

Beszédtechnológia orvosi alkalmazási lehetőségei

Németh Géza

Beszédkommunikáció és Intelligens Interakciók Laboratóriumok

BME Távközlési és Mesterséges Intelligencia Tanszék

SmartLab
Intelligent Interactions

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



 NVIDIA

GPU
EDUCATION
CENTER

BME TMIT SmartLabs: 3 labor

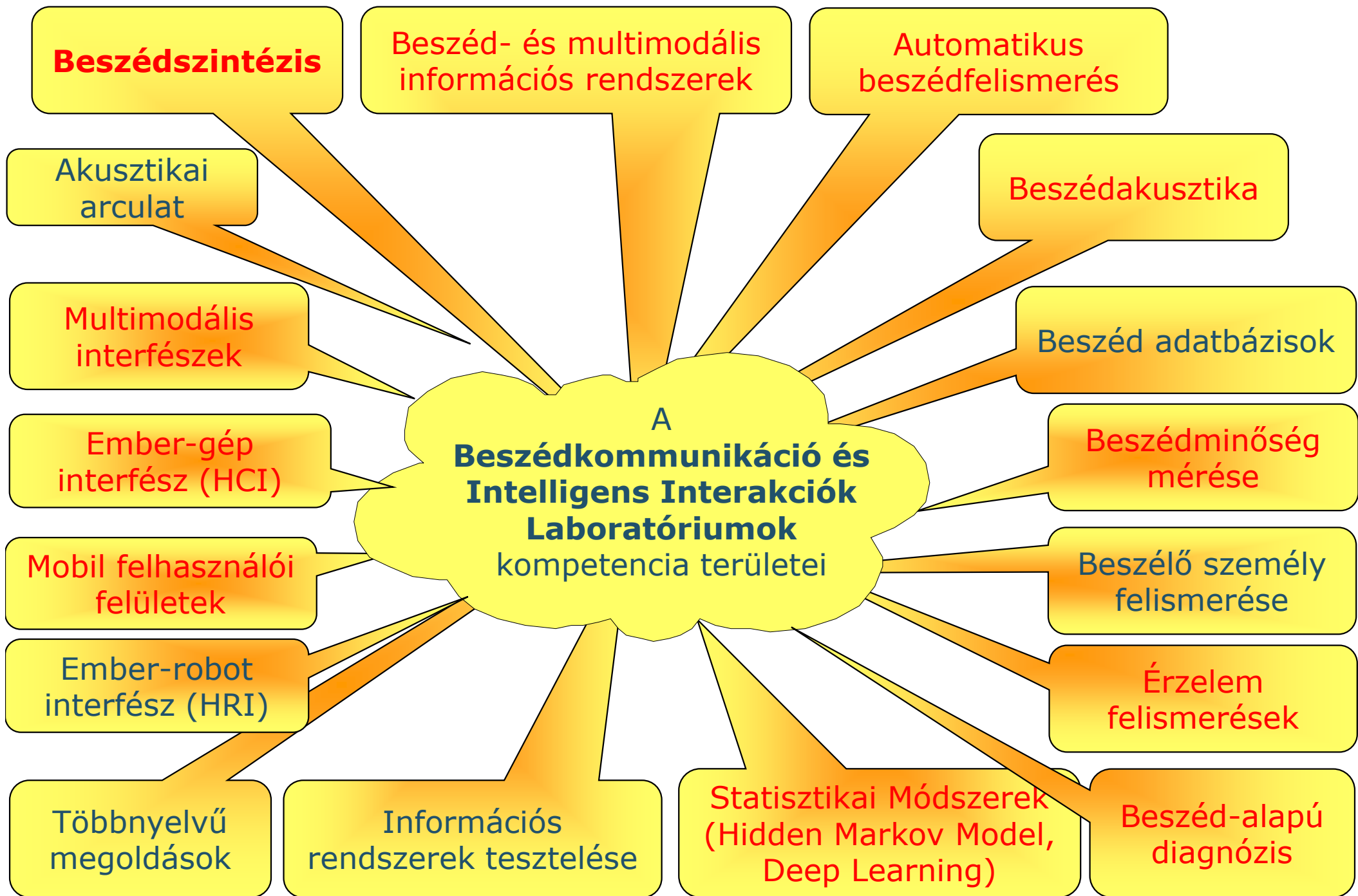
Beszédtechnológia és Intelligens Interakciók Labor (Németh Géza)

Beszédfelismerés és Hangbányászat Labor
(Mihajlik Péter)

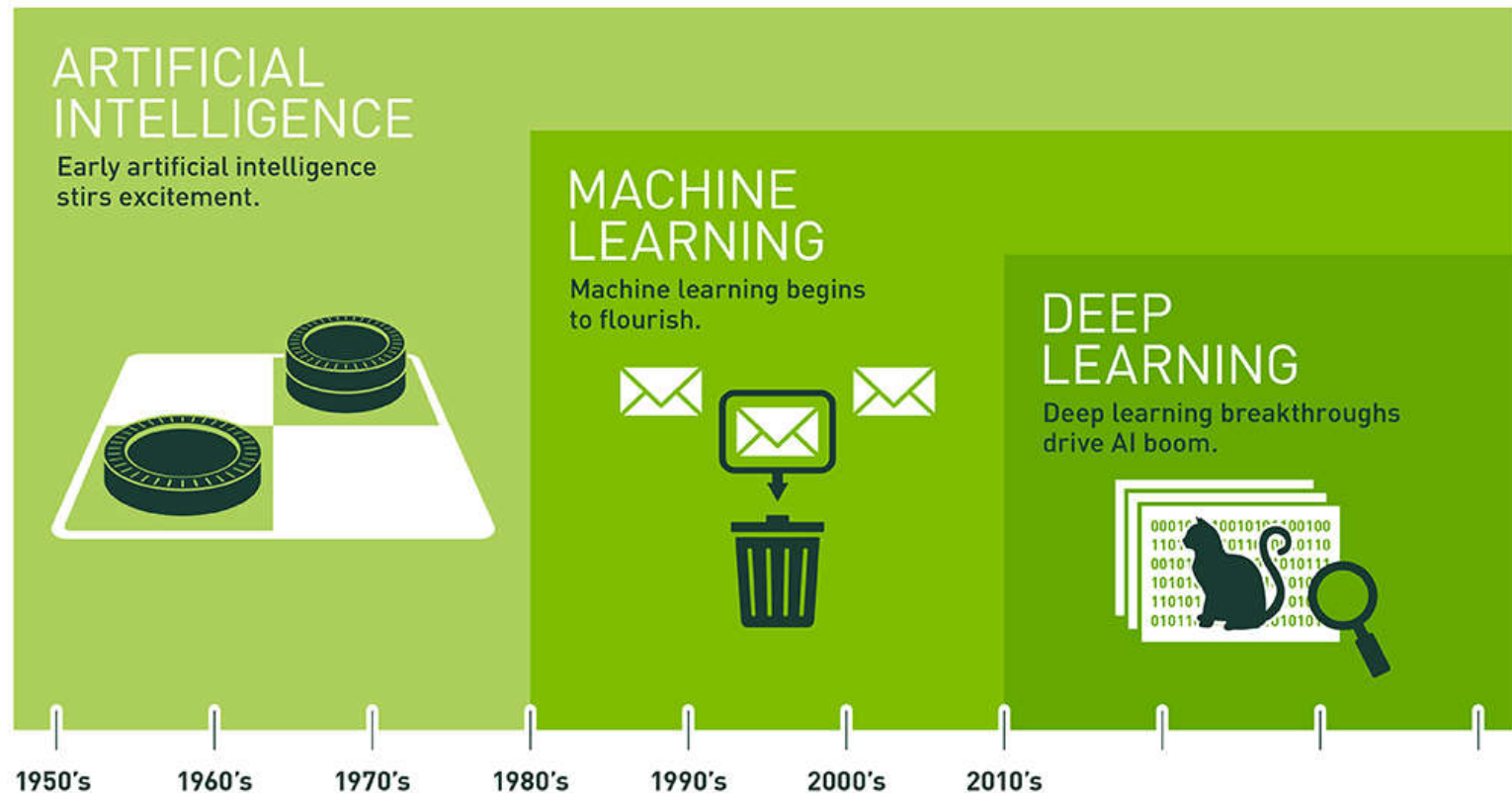
Beszédakusztikai Labor
(Sztahó Dávid)

~20 munkatárs, 5 állami
finanszírozású (3 DSc, 9 PhD)





A mesterséges intelligencia fejlődése



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.

A mesterséges intelligencia definíciói

Artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience.

