

Az SZTE Kutatóegyetemi Kiválósági Központ tudásbázisának
kiszélesítése és hosszú távú szakmai fenntarthatóságának megalapozása
a kiváló tudományos utánpótlás biztosításával”



Eötvös Loránd Kollégium Informatika műhely

2013. 04. 24.

Beszédtechnológia, beszédfelismerés
Dr. Tóth László



TÁMOP-4.2.2/B-10/1-2010-0012 projekt



Beszédtechnológiai témakörök



■ Beszédfelismerés

„Tiszta” felismerés = diktálás (beszéd→írás)

Más rendszer részeként: beszédinput (beszéd→akció)

Szorosan összefonódik a nyelvtechnológiával (nyelvi szinten is modellezni, elemezni kell)

■ Beszélőfelismerés

Verifikáció (igen/nem) vagy azonosítás (több beszélő közül)

■ Beszélőspecifikus információk kinyerése

Nem, életkor, érzelmi töltet, akcentus, hangszalagok eü. diagnózisa...

■ Beszédszintézis (text-to-speech)

Problémák: hangminőség, prozódia, érzelmi töltet



Beszédtechnológiai témakörök 2.



■ Beszédtárolás és -továbbítás

Beszédtömörítés (speech coding)

Minőség javítása (pl. telefonos torzítás) (speech enhancement)

■ Beszédatbázisok készítése

A statisztikai alapú algoritmusok tanítására/tesztelésére

Feldolgozás: alapos lehallgatás, szószintű átírat, zajok feljegyzése, szóhatárok (esetleg hanghatárok) bejelölése

■ Felismeréshez kapcsolódó nyelvtechnológiai problémák

Valószínűségi alapú nyelvi modellezés

Dialógusmodellezés

■ Kiértékelés

A beszédtech. alkalmazások hatásfokának objektív mérése



Beszédfelismerési alkalmazások



- **A teljesen általános felismerés nem megy!**
Valamilyen módon „szűkített” feladatokat keresünk, pl.
 - Szótár v. nyelvtan megszorítható (kicsi-közepes-nagy szótár)
 - A hangminőség jobb az átlagosnál
 - Beszédmód: izolált szavas – olvasott – folyó – spontán
 - Zaj jellege: nincs – modellezhető – változó
 - Beszélők: egyetlen beszélő sokáig (adaptáció) – változó beszélők
- **Klasszikus alkalmazások:**
 - Parancsvezérlés (kis szótár, izolált szavak → zaj esetén is OK)
 - Diktálórendszerek (nagy szótár, de tiszta jel, beszélőadaptáció)
 - Dialógusrendszerek (kis szótár, kötött nyelvtan, ált. telefonos)



Beszédfelismerési alkalmazások 2.



■ Jelenleg „divatos” alkalmazások:

Multimodalitás (szájról olvasás, arckifejezés, kézmozdulat felismerése, kombinálása a beszédfelismerővel)

Multilingvalitás: nyelv felismerése, akcentus felismerése és kezelése, fordítás

Hang-adatbányászat: audio indexing & retrieval, diarization, summarization

■ Néhány további érdekes téma:

Egyidejű beszédjelek szétválasztása

Audio forensics

Beszédterápia, olvasástanítás, nyelvtanulás segítése

Néma beszéd felismerése EMG-ből



Demonstráció – magyar rendszerek

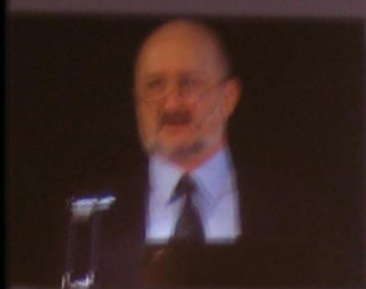


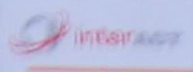
- **Izolált szavas felismerés (saját fejlesztés) :**
Kb. 1500 városnév, 98% körüli pontosság
- **Diktálórendszer közepes szótárral (saját fejlesztés) :**
Pajzsmirigy-leletek, kb. ezres szótár, erősen kötött nyelvtan
Szószintű pontosság 95% körül
- **Nagyszótáras felismerés (BME-AITIA):**
Százezres nagyságrendű szótár (morféma-tár)
Szószintű pontosság 80% körül
- **Dialógusrendszer**
Magyarra nem tudok ilyenről (folyamatos beszéddel)
- **Beszéd-beszéd fordítás (nemzetközi projekt)**
Felismerési pontosság 92%, fordítási pontosság 50%



