Kempelentől a Takotronig és azon túl: gépi beszédelőállítás

Németh Géza és Olaszy Gábor

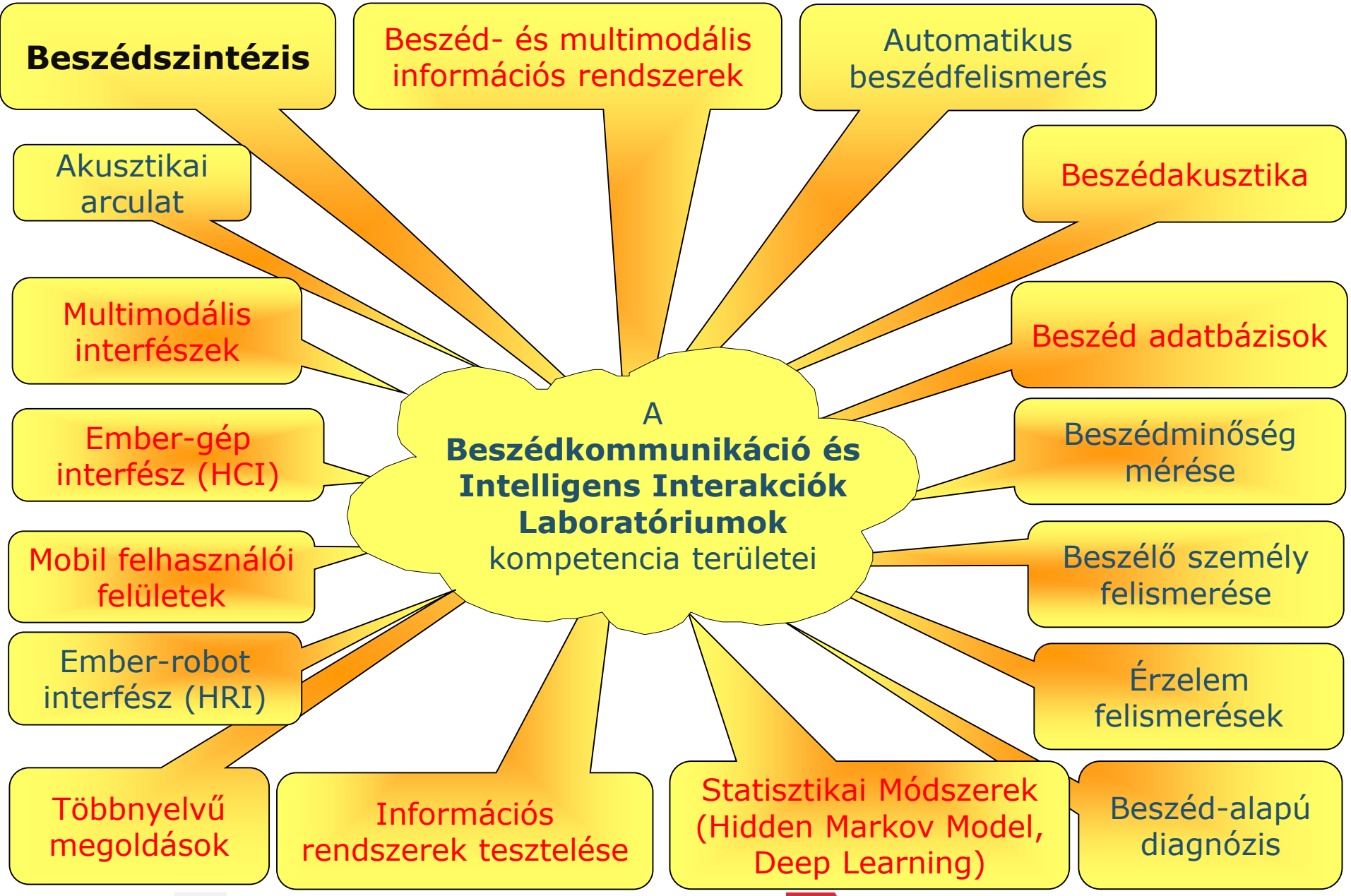
SmartLabs

Beszédkommunikáció és Intelligens Interakciók Laboratóriumok

2019. 12.12.

SmartLab
Intelligent Interactions
http://smartlab.tmit.bme.hu

NVIDIA GPU EDUCATION CENTER


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

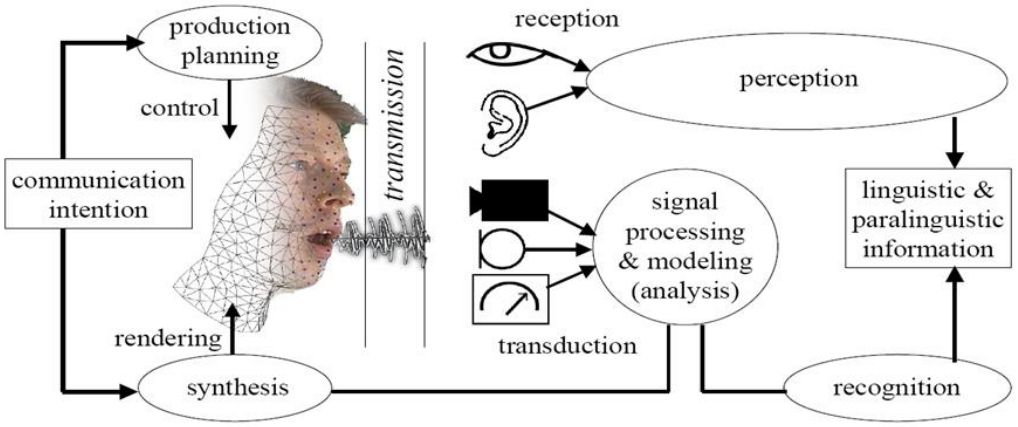
Laczkó Klára
asszisztens

Sevinj Yolchuyeva
PhD hallgató

SmartLab kutatási területek

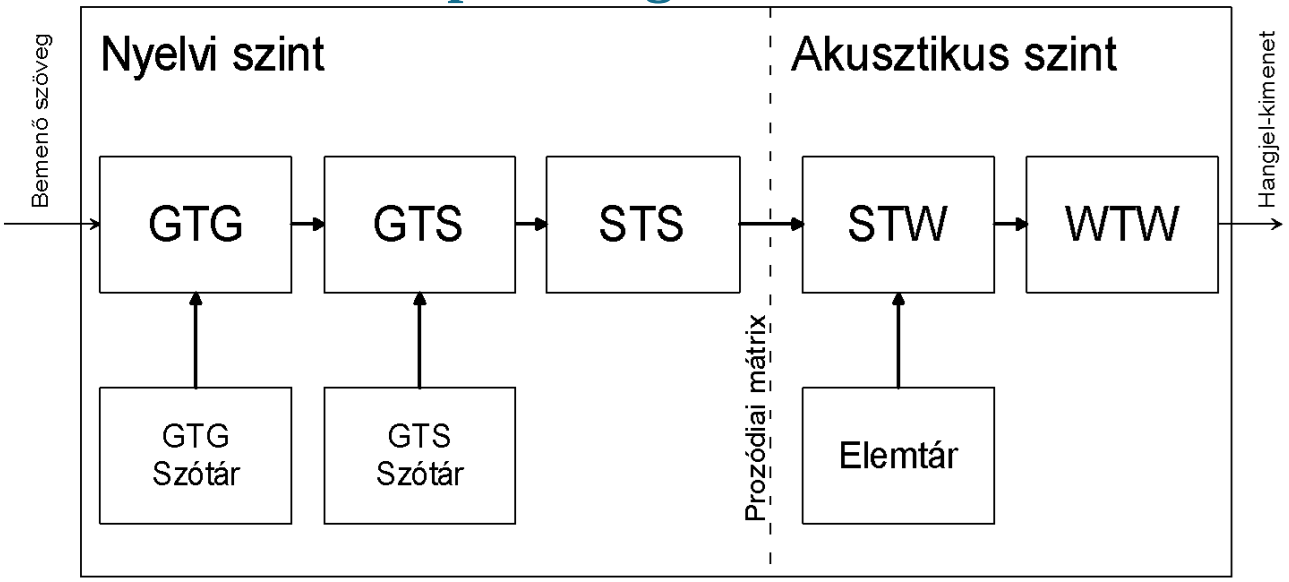
- Gépi szövegfelolvasás (text-to-speech, TTS)
  - Elemösszefűzéses és korpusz-alapú
  - Gépi tanulás alapú (Deep Learning, Hidden Markov-model)
- Beszédszintézis részproblémái
  - Parametrikus kódolás, gerjesztési modellek
  - Intonációs modellek
  - Rövid- és kérdő mondatok prozódiája
  - Kommunikációs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