(source: www.britannica.com)

A számomra kedves (saját) MI (magyar) / AI (angol) definíció:

Az emberi kognitív funkciók gépi megvalósítása és/vagy kiterjesztése

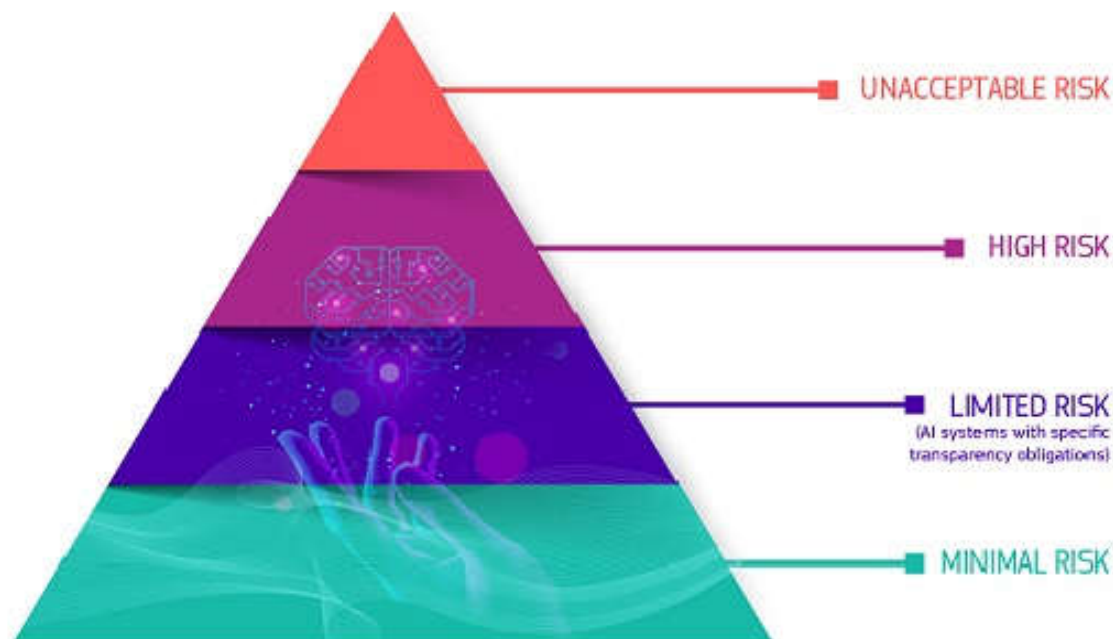


MESTERSÉGES INTELLIGENCIA
Nemzeti Laboratórium

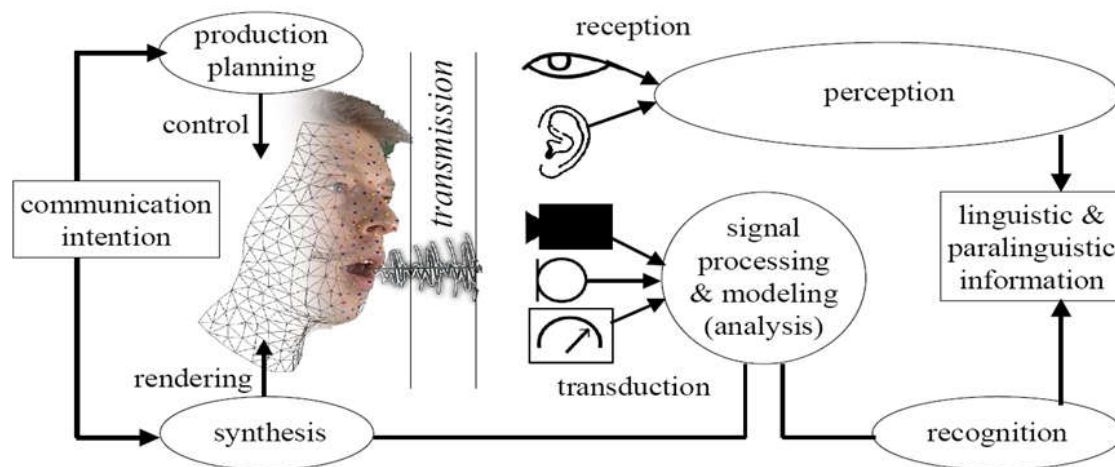
Természeti intelligencia kell az M.I. mellé

Az AI Act (459 oldal!!!) alapvető hatású és értelmezni kell:

<https://eur-lex.europa.eu/legal-content/EN-HU/TXT/?fromTab=ALL&from=EN&uri=CELEX%3A52021PC0206>

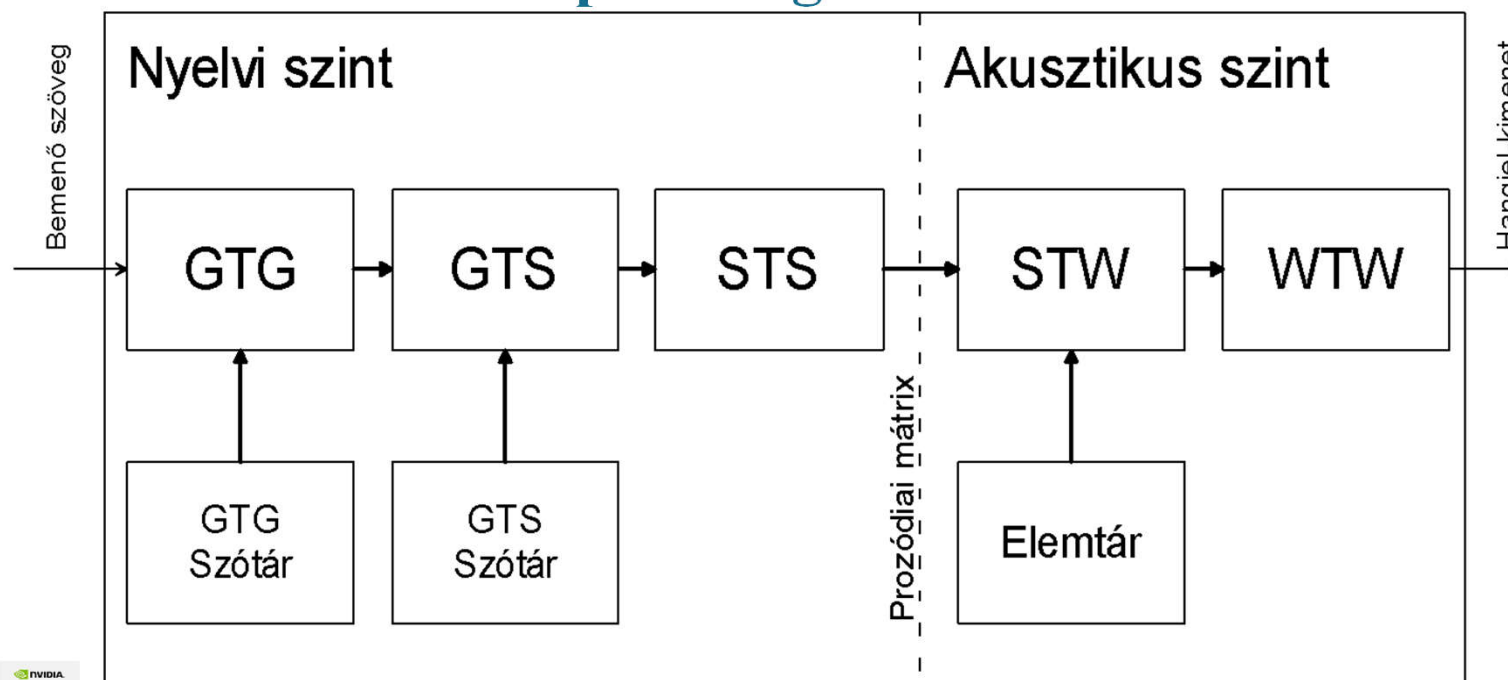


Mi is a beszédtechnológia?



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

Gépi szövegfelolvasás



Történelem

közeledés és beszédtechnológia



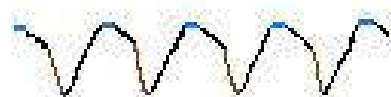
1791



2015



Alapkutatás (formáns 1791-) ₁



zöngé

+

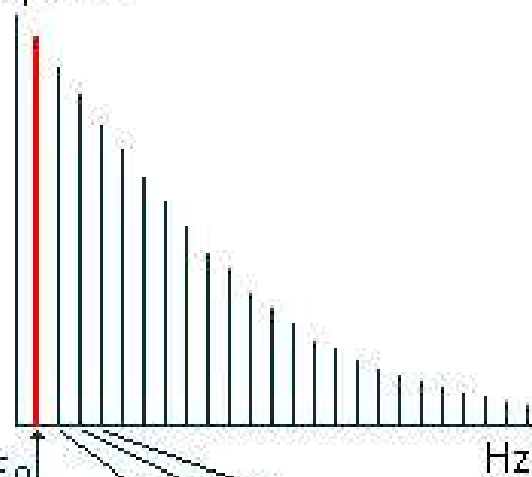
artikulációs csatorna

=

zöngés beszédhang



Amplitúdó



F_0

alaphang
(zöngé)

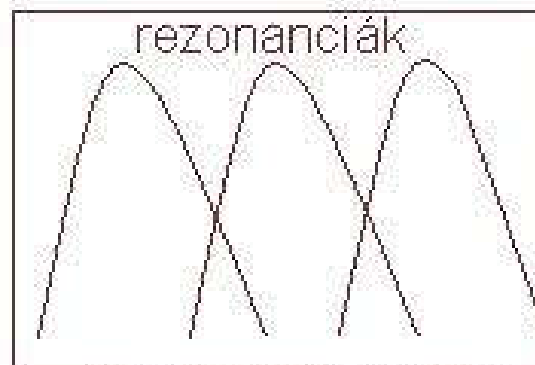
1.

2.

3.

