

Csapó Tamás Gábor

# Változatos prozódia megvalósítása szövegfelolvasó rendszerekben

Konzulensek:

Dr. Németh Géza, Dr. Fék Márk

Budapesti Műszaki és Gazdaságtudományi Egyetem  
Távközlési és Médiainformatikai Tanszék

Diplomavédés

2008. június 17.



BME TMIT

---

# Tartalom

---

- A diplomaterv célja
- Elméleti bevezetés
- Rendszerterv
- A megvalósítás részletezése
- A prototípus rendszer vizsgálata
- További kutatási irányok

# A diplomaterv célja

---

## Emberi beszédben:

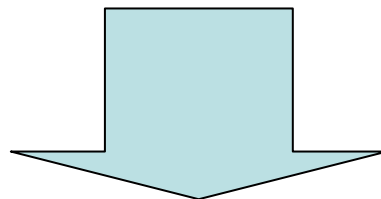
- ugyanaz a mondat többször kiejtve kicsit máshogy hangzik
- változatosság a dallamban, hangsúlyozásban, ritmusban

## Szövegfelolvasókban eddig:

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



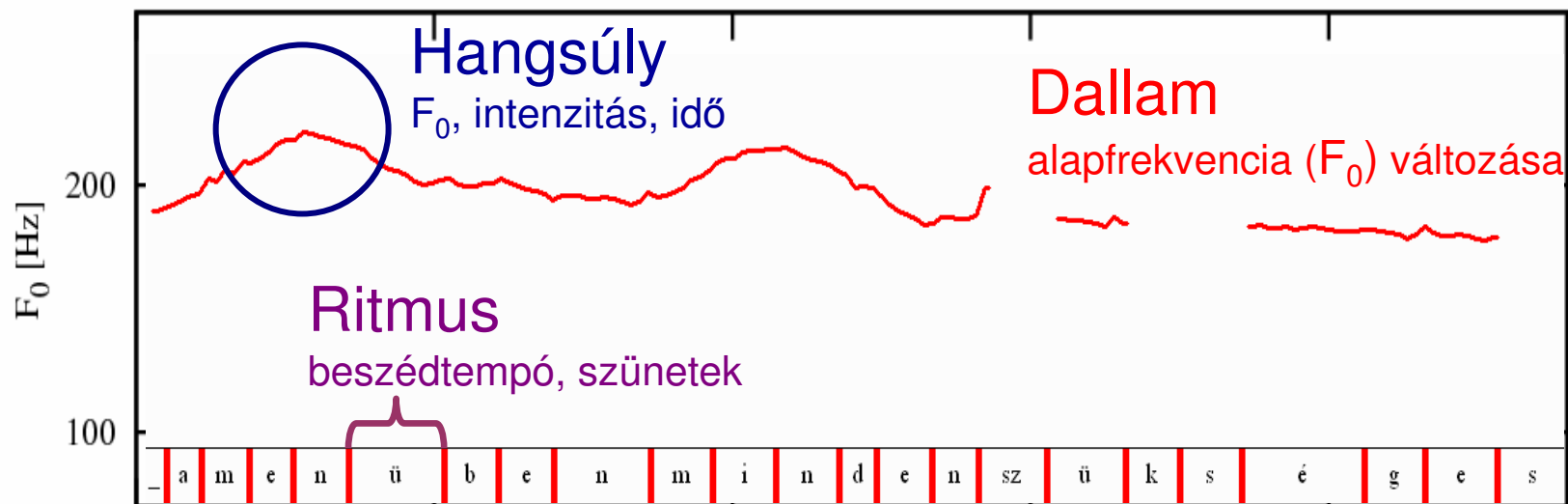
*„Párás, ködös az idő.”*



Cél: változatos beszéd megvalósítása a szövegfelolvasókban

# Elméleti háttér (1) - prozódia

- Beszéd (prozódia) 3 összetevője:



