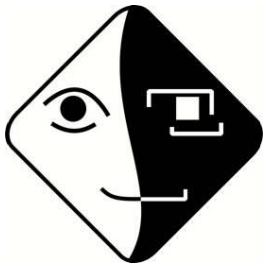
Prozódiai változatosság rejtett Markov-modell alapú szövegfelolvasóval

Csapó Tamás Gábor, Németh Géza

{csapot, nemeth}@tmit.bme.hu

BME Távközlési és Médiainformatikai Tanszék

VIII. Magyar Számítógépes Nyelvészeti Konferencia,
2011. december 2.



Prozódiai változatosság

Emberi beszédben:

- ugyanaz a mondat többször kiejtve kicsit máshogy hangzik
- változatosság a dallamban, hangsúlyozásban, ritmusban

Szövegfelolvasókban eddig:

- a legtöbb rendszerben egy adott mondat mindig ugyanúgy hangzik
- többnyire determinisztikus rendszer
- nincs jól észlelhető változatosság

Cél: szövegfelolvasóban is változatos dallam

Rejtett Markov-modell (HMM) alapú szövegfelolvasó

- Statisztikai alapú, parametrikus működés
- HTS nyílt forráskódú, HMM alapú szövegfelolvasó magyar nyelvű változata
- Beszéd paramétereinek gépi tanulása tanítóadatbázis alapján
- 1 női beszélőtől 2,05 óra felvétel, 1940 mondat

Prozódiai változatosság megvalósítása

- Tanítóadatbázis több részre bontása
- HTS rendszer tanítása külön-külön a beszéd alapfrekvenciájára (F0)
- Mennyire lesznek eltérőek a szintetizált mondatok (dallam szempontjából)?
- Szétbontás: véletlenszerű megfelelő, vagy célzott megoldás kell?

Mondatdallam távolságmértékek

- Objektív különbség két mondat F0-ja között

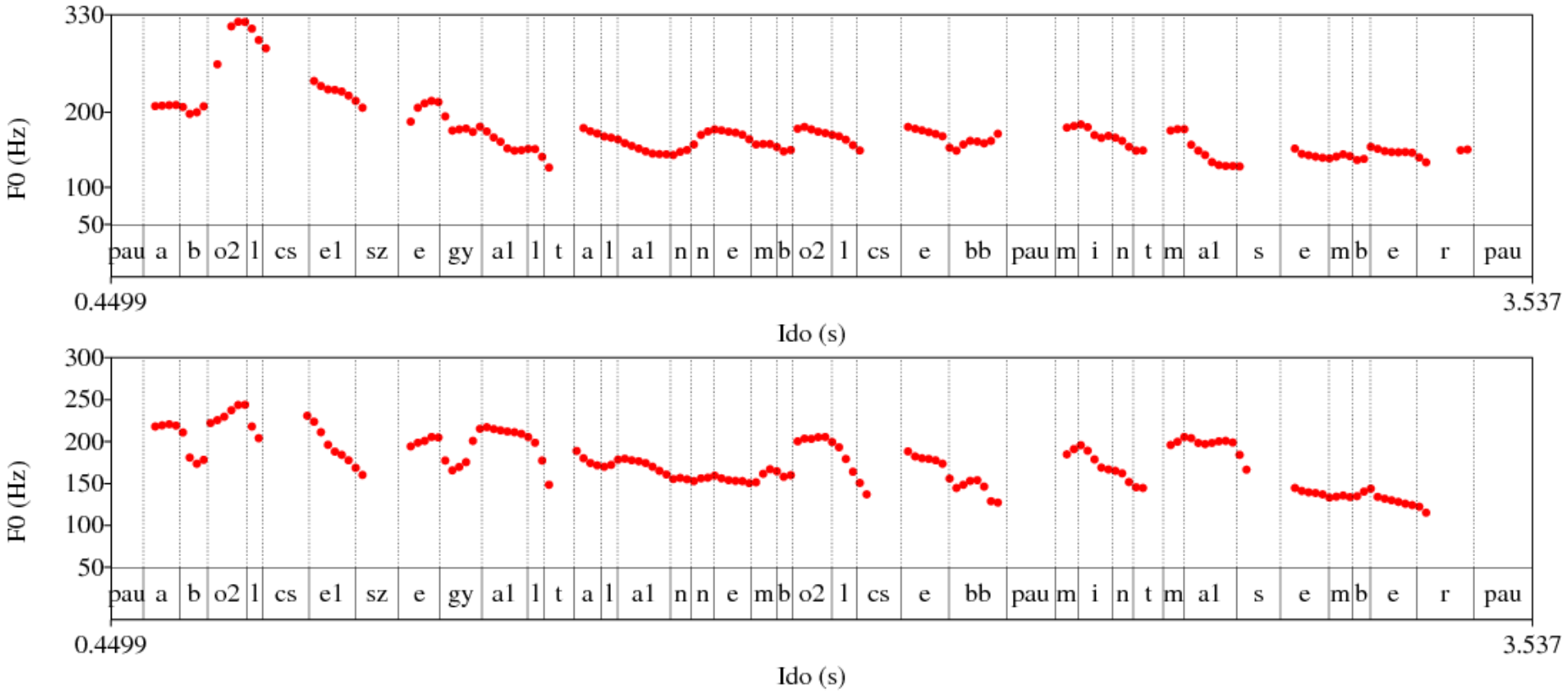
$$RMSE_{f1,f2} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left(f_1(i) - f_2(i)\right)^2\right)}$$

