Gyires-Tóth Bálint Pál
PhD 2013



Sevinj Yolchuyeva
PhD hallgató



Csapó Tamás Gábor
PhD 2014



Bartalis Mátyás
Kutatómérnök



Nagy Péter
PhD jelölt



Laczkó Klára
Kutatási asszisztens



Mohammed Al-Radhi
PhD hallgató



Hajgató Gergely
PhD hallgató



Hamdi Abed
PhD hallgató



Moni Róbert
PhD hallgató

SmartLab kutatási területek

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

Történelem

közlekedés és beszédtechnológia

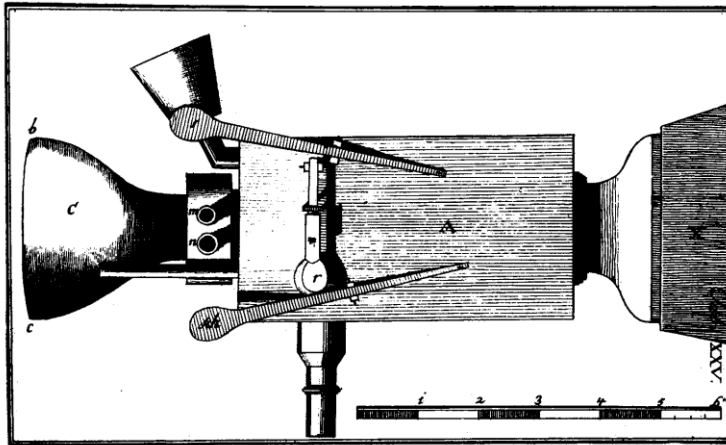


1791



2019

Lépünk vissza az időben 210 + 18 évet



Kempelen eredeti gépe

1791

2001

Pontosan elkészített, működő másolat

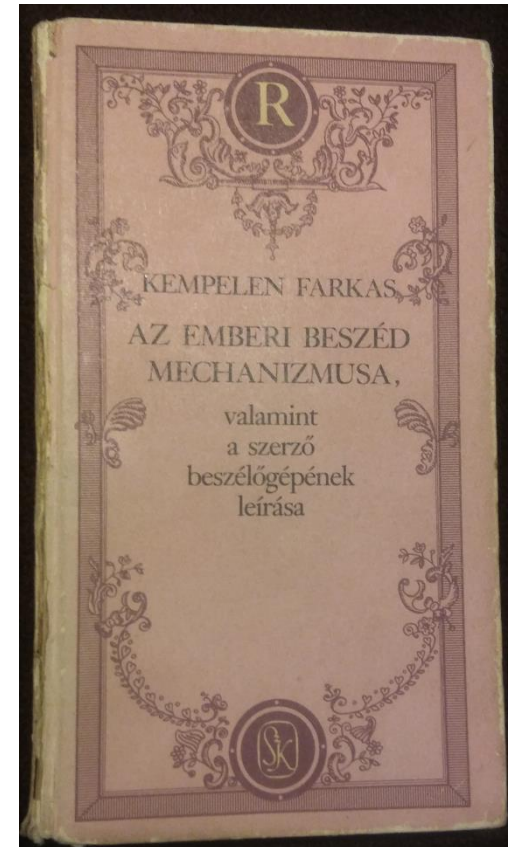


Kempelen könyve és a magyar fordítása



Mechanismus der menschlichen Sprache

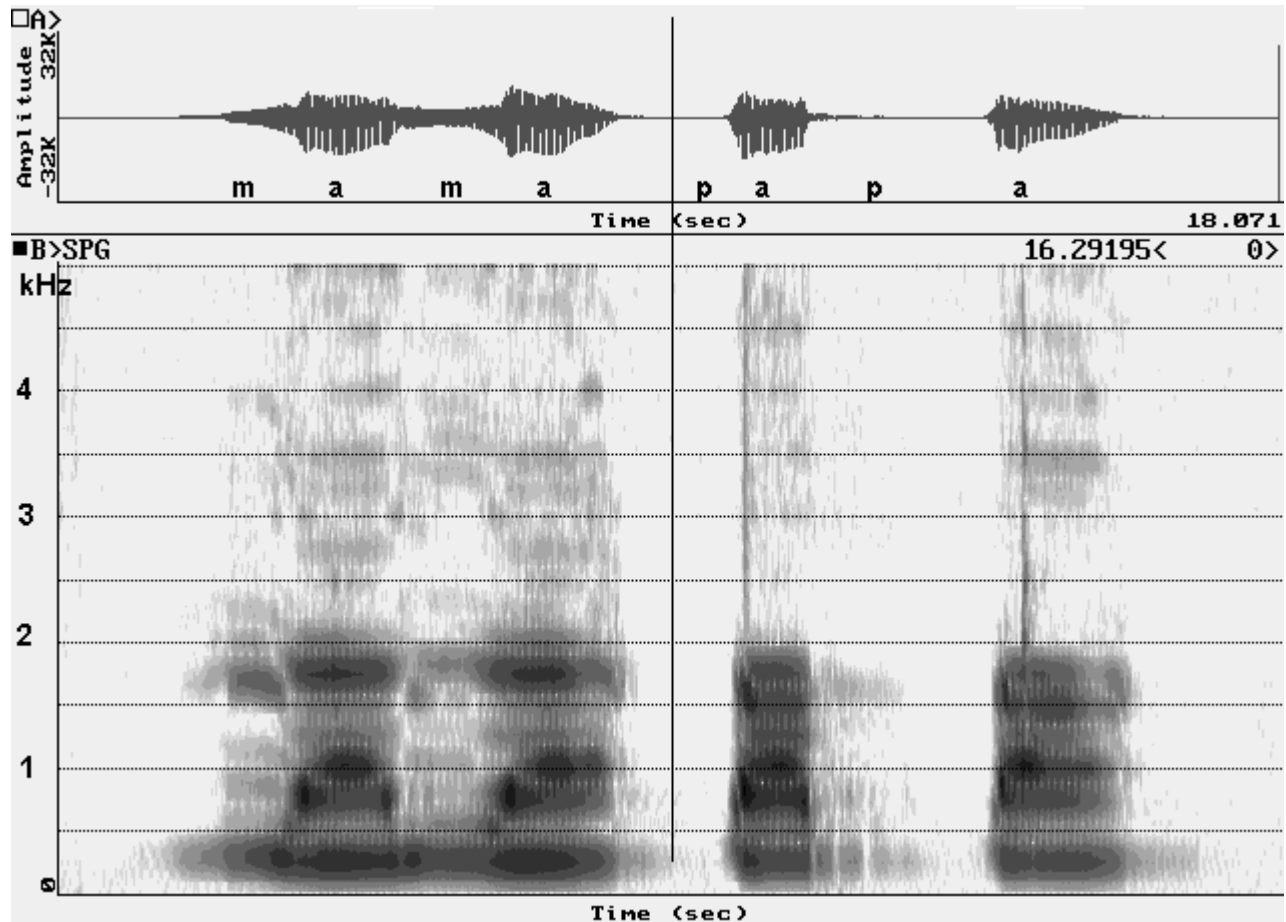
1791



Mollay Károly fordítása
1989

A rekonstruált gép hangja

2001



Kempelen gépe volt az első nyelvfüggetlen artikulációs beszédkeltő szerkezet



Es war.

I go.

Je t'aime.



Kempelen replikák találkozója 2019. Bécs



Exakte Nachbildung der Sprachbox der ersten sprechenden Maschine der Welt.