alapfrekvencia: fontos paraméter

⇒ ezzel foglalkoztunk

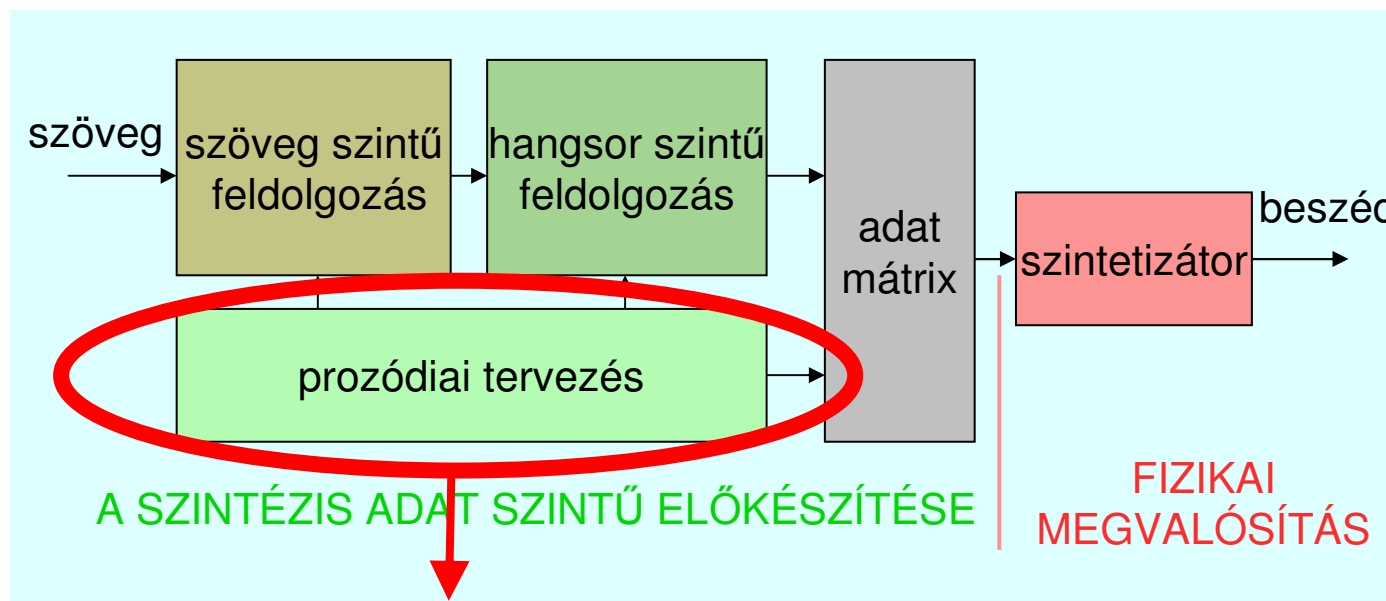
$F_0$  modellezésével



$F_0$  modellezésével

# Elméleti háttér (2) - beszédszintézis

- Beszédszintézis: szöveg → beszéd  
(Text-To-Speech = TTS)



## Dallammenet módosítása

Forrás: Magyar nyelvi  
beszédtechnológiai  
alapismeretek CD

## Elméleti háttér (3) - prozódia modellezése

- **Prozódiai modellek**
  - Szabály alapú, szuperpozíciós modellek
    - BME TMIT: Profivox
  - Korpusz alapú modellek
    - természetes beszédből származó minták alapján
    - különbségek a rendszerek között:
      - egyszerű / többszintű (szuperpozíciós)
      - felhasznált elem mérete (hang - ... - mondat)
      - módszer (statisztikai / CBR / szabályok)
      - prozódia mely eleme (főleg  $F_0$ , esetleg  $t$ )

# Rendszerterv

---

- Követelmények
  - lefedettségi arány (kb. 50%)
  - változatok száma (2-3)
  - futásidő (közel valós idő)
- Megvalósítási lehetőségek
  - beszéd szintetizátor-technológia (elemösszefűzés)
  - prozódiai modell (korpusz alapú)
- Kiválasztott rendszer terve
  - prozódia-minta adatbázis létrehozása
  - módszer működése
  - szövegfelolvasóhoz illesztés