$$Hermes_{f1,f2} = \frac{\sum_i w(i)(f_1(i)-m_1)(f_2(i)-m_2)}{\sqrt{\sum_i w(i)(f_1(i)-m_1)^2 \sum_i w(i)(f_2(i)-m_2)^2}}$$

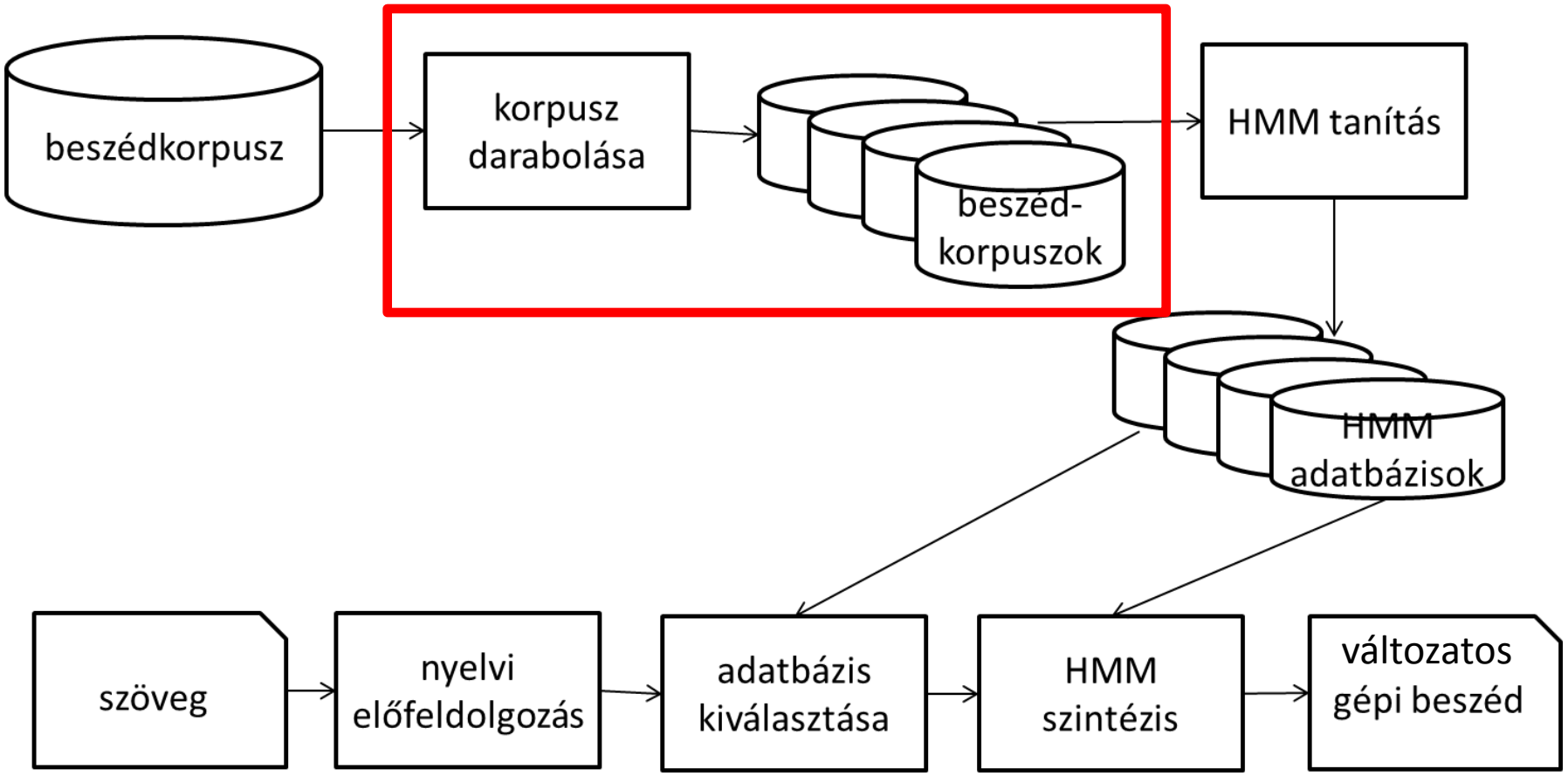
- **f1/f2**: egyik/másik mondat F0-menete, szótagonként
- **m1/m2**: egyik/másik mondat F0 átlaga
- **w**: súlyozó faktor

