

Csapó Tamás Gábor
IV. évf.

Szintetizált beszéd természetesebbé tétele

Konzulensek:
Dr. Németh Géza, Dr. Fék Márk

Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék

OTDK előadás
2007. április 26.



BME TMIT

Tartalom

- A kutatás célja
- Elméleti háttér
- Kísérleti munka
- Meghallgatásos vizsgálatok
- Továbbfejlesztési lehetőségek

A kutatás célja (1)

- Beszédszintetizátorok generációi:

- Formánsszintézis



- Elemösszefűzés



- Korpusz alapú elemkiválasztásos



„A hűvös, párás hajnalt követően változóan felhős lesz az ég, több-kevesebb napsütés várható.”

A kutatás célja (2)

Emberi beszédben:

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

A kutatás célja (2)

Emberi beszédben:

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

Szövegfelolvasókban eddig:

- egy adott mondat szinte mindig ugyanúgy hangzik
- többnyire determinisztikus rendszer
- nincs jól észlelhető változatosság

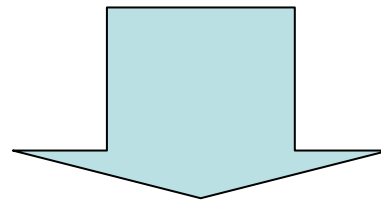
A kutatás célja (2)

Emberi beszédben:

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

Szövegfelolvasókban eddig:

- egy adott mondat szinte mindig ugyanúgy hangzik
- többnyire determinisztikus rendszer
- nincs jól észlelhető változatosság



Cél: változatos beszéd megvalósítása a szövegfelolvasókban

Elméleti háttér (1)

- Beszéd (prozódia) összetevői:

Szubjektív

- Dallam
- Hangsúly
- Ritmus

Objektív

- alapfrekvencia (F_0) változtatás
- F_0 , intenzitás emelés, időtartam módosítás
- beszédtempó változtatása, szünetek

„A menüben minden szükséges információ elhangzik.”

prozódia nélkül



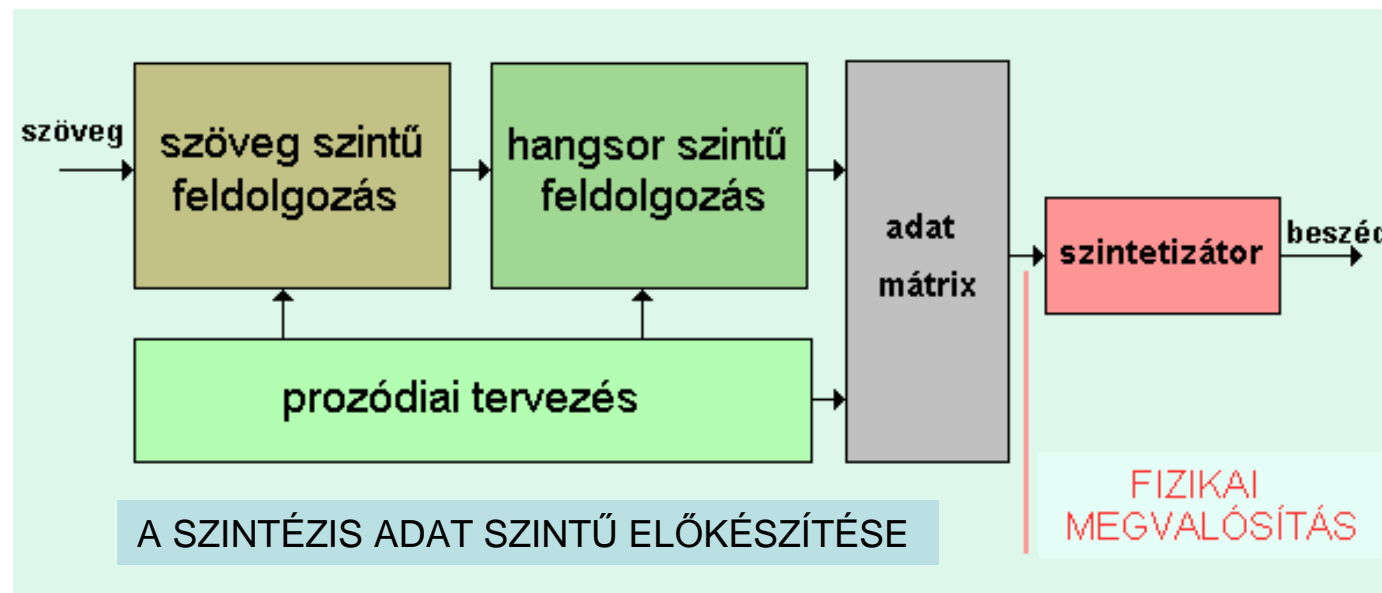
prozódiával



- Mivel a dallam a meghatározó, a továbbiakban ezzel foglalkozunk

Elméleti háttér (2)

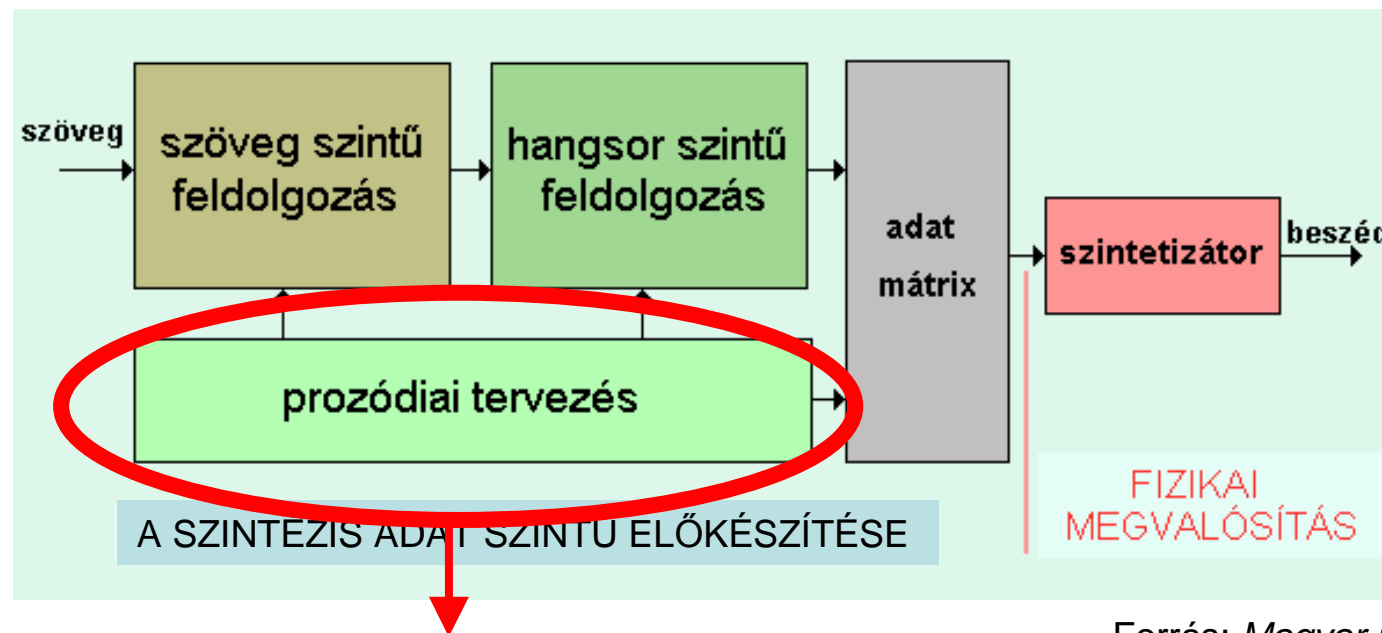
- Beszédszintézis: szöveg → beszéd



Forrás: Magyar nyelvi
beszédtechnológiai
alapismeretek CD

Elméleti háttér (2)

- Beszédszintézis: szöveg → beszéd



Dallammenet módosítása

Forrás: Magyar nyelvi
beszédtechnológiai
alapismeretek CD

Elméleti háttér (3)

- Prozódiai modellek
 - Leíró jellegű (ToBi, GToBi, IViE)
 - Szabály alapú
 - Adatvezérelt (CART-tal, neurális hálóval)
 - Szabály alapú, szuperpozíciós
 - BME TMIT: Profivox, elemösszefűzéses