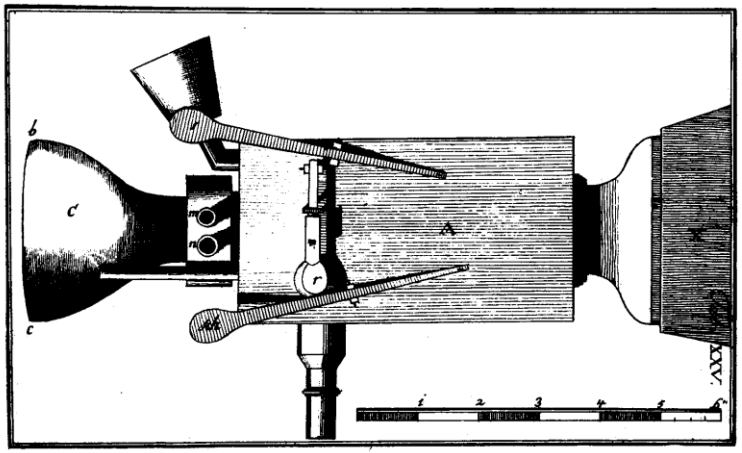


Történelem
közlekedés és beszédtechnológia

1791

2019

Lépjünk vissza az időben
210 + 18 évet



Kempelen eredeti gépe

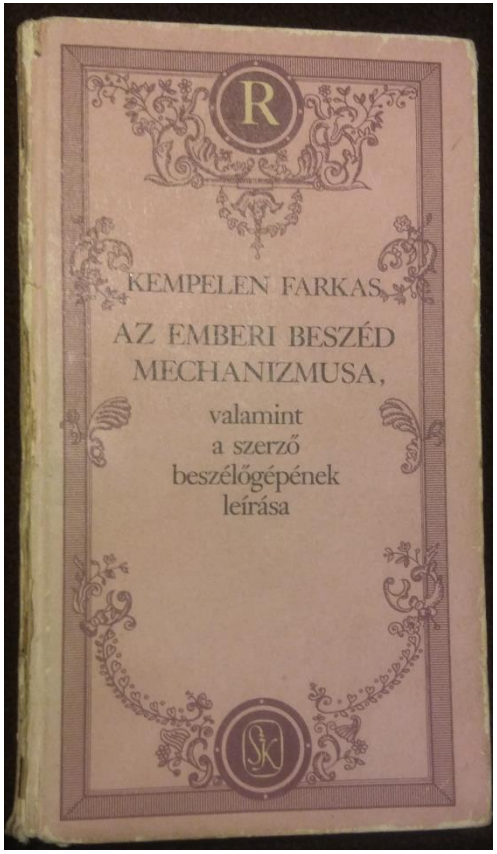
1791

2001

Pontosan elkészített, működő másolat

Kempelen könyve és a magyar fordítása

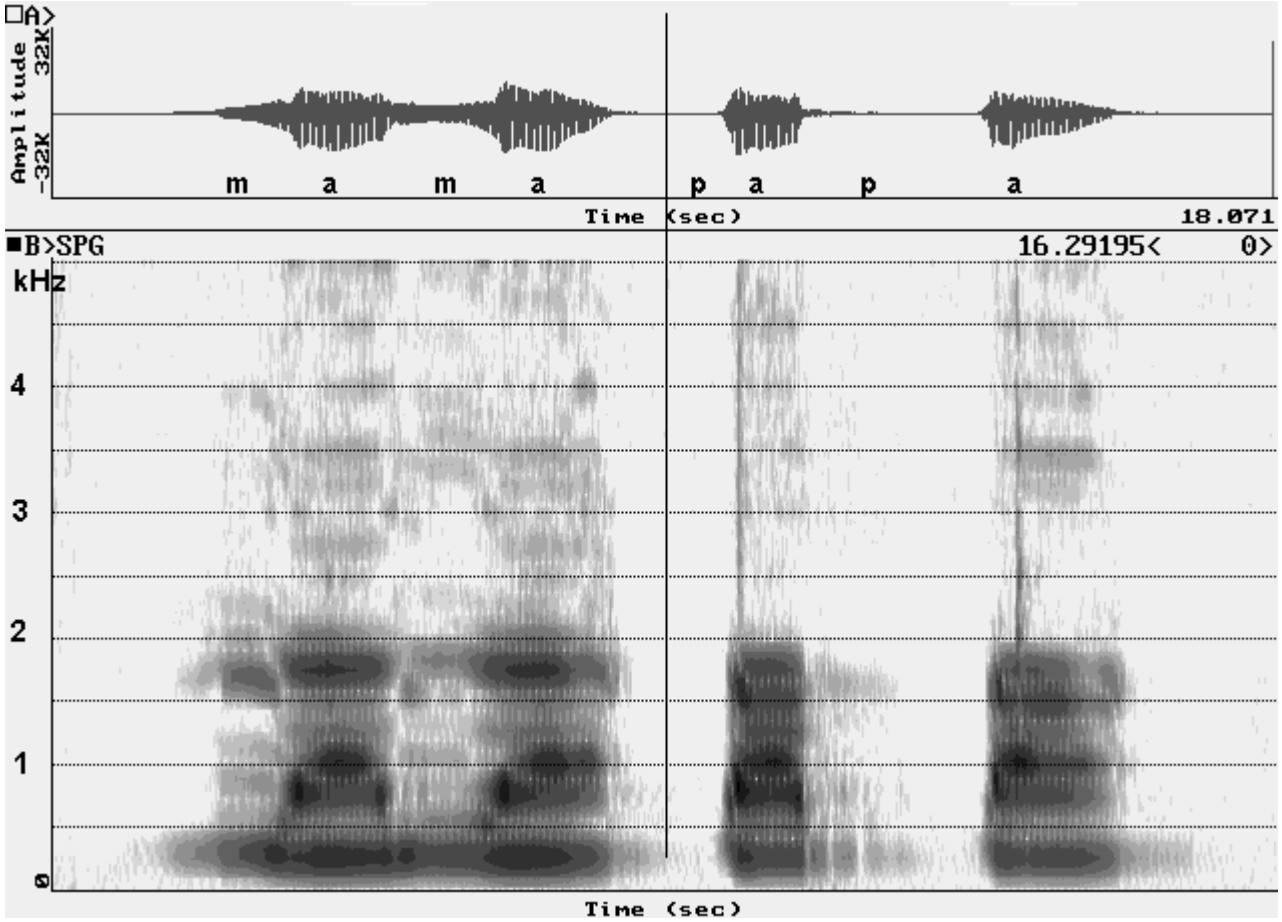
Mechanismus der menschlichen Sprache
1791



Mollay Károly fordítása
1989

Tarnóczy Tamás könyvtárából

A rekonstruált gép hangja 2001



Kempelen gépe volt az első nyelvfüggetlen artikulációs beszédkeltő szerkezet

Es war.

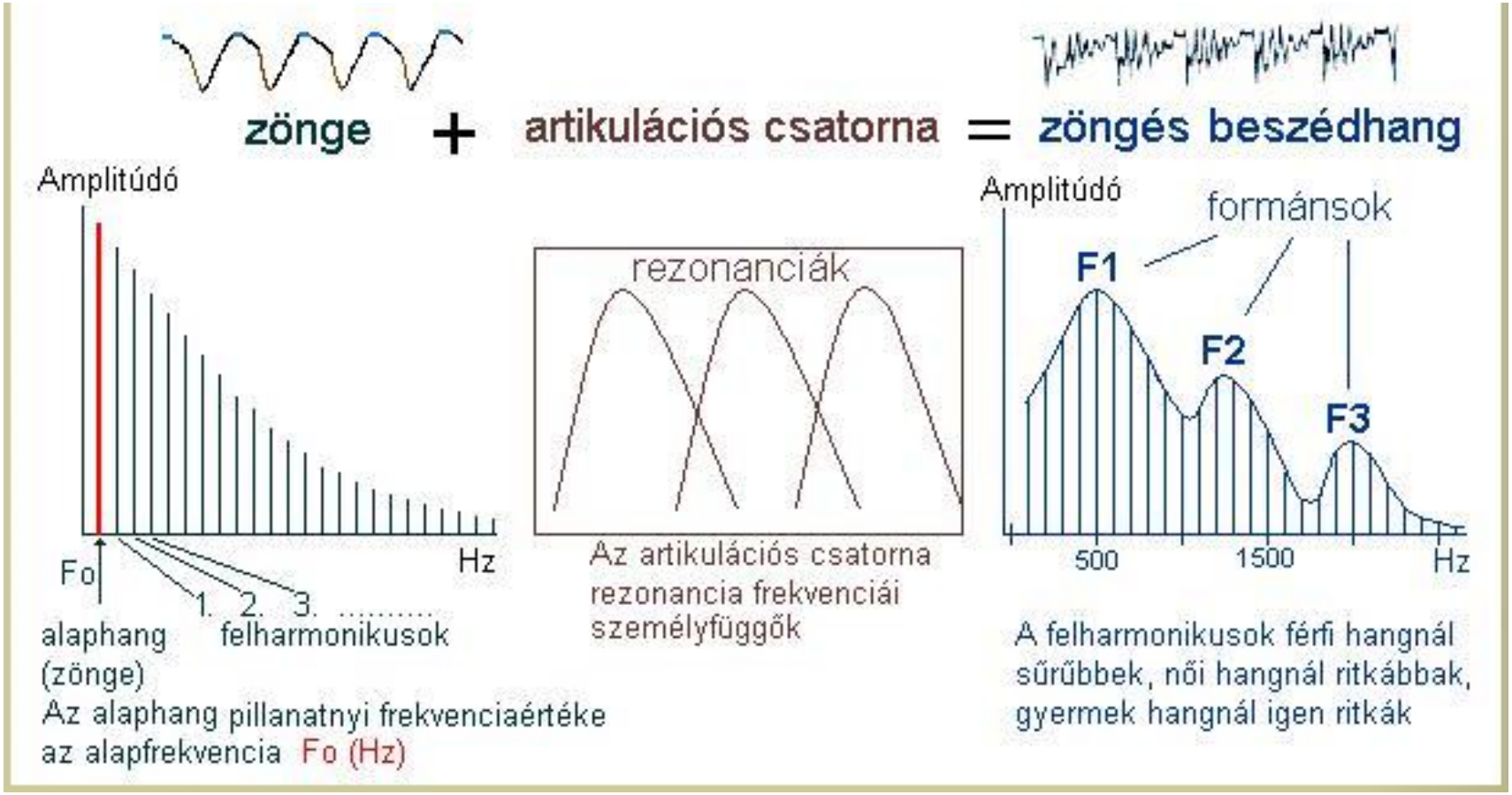
I go.

Je t’aime.