Mondatdallam távolságmérték példa

- RMSE távolság: **0,1619**
- Hermes korreláció: **0,6337**



Változatos dallamú szövegfelolvasás



Tanítóadatbázis véletlen felbontása

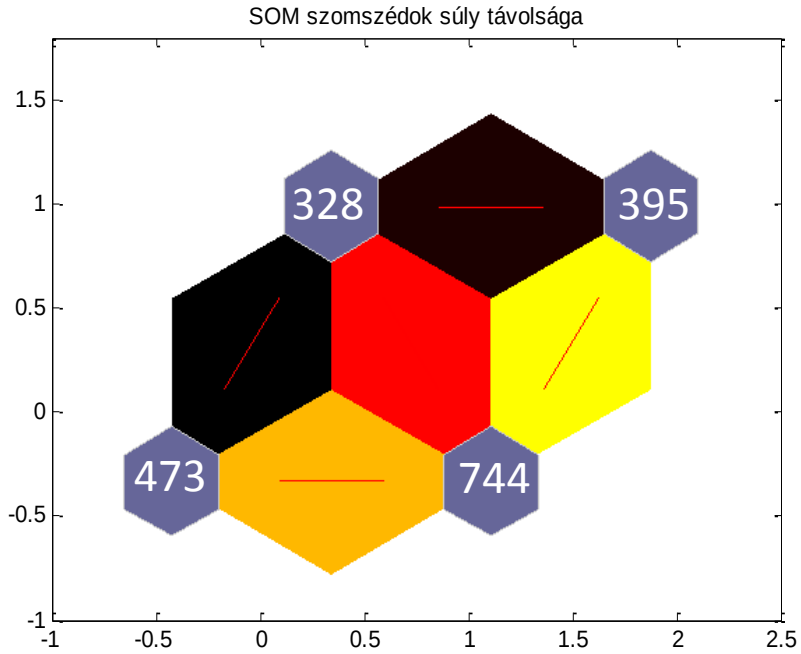
- Tanítóadatbázis felbontása 2/4/8 különálló részre, véletlenszerűen
- F0-modell tanítása mindegyik részkorpuszon
- Eredmény:
  - Mondatváltozatok nem különböztek lényegesen
  - Magas F0 korreláció
  - Nincs érezhető dallam-különbség

Tanítóadatbázis SOFM alapú felbontása (1)

- Self-Organizing Feature Map
- Felügyelet nélküli gépi tanulási módszer
- Klaszterezés adott számú csoportra
- Tanítóadatbázis felbontása 4 különálló részre
- Bemenet mondatonként:
  - F0 minimum
  - F0 maximum
  - F0 átlag
  - F0 szórás

Tanítóadatbázis SOFM alapú felbontása (2)