.....
felharmonikusok

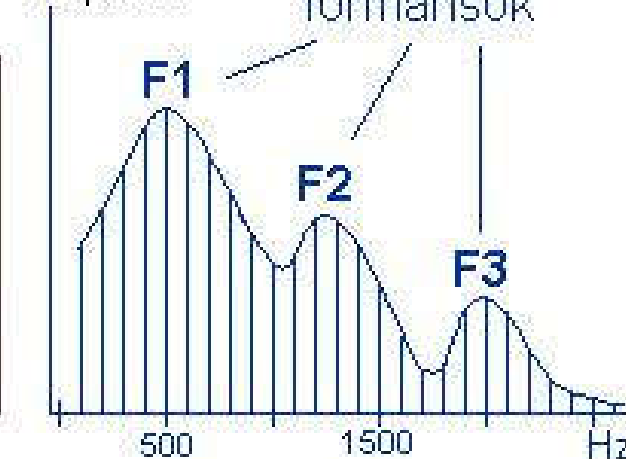
Az alaphang pillanatnyi frekvenciaértéke
az alapfrekvencia F_0 (Hz)



rezonanciák

Az artikulációs csatorna
rezonancia frekvenciái
személyfüggők

Amplitúdó



formánsok

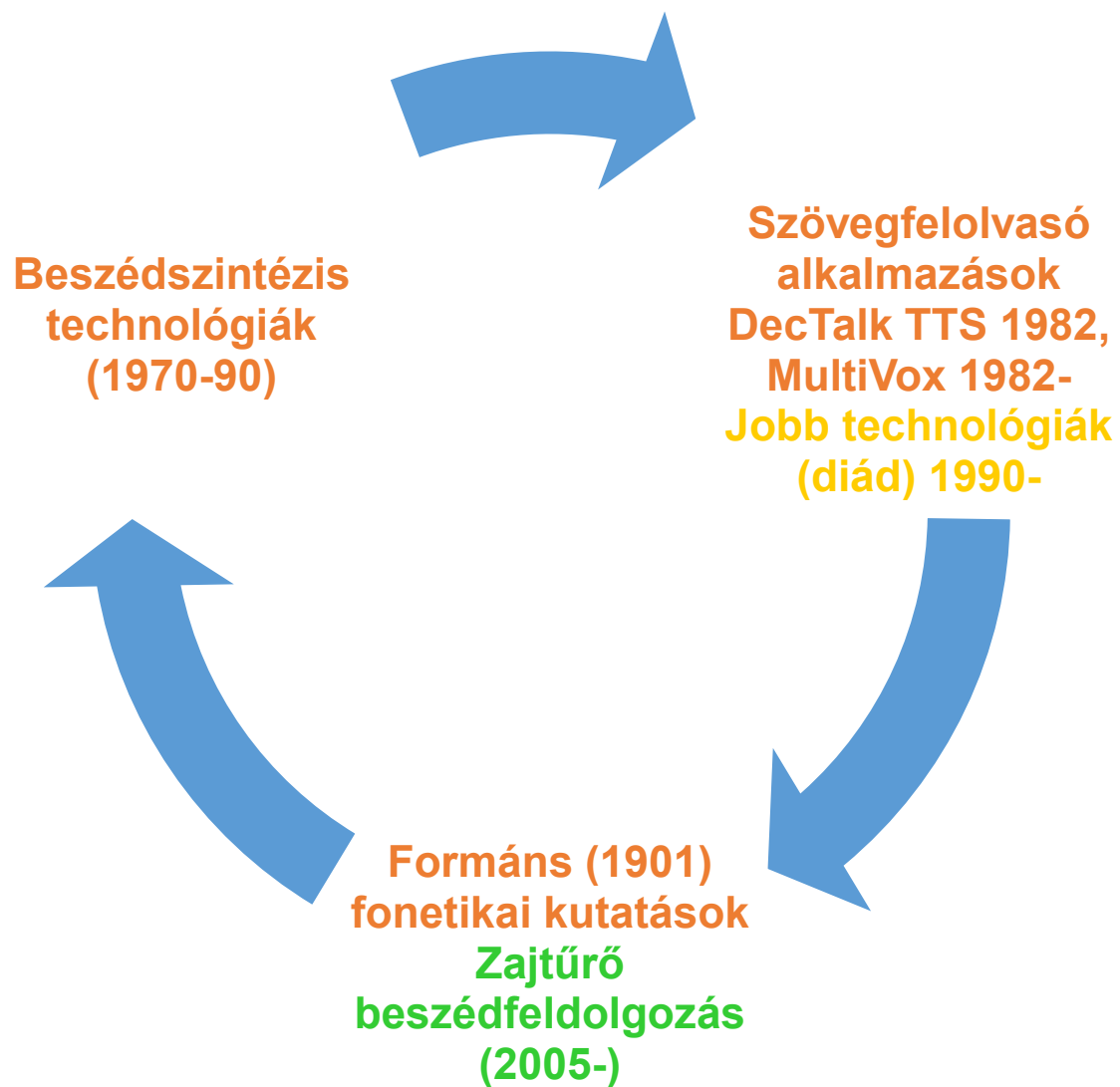
F_1

F_2

F_3

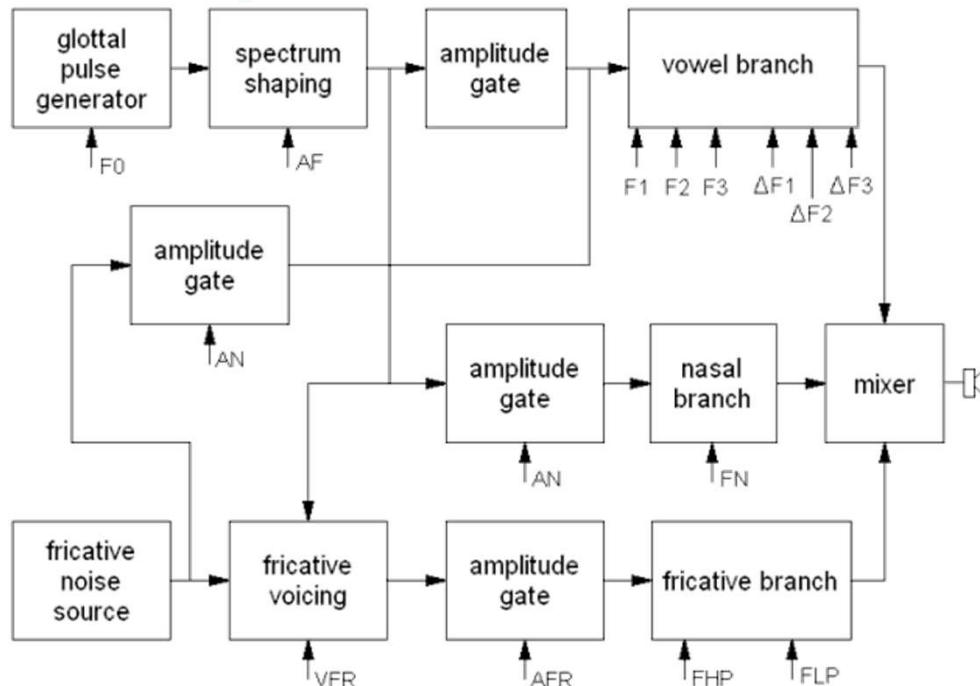
A felharmonikusok férfi hangnál
sűrűbbek, női hangnál ritkábbak,
gyermek hangnál igen ritkák

Alapkutatás (formáns) ₂



Forrás- szűrő modell (formánsok, érthetőség)