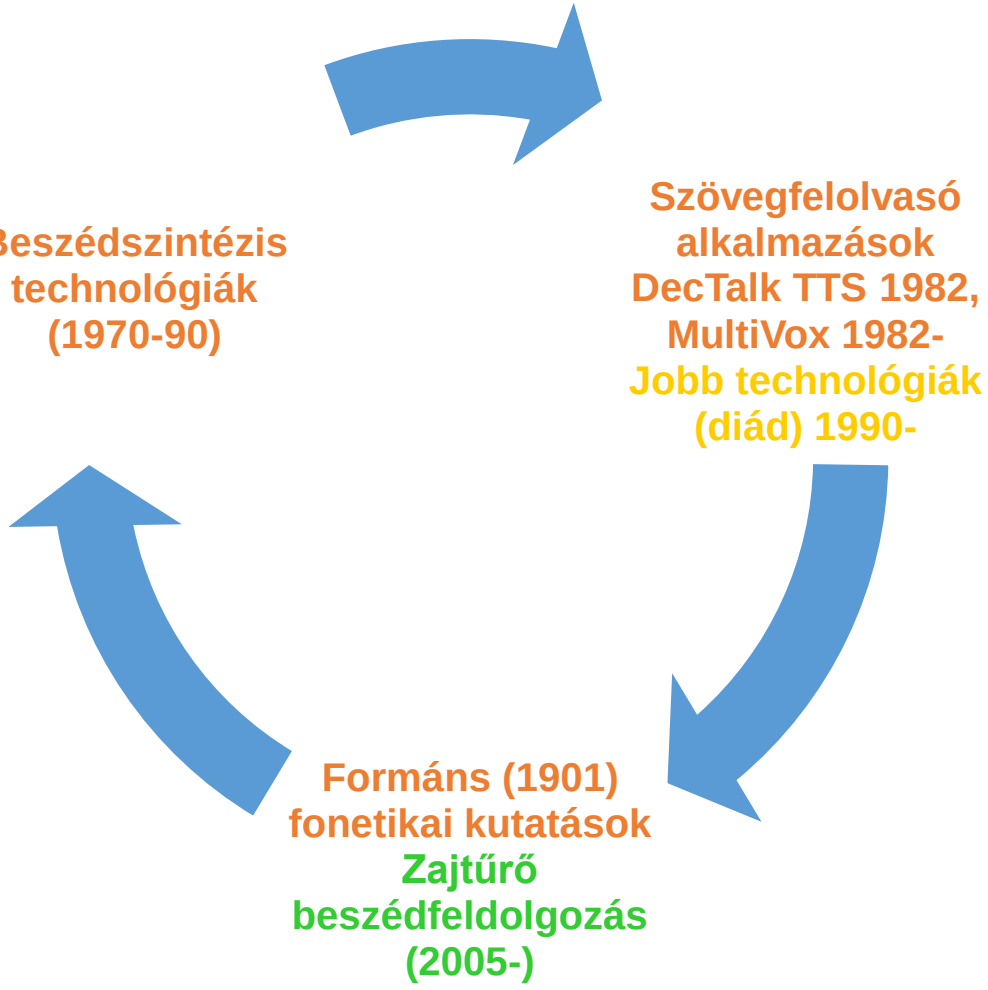
Kempelen replikák találkozója 2019. Bécs

A Kempelen gépek „kórusa” 2019. Bécs

Alapkutatás (formáns 1791-) [1]

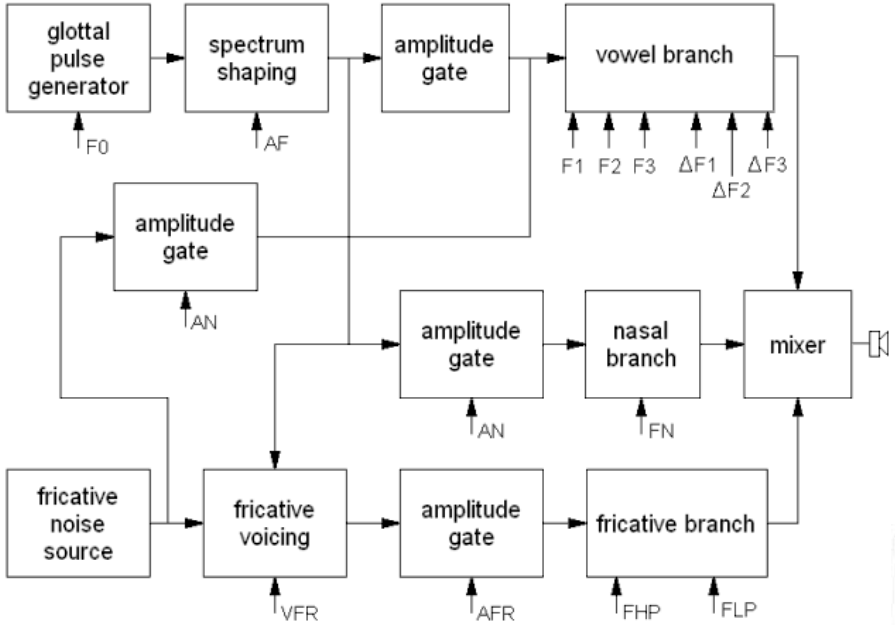


Alapkutatás (formáns) 2



Forrás- szűrő modell (formánsok, érthetőség)

Kempelen Farkas 1791



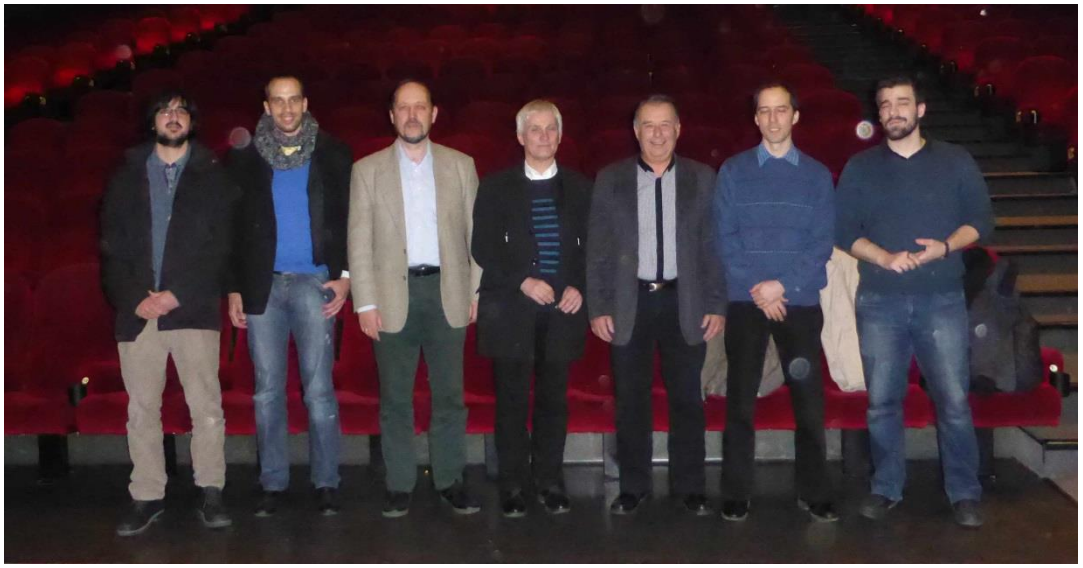
HungaroVox 1982

MultiVox 1986-2002

Olaszi P. – Olaszy 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 film magyar szinkronhangja

A fejlődés útja

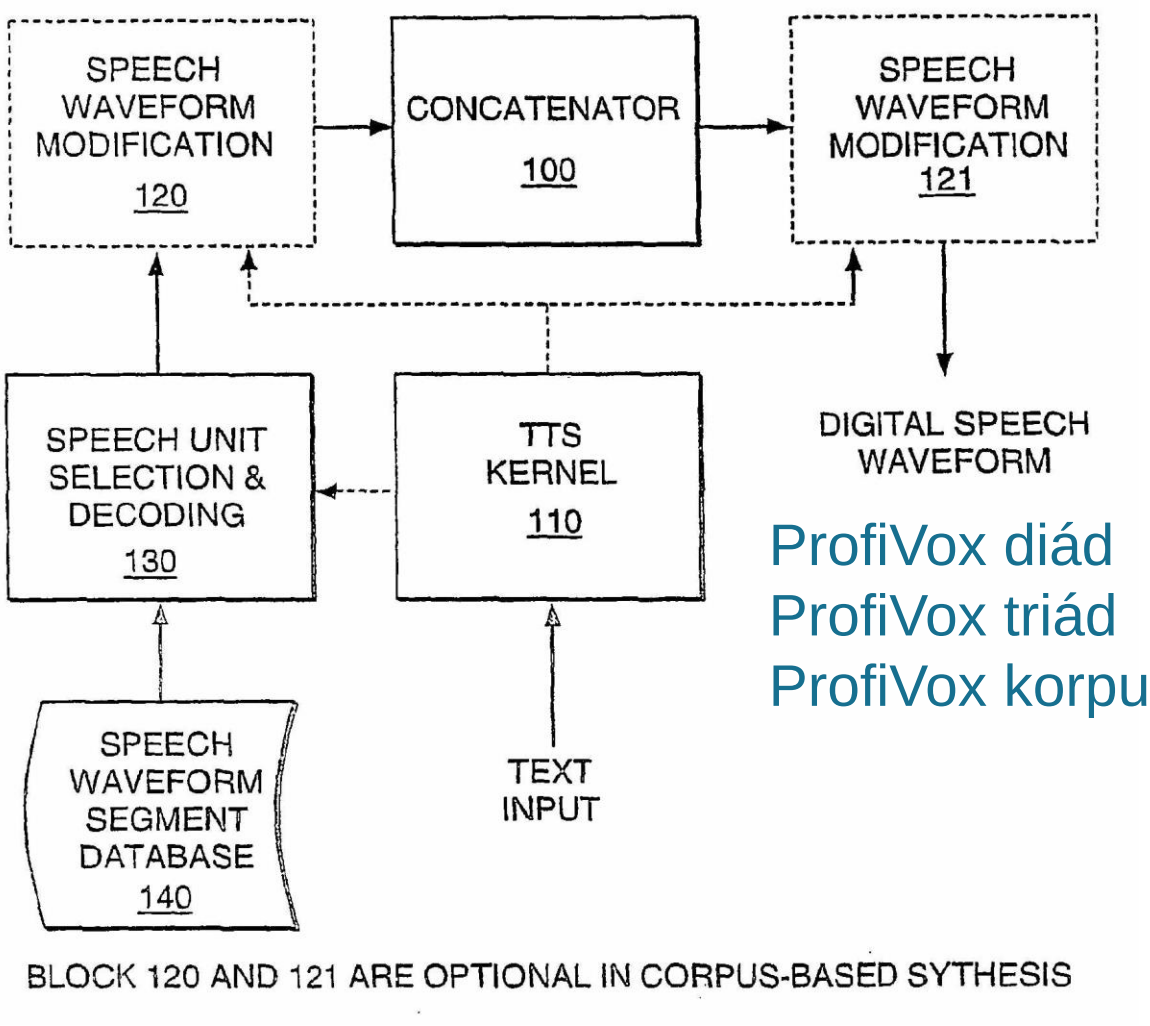
A szabály-alapú modellek
(artikulációs csatorna, prozódia)