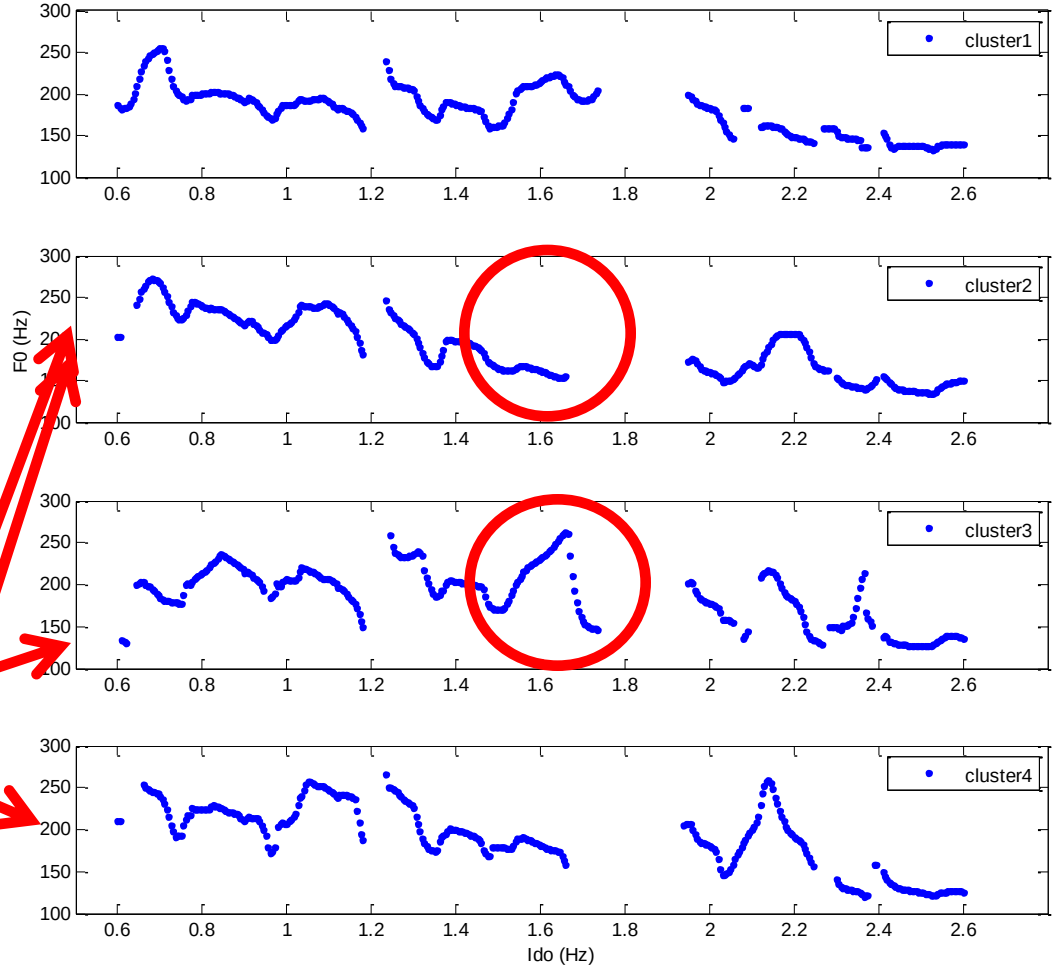
- F0-modell tanítása mindegyik részkorpuszon
- Eredmény:
  - Mondatváltozatok különböznek
- További vizsgálatok
  - Objektív elemzés
  - Szubjektív teszt



SOFM alapú felbontás eredménye

- Objektív különbségek

	V 1	V 2	Hermes korreláció
#1625	1	2	0,7812
#1625	1	3	0,8299
#1625	1	4	0,8523
#1625	2	3	0,6547
#1625	2	4	0,9233
#1625	3	4	0,8081