Szintetizált beszéd természetesebbé tétele

Beszéddallam-adatbázis

Kísérleteinkben

200 mondatos hangkorpusz
(időjárásjelentés)

5000 mondatos hangkorpusz
(időjárásjelentés)

50 mondatos hangkorpusz
(„harangok története”)



BME TMIT

Szintetizált beszéd természetesebbé tétele

Beszéddallam-adatbázis

Kísérleteinkben

200 mondatos hangkorpusz
(időjárásjelentés)

5000 mondatos hangkorpusz
(időjárásjelentés)

50 mondatos hangkorpusz
(„harangok története”)

- Tartalma:
 - Hullámforma fájl
 - Szöveges átírás
 - Fonetikus átírás
 - Hang-, szünethatárok
 - Zöngeperiódus-határok



BME TMIT

Beszéddallam-adatbázis

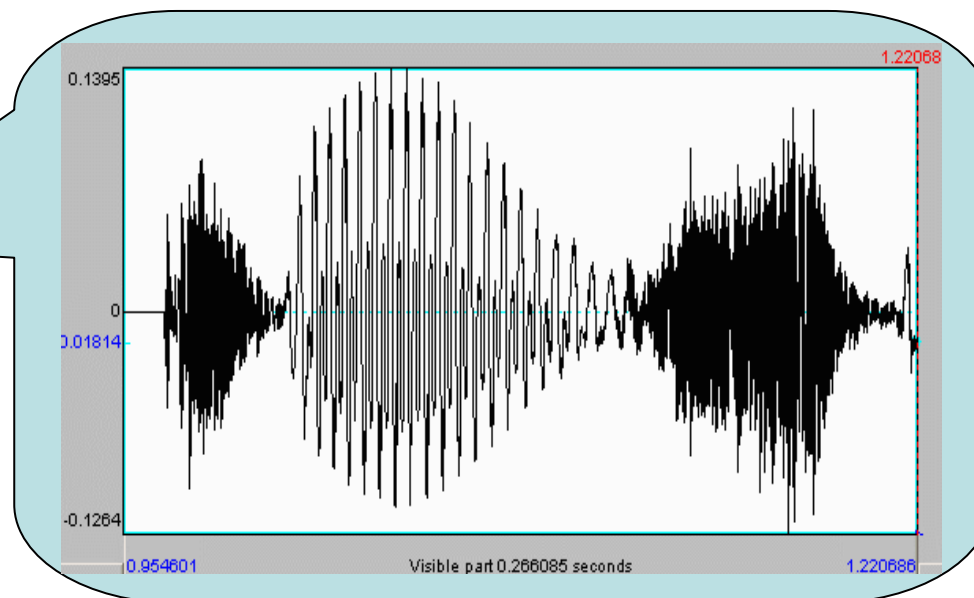
Kísérleteinkben

200 mondatos hangkorpusz
(időjárásjelentés)

5000 mondatos hangkorpusz
(időjárásjelentés)

50 mondatos hangkorpusz
(„harangok története”)

- Tartalma:
 - Hullámforma fájl
 - Szöveges átírás
 - Fonetikus átírás
 - Hang-, szünethatárok
 - Zöngeperiódus-határok



Beszéddallam-adatbázis

Kísérleteinkben

200 mondatos hangkorpusz
(időjárásjelentés)

5000 mondatos hangkorpusz
(időjárásjelentés)

50 mondatos hangkorpusz
(„harangok története”)

- Tartalma:
 - Hullámforma fájl
 - Szöveges átírás
 - Fonetikus átírás
 - Hang-, szünethatárok
 - Zöngeperiódus-határok

„Többfelé lesznek
csúszós útszakaszok.”

Beszéddallam-adatbázis

Kísérleteinkben

200 mondatos hangkorpusz
(időjárásjelentés)

5000 mondatos hangkorpusz
(időjárásjelentés)

50 mondatos hangkorpusz
(„harangok története”)

- Tartalma:
 - Hullámforma fájl
 - Szöveges átírás
 - Fonetikus átírás
 - Hang-, szünethatárok
 - Zöngeperiódus-határok

„töpfelé lesznek
csúszós úcakaszok”

Beszéddallam-adatbázis

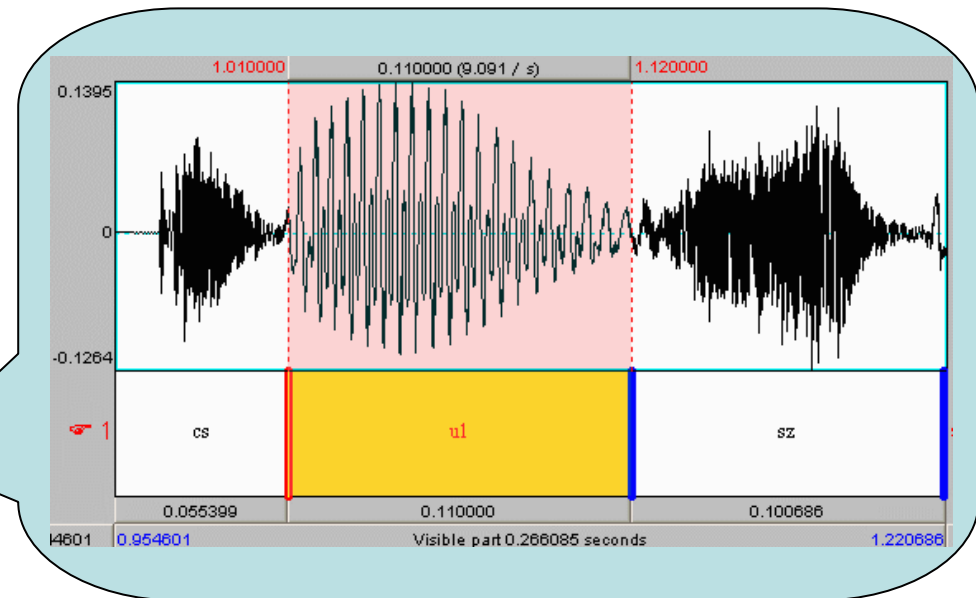
Kísérleteinkben

200 mondatos hangkorpusz
(időjárásjelentés)

5000 mondatos hangkorpusz
(időjárásjelentés)

50 mondatos hangkorpusz
(„harangok története”)

- Tartalma:
 - Hullámforma fájl
 - Szöveges átírás
 - Fonetikus átírás
 - Hang-, szünethatárok
 - Zöngeperiódus-határok



Beszéddallam-adatbázis

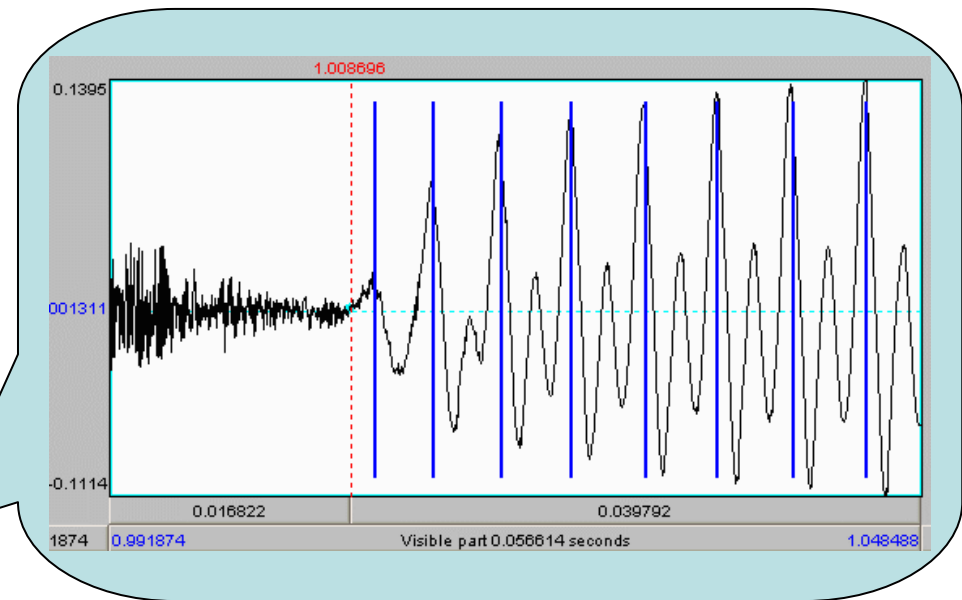
Kísérleteinkben

200 mondatos hangkorpusz
(időjárásjelentés)