mellett és helyett

Természetes elemek
- egyre nagyobb halmaza
- statisztikai modellépítés
- minimális jelfeldolgozás

Egységes(re törekvő) kiértékelés

Hullámforma összefűzés (természetesség, 1916-)

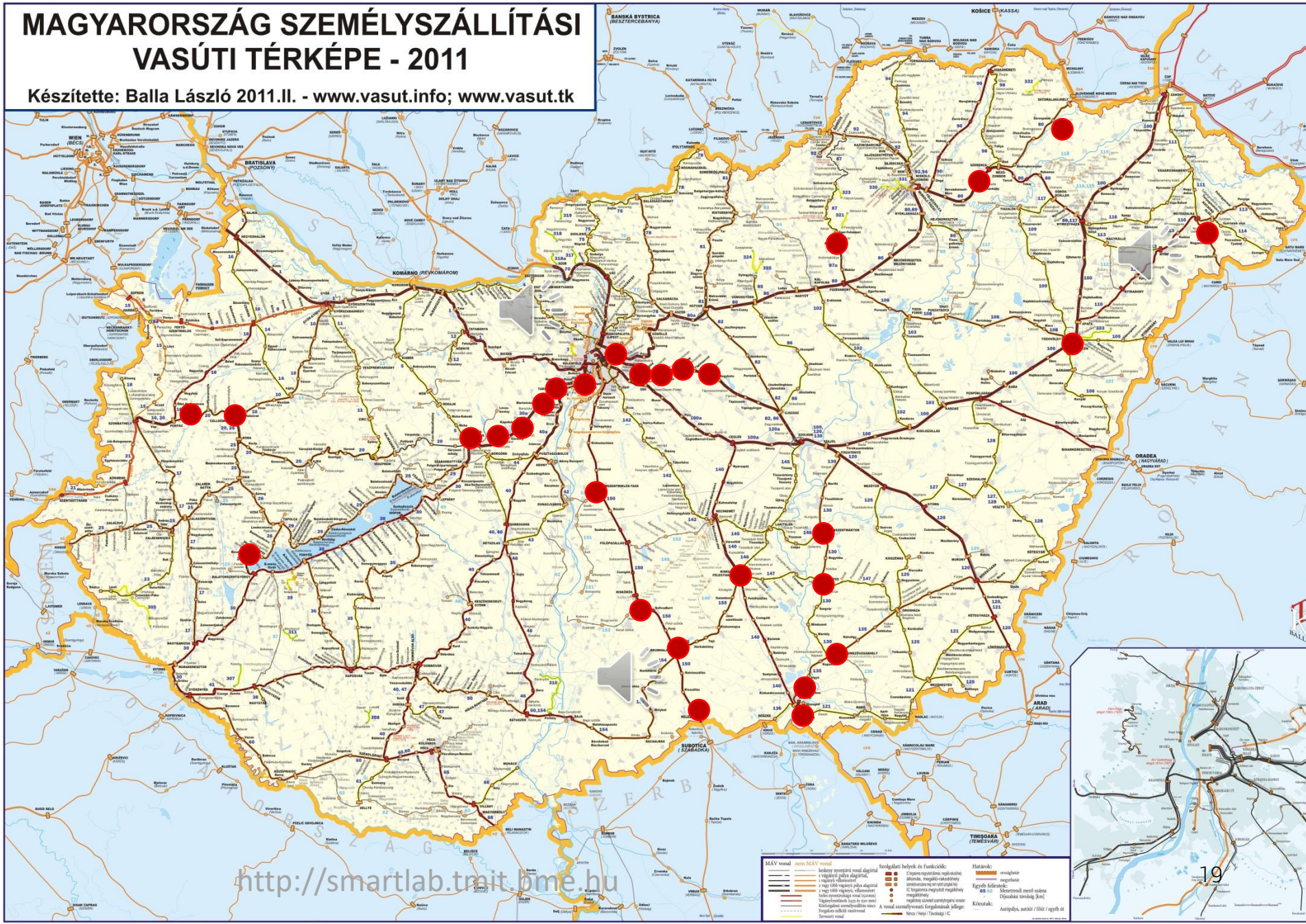


ProfiVox diád 1995-

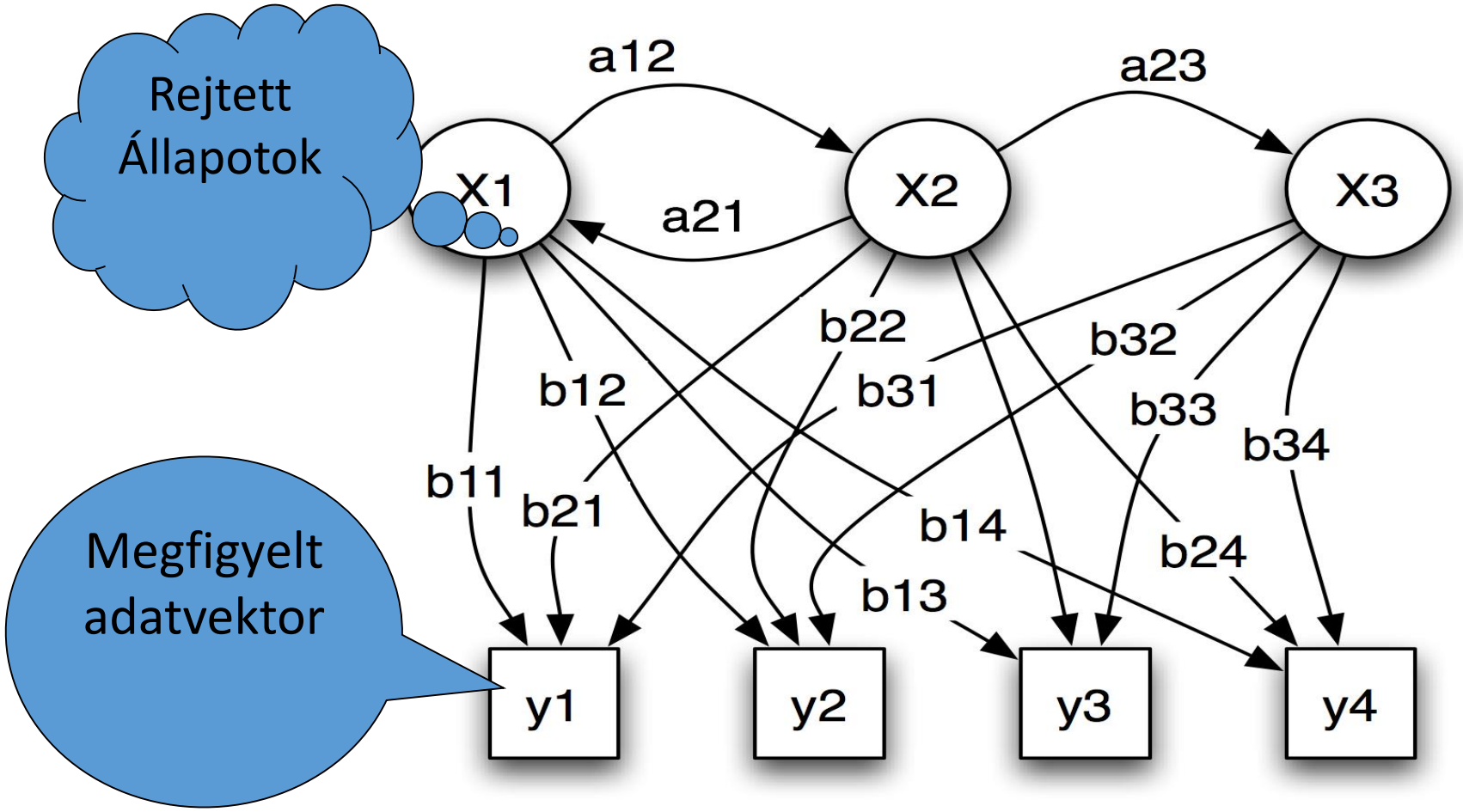