Kempelen Farkas 1791



HungaroVox 1982



MultiVox 1986-2002



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

Hawking gépi hangja angolul és magyarul

Dectalk 1982



ProfiVox 2000 – 2014

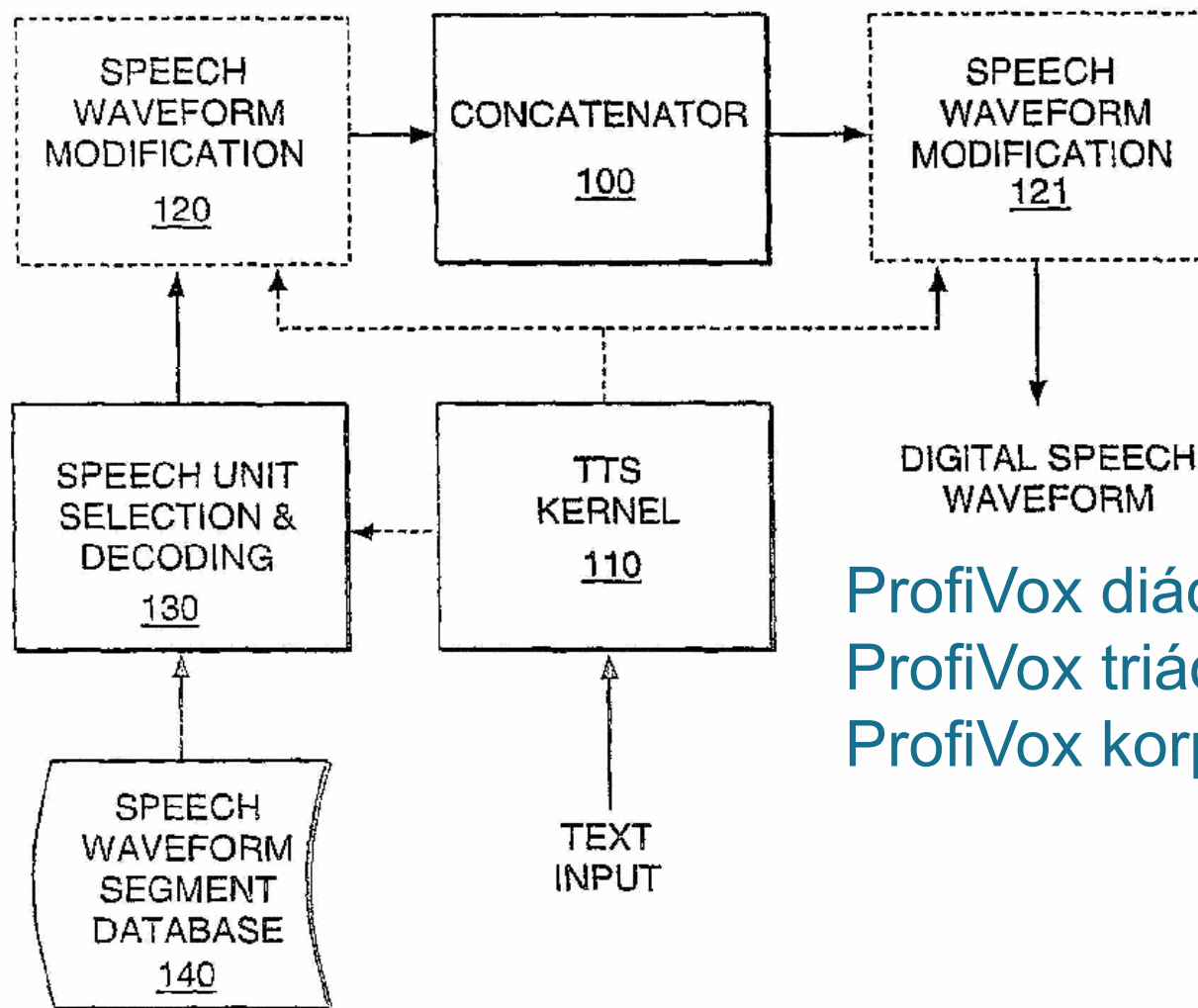


Mindenség elmélete magyar szinkronhangja

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

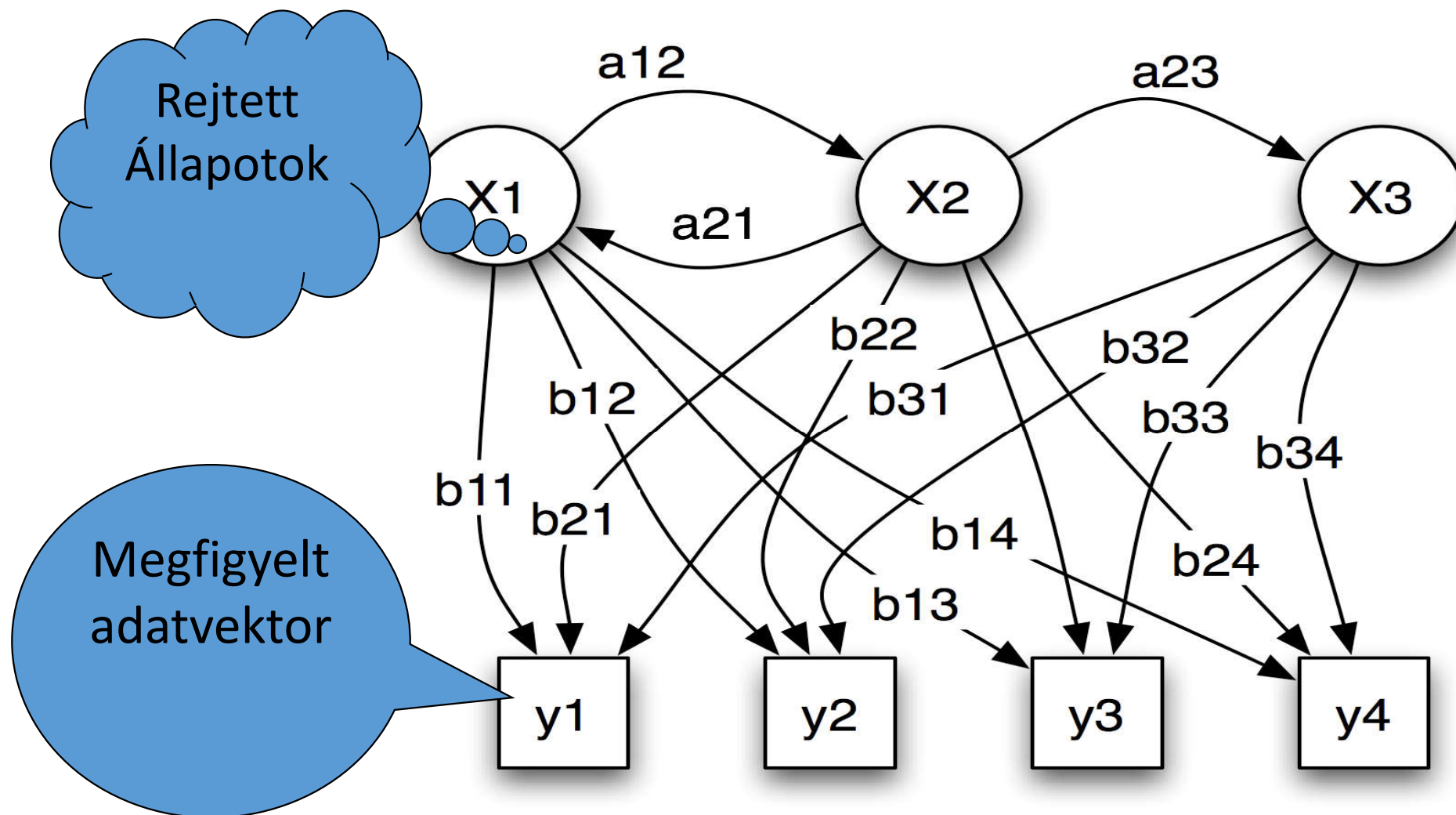