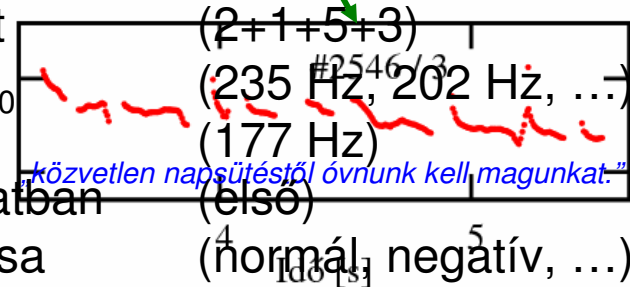
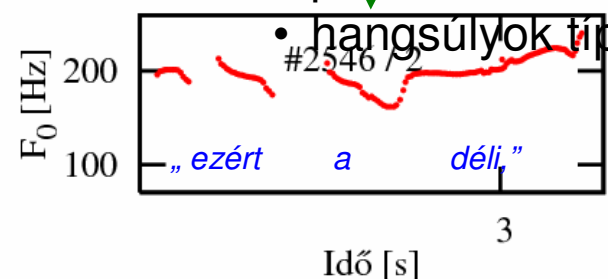
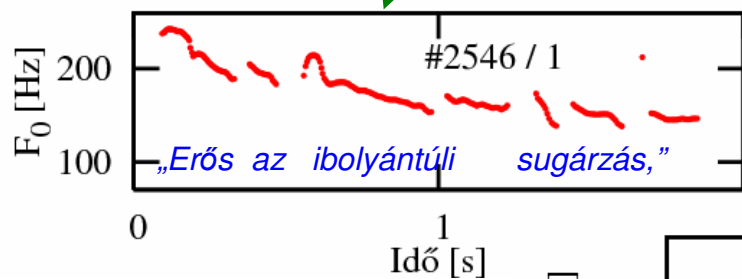
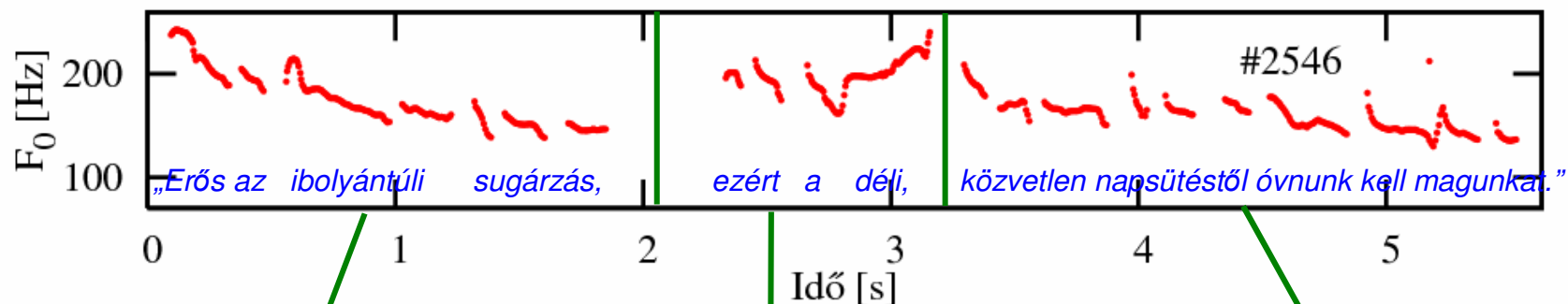
# Megvalósítás (1) - beszédkorpuszok

- 5 kötött témakörű beszédkorpusz vizsgálata  
(Időjárás-előrejelzés, Harangok története, Árlista, Prompt, Vonat)
- Kb. 13 ezer magyar nyelvű kijelentő mondat
  - 180 ezer szó, 21 óra
- Tartalmuk:
  - hullámforma
  - szöveges és fonetikus átírás
  - hangidőtartamok
  - $F_0$ -menet



# Megvalósítás (2) - beszédkorpuszok

- Feldolgozás: prozódiai egységekre bontás



Eltárolva:

- szótagszerkezet (2+1+5+3)
- szótagonkénti  $F_0$  (235 Hz, 202 Hz, ...)
- átlagos  $F_0$  (177 Hz)
- pozíció a mondatban (első)
- hangsúlyok típusa (normál, negatív, ...)