1791

2001

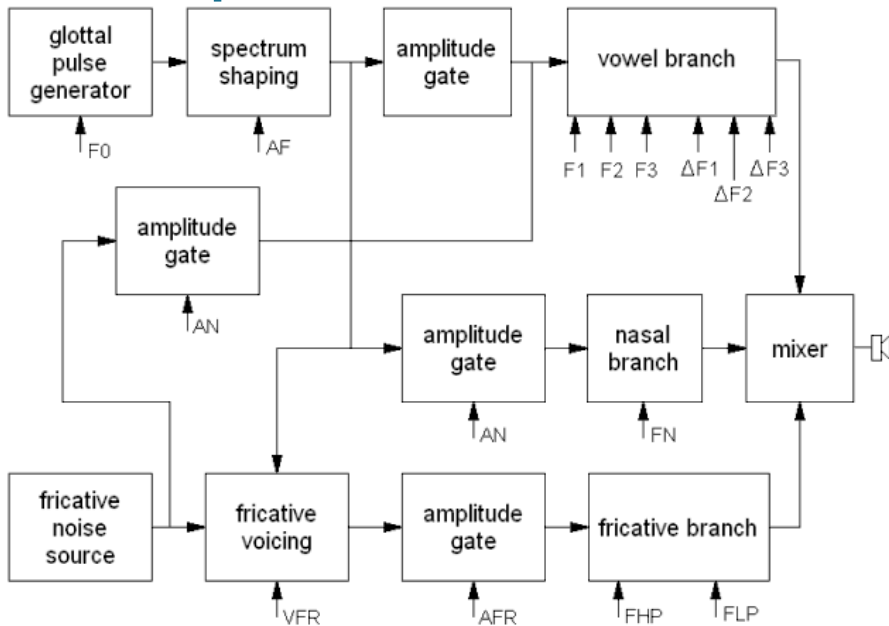
Die ursprüngliche Beschreibung wurde genau befolgt

A Kempelen gépek „kórusa” 2019. Bécs



Forrás- szűrő modell (formánsok, érthetőség)

Kempelen Farkas 1791



HungaroVox 1982



MultiVox 1986-2002



Olaszi P. – Olasz G. – Kálmán Zs.: A Blissvox-beszélő kommunikációs rendszer. Beszédkutató'94, MTA Nyelvtudományi Intézete, Budapest, 1994. 228-236.

Hawking gépi hangja angolul és magyarul

Dectalk 1982



ProfiVox 2000 – 2014



Mindenség elmélete film magyar szinkronhangja

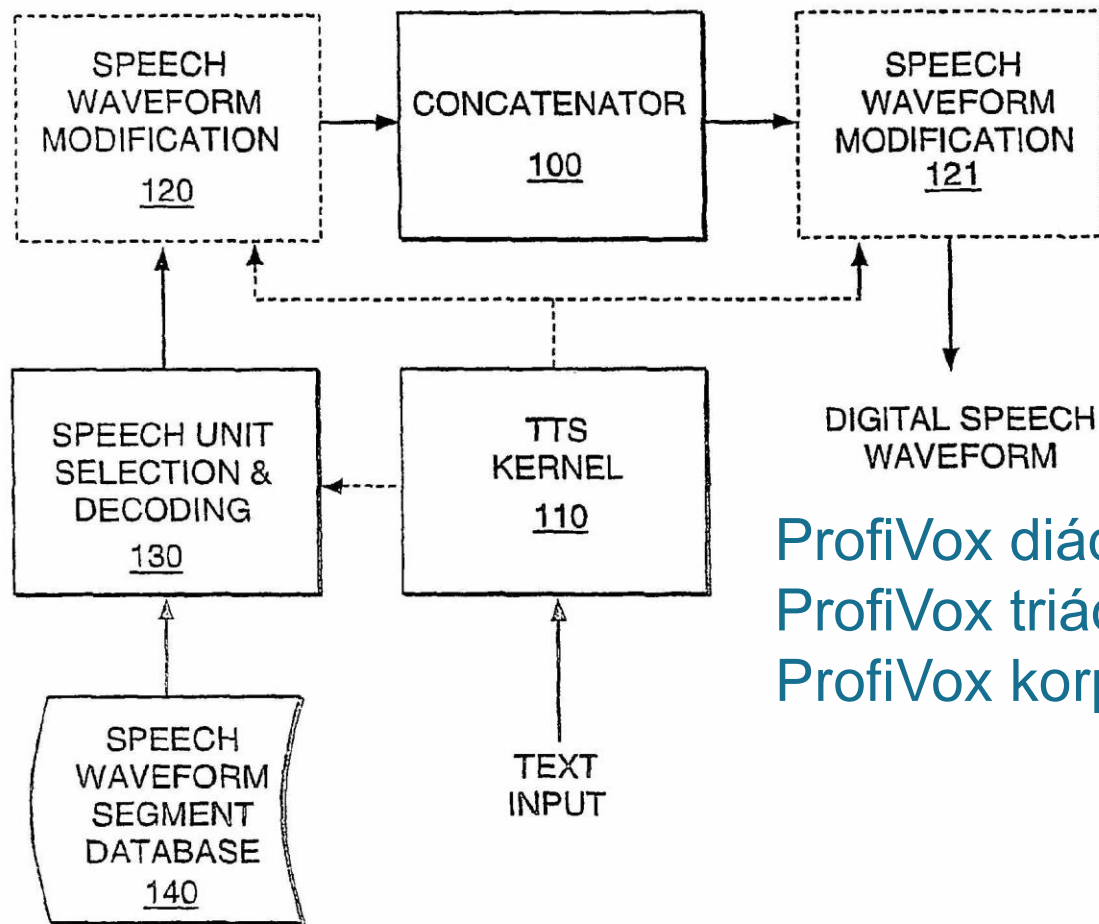
A fejlődés útja

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 (természetesség, 1916-)