ProfiVox triád 2000-

ProfiVox korpusz 2002-

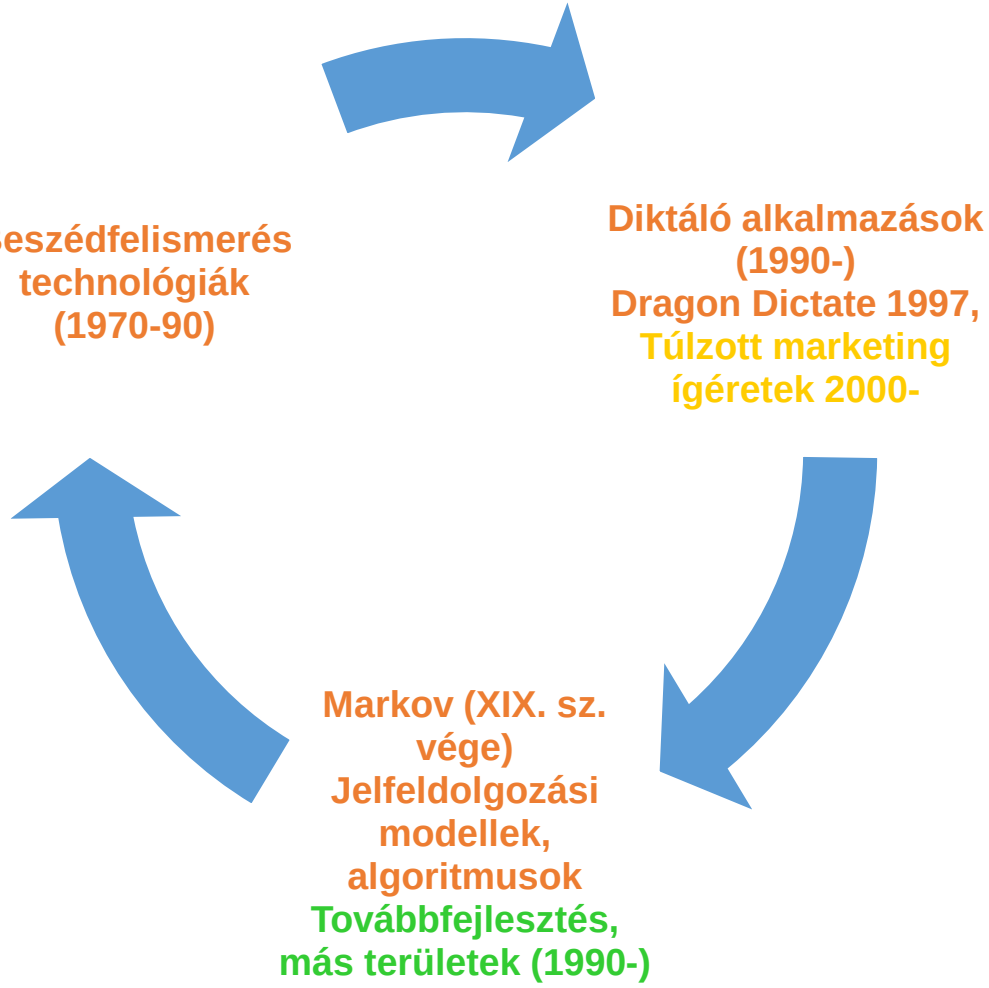
Új hang a magyar vasútállomásokon



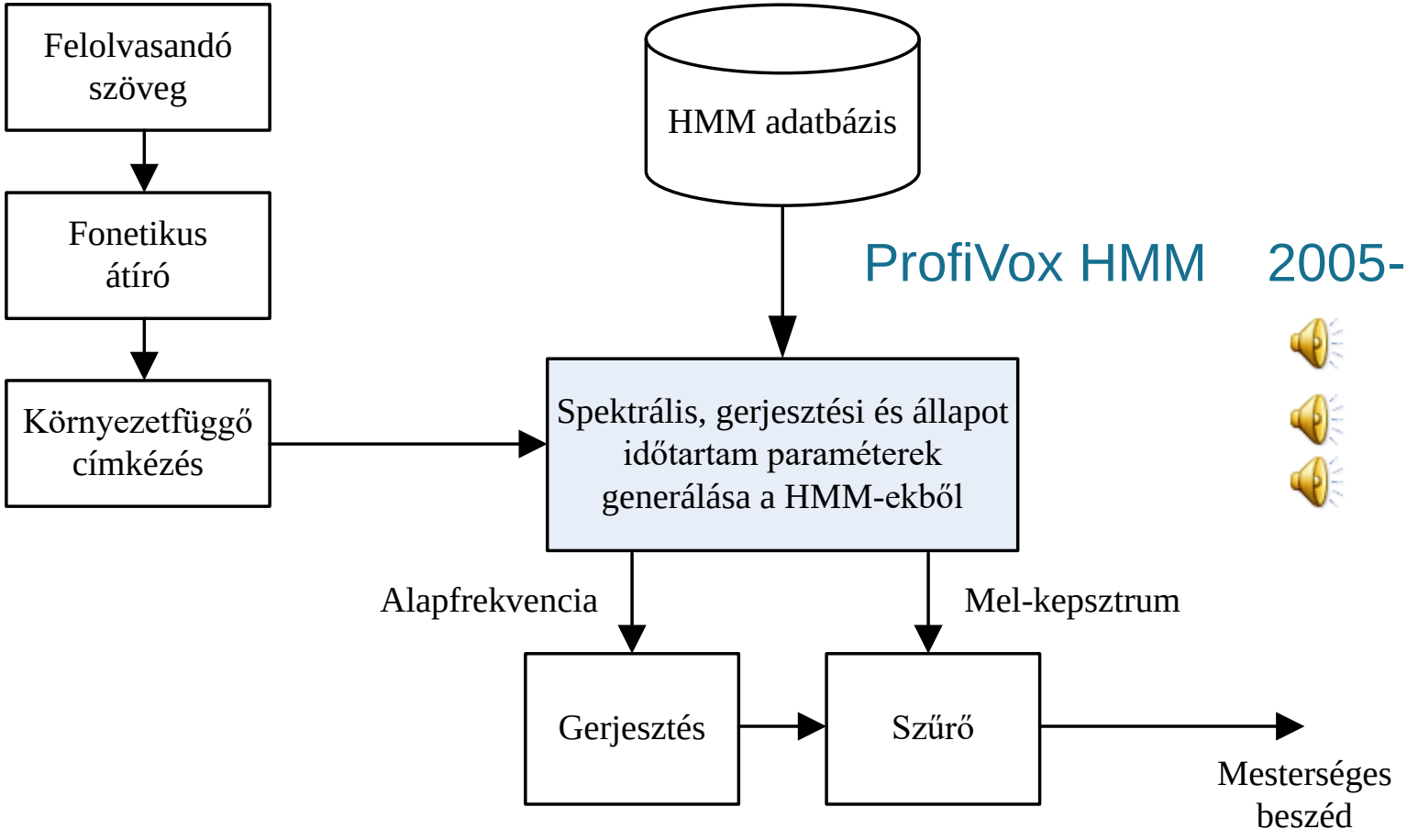
Alapkutatás (Hidden Markov Model, HMM 1970-) 1



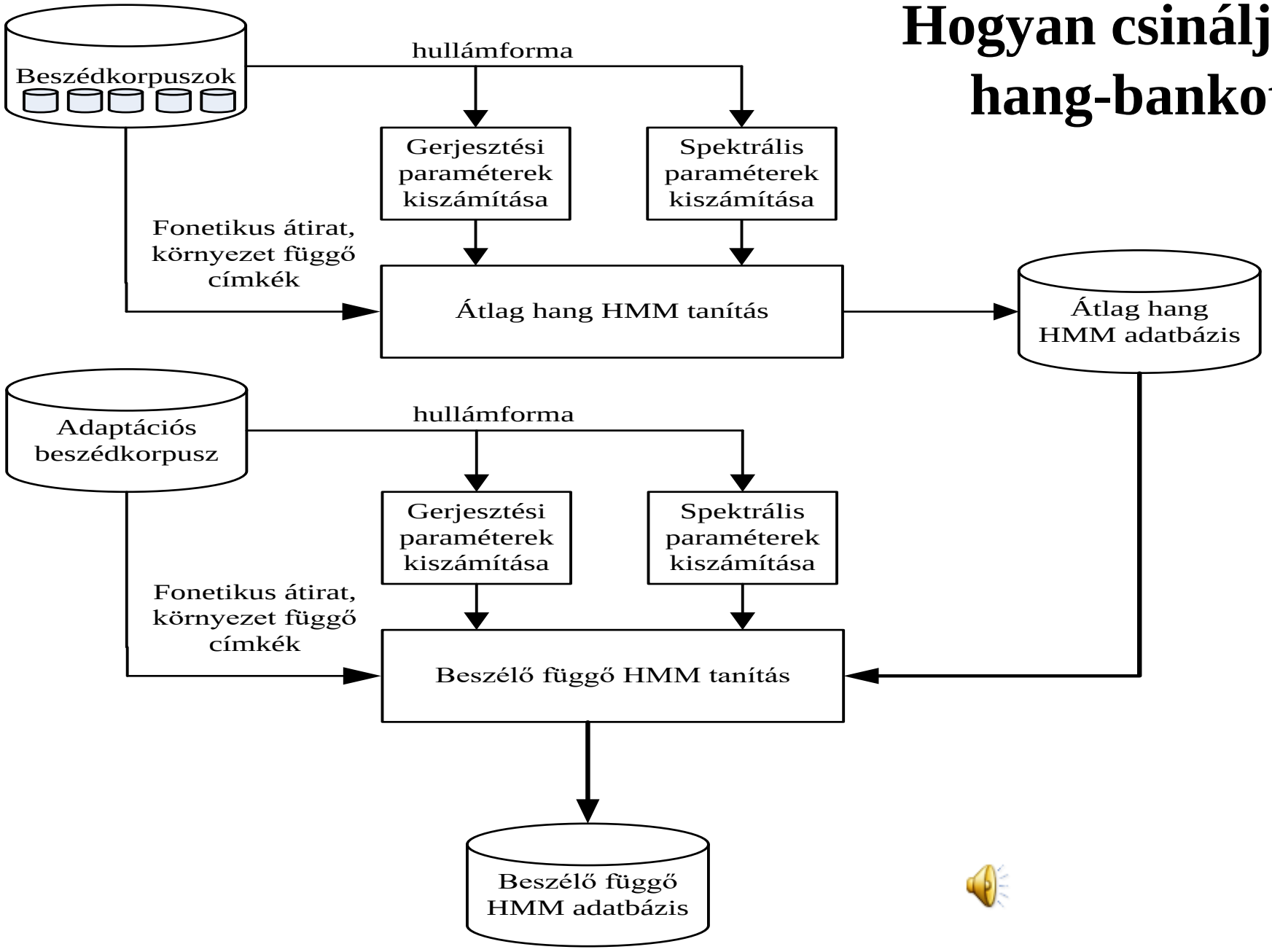
Alapkutatás (Rejtett markov modell, HMM) 2