BME TMIT

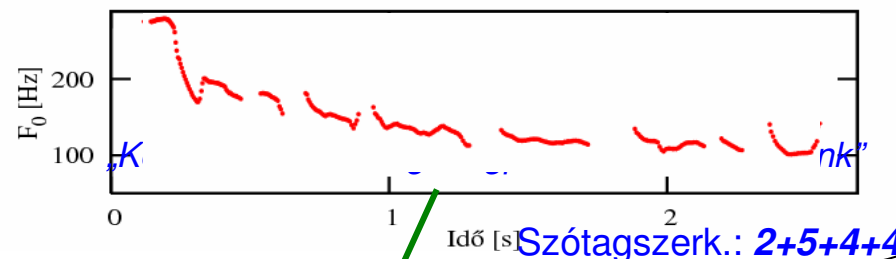
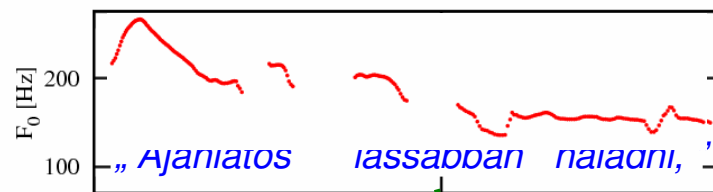
# Megvalósítás (3) - dallammásolás