Objektív elemzés eredménye

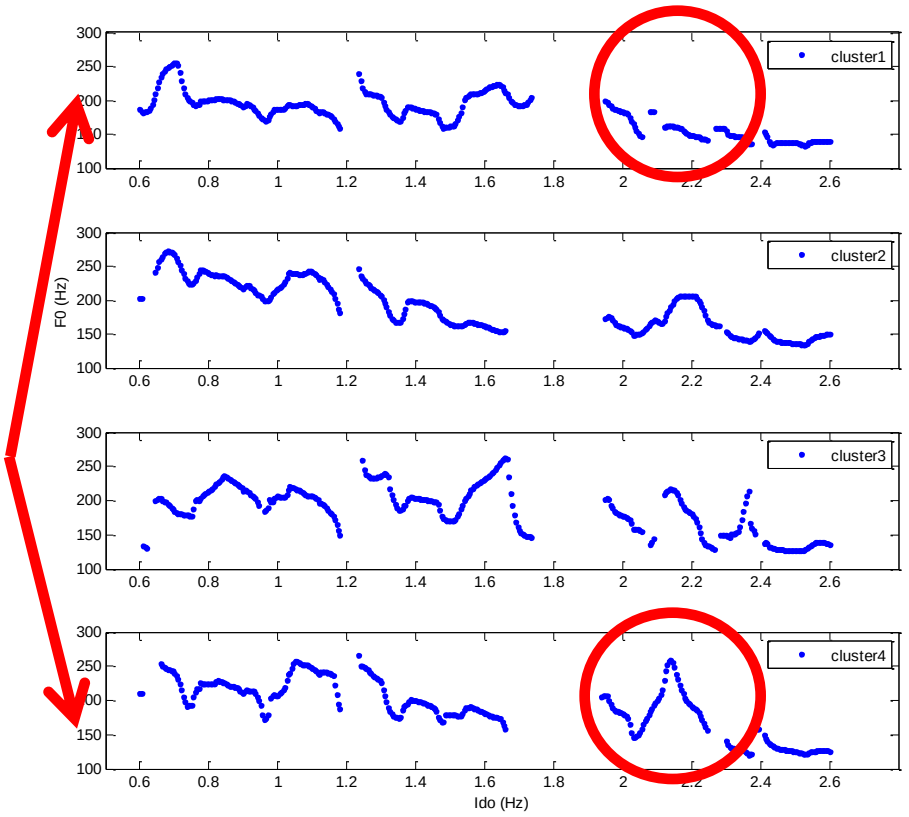
- 2000 mondat szintetizálása, 4-4 változat
- Páronkénti összehasonlítás F0-menet alapján
  - Hermes korreláció számítása páronként
  - 465 mondat esetén változatok közötti korreláció < 0,9
- Hermes korreláció szerint leginkább különböző 10 mondat kiválasztása

Meghallgatásos teszt

- 10 mondat, 4-4 változat, páros összehasonlítás
- 60 mondatpár
- Internetes teszt
- „Hallasz-e különbséget a két mondat dallama között? Igen – Nem”
- 9 tesztelő, főleg beszédtechnológiai szakértők és fonetikusok

Szubjektív különbségek

	V 1	V 2	Szubjektív „Igen” arány	Hermes korreláció
#1625	1	2	44,44%	0,7812
#1625	1	3	44,44%	0,8299
#1625	1	4	77,78%	0,8523
#1625	2	3	77,78%	0,6547
#1625	2	4	66,67%	0,9233
#1625	3	4	66,67%	0,8081



Szubjektív teszt eredménye

- Mondatpárok 58%-ában hallottak eltérést a tesztelők
- Ahol hallottak eltérést
  - a legnagyobb F0 különbség akár a 70 Hz-et is elérte
  - a mondat hangsúlya is másik szóra került (nem mindig jó pozícióra)
- Ahol nem hallottak eltérést
  - a mondatváltozatok közötti szótagonkénti átlagos F0 különbsége legfeljebb 10-20 Hz

Összefoglalás

- Következtetések
  - SOFM alkalmas a beszédkorpusz szétbontására
  - Objektív F0 korreláció és szubjektív dallam különbség nem mindig egyezik
  - Egyes mondatoknál sikerült dallamváltozatokat létrehozni
  - Még nem általános megoldás
- Alkalmazhatóság
  - Hosszabb szöveg felolvasása: könyvfelolvasó, e-levél felolvasó
- További kutatási irányok
  - Beszélők közötti dallam adaptálás
  - Prozódia más részei (pl. ritmus)

Köszönöm a figyelmet

- csapot@tmit.bme.hu

- A kutatást részben a TÁMOP-4.2.1/B-09/1/KMR-2010-0002 projekt támogatta.