BLOCK 120 AND 121 ARE OPTIONAL IN CORPUS-BASED SYTHESIS

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



Új hang a magyar vasútállomásokon

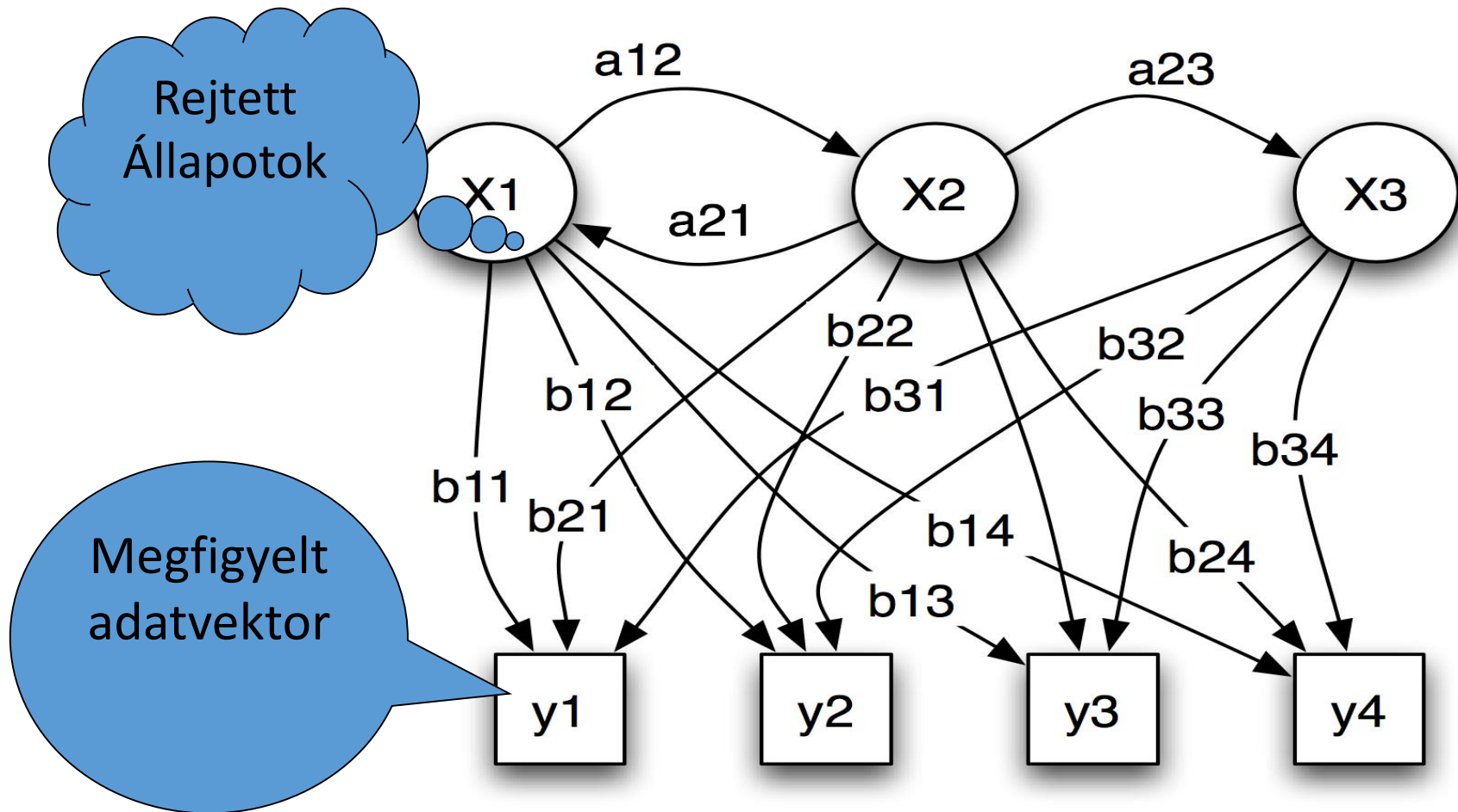
MAGYARORSZÁG SZEMÉLYSZÁLLÍTÁSI VASÚTI TÉRKÉPE - 2011

Készítette: Balla László 2011.II. - www.vasut.info; www.vasut.tk

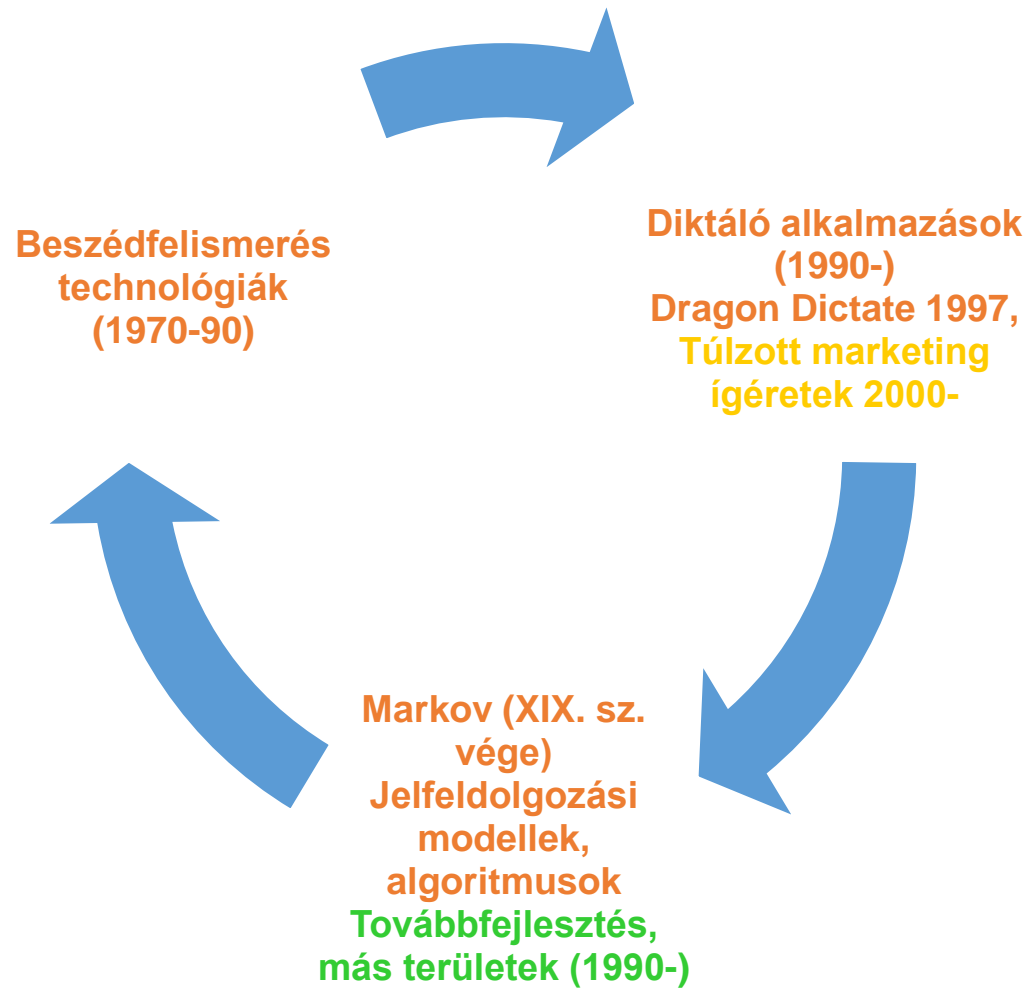


<http://smartlab.tmit.bme.hu>

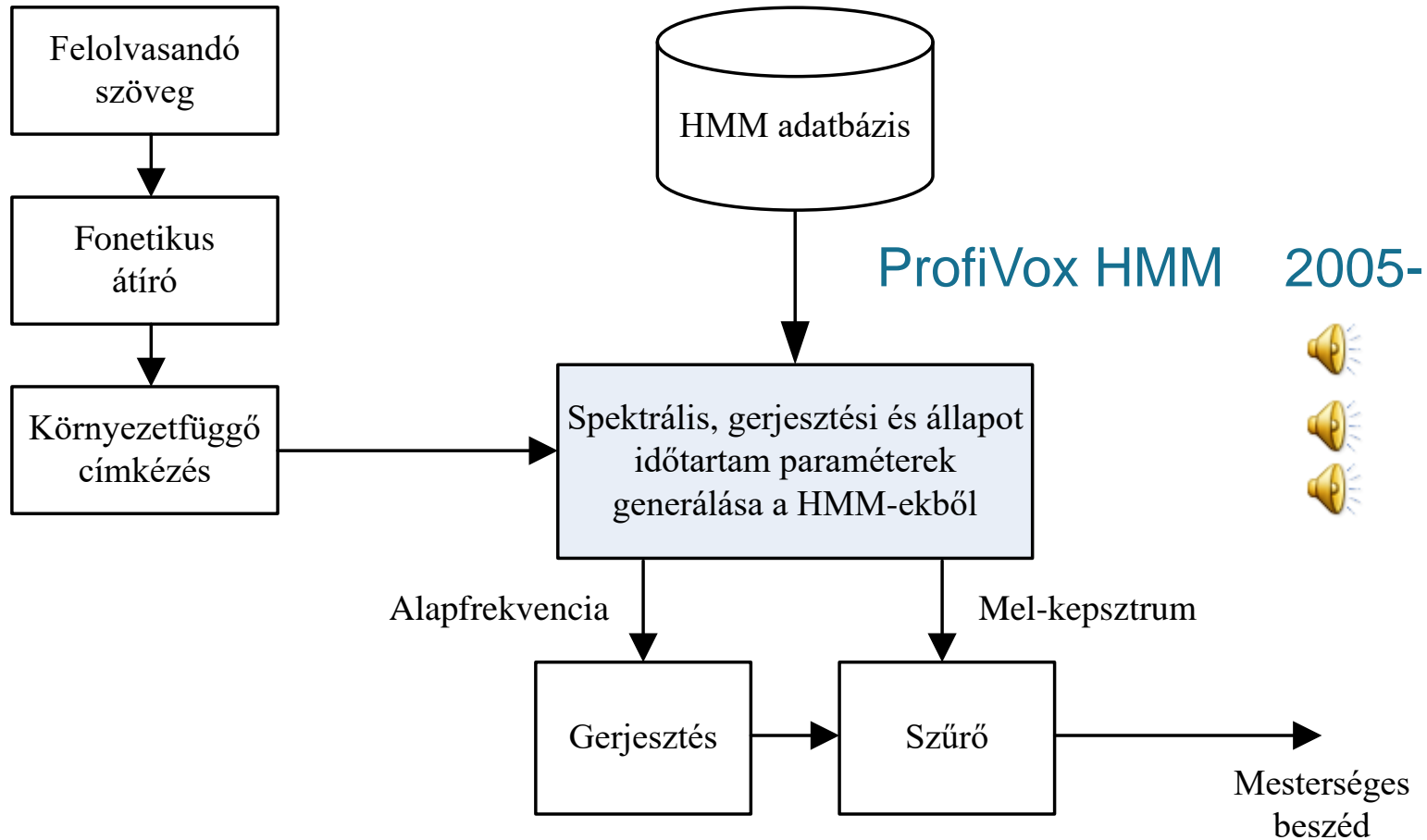
Alap kutatás (Hidden Markov Model, HMM 1970-) ¹