- **Prozódiai egység alapján**  
– keresés: szótagszerkezet szerint

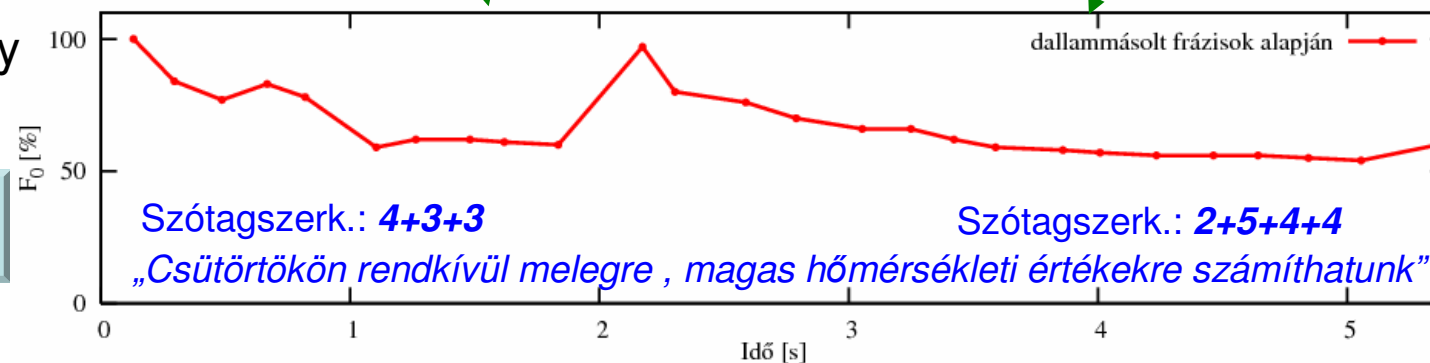
#1091 /1

#2514 /1

F<sub>0</sub>-minta adatbázis prozódiai egységekkel



szabály alapú

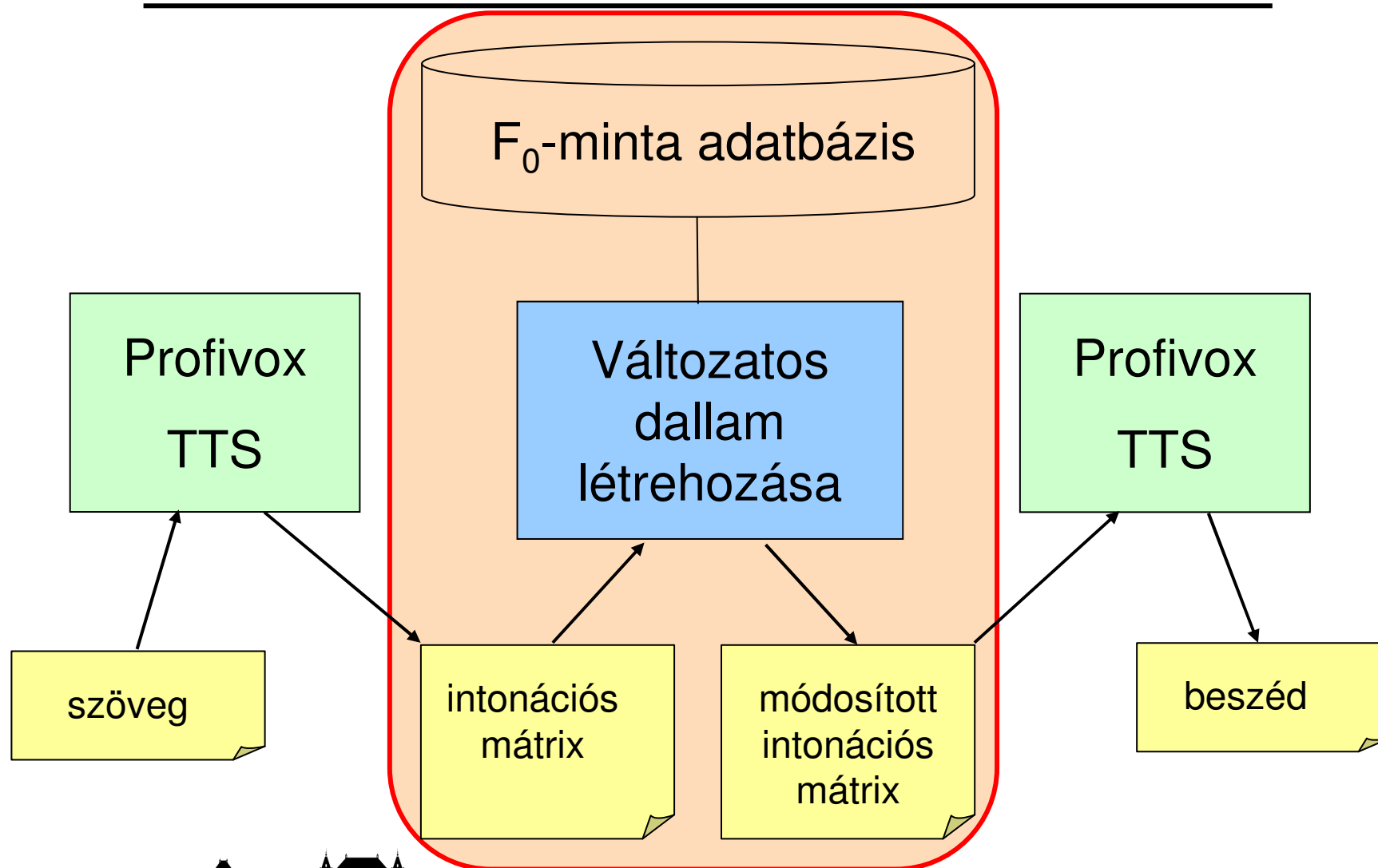


dallam-  
másolt



BME TMIT

# Megvalósítás (4) - Profivox szövegfelolvasóhoz illesztés



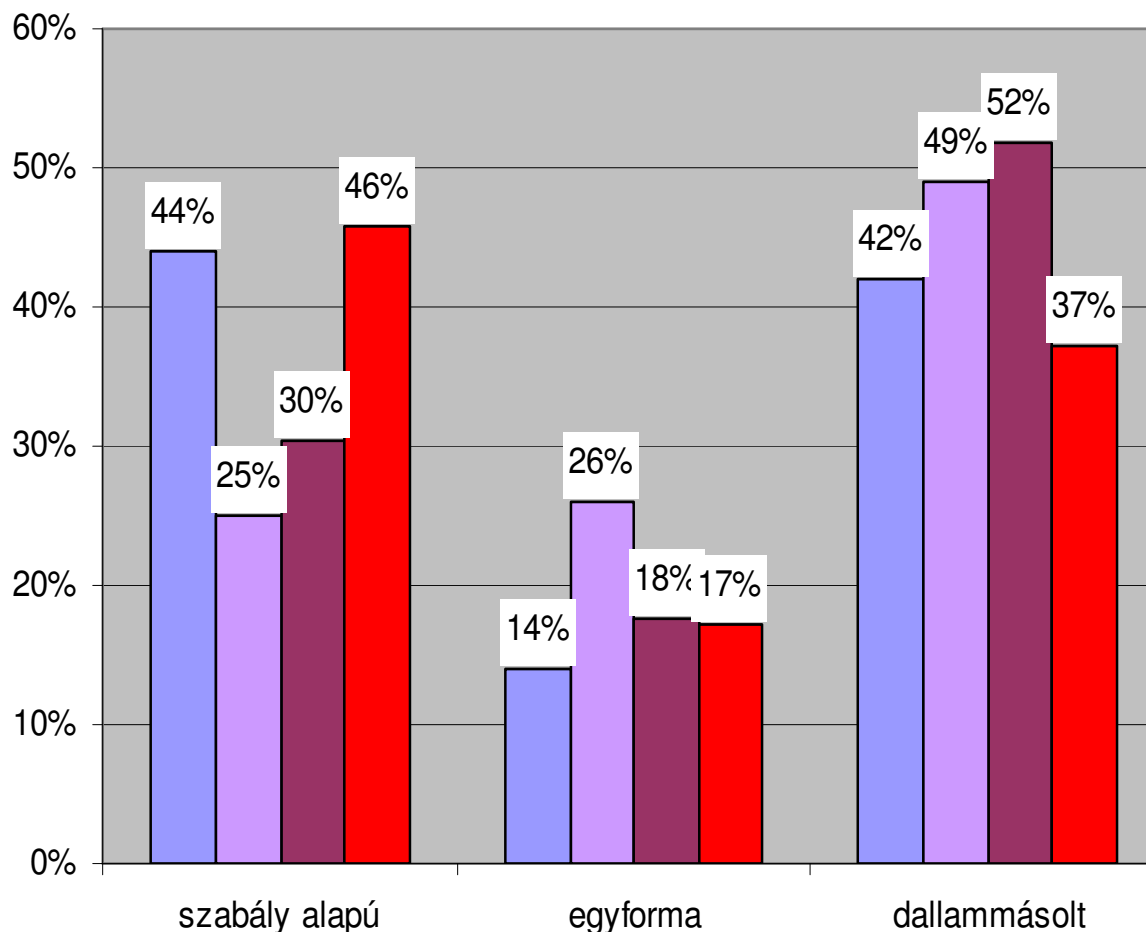
# Elemzés (1) - tesztek

---

- Lefedettség elemzése
  - felhasznált adatbázis minőségétől, méretétől függ
  - találati arány kb. 50%
- Megvalósítható változatosság vizsgálata
  - 2-3 változat / mondat
- Futásidő vizsgálata
  - adatbázis méretétől függ
  - kb. 25 ms / mondat
- Szubjektív tesztek
  - meghallgatásos vizsgálatok weben
  - különböző változatok összehasonlítása  
(szabály alapú dallam ⇔ dallammásolt)
  - átlagosan 20 tesztelő mondatpáronként

## Elemzés (2) - szubjektív tesztek eredményei

- 2006
  - dallammásolás manuális módszerrel
  - rövid mondatok
- 2007/1
  - hosszabb mondatok
- 2007/2
  - dallammásolás automatikus módszerrel
- 2008
  - Profivox szövegfelolvasóba illesztve
  - hosszabb szövegek (bekezdések) vizsgálata



# Előadás összefoglalása

---

- Elméleti bevezetés
  - prozódia, modellek, beszéd-szintézis
- Tervezés