5000 mondatos hangkorpusz
(időjárásjelentés)

50 mondatos hangkorpusz
(„harangok története”)

- Tartalma:
 - Hullámforma fájl
 - Szöveges átírás
 - Fonetikus átírás
 - Hang-, szünethatárok
 - Zöngeperiódus-határok



Mondatok vizsgálata - 2006

- Hasonlóságok keresése mondatok között
- Hasonlóság alapja:
 - időtartam
 - szótagszám
 - teljes mondat
 - egyes szavak
- Szótagszám \implies a mondatok dallama feltételezhetően hasonló

mondat azonosító	szótag-szám	szó-szám	szavak szótagszáma
#2203	10	4	6-1-1-2
#2563	10	4	3-4-2-1
#3056	10	3	4-3-3
#2019	11	5	1-1-3-3-3
#2150	11	4	4-1-5-1
#2565	11	4	3-2-2-4
#3024	11	4	4-1-3-3
#2031	12	4	3-4-2-3
#2380	12	5	3-4-1-2-2
#3031	12	7	3-1-1-1-3-1-2

Mondatok vizsgálata - 2006

- Hasonlóságok keresése mondatok között
- Hasonlóság alapja:
 - időtartam
 - szótagszám
 - teljes mondat
 - egyes szavak
- Szótagszám \implies a mondatok dallama feltételezhetően hasonló

mondat azonosító	szótag-szám	szó-szám	szavak szótagszáma
#2203	10	4	6-1-1-2
#2563	10	4	3-4-2-1
#3056	10	3	4-3-3
#2019	11	5	1-1-3-3-3
#2150	11	4	4-1-5-1
#2565	11	4	3-2-2-4
#3024	11	4	4-1-3-3
#2031	12	4	3-4-2-3
#2380	12	5	3-4-1-2-2
#3031	12	7	3-1-1-1-3-1-2

Mondatok vizsgálata - 2006

- Hasonlóságok keresése mondatok között
- Hasonlóság alapja:
 - időtartam
 - szótagszám
 - teljes mondat
 - egyes szavak
- Szótagszám \implies a mondatok dallama feltételezhetően hasonló

mondat azonosító	szótag-szám	szó-szám	szavak szótagszáma
#2203	10	4	6-1-1-2
#2563	10	4	3-4-2-1
#3056	10	3	4-3-3
#2019	11	5	1-1-3-3-3
#2150	11	4	4-1-5-1
#2565	11	4	3-2-2-4
#3024	11	4	4-1-3-3
#2031	12	4	3-4-2-3
#2380	12	5	3-4-1-2-2
#3031	12	7	3-1-1-1-3-1-2

Mondatok vizsgálata - 2006

- Hasonlóságok keresése mondatok között
- Hasonlóság alapja:
 - időtartam
 - szótagszám
 - teljes mondat
 - egyes szavak
- Szótagszám \implies a mondatok dallama feltételezhetően hasonló

mondat azonosító	szótag-szám	szó-szám	szavak szótagszáma
#2203	10	4	6-1-1-2
#2563	10	4	3-4-2-1
#3056	10	3	4-3-3
#2019	11	5	1-1-3-3-3
#2150	11	4	4-1-5-1
#2565	11	4	3-2-2-4
#3024	11	4	4-1-3-3
#2031	12	4	3-4-2-3
#2380	12	5	3-4-1-2-2
#3031	12	7	3-1-1-1-3-1-2

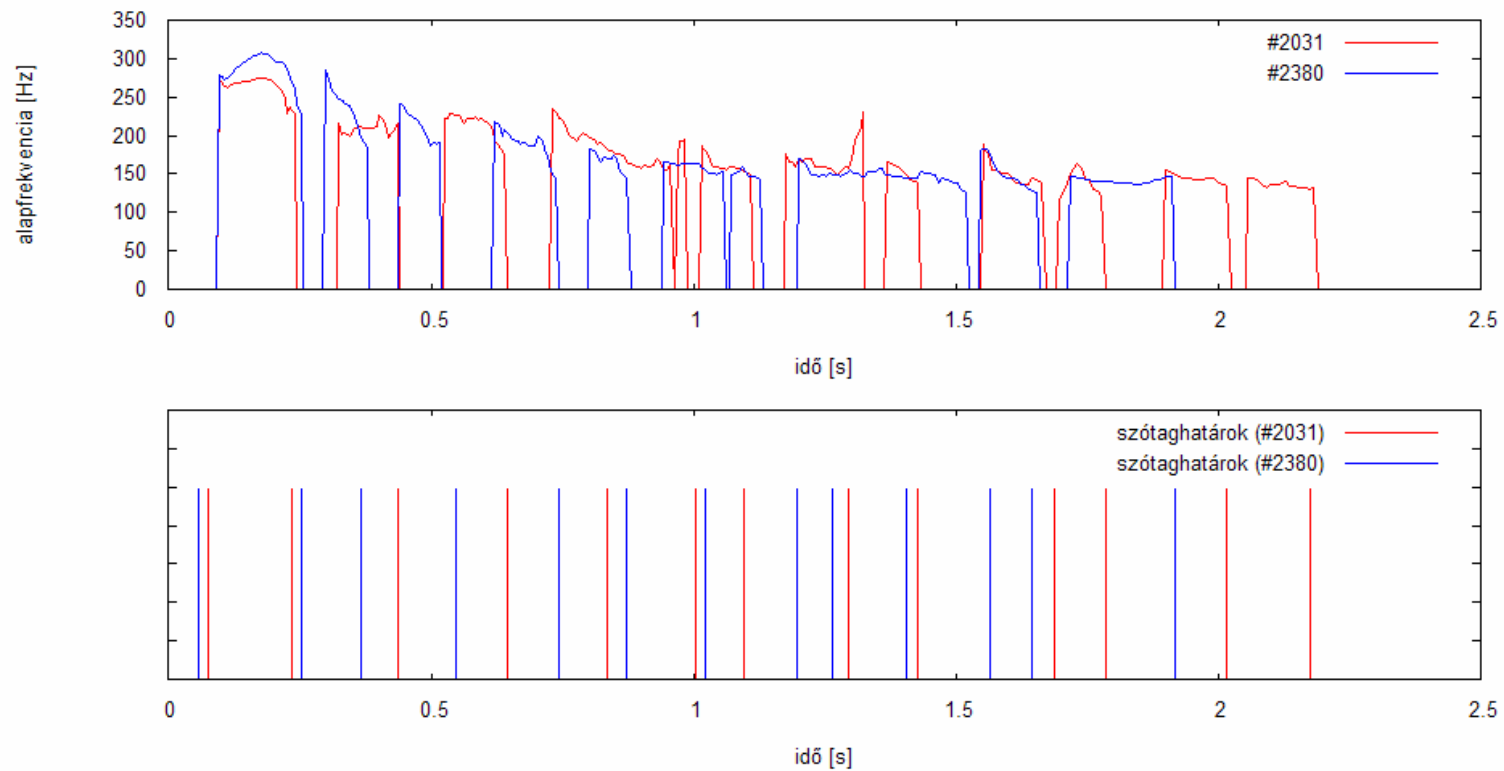
Mondatok vizsgálata - 2007

- Hasonlóság keresése nagyobb adatbázisban: 5000 mondat
- Hasonlóság alapja:
 - egyező szótagszerkezet

mondat azonosító	szótag-szám	szó-szám	szavak szótagszáma
#3053	7	3	3-1-3
#3614	7	3	3-1-3
#3855	7	3	3-1-3
#3056	10	3	4-3-3
#3373	10	3	4-3-3
#1773	11	4	3-2-2-4
#2565	11	4	3-2-2-4
#3517	17	6	3-4-1-4-2
#3953	17	6	3-4-1-4-2
#2551	17	6	1-5-3-5-2-1
#3966	17	6	1-5-3-5-2-1