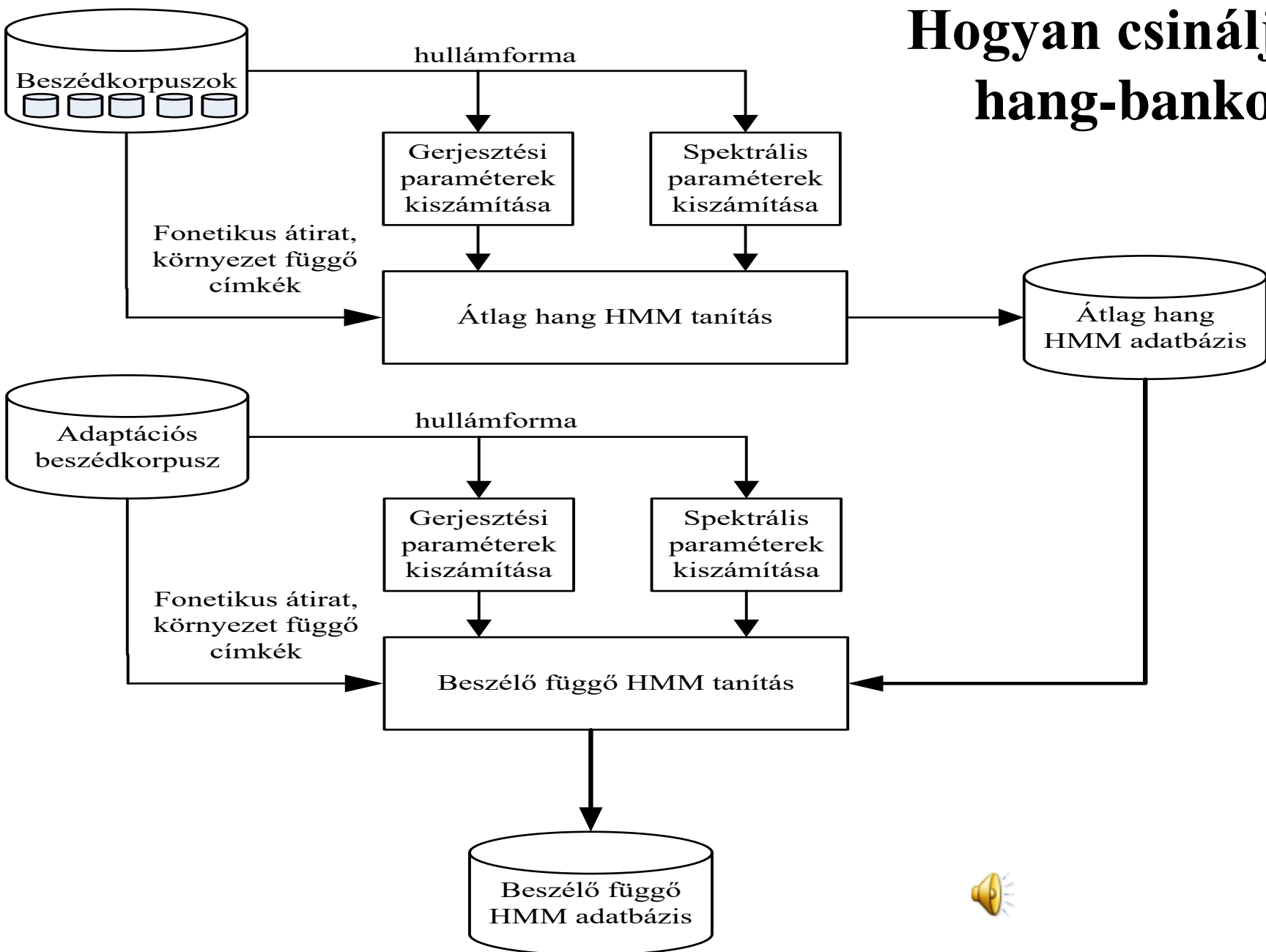
Alap kutatás (Rejtett markov modell, HMM) ²



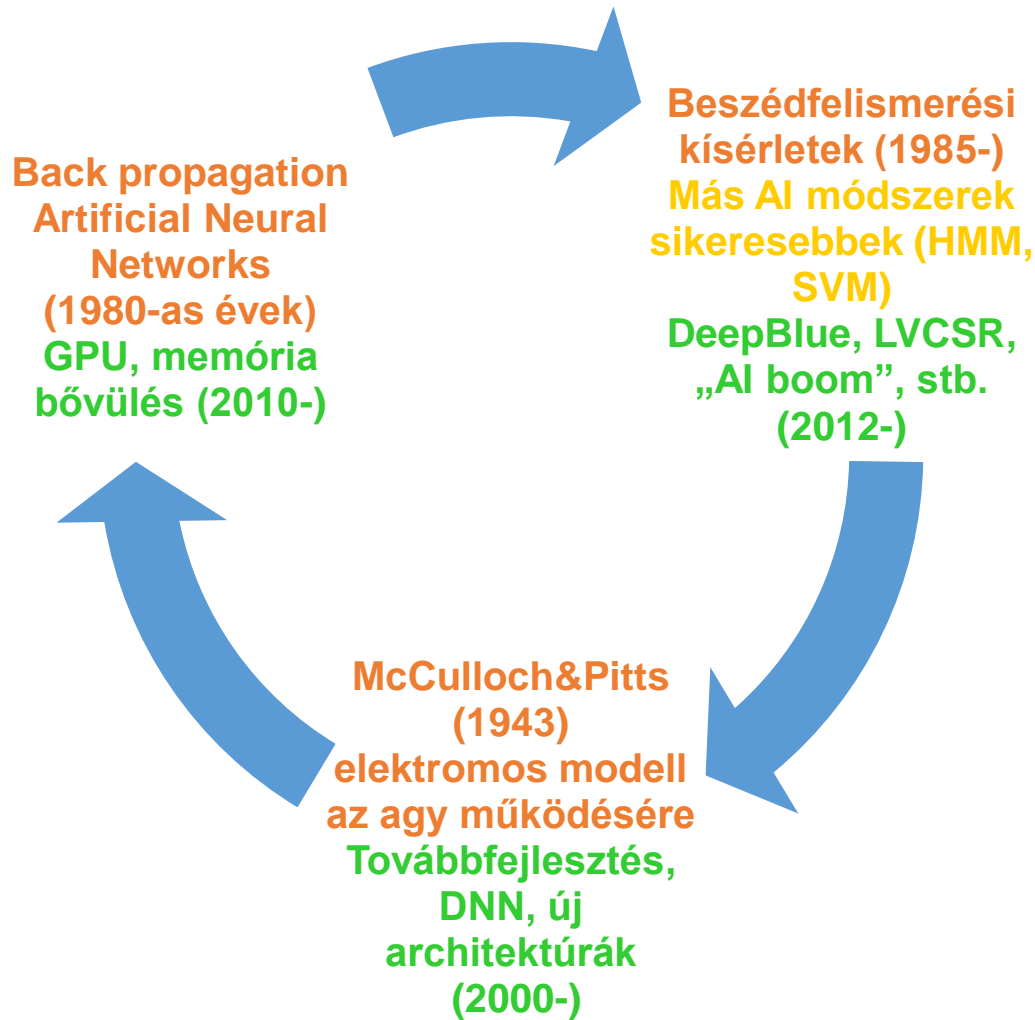
Technológia fejlesztés HMM-alapokon (rugalmasság, 200x-)