MAGYARORSZÁG SZEMÉLYSZÁLLÍTÁSI VASÚTI TÉRKÉPE - 2011

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

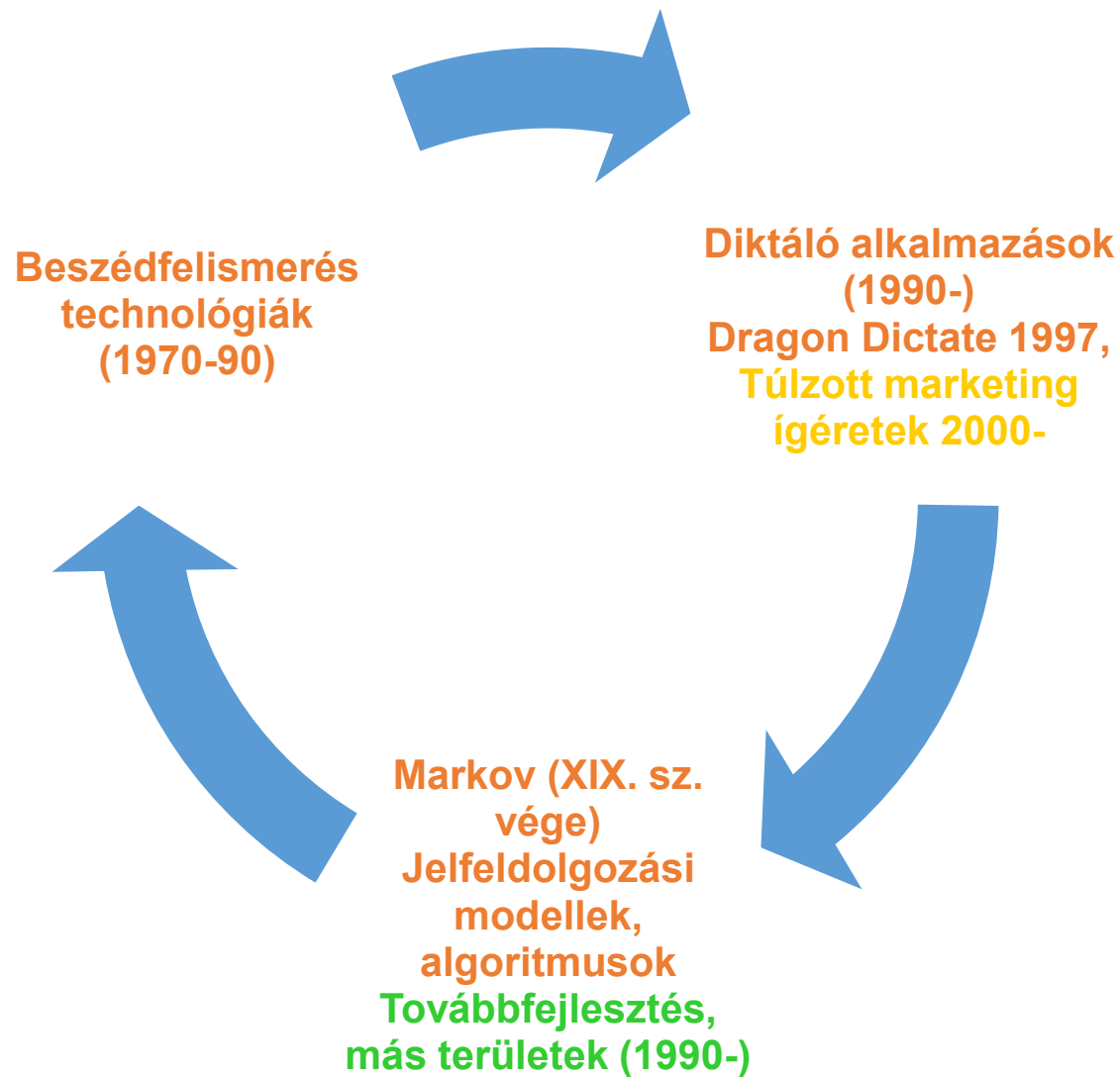


<http://smartlab.tmit.bme.hu>

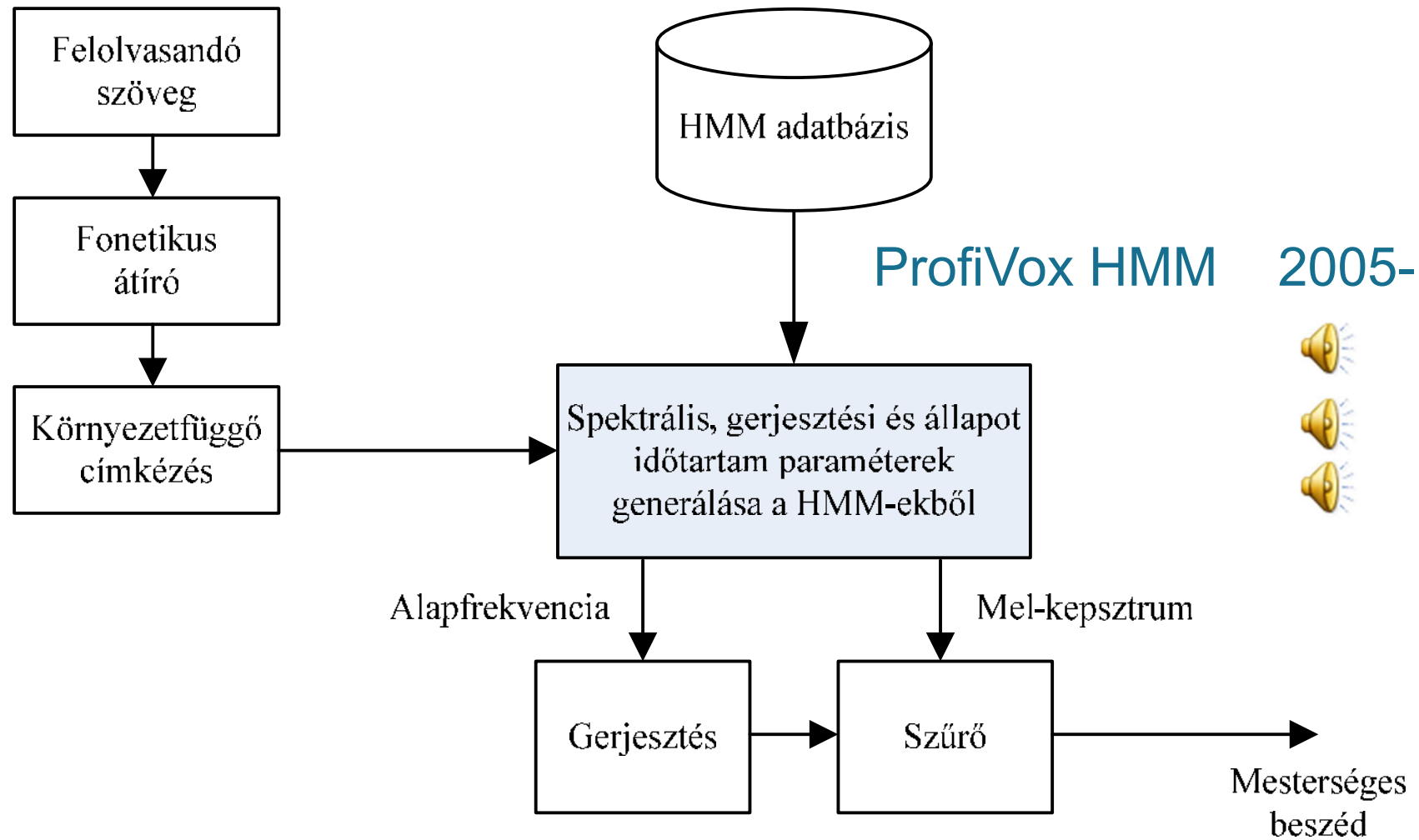
Alap kutatás (HMM 1970-) ₁



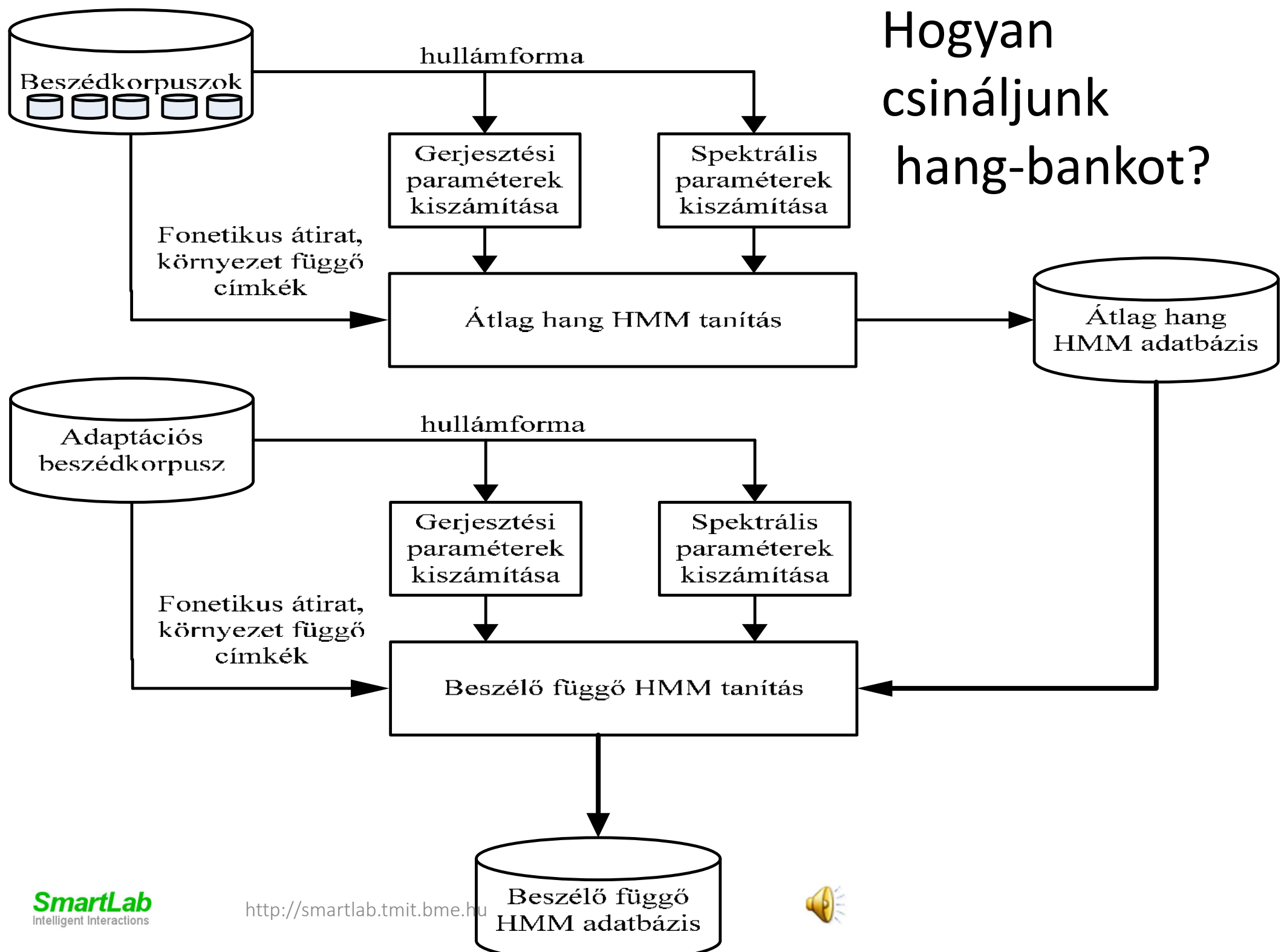
Alap kutatás (HMM) ₂



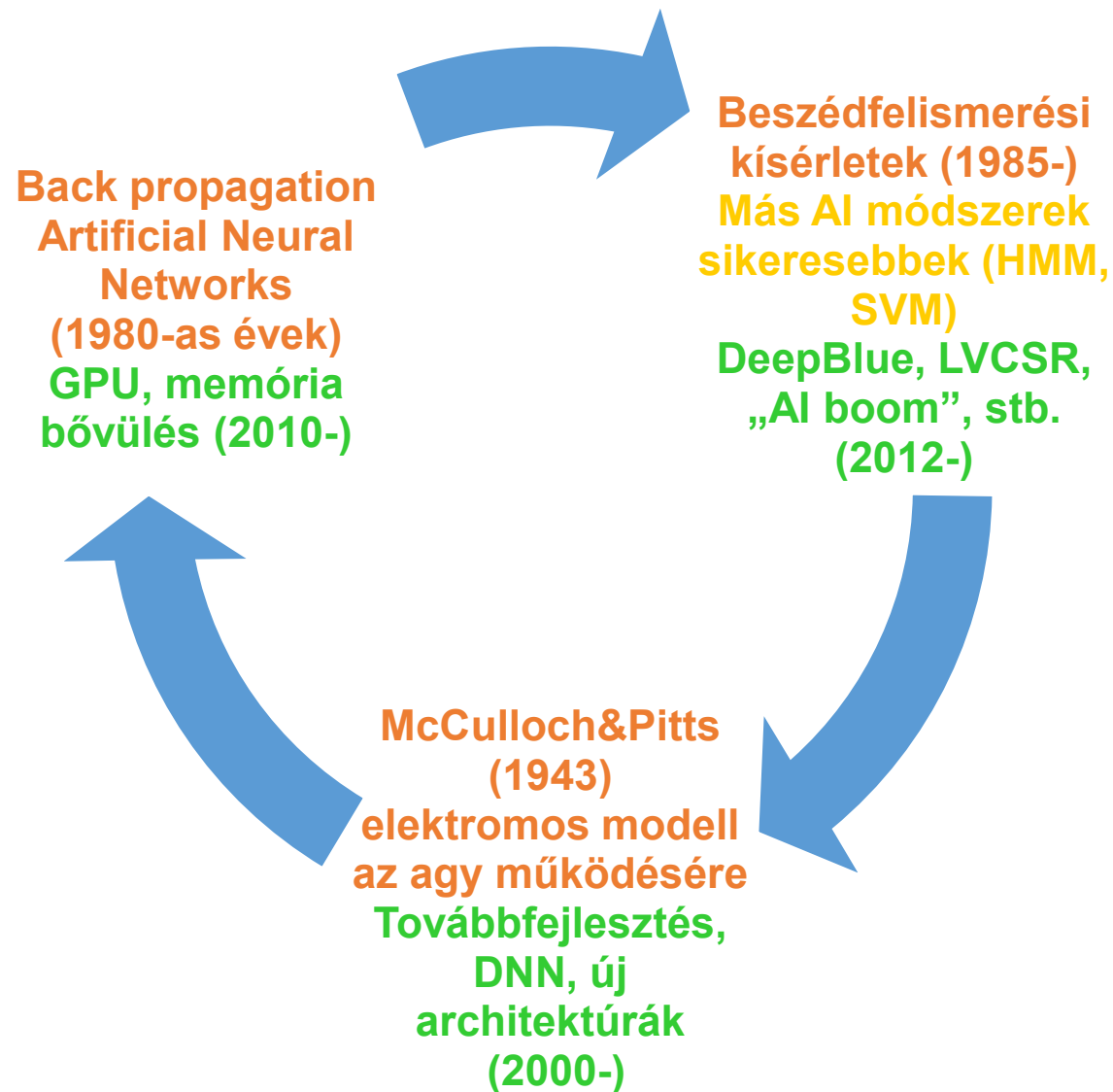
Technológia fejlesztés HMM-alapokon (rugalmasság, 200x-)