Mondatok vizsgálata ⁽²⁾ – 2006

- Dallammenetek összehasonlítása



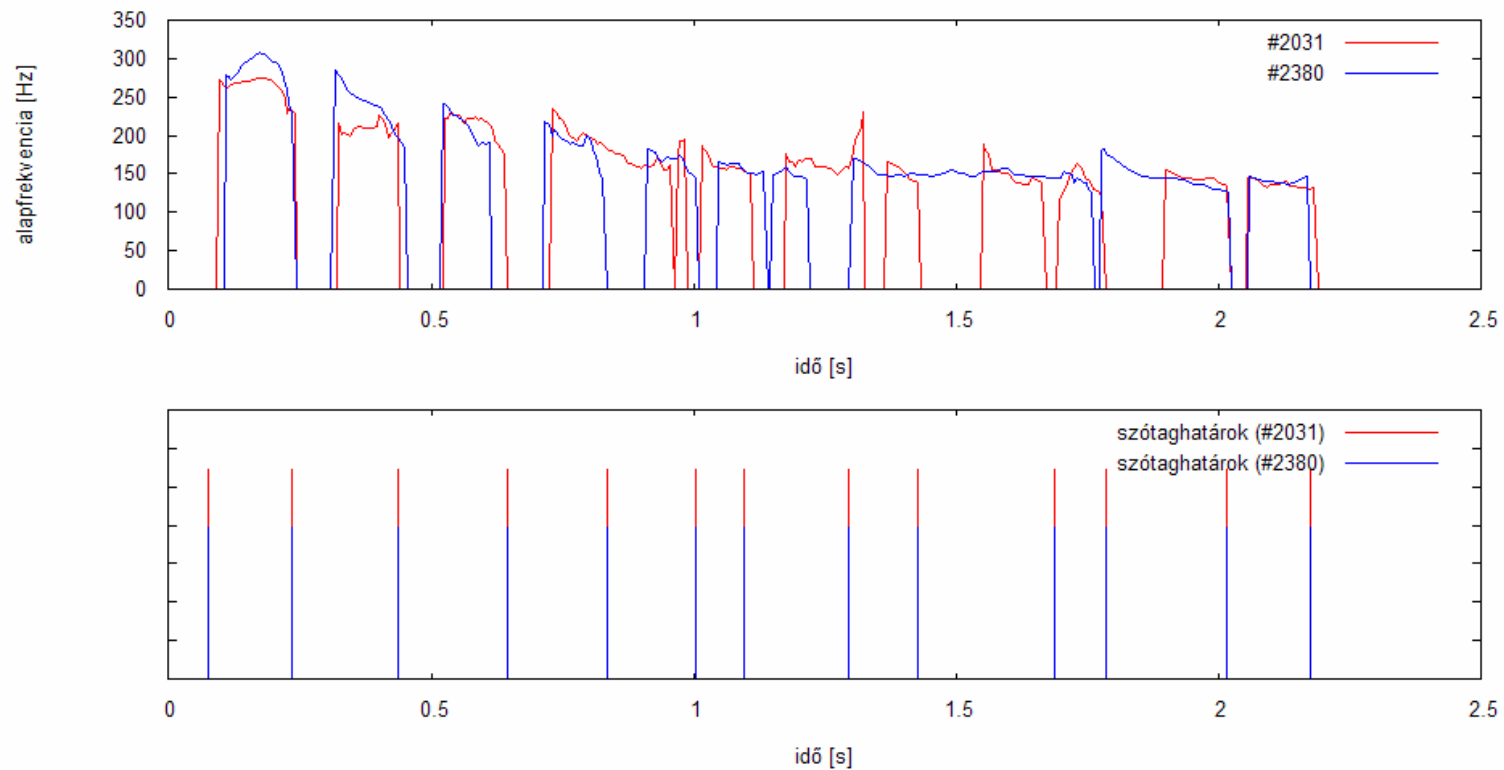
Mondatok vizsgálata (2) – 2006

- Dallammenetek összehasonlítása



Mondatok vizsgálata (2) – 2006

- Dallammenetek összehasonlítása idővetemítéssel



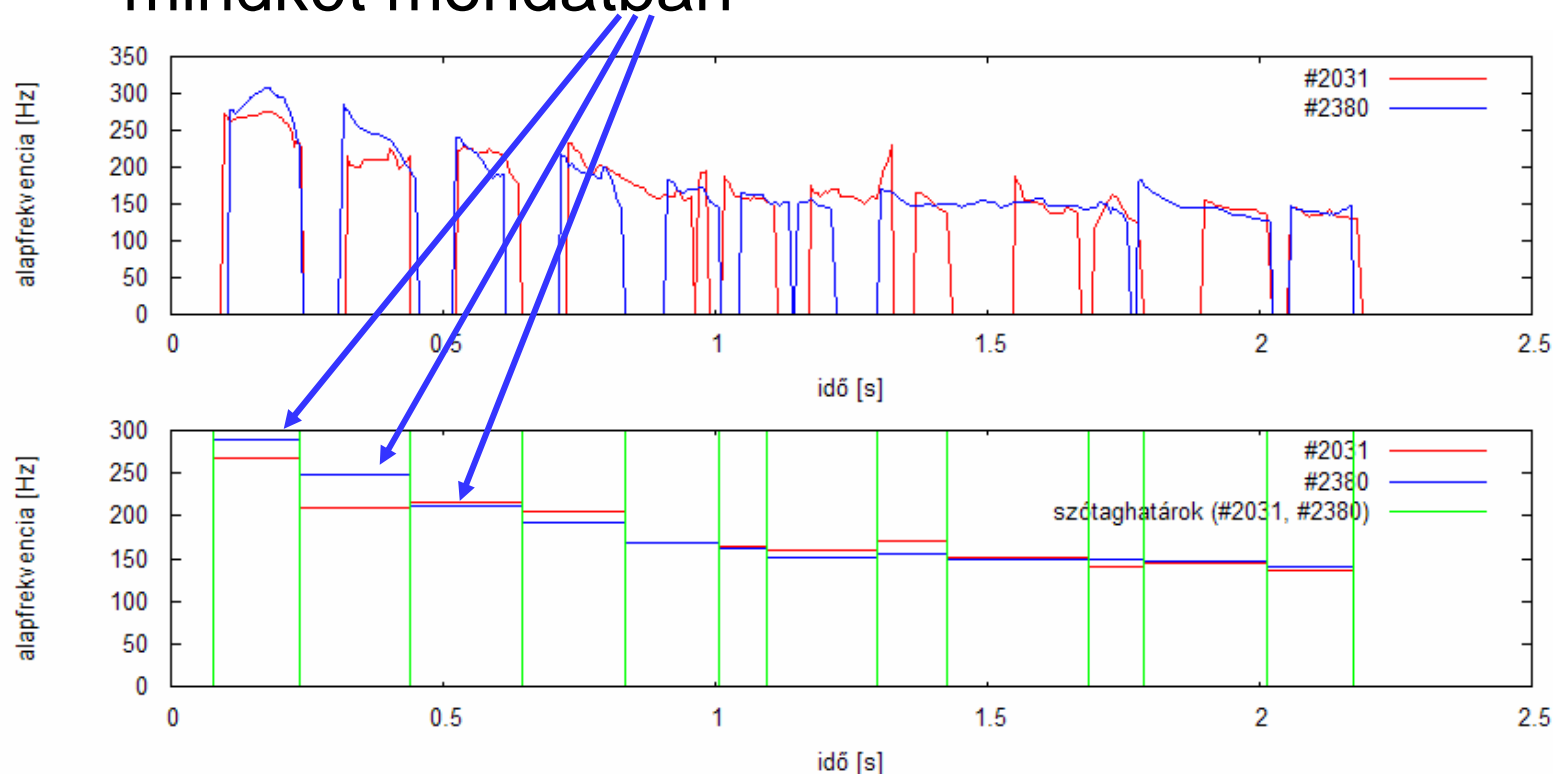
Mondatok vizsgálata ⁽²⁾ – 2006

- Dallammenetek összehasonlítása idővetemítéssel



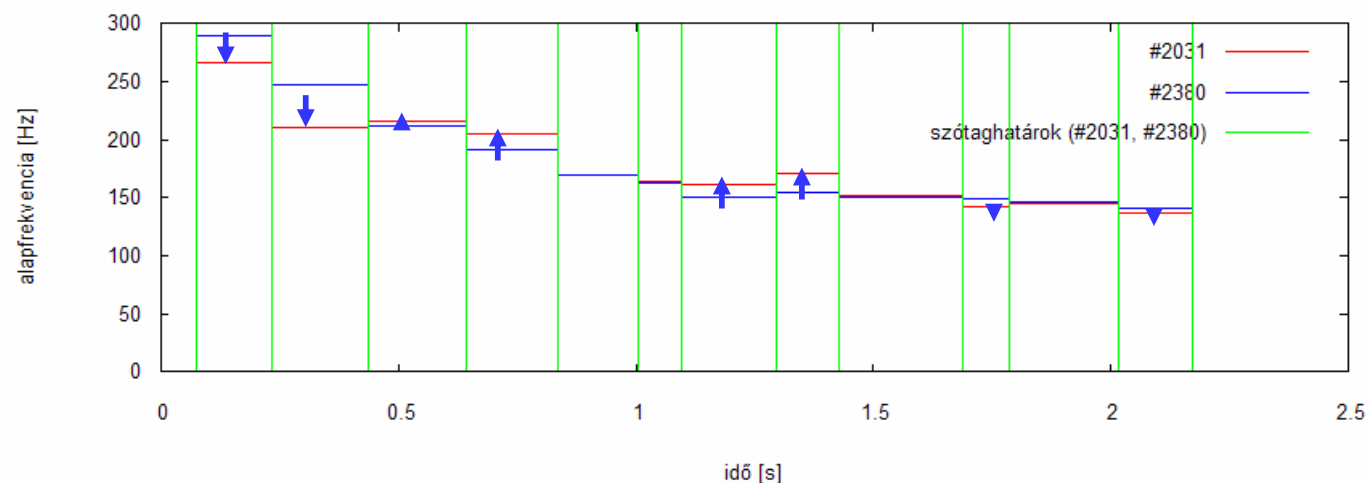
Dallamcsere - 2006

- Természetes bemondások között: Praat [www.praat.org]
 - Szótagonkénti átlagos alaphfrekvencia kiszámítása mindkét mondatban



Dallamcsere - 2006

- Természetes bemondások között: Praat [www.praat.org]
 - Szótagonkénti átlagos alapfrekvencia kiszámítása mindkét mondatban
 - Ez alapján dallam emelés / csökkentés szótagonként



Dallamcsere - 2006