Hogyan csináljunk hang-bankot?



Alap kutatás (Neurális hálózatok)




Deep Learning a SmartLab-ban

- Elsődleges kutatási területek
 - Beszédszintézis
 - Idősor modellezés
 - Audió osztályozás
 - Természetes nyelv megértése
 - Kép- és video szegmentálás és osztályozás
 - Pénzügyi alkalmazások
 - Áramlástan modellezés
- Oktatás
 - Hivatalos NVidia GPU Education Center
 - Deep Learning a gyakorlatban Python és LUA alapokon
 - 2x2 előadás / hét, projekt feladat, 120 hallgató ebben a félévben

Beszédészintézis technológiák

- Hullámforma összefűzés

- Elemösszefűzés (diád, triád) 

- Kis erőforrás, pl. beágyazott robot, KKV-IVR

- Elemkiválasztás (korpusz alapú) 

- Közel emberi minőség korlátozott tematikán
 - PC, felhő



- Statisztikus parametrikus

- Rejtett Markov modell alapú 

- Általános szövegre is alkalmazható
 - Egyedi személy hangjára is elkészíthető
 - Android, PC

- WaveNet (end-to-end DNN, fejlesztés alatt) 

- Bármely személy hangjára elkészíthető
 - Átlag hang képezhető

Miért nem általános „Dobozos” TTS?

- Nem elég emberi minőségű hangzás
- „Gyári” –többnyire egyetlen- hang,
 - Nem vagy nagy költséggel cserélhető
 - Gyakorlatilag nem javítható
- Természetellenes hangsúlyozás
 - Nem megfelelő nyelvi modul
- Nehézkes vagy lehetetlen témakör adaptáció
- Korlátozott rendszertámogatás

Ipari TTS referenciák

- 1999 Westel: e-levél felolvasó (MailMondó)
- 2002 Digitania Rt.- Origo: SMS felolvasás
- 2002- Jaws for Windows (USA): képernyőolvasó magyar változat
- 2003 T-mobile: SMS felolvasó mobilba (SMSMondó)
- 2004- T-mobile, Vodafone: automatikus számszerinti tudakozó
- 2004- Volksbank, Wincor Nixdorf, NCR, K&H, OTP: beszélő bankautomata
- 2008- Gyógyszervonal, automatikus telefonos információs rendszer
- 2008- Automatikus mobiltelefon-árlista felolvasó
- 2012- Menedzselt mobiltelefon felügyelete Androidon
- 2014- Hangos utastájékoztatás MÁV

SmartLab projektek mostanában

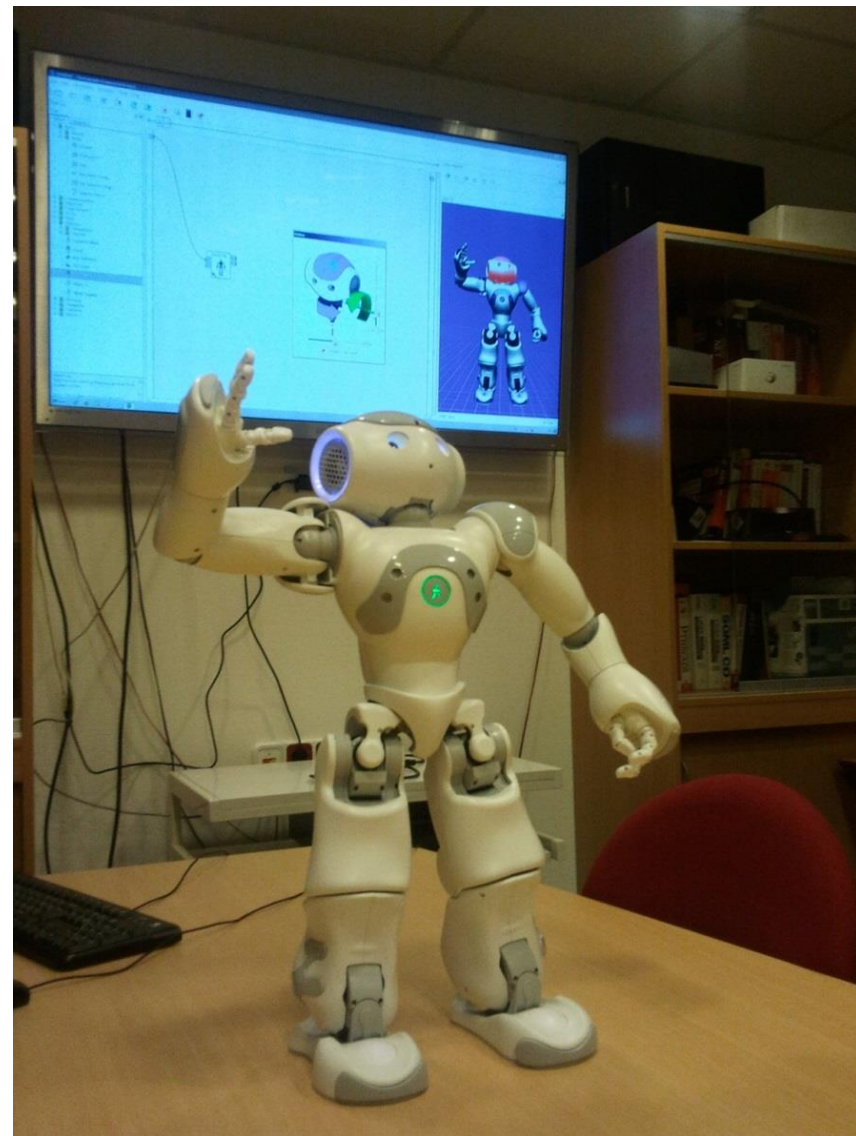
- Vasútállomási hangos utastájékoztató
- Navigáció vak emberek számára
(Visionless Supporting Framework, EU AAL, 2016-2018)
- SmartPhone / Tablet demó alkalmazások
 - Weather for All – Windows 8; > 40.000 letöltés
 - NAO humanoid beszélő robot
- Mobil szenzor analitika (SoleCall spinoff, APH-ALARM EU AAL 2019)

- Szolgáltatás: termék- és rendszer támogatás
 - Automatikus szám-szerinti tudakozó
 - Jaws for Windows képernyőolvasó