S: Welcome to British Airways flight information service. How can I help you?
C: Yes, I'd like the departure time of flight BA four one eight from Heathrow Airport.
S: Do you know the flight number?
C: The flight number is BA four one eight.
S: Please tell me the flight number.
C: BA four one eight.
S: Was that flight number three eight four?
C: BA four one eight.
S: Please tell me the name of the airline.
C: British Airways.
S: Did you say British Airways?
C: Yes.
S: Was that British Airways?
C: Yes.
S: Sorry, I missed what you said. Did you say British Airways?
C: Yes, British Airways.
S: I'm sorry. I don't know of any such flight. Do you have another inquiry?
C: Yes.
S: Thank you for calling. Good bye.







Hearing Personal Translations

- Technology: Targeted Audio
 - Research under DARPA Project CHL (Built-In-Hearable Computer Services)
 - Project Partner: Sennheiser/Neodym
 - Array of Ultra-Sound Speakers
- Result: Narrow Sound Beam
 - Audible by one individual only
 - Others not disturbed
 - Multiple arrays could provide multiple languages
 - Steerable
 - Recognize/Track Individual Listener and Keep Language Beam on Target



Copyright © 2010



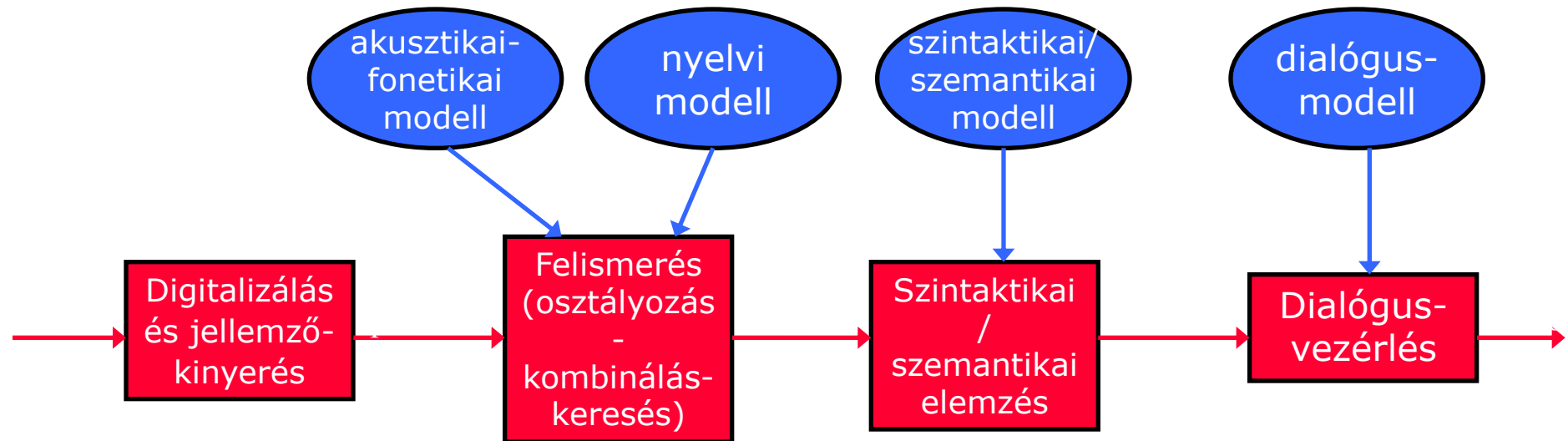
bean that can deliver acoustic → output in a particular direction



acústico haba que pueden entregar acústicas → producción en una dirección particular