- Természetes bemondások között: Praat [www.praat.org]
 - Szótagonkénti átlagos alapfrekvencia kiszámítása mindkét mondatban
 - Ez alapján dallam emelés / csökkentés szótagonként

#2031: „*Elszórtan számíthatunk kisebb esőre.*”

#2380: „*Péntekig folytatódik a meleg idő.*”

eredeti
(#2031)



referencia
(#2380)



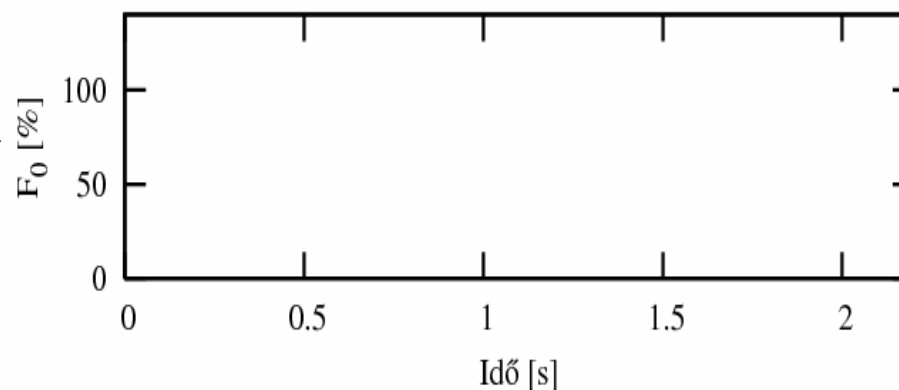
módosított
(#2031)



Dallamcsere - 2006, 2007

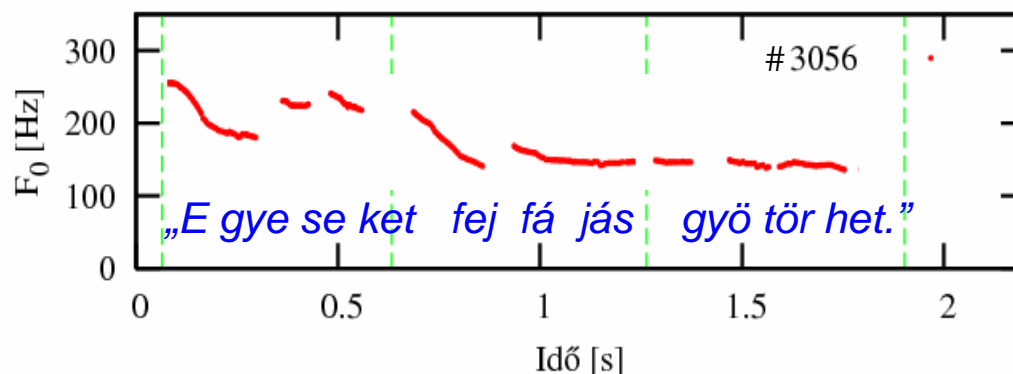
- Szintetizált mondatok között: szabály alapú szövegfelolvasóval (Profivox)

„*Felhősödés estétől várható.*”

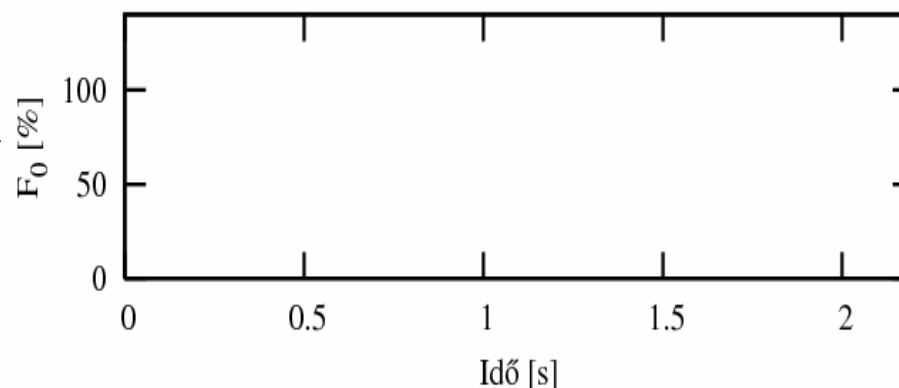
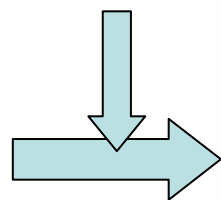


Dallamcsere - 2006, 2007

- Szintetizált mondatok között: szabály alapú szövegfelolvasóval (Profivox)

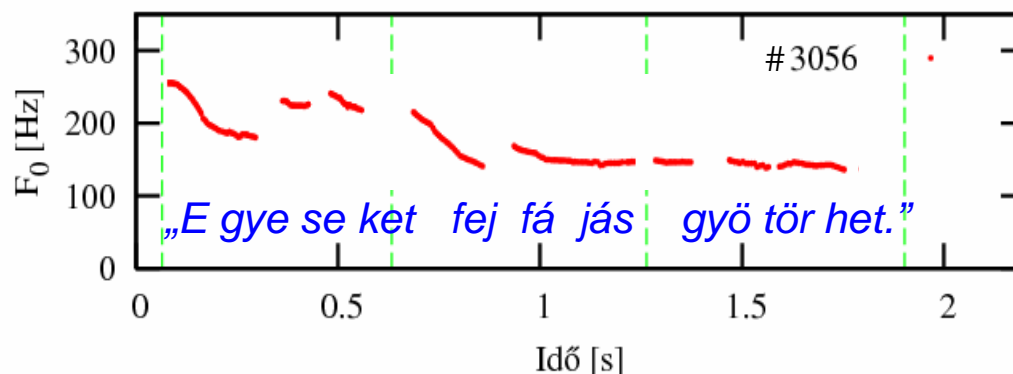


„Felhősödés estétől várható.”