NAO humanoid robot

- Magyar TTS a NAO-ra
- Játékok gyerekeknek
- Kórházi alkalmazás

- Ember-robot interfész választható tárgy



SoleCall Demo



Céges kapcsolatok



THINKTech

Speechtex The Speech Technology Expert



IT.DOT Kft

Morgan Stanley



AVAYA



otpbank

Continental



NET



algotech

DIGITAL NATIVES

RoboBraille.org



VOLKSBANK



WINCOR NIXDORF



Informatika a látássérültekért Alapítvány

E-GROUP
SOFTWARE & BEYOND

MOHAnet
Integrált Mobil Alkalmazások



mobilitás és multimédia klaszter

NEMZETI KUTATÁSI, FEJLESZTÉSI ÉS INNOVÁCIÓS HIVATAL

NKTH
Nemzeti Kutatási és Technológiai Hivatal

SmartLab
Intelligent Interactions

nvidia
GPU EDUCATION CENTER

Beszéd felismerés és Hangbányászat Labor

1994-

Múlt és jelen

Gordos Géza †



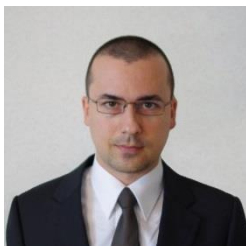
1956

Elődök:

Tatai Péter
(AITIA)



Szarvas Máté
(NVIDIA)



Fegyő Tibor



Mihajlik Péter



SmartLab
Speech Recognition

*Hallgatók:
MSc, BSc
kb. 6 fő*

*Nálunk végzett:
kb. 40 fő*



Tarján Balázs



Juhász Krisztián

Tüske Z.
(IBM Research)



Varga Á.
(Continental)



Szaszák Gy.
(Avaya)



Tevékenységeink

Beszéd- és mintázatfelismerő technikák
kutatás-fejlesztése, oktatása,
partnereinkkel
kapcsolódó alkalmazások fejlesztése.

Referenciák / főbb felhasználási területek



Telefonos ügyfélszolgálatok

Médiafeliratozás

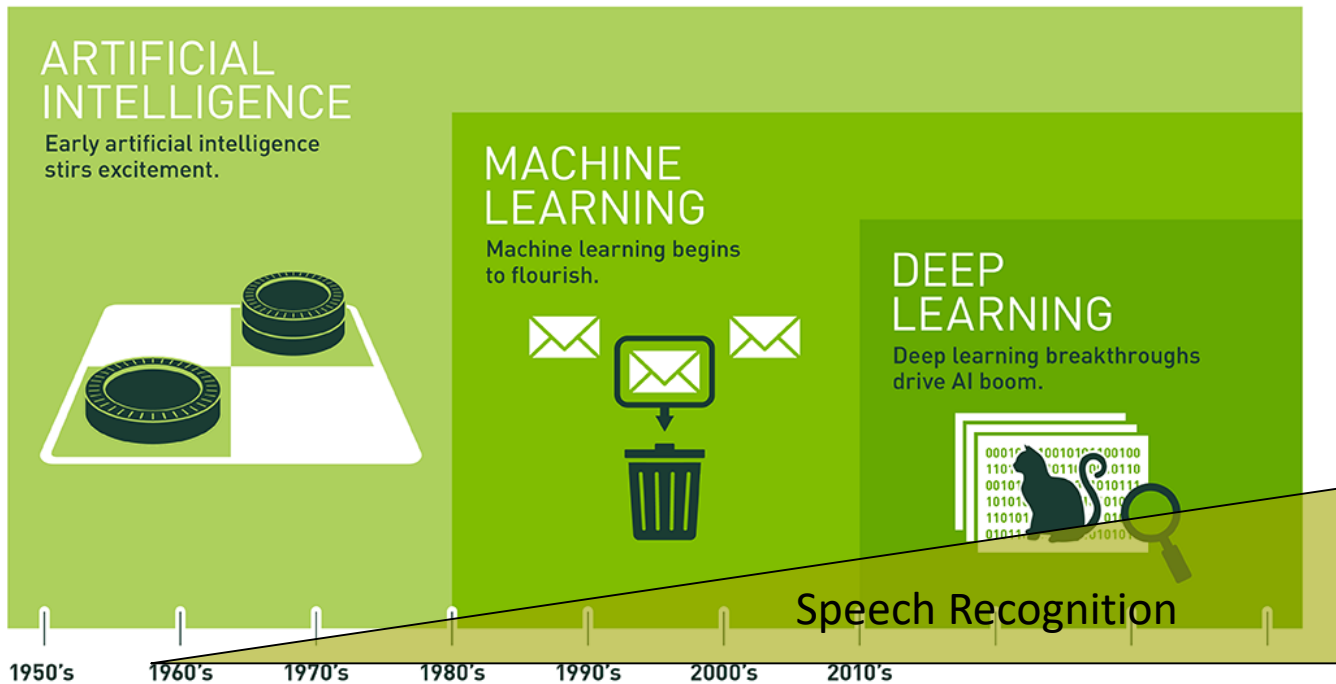


Diktálás

Jegyzőkönyvkészítés

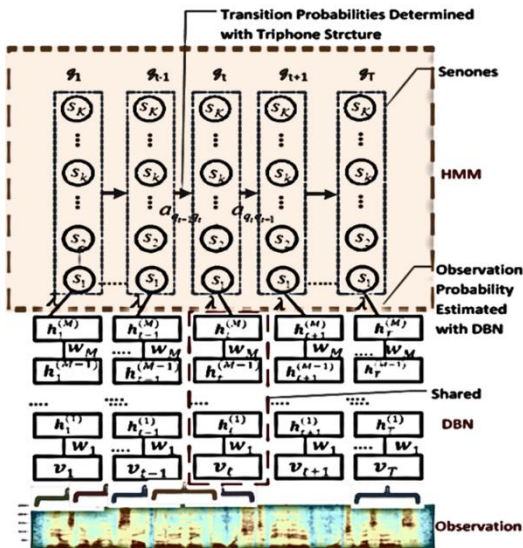


96- gépi tanulás az alaptéchnológia (HMM)



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

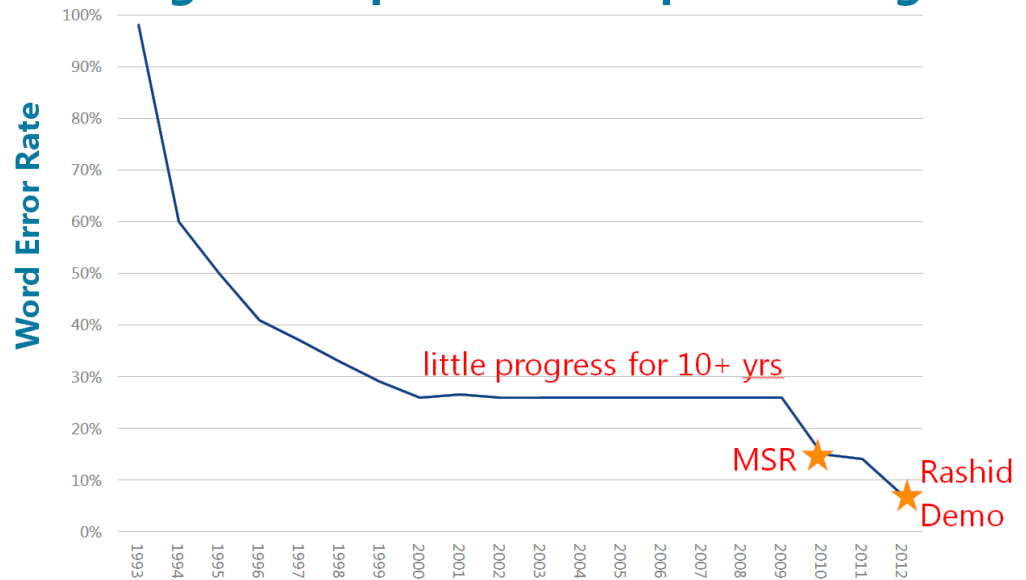
Beszédfelismerés és deep learning (2011-)



Dahl, Yu, Deng, and Acero, "Context-Dependent Pre-trained Deep Neural Networks for Large Vocabulary Speech Recognition," *IEEE Trans. ASLP*, Jan. 2012 (also ICASSP 2011)

Seide et al, Interspeech, 2011.