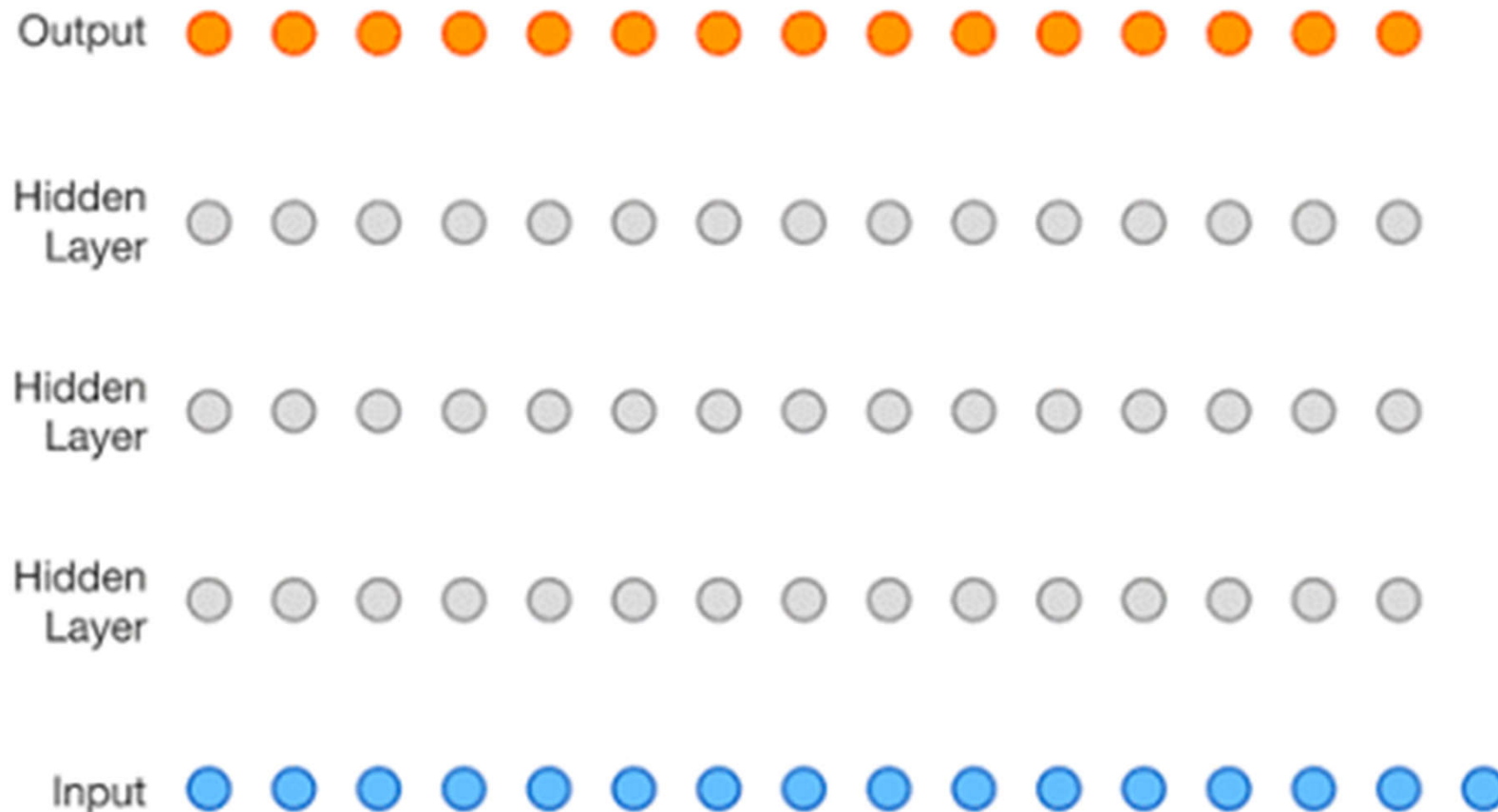
Hogyan csináljunk hang-bankot?



Alap kutatás (Neurális hálózatok)

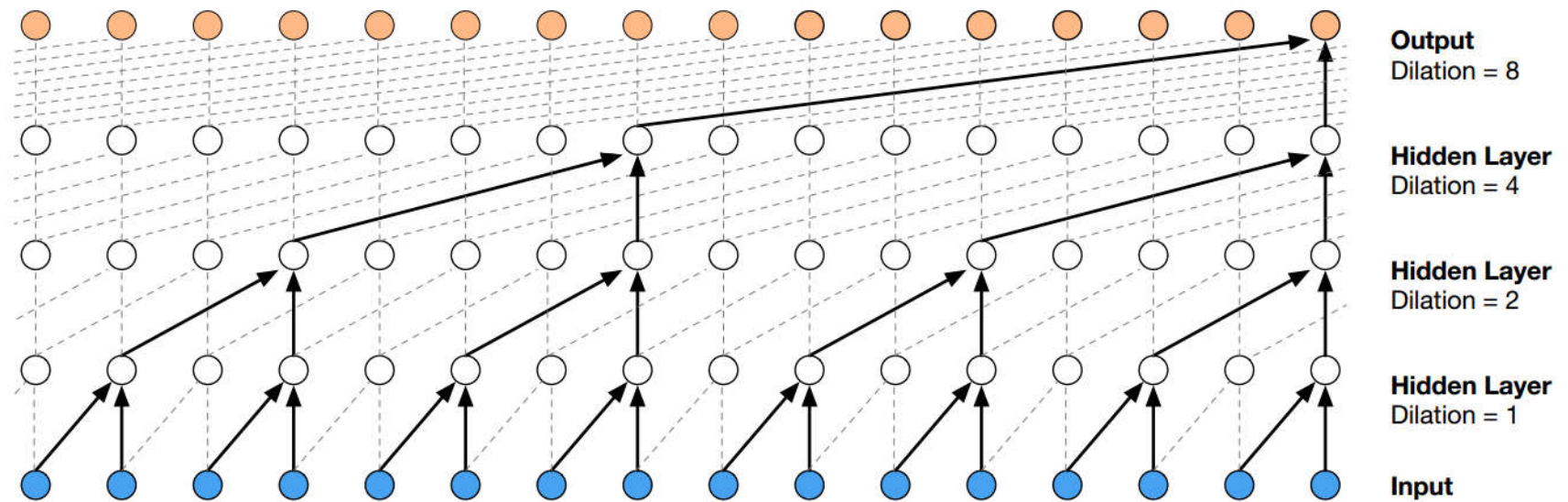


Wavenet (2016. szept.-)