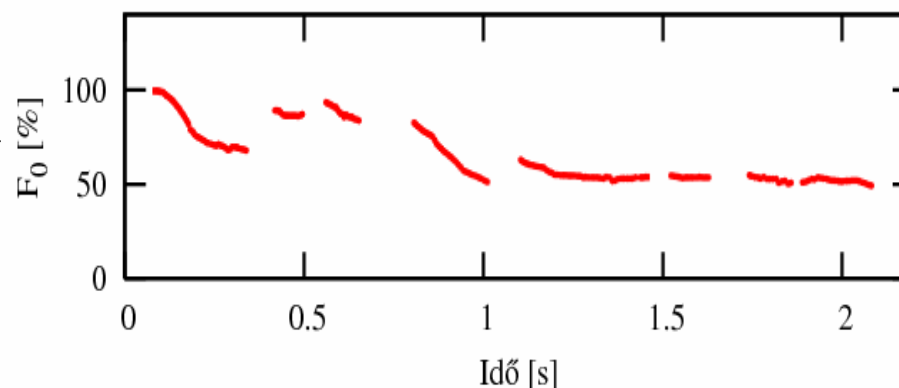
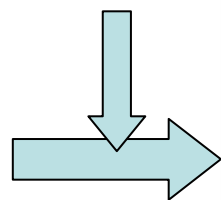


Dallamcsere - 2006, 2007

- Szintetizált mondatok között: szabály alapú szövegfelolvasóval (Profivox)



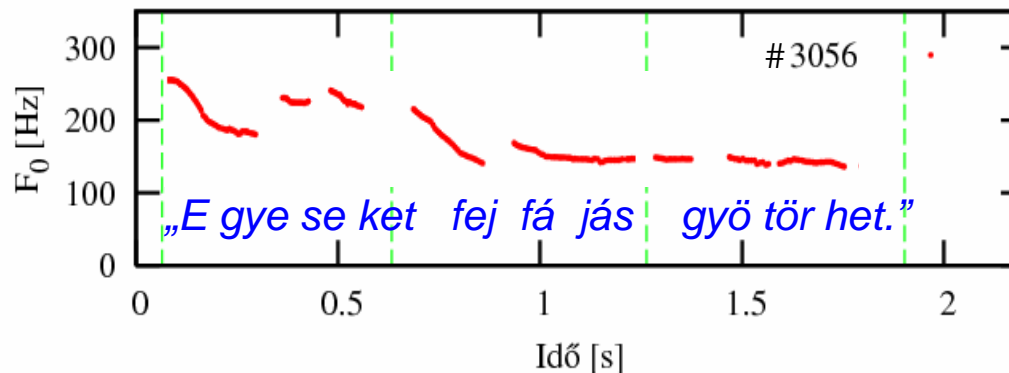
„Felhősödés estétől várható.”



Dallamcsere - 2006, 2007

- Szintetizált mondatok között: szabály alapú szövegfelolvasóval (Profivox)

referencia
(#3056)

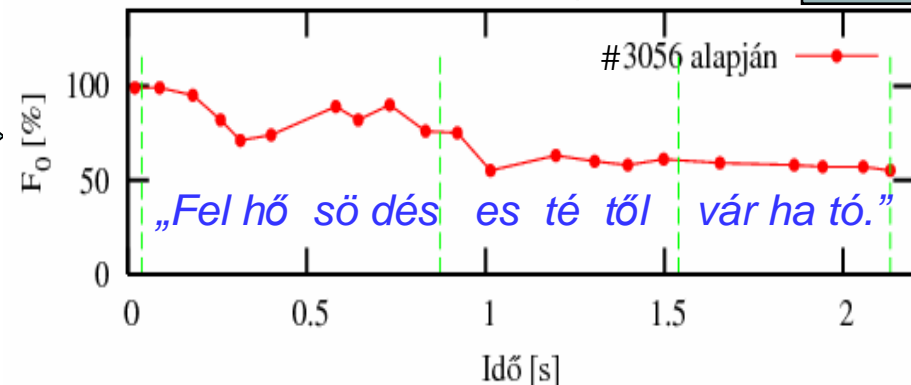
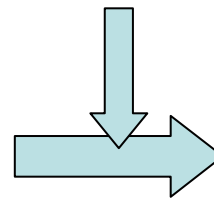


módosított
(#3373)



„Felhősödés estétől várható.”

Profivox szabály alapú
(#3373)

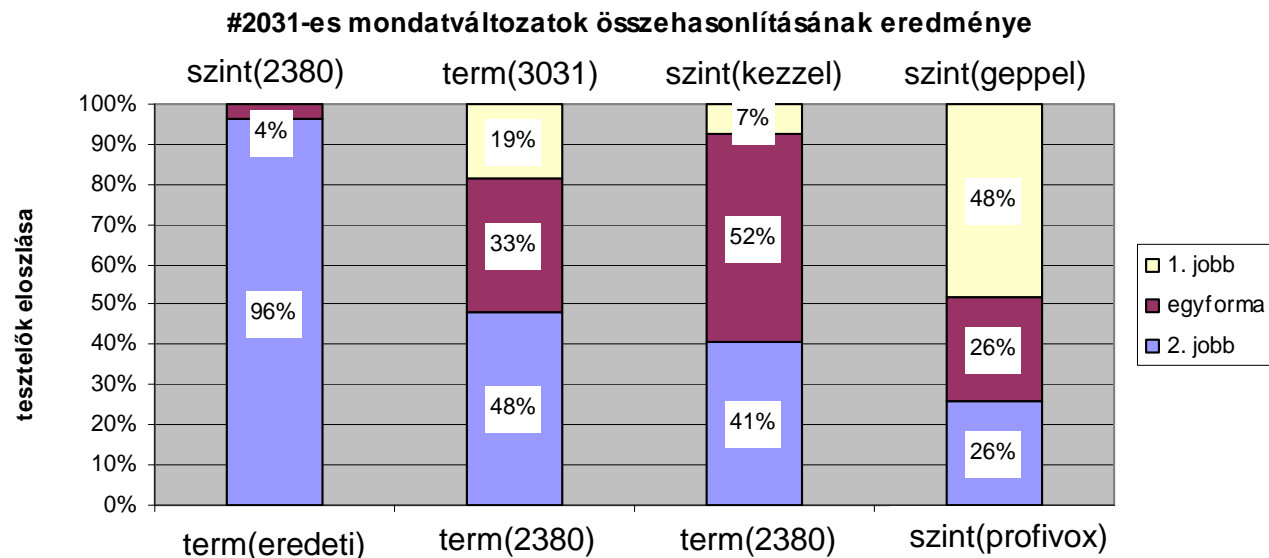


Vizsgálatok - 2006

- Meghallgatásos vizsgálatok weben
- 28 mondatpár, egy-egy mondat több változata:
 - eredeti, eredeti módosítva, szintetizált, szintetizált módosítva
- 26 tesztelő, el kellett dönteniük, hogy melyik változat „természetesebb” (1. jobb / 2. jobb / egyformák)
- Mit akartunk megtudni?
 - Vajon a megváltoztatott prozódia észlelhető-e?
 - Ha igen, elviselhető-e a változás?

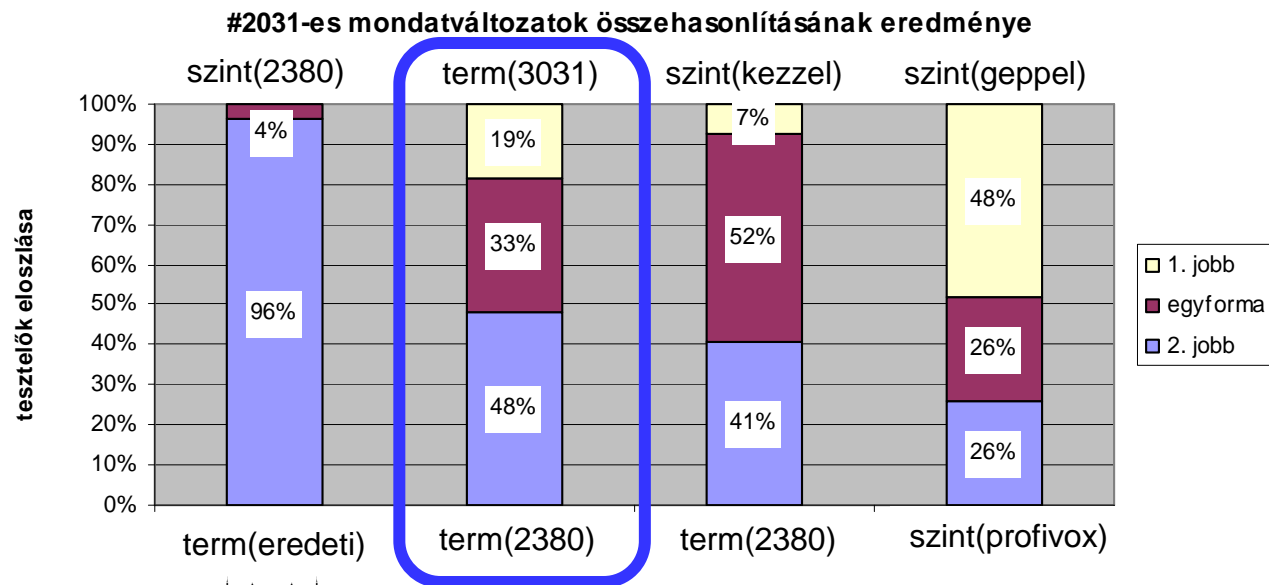
Kísérletek eredménye - 2006

- Az eredeti változat jobb, mint a módosított
 - Különböző referenciával módosított változatok között hallható a különbség, minőségük hasonló
- ➔ sikerült elérni a prozódiai változatosságot