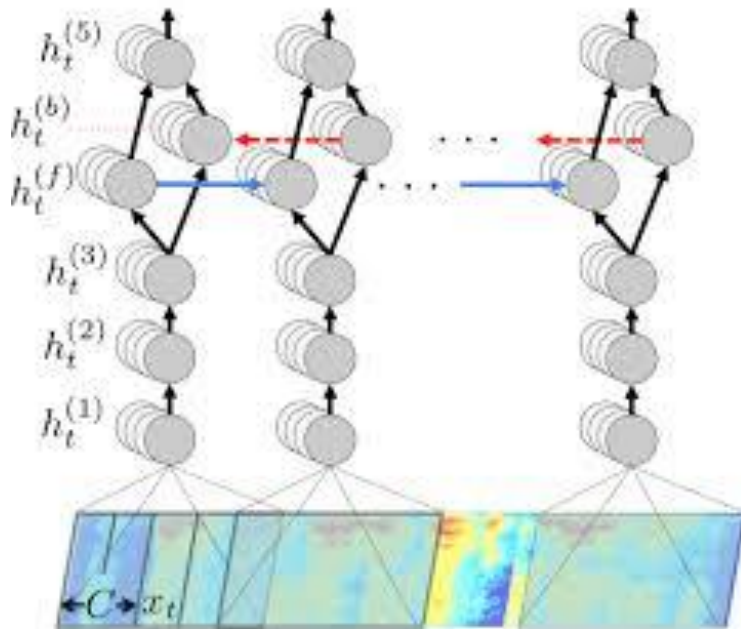
Progress of spontaneous speech recognition



After no improvement for 10+ years by the research community...

...MSR reduced error from ~23% to <13% (and under 7% for Rick Rashid's S2S demo in 2012)!

A jövő: End-to-end (NN) speech recognition



RNN+CTC
(Recurrent NN
+
Connectionist Temporal Classification)

Újabb és újabb nyílt forráskódú eszközök,
modellek (főleg Python alapon)

**NEM utópisztikus cél:
Az emberinél jobb pontosság!**

Demók: Jobb-e már a gép mint az ember?

- Telefonos ügyfélszolgálat ([videó](#))
- MTVA teletext ([videó](#))
- Fordítástámogatás ([videó](#))



Főbb referenciák

- Televíziós műsorok feliratozása
- Telefonos ügyfélszolgálati megoldások
- Diktálás, szakfordítás
- Önkormányzati ülések jegyzőkönyvezése





Beszédakusztikai laboratórium

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

Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Távközlési és Médiainformaticai Tanszék

Munkatársak

Sztahó Dávid

PhD (2014)



Laborvezető

- Parkinson-kór detektálása
- Kiváltott agyi jelek (EEG)

Vicsi Klára

DSc (2005)

Habil (2007)



- Audiológiai kutatások

Szaszák György

PhD (2009)



- Prozódia felismerés
- (Beszéd felismerés)

Kiss Gábor

PhD jelölt



- Depresszió felismerése
- Hipoxia detektálás

Tulics Miklós

Gábrriel

PhD hallgató



- Gégészeti-patológiás elváltozások detektálása

Tündik Máté Ákos

PhD jelölt



- Prozódia felismerés

Kovács Annamária

PhD hallgató



- Beszéd és agyi jelek kapcsolata

Részmunkaidős

hallgatók



Lavati Livia

- Mészáros Zoltán

Nagy Ildikó

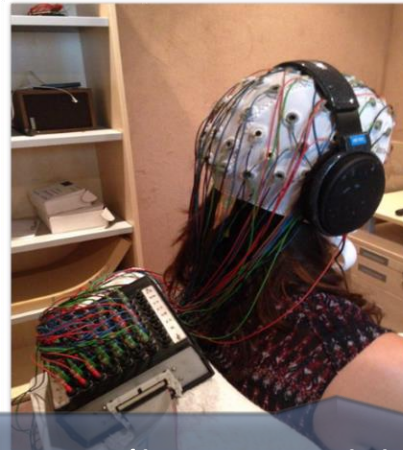
- Gazdasági ügyintéző



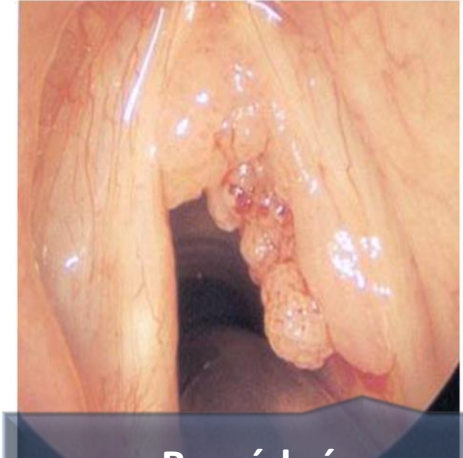
Oktatás

Részvétel a szakok oktatási tárgyaiban

- **Mérnök-informatikus**
- **Villamosmérnök**
- **Egészségügyi mérnök**



Kiváltott agyi jelek
informatikai feldolgozása



Beszéd- és
hallásdiagnosztika

Kutatási témák

Egészségi állapotok hangalapú felismerése

- diszfóniás beszéd
- Parkinson-kór
- depresszió

FORENSICspeech:

- Hangalapú automatikus beszélő azonosító és verifikáló rendszer bírósági és nyomozási folyamatokhoz

Érzelem, szándék felismerés

Prozódia felismerése

Kiejtés oktatás

Hangdiagnosztikai eszközök fejlesztése

automatikus diagnosztizálási lehetőségek

A beszélő **fiziológiai, pszichológiai, neurológiai állapota** hatással van a létrehozott beszéd fizikai paramétereire

gégészeti diagnosztikai rendszer

- hangképzőszervek kóros elváltozásának a beszédre gyakorolt hatás-vizsgálatával

depresszió detektálás

- pszichológiai kóros elváltozások beszédre gyakorolt hatásának vizsgálatával

Parkinson kór detektálása

- neurológiai kóros elváltozások beszédre gyakorolt hatásának vizsgálatával

Depresszió súlyosságának detektálása

DEMO

Psychological State Assessment

Language: hungarian
Gender: male

LOAD RECORDING START RECORDING

Microphone volume

ASSESSMENT

Népmese: Az északi szél, és a Nap

Az Északi Szél nagy vitában volt a Nappal, hogy kettejük közül melyiknek van több ereje. Egyszer csak egy utast pillantottak meg, amint köpenybe burkolódzva közeledett. Elhatározták, hogy a vitát az nyeri meg, amelyik előbb veszi rá az utast, hogy kabátját levegye. Az Északi Szél összeszedte egész erejét, és fújni kezdett, de minél erősebben fújt, az utas annál szorosabbra fogta össze a kabátját. Az Északi Szél végre feladta a harcot. Ekkor a Nap küldte meleg sugarait az utasra, aki rövidesen levette a kabátját. Az Északi Szélnek tehát el kellett ismernie, hogy kettejük közül a Nap az erősebb.