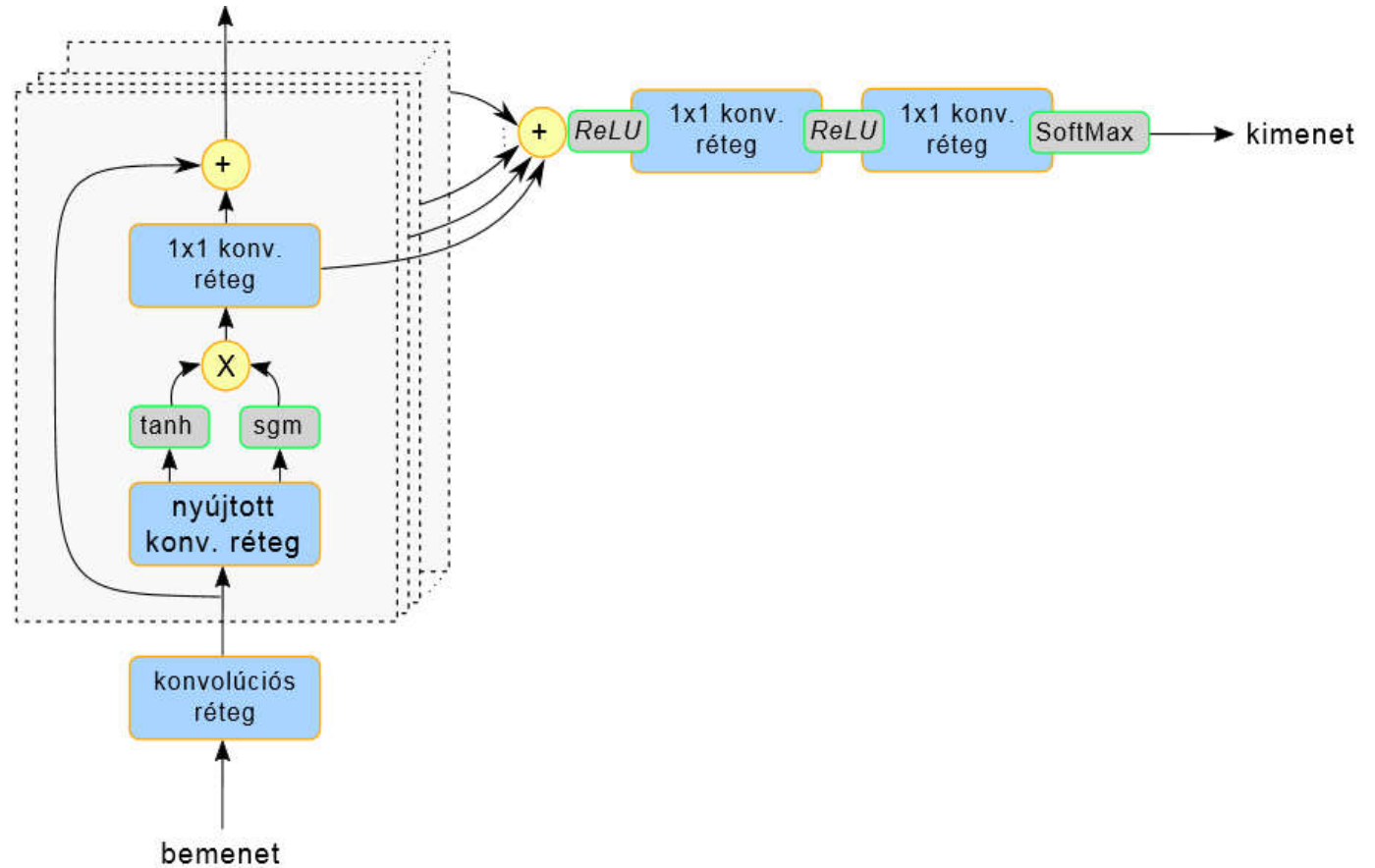
Ábra forrása: <https://deepmind.com/blog/wavenet-generative-model-raw-audio/>

Generálás



Wavenet-alapú 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.

Január 

Szeptember 

Albert Einstein

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

Beszédsszintézis 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 és társai (end-to-end DNN, fejlesztés alatt)

- Bármely személy hangjára elkészíthető
 - Átlag hang képezhető



Miért nem általános „Dobozos” TTS?

- Nem mindig 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

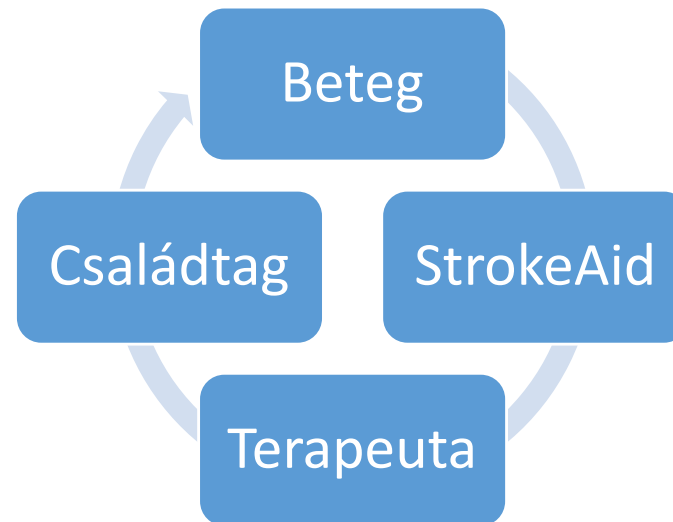
A fejlődés egy mértéke

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)
				+ 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
2017-19	4*	3,3*	0,7*	6,5 órányi angol hangoskönyv (56db) gyermekeknek (változatos stílus)*

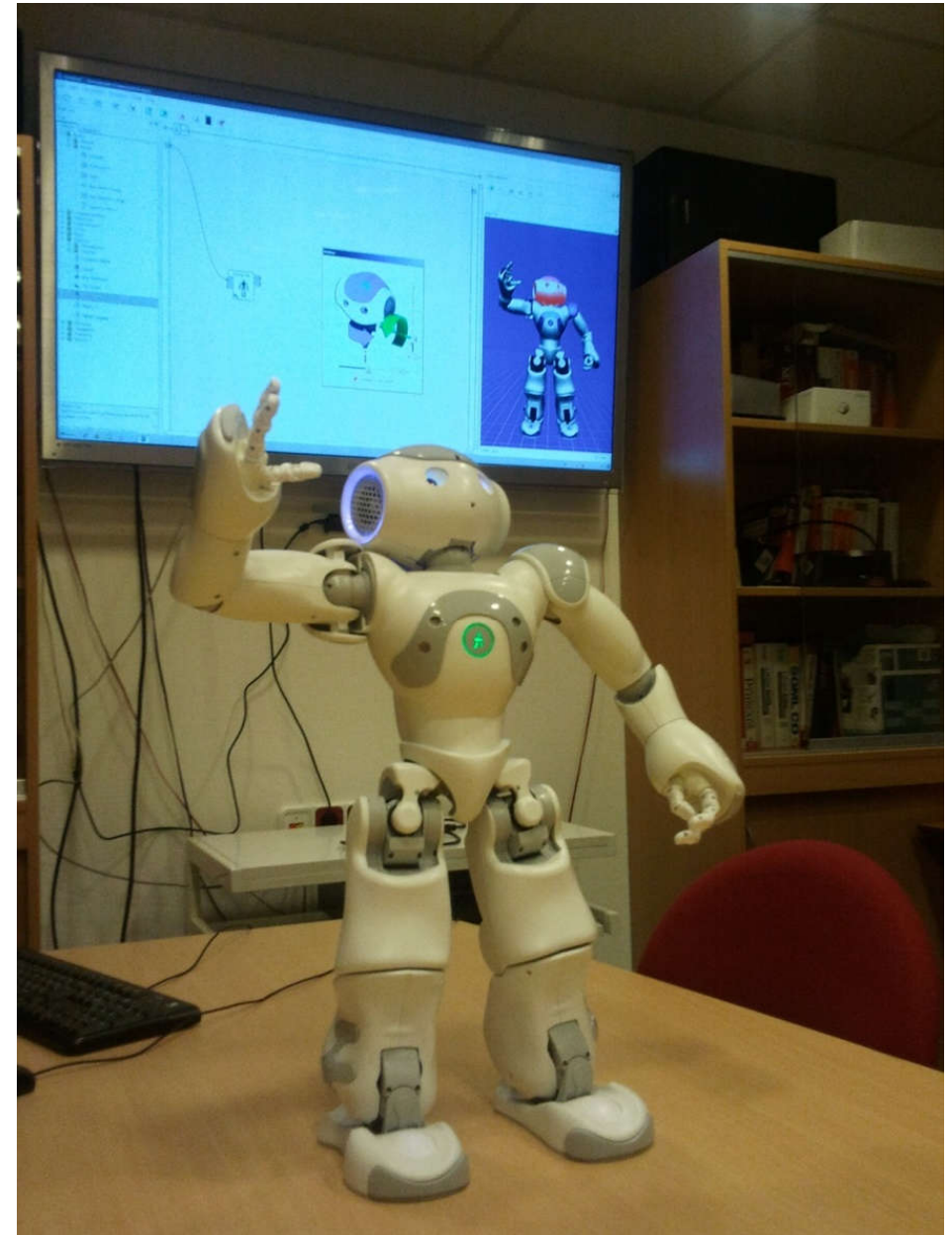
SmartLab VoxAid/StrokeAid/Monddki alkalmazás

- VoxAid/StrokeAid/Monddki – Windows és Android alkalmazás
 - Beszédkommunikációs segédeszköz
 - Kötött, félig kötött és szabad szöveg
- A hagyományos beszédterápiás eszközök és eljárások sok szempontból elavultak és sok időt igényelnek
- Az új (25 éves) megoldással a terapeuta idejét felszabadíthatjuk -> javulhat a beteg/terapeuta arány
- Integrált visszacsatolás növeli a hatékonyságot és csökkenti a költségeket

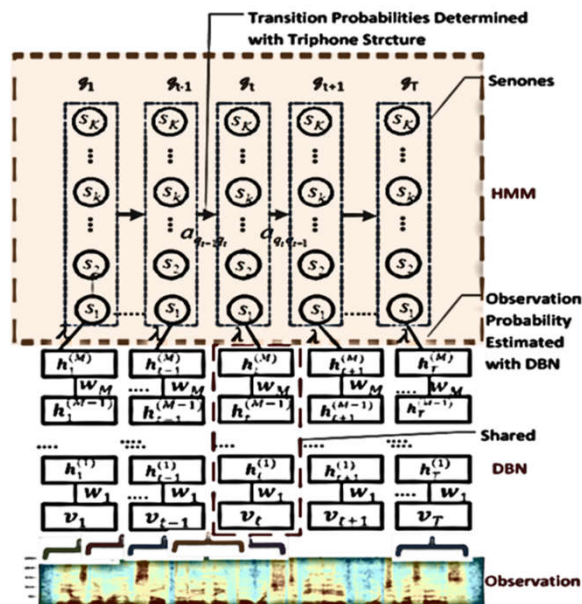


NAO humanoid robot

- Magyar TTS a NAO-ra
- Játékok gyerekeknek
- Kórházi alkalmazás
- Depresszió ellen
- Ember-robot interfész választható tárgy



Beszédfelismerés és deep learning (2011-)