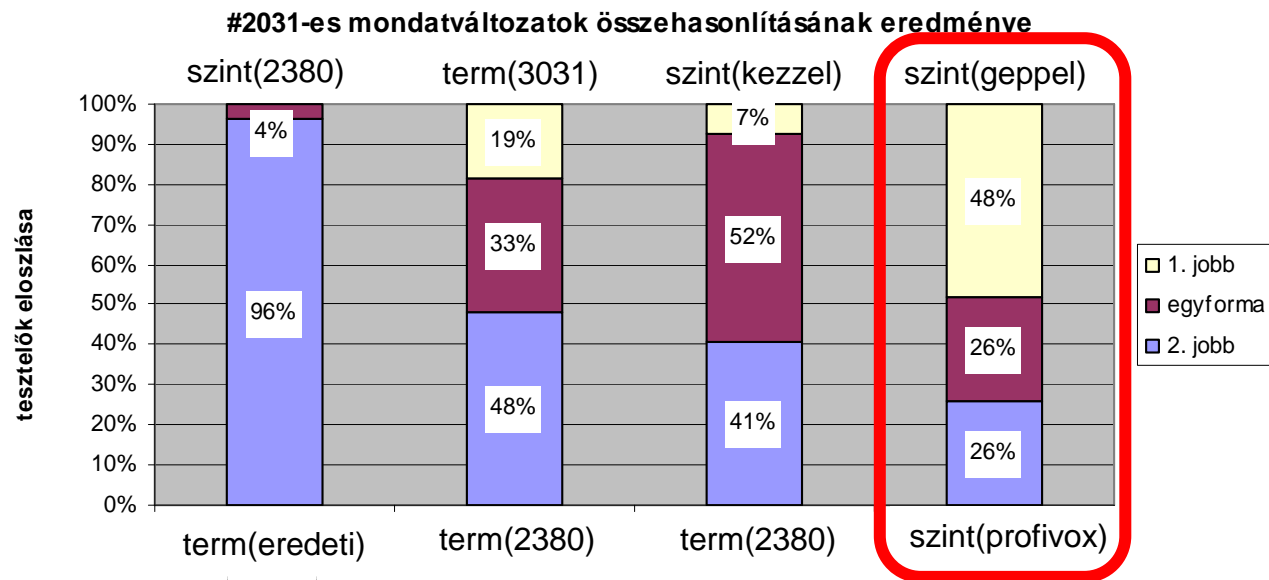
Kísérletek eredménye - 2006

- Az eredeti változat jobb, mint a módosított
 - Különböző referenciával módosított változatok között hallható a különbség, minőségük hasonló
- ➔ sikerült elérni a prozódiai változatosságot



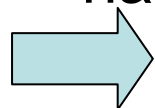
Kísérletek eredménye - 2006

- Az eredeti változat jobb, mint a módosított
 - Különböző referenciával módosított változatok között hallható a különbség, minőségük hasonló
- ➔ sikerült elérni a prozódiai változatosságot

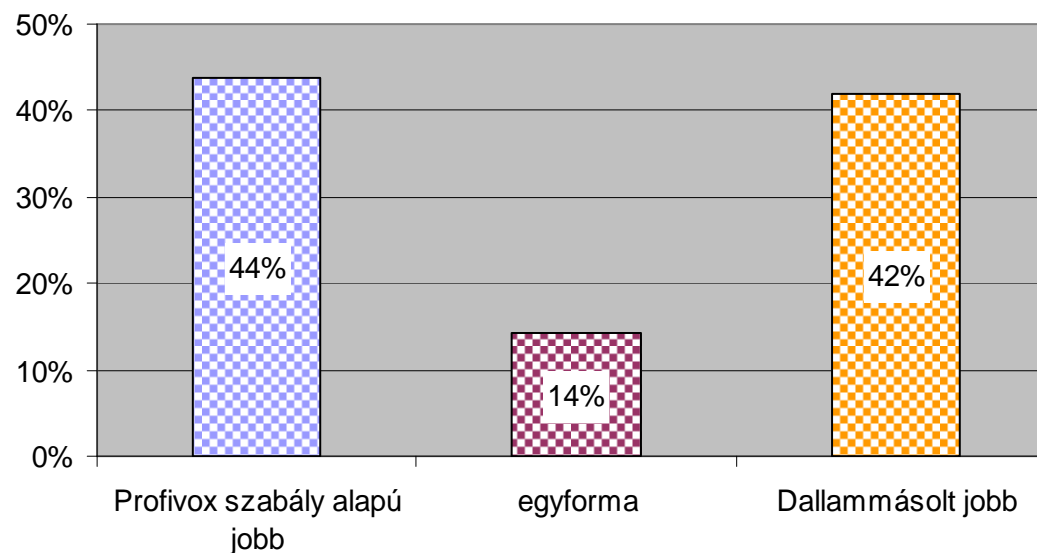


Kísérletek eredménye - 2006

- Az eredeti változat jobb, mint a módosított
- Különböző referenciával módosított változatok között hallható a különbség, minőségük hasonló



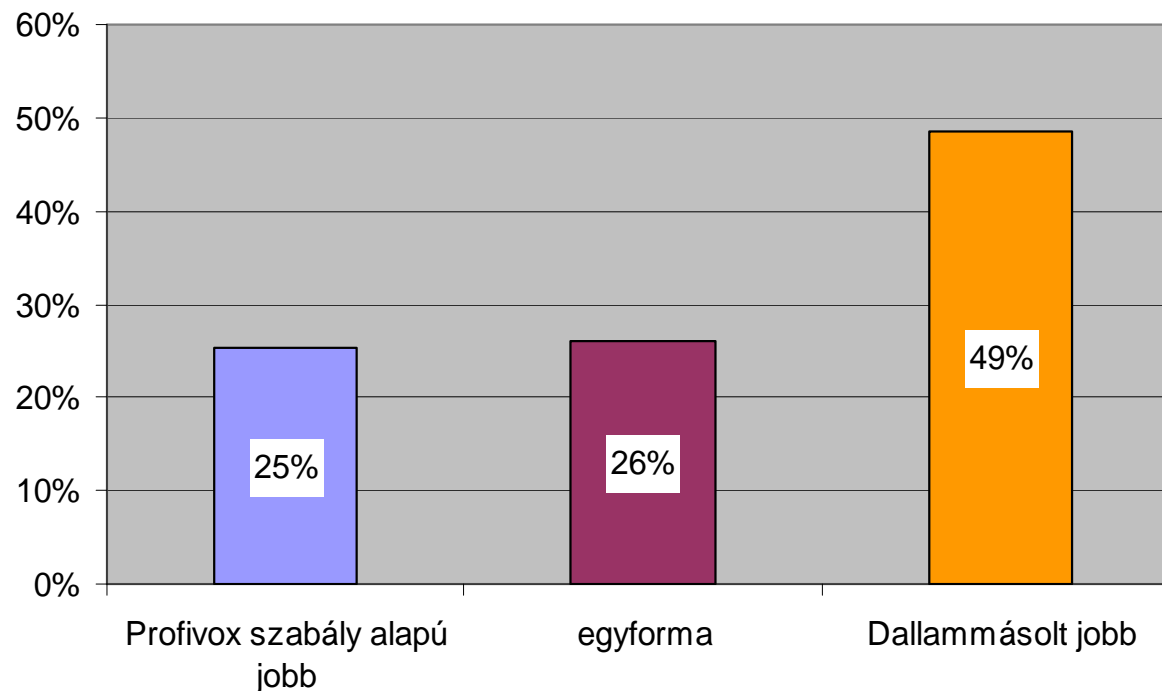
sikerült elérni a prozódiai változatosságot