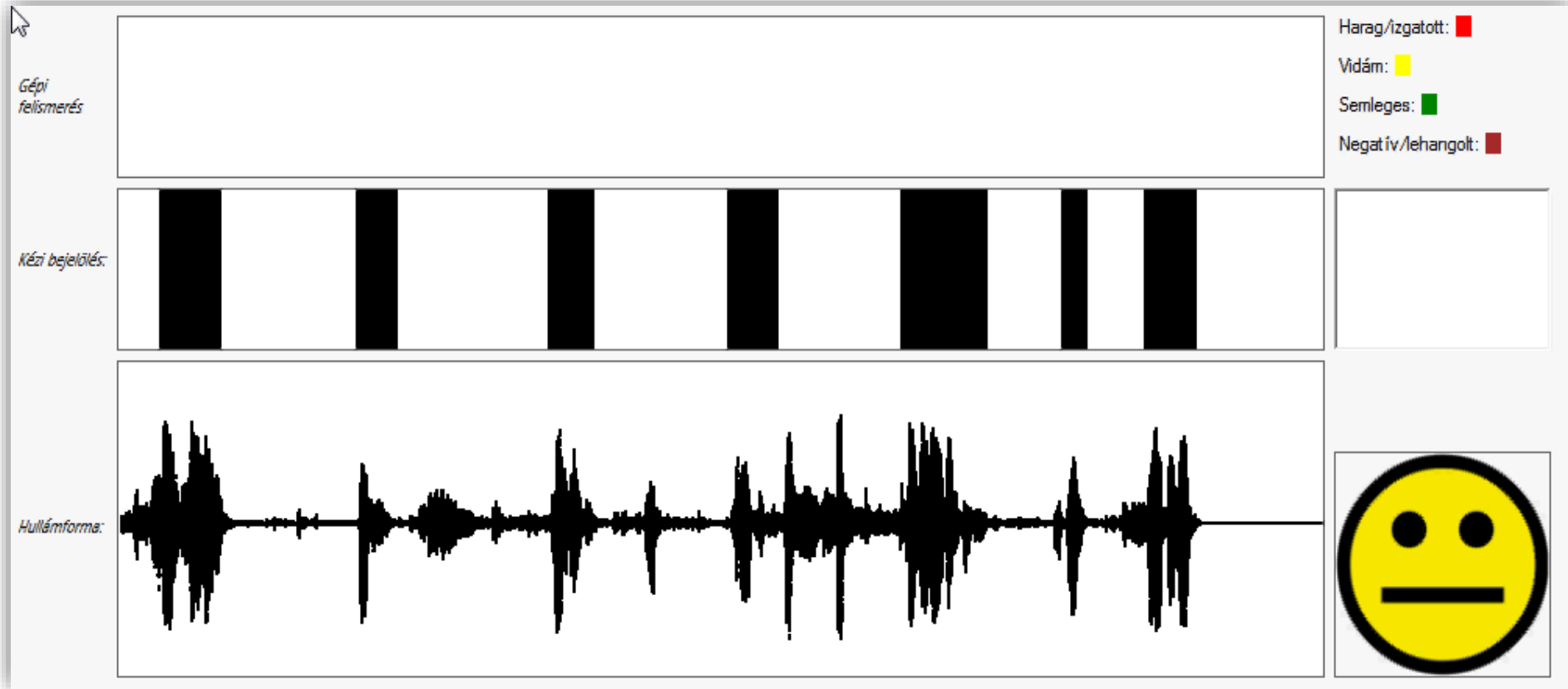
COALA-Phonetics

Psychological State Assessment

Laboratory of Speech Acoustics (LSA) - BME TMIT

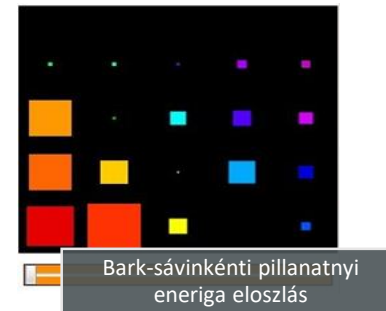
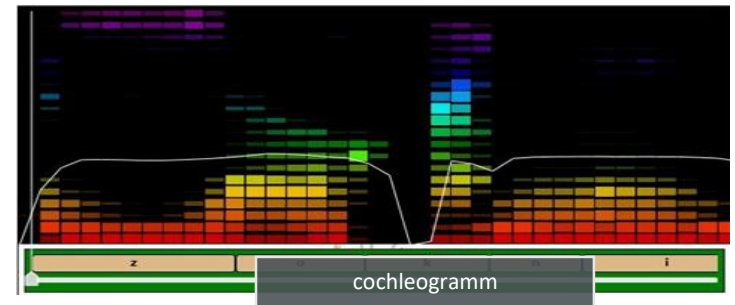
Érzelem felismerés

demo



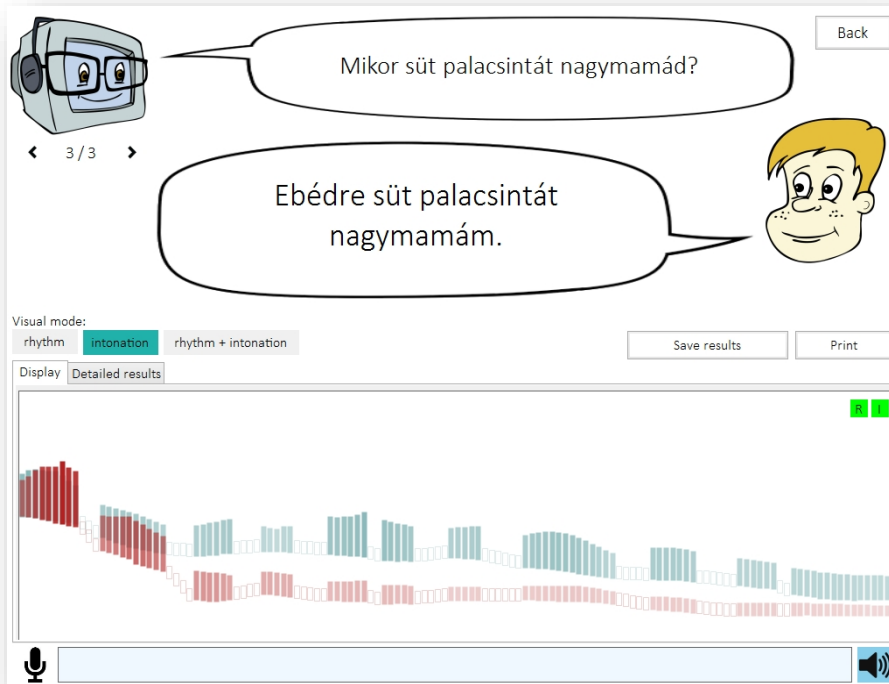
Beszéd kiejtés oktatósi rendszer siket és nagyothalló, valamint kochleáris implantátummal rendelkező gyerekek számára

Speech Assistant



TÁMOP 4.2.2.C FuturICT.hu Project

Az intonáció és hangsúly vizuális megjelenítése és automatikus kiértékelése



<http://lsa.tmit.bme.hu/pop/>

Ipari kapcsolatok

Auxilis Pharma Kft.

- Csecsemő kóros sírás, köhögés és szívhang detektálása



Kürt Zrt.

- Prozódia oktatás



Robot Control Kft.

- Kiejtés oktatás

Számos európai kutatási hálózat tagja a SmartLabs

- **AI4EU – Artificial Intelligence for EU** (www.ai4eu.org)
 - Németh Géza magyar NCP
- Mesterséges Intelligencia Koalíció (www.miagyakorlatban.hu)
- CLARIN - Common Language Resources and Technology Infrastructure
- META_NET – Network of Excellence
- CESAR (CT PSP CALL 4) – Central and South Eastern European Resources
- FLaReNet – Worldwide Map of the Contact Members in Language Resources and Technologies



SmartLabs

Kapcsolat:

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

E-mail: smartlab@tmit.bme.hu

Tel: (+36 1) 463-38-83 (Németh Géza)

SmartLab
Intelligent Interactions

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

SmartLab
Intelligent Interactions



GPU
EDUCATION
CENTER