Beszéd felismerő komponensei



beszéd → szöveg

beszéd → tartalom

beszéd → akció



TÁMOP-4.2.2/B-10/1-2010-0012 projekt



Jellemzőkinyerés

- **Cél: a lényegtelen információ kidobása, a lényeges tömör reprezentálása**

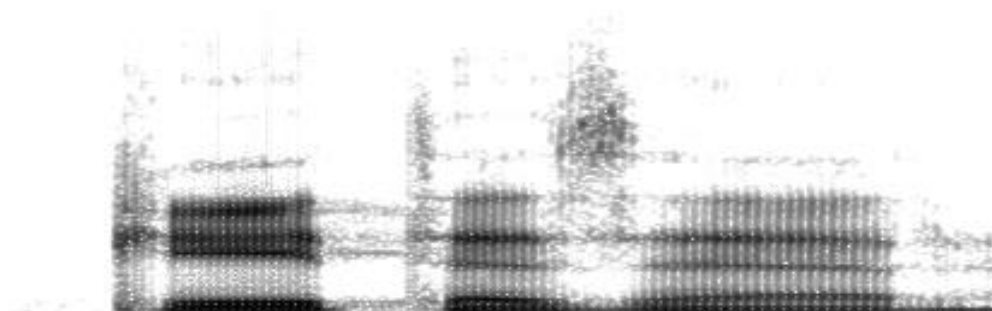
„lényeg”: nyelvi tartalom (ld. fonetika)

Klasszikus módszerek: tisztán matematikai megoldások

(ld.: digitális jelfeldolgozás, beszédfeldolgozás)

Újabb módszerek: az emberi fül/hallás (durva) modellezése (ld.: pszichoakusztika, neurofiziológia)

Mindig valamilyen spektrális vektorsorozat a kimenet, felbontás: 100 vektor/sec, 40-50 komponens/vektor



TÁMOP-4.2.2/B-10/1-2010-0012 projekt



A felismerés szintjei

- **A beszéd (és nyelv) specialitása, hogy hierarchikus:**

Spektrális vektor („adatkeret”): 25-30 ms-os részletet ír le

Beszédhang: a magyarban kb. megfelel a betűknek

Szó: angolban ~felsorolhatók, magyarban toldalékolás is van!

Mondat: nyelvtani szabályokkal próbáljuk leírni (szintaxis)

Szöveg ill. dialógus: szemantikai tartalom ill. beszédaktus

- **A színtezettségek kezelése:**

Alapvetően bottom-up módon:

Az alacsonyabb szinteken a felismerés nem megy (ld. későbbi példák!), ezért a döntést késleltetjük

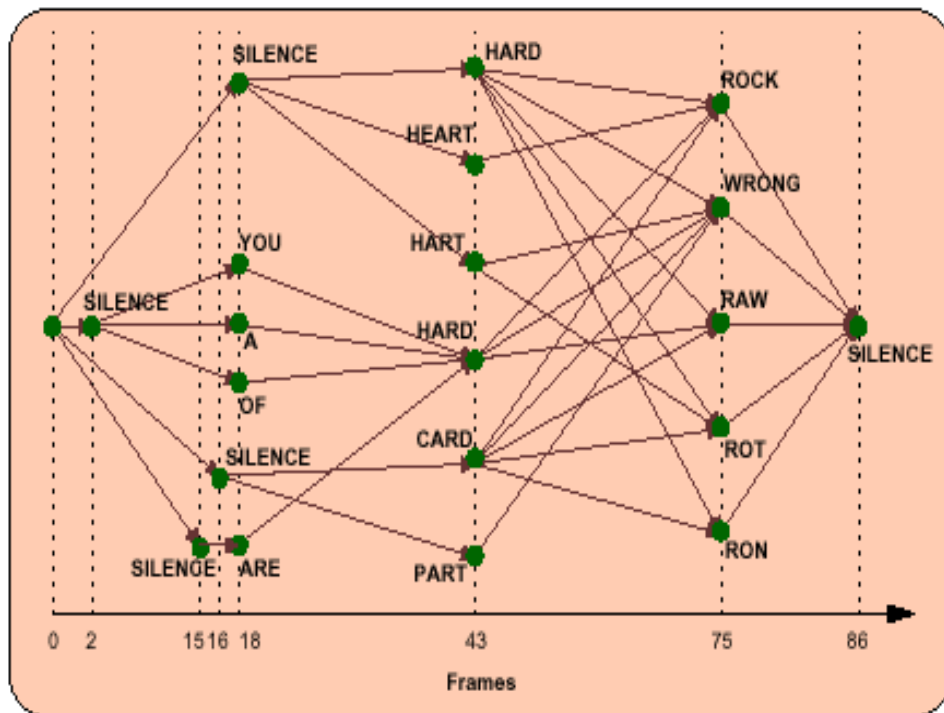
Valószínűségek hozzárendelése, és több megoldási javaslat továbbítása a következő elemzési szintnek