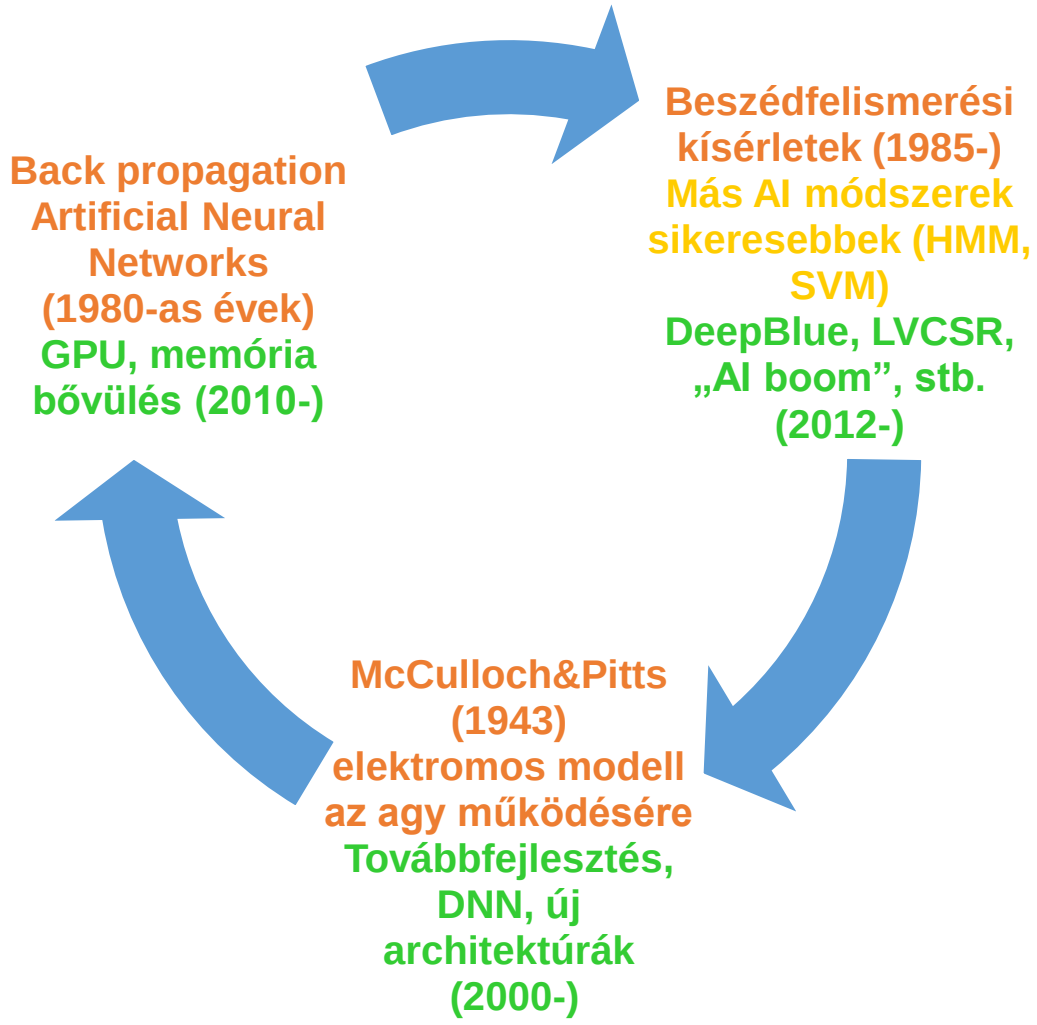
Technológia fejlesztés HMM-alapokon (rugalmasság, 200x-)



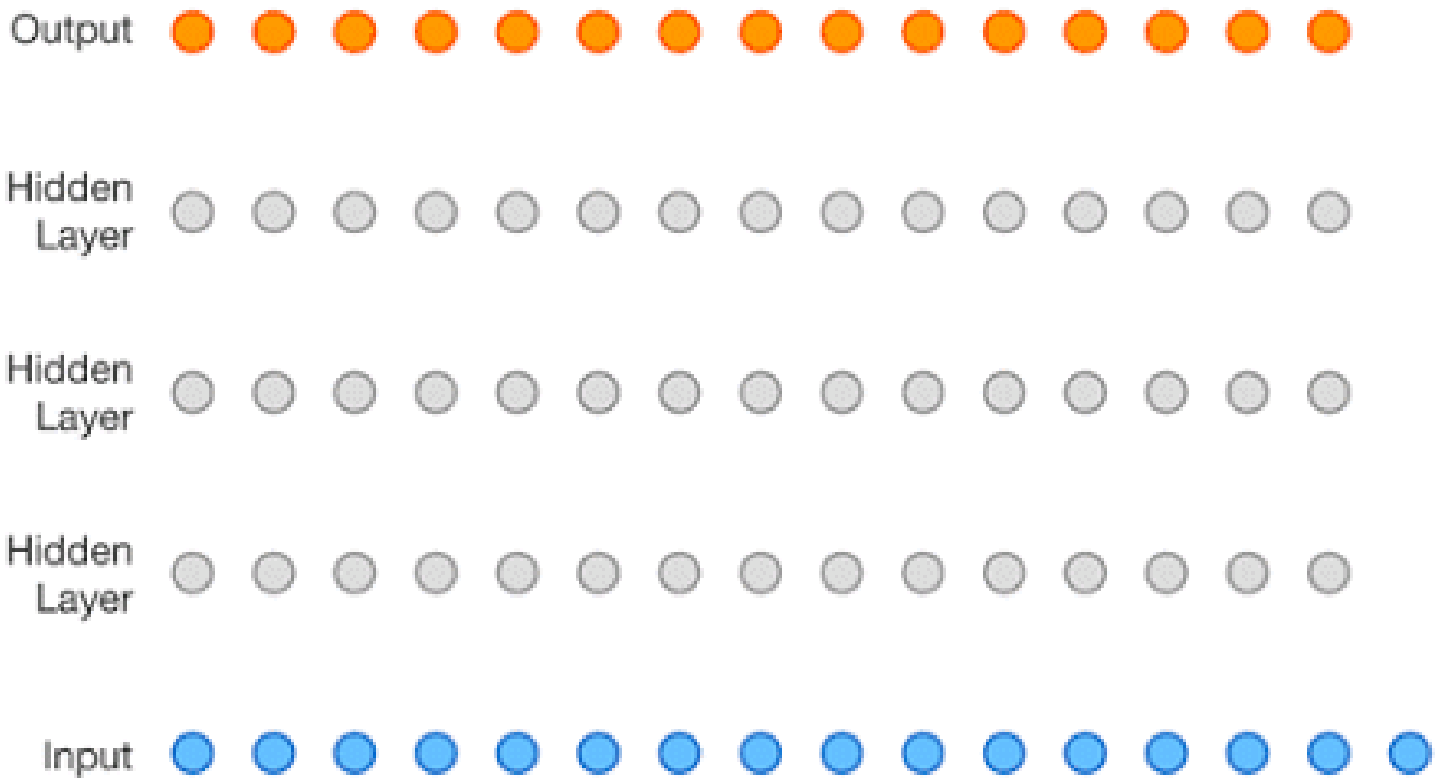
Hogyan csináljunk hang-bankot?



Alapkutatás (Neurális hálózatok)

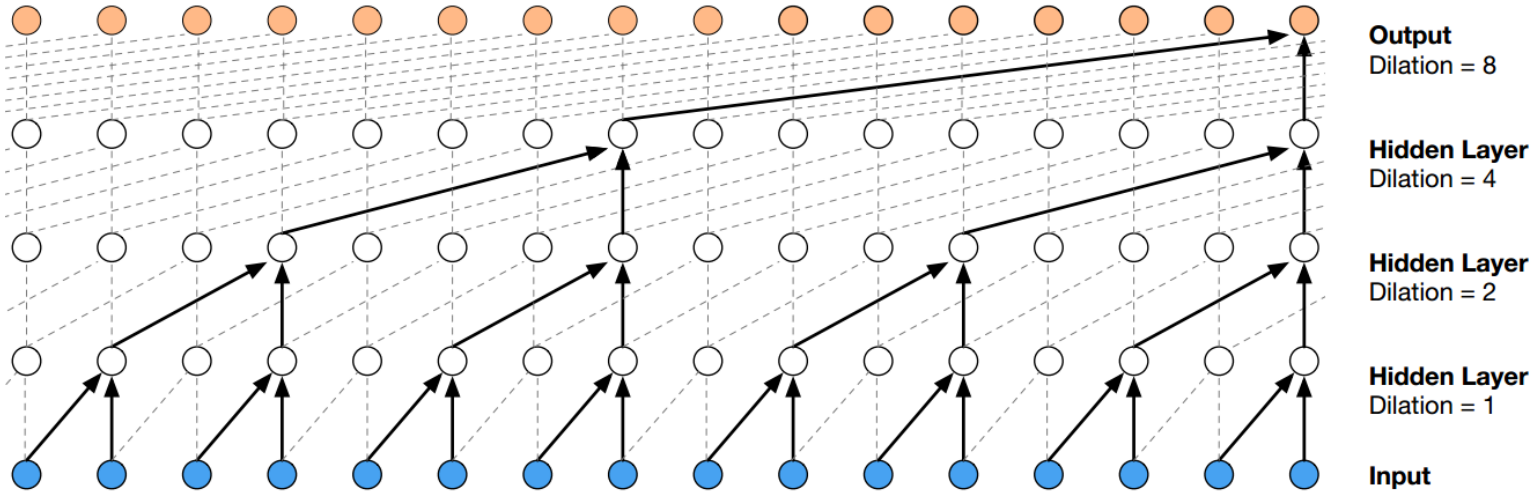


Wavenet (2016. szept.-)