Vizsgálatok, eredmények - 2007

- Újabb meghallgatásos vizsgálatok weben

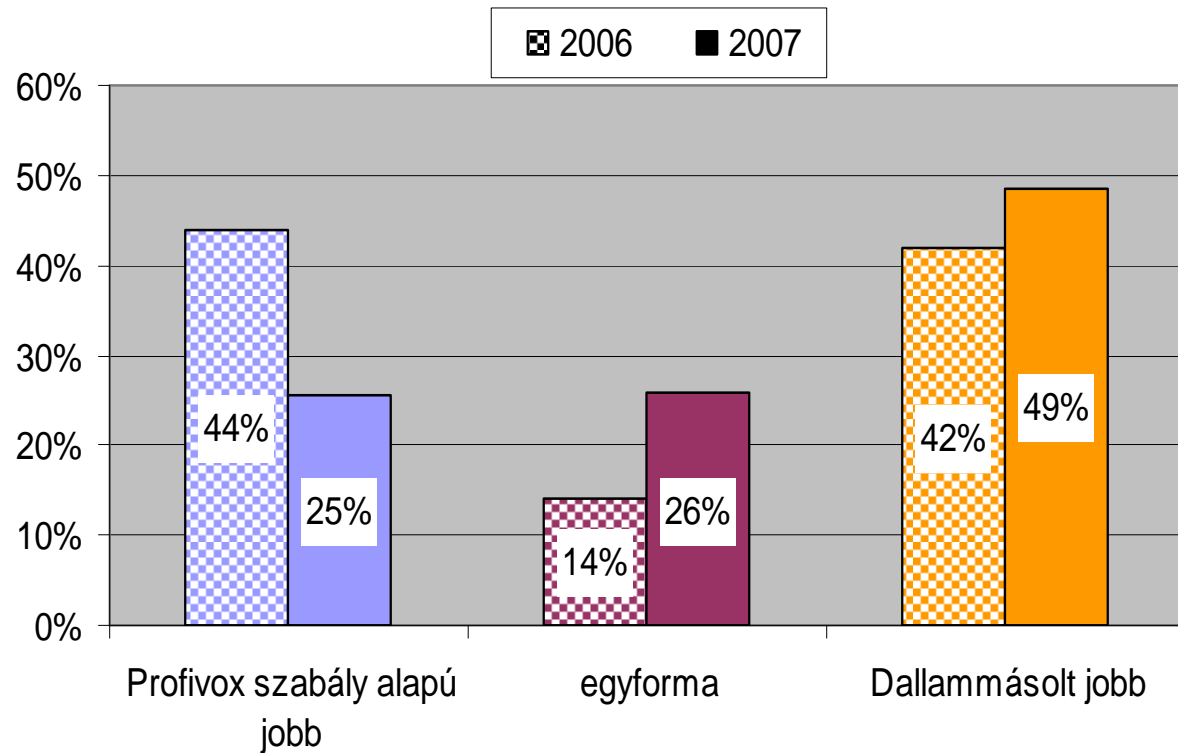
- 48 mondatpár
- összesen 194 tesztelő
- 21 tesztelő / mondatpár



Vizsgálatok, eredmények - 2007

- Újabb meghallgatásos vizsgálatok weben

- 48 mondatpár
- összesen 194 tesztelő
- 21 tesztelő / mondatpár



Eredményeim összegzése

- Szakirodalom áttekintése (összesen 17 tétel, ebből 5 magyar, 12 angol)
 - Adatbázisok megismerése, címkézési munka végzése
 - Mondatdallamok összehasonlítása idővetemítéssel
 - Mondatok dallamának cseréje
 - Meghallgatásos vizsgálat 26 tesztelővel
 - Bebizonyosodott, hogy a prozódiai változatosság megvalósítható
-
- Nagyobb adatbázis, algoritmusok javítása
 - Újabb meghallgatásos vizsgálat 194 tesztelővel
 - Interspeech 2007 konferenciára beadott cikk

Továbbfejlesztési lehetőségek

- „hasonló” mondatok keresése pontosabban
- nagyobb beszéd-adatbázis:
 - NAVA (Nemzeti Audiovizuális Archívum)
 - NDA (Nemzeti Digitális Adattár)
- prozódiai változatosságért felelős modul beépítése a Profivox-ba

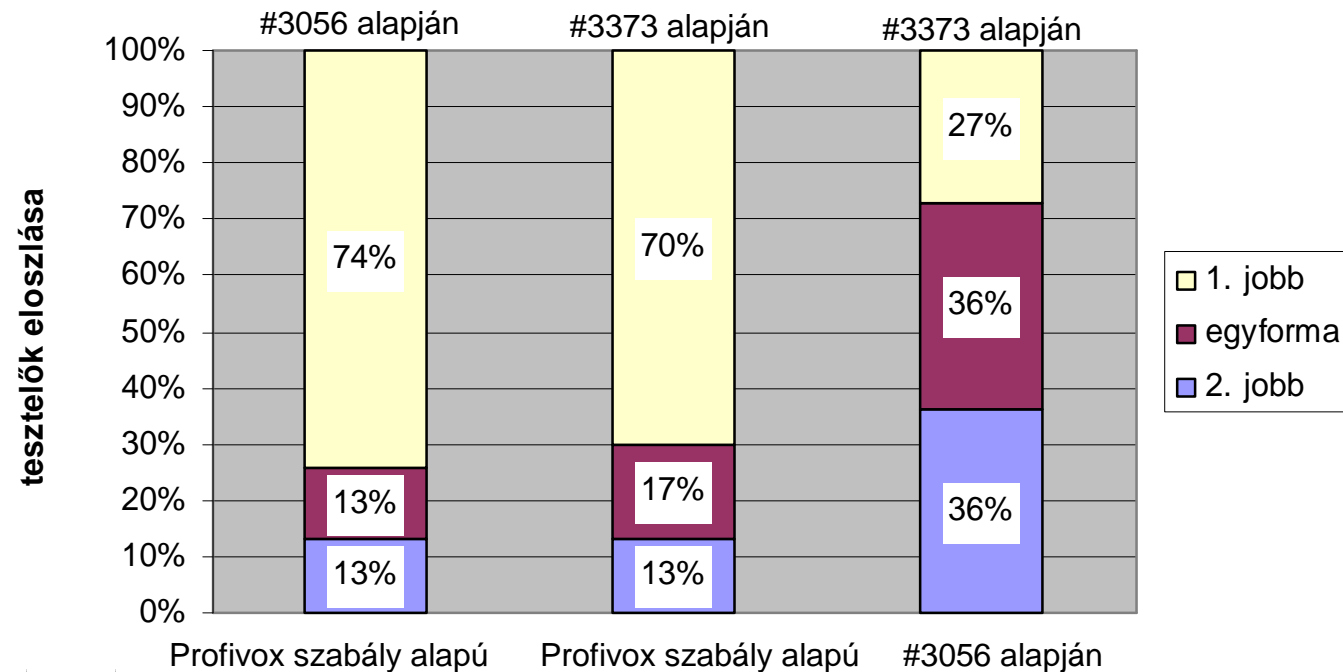
Bírálatra reagálva (1)

- Beszéddallam-adatbázis mérete:
- *„a vizsgálatok nagyon kis méretű adatbázisokon történtek, és nem világos, hogy mekkora nagy adatbázis lenne elégséges a módszer általánosításához”*
 - Cél a koncepció ellenőrzése volt -> kis adatbázis elég
 - Pontosabb mérethez hosszabb kutatás szükséges
 - Tárolás: megoldott (csak dallam- és időzítéscímke, hullámforma nem)
 - Hosszabb szöveg szintetizálása -> elég néhány változatos mondat (pl. ismétlődés helyett)

Bírálatra reagálva (2)

- Egy mondat különböző dallamú változatai:
- „a 7. táblázatból pont a legizgalmasabbnak tűnő kombinációk hiányoznak”

#3056-os mondat változatai



Szintetizált beszéd természetesebbé tétele

Köszönöm a figyelmet!

Csapó Tamás Gábor

csapszi@sch.bme.hu



BME TMIT