Példa a legv. hipotézisek átadására

- **N-best list és word lattice mint a felismerés kimenete:**

Rank	Hypotheses	Likelihood
1	SILENCE HARD ROCK SILENCE	-5880.11
2	SILENCE HARD WRONG SILENCE	-5905.17
3	SILENCE HARD RAW SILENCE	-5906.32
4	SILENCE A HARD ROCK SILENCE	-5920.68
5	SILENCE HARD ROT SILENCE	-5922.05
6	SILENCE HARD RON SILENCE	-5923.69
7	SILENCE CARD WRONG SILENCE	-5924.51
8	SILENCE CARD RAW SILENCE	-5925.66
9	SILENCE YOU HARD ROCK SILENCE	-5928.95
10	SILENCE HART WRONG SILENCE	-5929.97
11	SILENCE HEART WRONG SILENCE	-5930.42
12	SILENCE ARE HARD ROCK SILENCE	-5936.11
13	SILENCE CARD ROCK SILENCE	-5936.86
14	SILENCE OF HARD ROCK SILENCE	-5937.56
15	SILENCE CARD ROT SILENCE	-5941.39
16	SILENCE CARD RON SILENCE	-5943.03
17	SILENCE A HARD WRONG SILENCE	-5945.74
18	SILENCE PART WRONG SILENCE	-5946.36
19	SILENCE HART ROT SILENCE	-5946.85
20	SILENCE A HARD RAW SILENCE	-5946.89



- **Ha van további szemantikai elemzés vagy dialógus-modellezés, akkor az az ide bekerült hipotéziseket vizsgálja tovább**

A felismerés



■ 1. részfeladat: osztályozás

Spektrális vektorok, esetleg beszédhangnyi szegmentumok azonosítása (milyen hangnak felel meg)

Statisztikai gépi tanulási megoldások: rejtett Markov-modell (HMM) vagy neuronháló

■ 2. részfeladat: kombinálás

A vektorokat kombinálni kell beszédhangokká, majd a beszédhangokat szavakká, a szavakat mondatokká

Valószínűségek kezelése: valszám. szabályai alapján

■ 3. részfeladat: keresés

Keressük a legnagyobb valószínűségű kombinációt

Din. prog., heurisztikák, vágás kell a hatékony bejáráshoz

A HMM mind a 3 lépést tudja egyben!

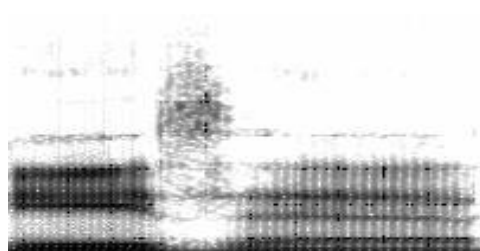


Szemléltetés



SZÉCHENYI TERV

- **1. Az időpillanatokhoz (spekt. vektorokhoz) osztályonkénti valószínűségeket rendelünk**
- **2. Legvalószínűbb hangsorozat megtalálása: maximumot adó út keresése dinamikus programozással**



- **Keresési megszorítások a szegmentumokra (hangokra):**

Hosszmodellez

Egymásra követ



hangoké)

- **A legvalószínűbb szó sorozat megtalálása is ugyanígy!**

bizonyos sorozatok engedélyezése/kizárása (szótár)

Egymásra következő valószínűség (szavaké) itt is segít



TÁMOP-4.2.2/B-10/1-2010-0012 projekt



A projekt az Európai Unió támogatásával,
az Európai Szociális Alap
társfinanszírozásával valósul meg.

Az egyes szintek kiértékelése



- **Spektrális vektorok felismerési pontossága:**
50-70% közötti pontosság érhető el csak!!
- **Hangsorozatok felismerési pontossága:**
55-80% közé szokott esni
Példák (74.4% esetén):
erejgyűlkörnyezet édeulebekregyutozás
Azegyisikereamásikösatásalet
miregalábosajaterületünkörendettaszt
- **Szószintű pontosság:**
Láttuk korábban, hogy nagyon függ a szótár és a nyelvtan kötöttségétől!!! (95% vs. 80%)
- **Dialógusrendszer pontossága:**
Egyáltalán nem triviális kiértékelni



Dialogusrendszer kiértékelése

"In the US (and some parts of Europe) it's quite common now to encounter a spoken dialogue system. Several of the cinema chains in the UK use such a system. However, the fact that they're frequently portrayed as a joke shows just how bad the systems are. In my lectures I always ask how many people have come across such systems in their daily life (answer - the majority), then I ask how many had a useful experience (answer - almost always, none!)."

Prof. Roger K. Moore, University of Sheffield