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

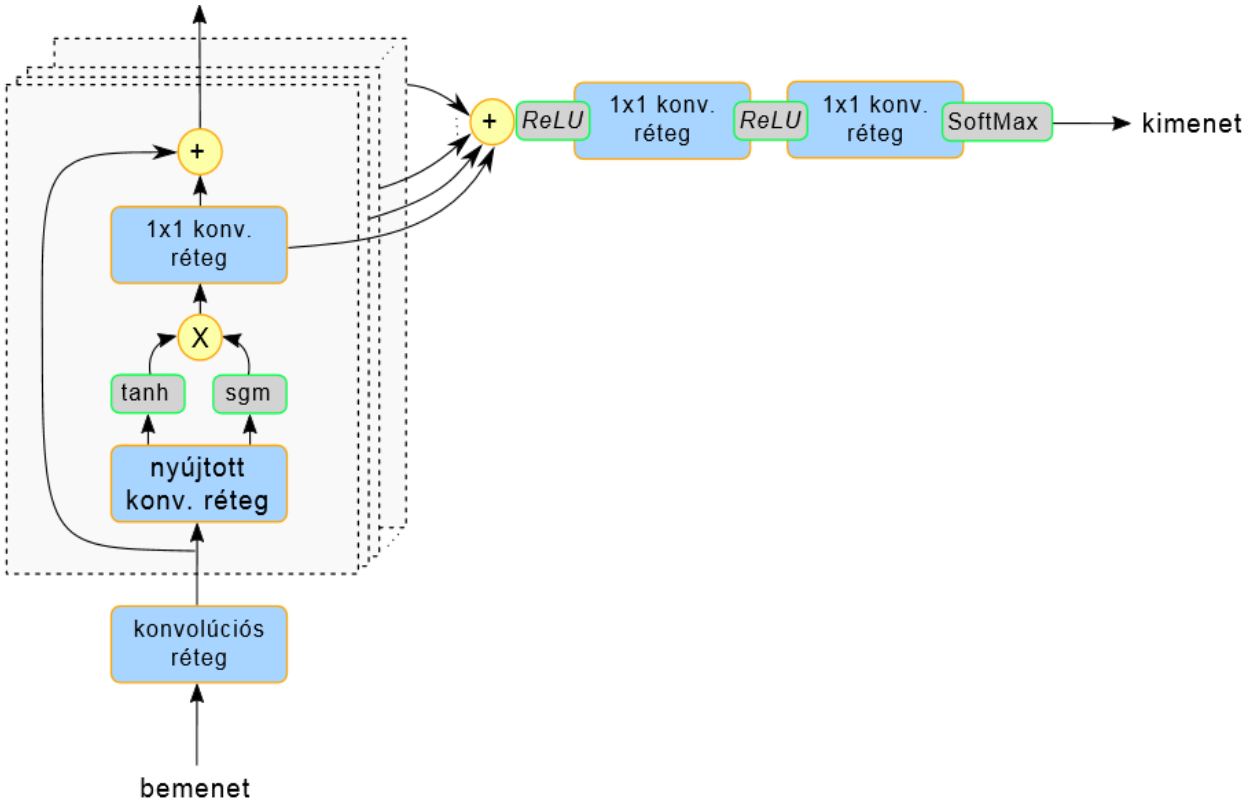
Generálás



Google 2017. okt
(US angol és kínai 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:

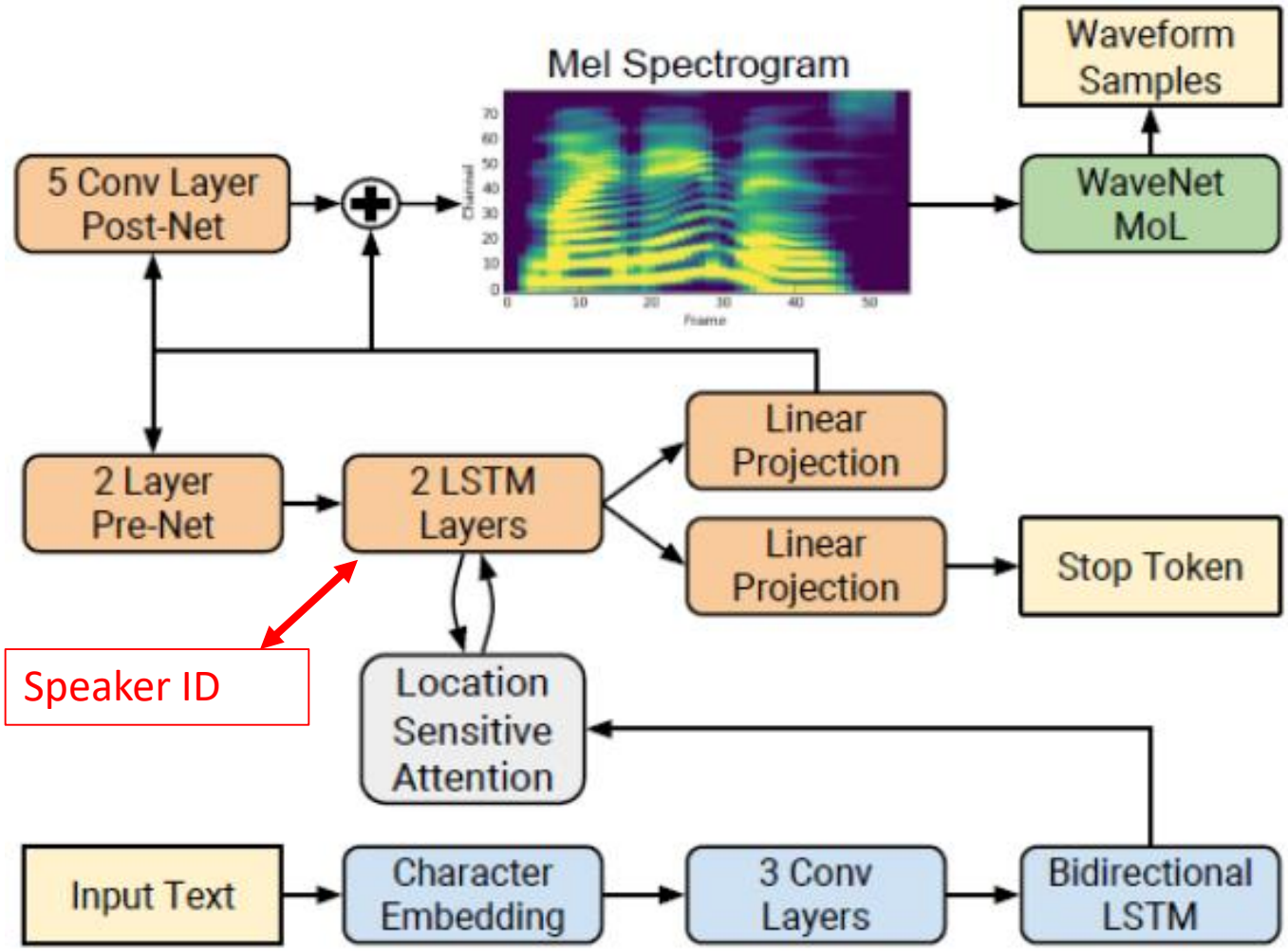
Amit nem tudsz egyszerűen elmagyarázni,

azt nem is érted egészen.