  - alternatívák közül egy kiválasztása
- Megvalósítás
  - beszédkorpusz prozódiai egységekre bontása
  - dallammásolás prozódiai egység alapján
  - Profivox szövegfelolvasóhoz illesztés
- Elemzés
  - objektív és szubjektív tesztek
- Eredmény
  - a dallammásolás a legtöbb esetben jó eredményt nyújt
  - még vannak hibák

## További kutatási irányok

---

- Prozódia többi komponense
  - időzítés
  - intenzitás
- Dallammásolás során alkalmazott hasonlósági mérték finomítása
  - hangsúlyok figyelembe vétele
- Más nyelvek
  - angol, német
- Változatosság megvalósítása elemkiválasztásos beszéd szintetizátorban

# Reagálás a bírálatra

---

- Beszédkorpusz címkézése
  - automatikusan, manuális korrekcióval (másik kutatás)
  - nagy méretű adat: 21 óra
  - pontosság megfelelő, de esetenként  $F_0$ -hiba
- $F_0$ -átlagolás a korpusz feldolgozásakor
  - csak a zöngés szakaszokon történt
- Adatbázis mérete és futásidő közti összefüggés
  - teljes rendszer vizsgálata volt a cél
  - leghosszabb esetben sem jelentős időtartam
- Összehasonlító felvételek csak 1-1 mondatot tartalmaznak
  - korábbi tesztek így végeztük
  - most az összbenyomás vizsgálata volt a cél



Köszönöm a figyelmet!

Csapó Tamás Gábor

csapo@tmit.bme.hu



BME TMIT

---