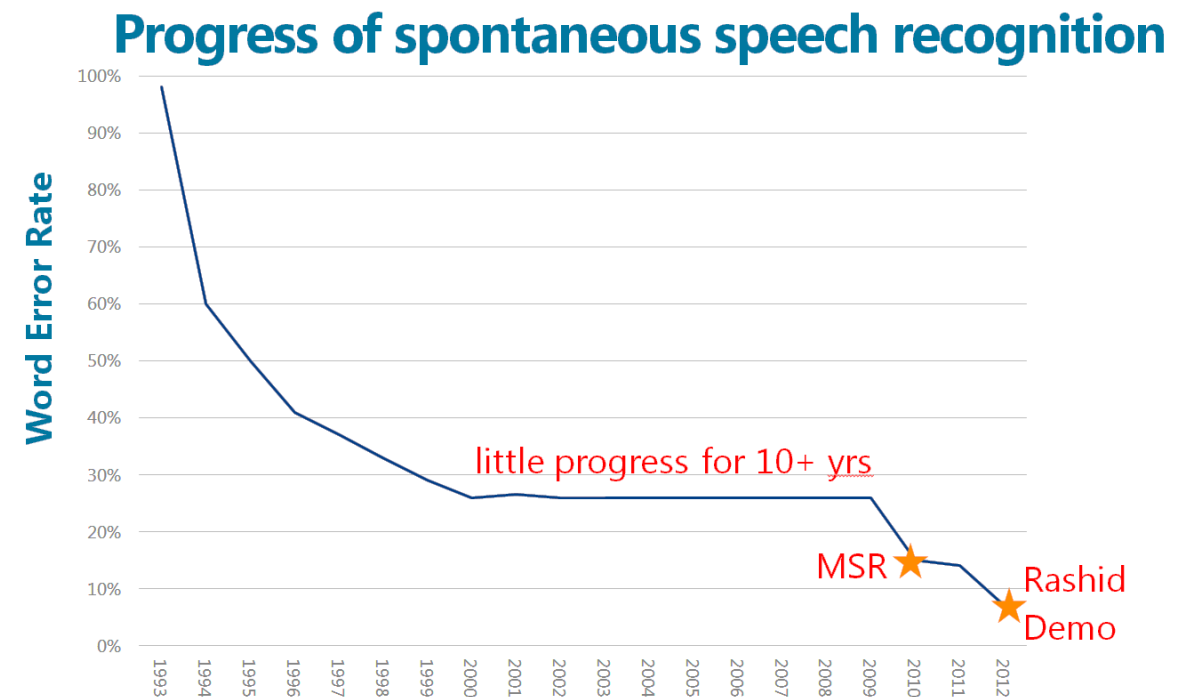


After no improvement for 10+ years by the research community...

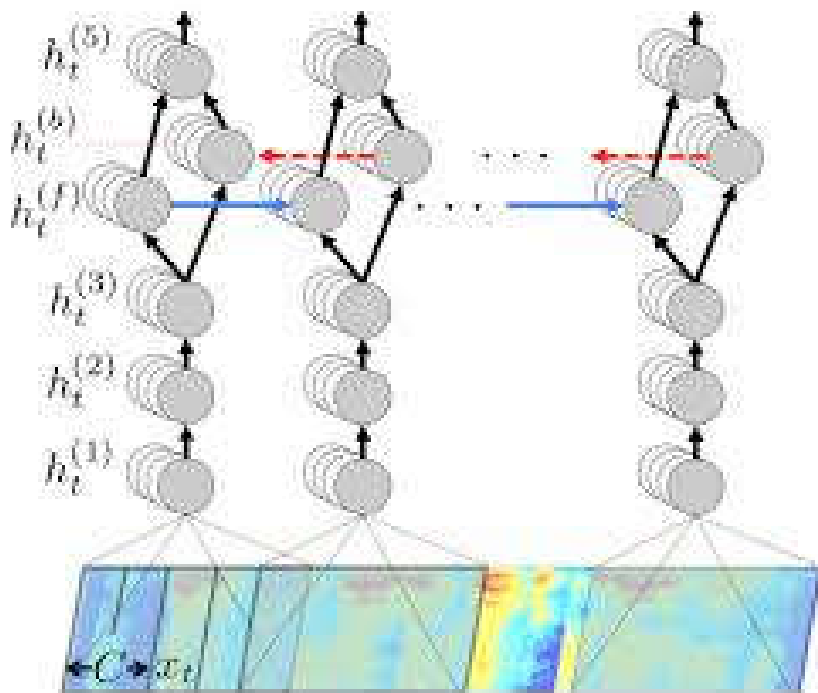
...MSR reduced error from ~23% to <13% (and under 7% for Rick Rashid's S2S demo in 2012)!

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.



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!**

Referenciák / főbb felhasználási területek



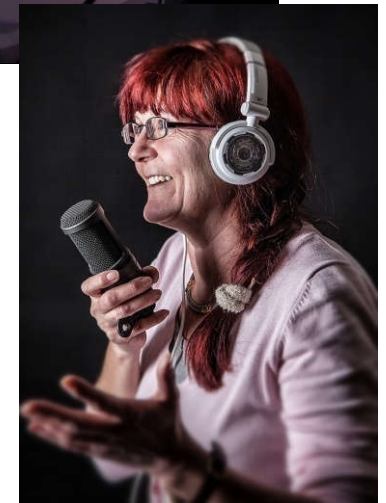
Telefonos ügyfélszolgálatok

Médiafeliratozás



Diktálás

Jegyzőkönyvkészítés



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

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



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:

Gender:

LOAD RECORDING **START RECORDING**

Microphone volume

ASSESSMENT



COALA-Phonetics

Psychological State 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.

Laboratory of Speech Acoustics (LSA) - BME TMIT

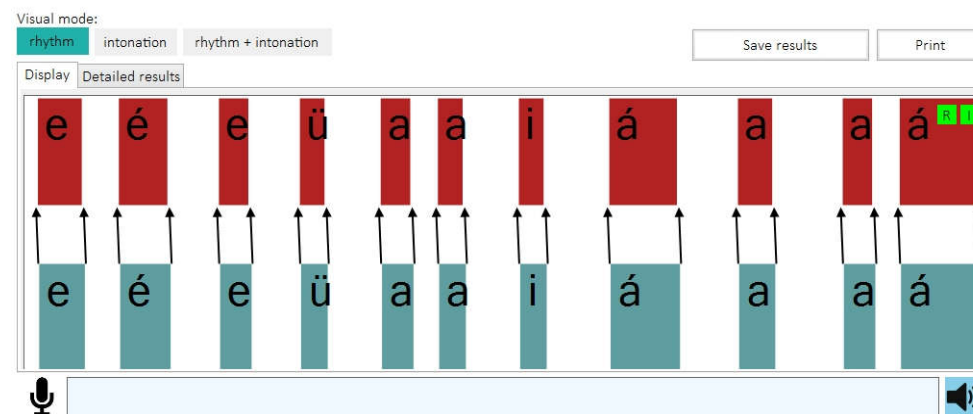
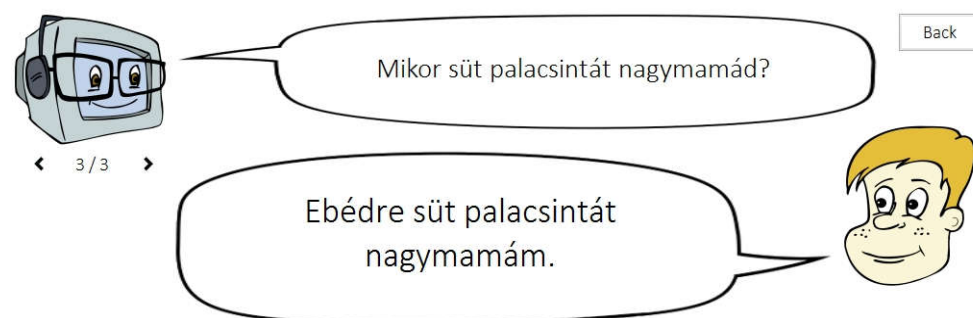
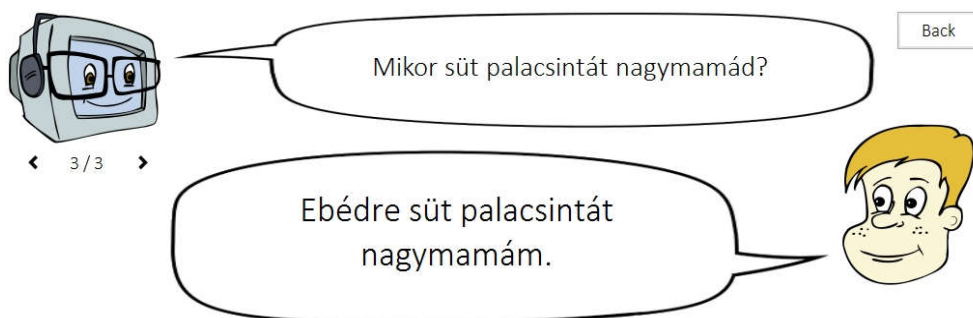
Beszéd kiejtés oktatósi rendszer siket és nagyothalló, valamint kochleáris implantátummal rendelkező gyerekek számára

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/>

Hozzászólások

Mélyebb érdeklődőknek:

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

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

<https://www.enfield-project.eu/>



Köszöniük



EUROPEAN
COMMISSION

támogatását.



(Teleauto, BelAmi, EtoCom -TÁMOP-4.2.2-08/1/KMR-2008-0001/- , DIVIDE KULCSKÖZVEZŐK -TÁMOP-4.2.1/B-09/1/KMR-2010-0002/- , CIP CESAR, AAL PAELIFE, VUK, Eureka DANSPLAT projektek)



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