2017. január november

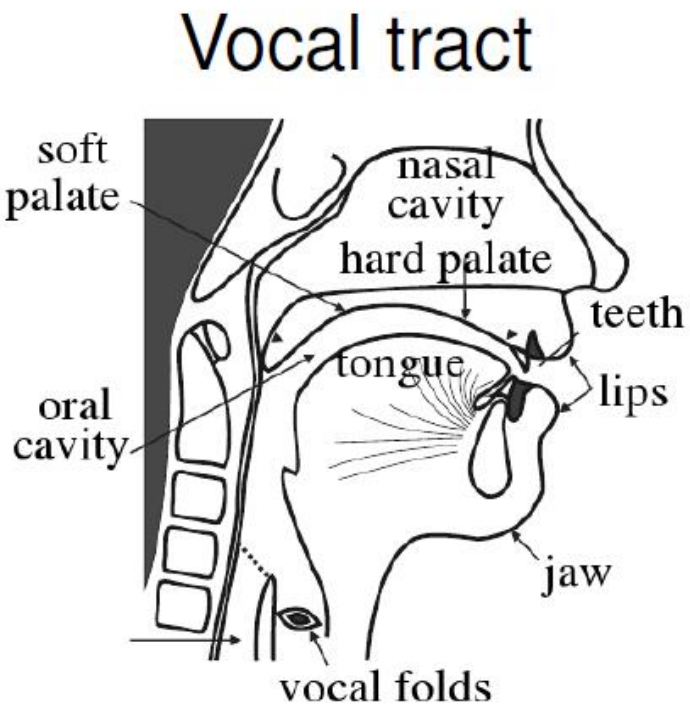
Albert Einstein

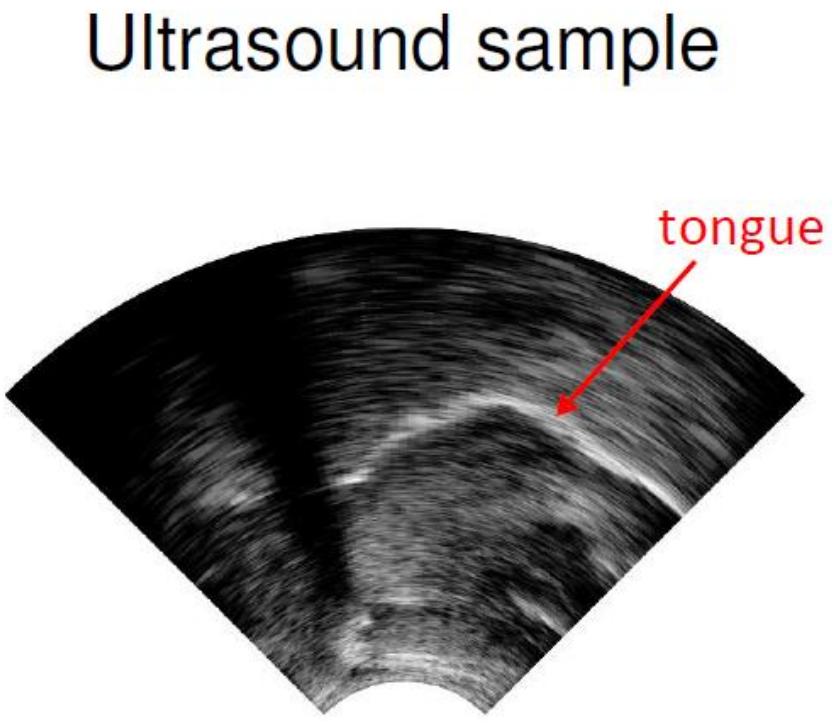
Tacotron + Wavenet/WaveGlow



Minták 10 beszélővel tanított Adatbázisból
2 óra/beszélő

Artikuláció-akusztika átalakítás ultrahangos nyelvpásztázás alapján („Silent Speech Interface”)





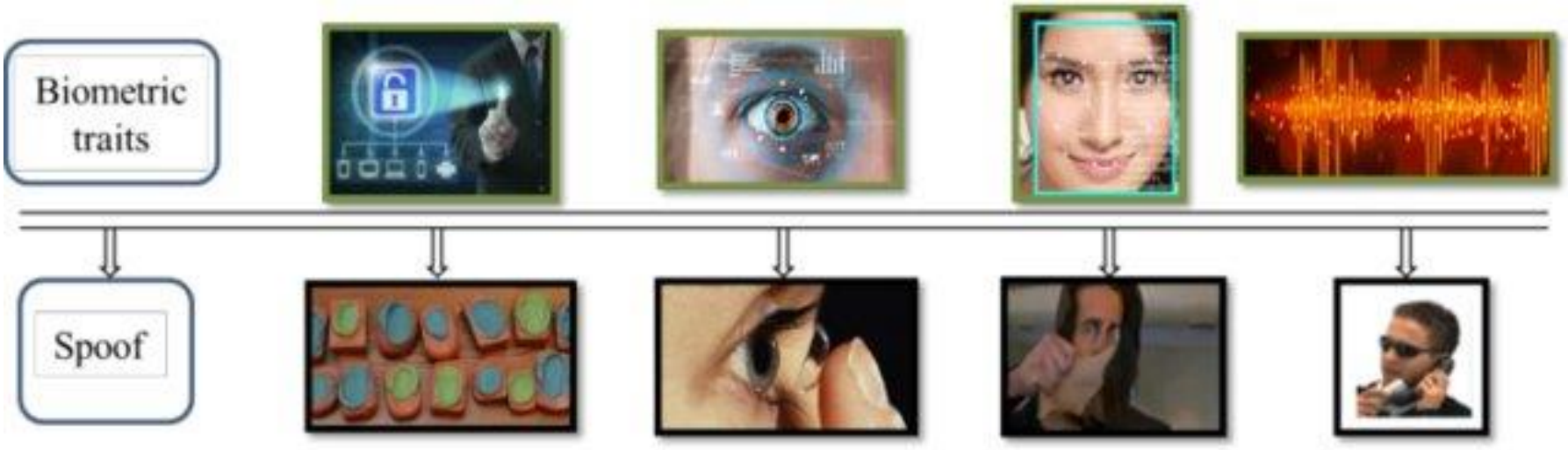
T. G. Csapó, T. Grósz, G. Gosztolya, L. Tóth, A. Markó, DNN-based Ultrasound-to-Speech Conversion for a Silent Speech Interface, Interspeech 2017, pp. 3672-3676.

Ultrahang képről beszéd előállítása (nincs szöveges bemenet!)

- Végső cél:
  - A csendes artikuláció hangos, érthető gépi beszéddé alakítása
  - Hasznos lehet a beszédsérültek számára
- Hangminták – Magyar („Az északi szél és a nap”)
- Női / 1 
- Női / 2 
- Férfi / 1 
- Férfi / 2 

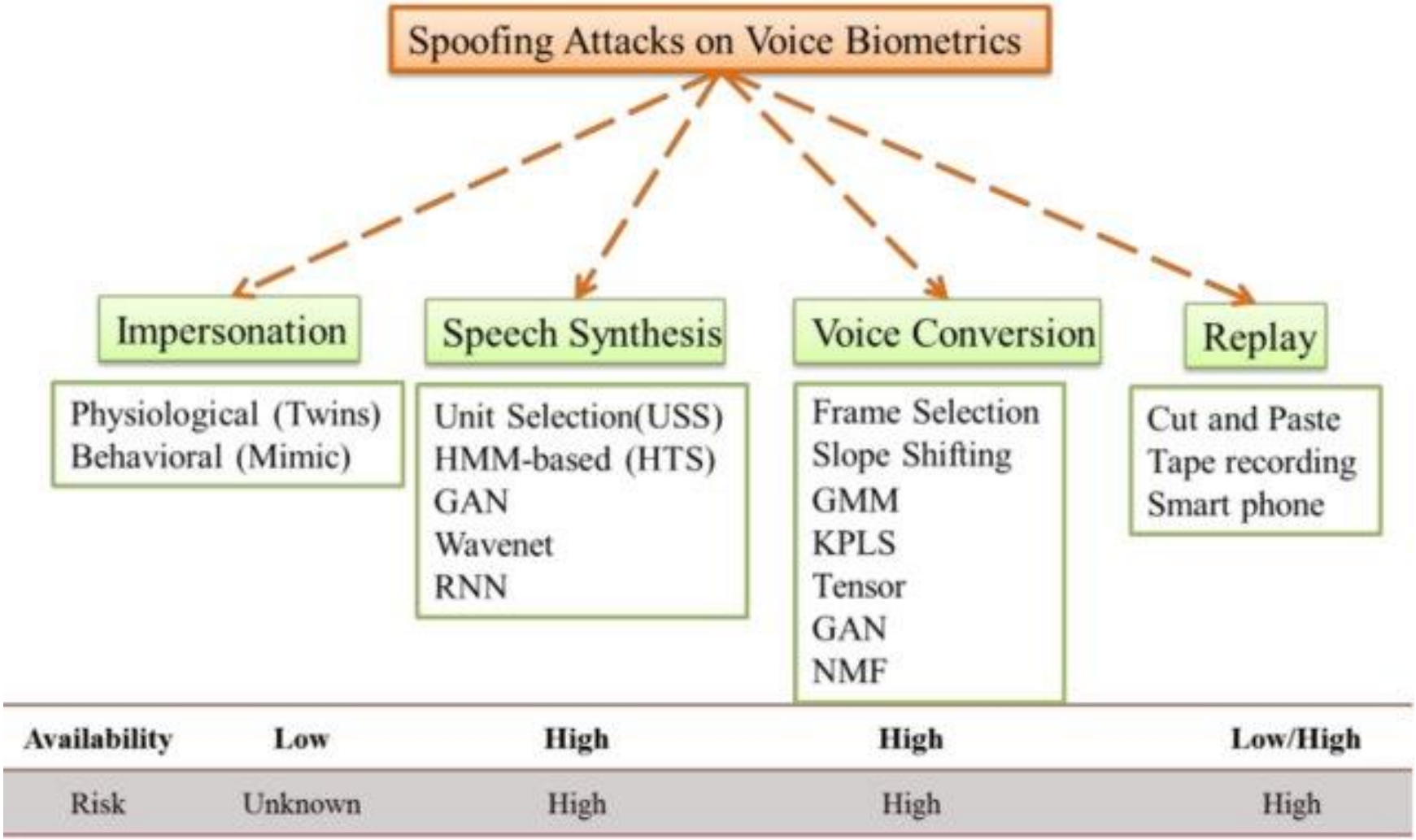
T. G. Csapó, T. Grósz, G. Gosztolya, L. Tóth, A. Markó, DNN-based Ultrasound-to-Speech Conversion for a Silent Speech Interface, Interspeech 2017, pp. 3672-3676.

Új technológiai és alkalmazási kihívások



- Source: Muhammad Jalaluddin Akbar: Overview of Spoof Speech Detection for Automatic Speaker Verification
- Obama deepface
- Automatic Speaker Verification Spoofing and Countermeasures Challenge, (https://www.asvspoof.org/ 2013, 2015, 2017, 2019)
- Nyelvfüggés -> nyelv/dialektus detekció/előállítás
- idegen/új szavak kiejtése (pl. Ansu Fati, Nagy, Gyöngyös angolul)

Speaker identification challenges



Availability	Low	High	High	Low/High
Risk	Unknown	High	High	High

Source: Muhammad Jalaluddin Akbar: Overview of Spoof Speech Detection for Automatic Speaker Verification

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)*

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 **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)