

Irreguláris beszéd regulárissá alakítása beszédkódoláson alapuló módszerrel

Csapó Tamás Gábor, Németh Géza

{csapot,nemeth}@tmit.bme.hu

BME Távközlési- és Médiainformatikai Tanszék



Beszédkutatás,
2013. november 15.

Tartalom

- Irregularis beszéd feldolgozása
- Beszédkódolás
- Új transzformációs eljárás
- Percepciós és akusztikai tesztek

Bevezetés: irreguláris beszéd feldolgozása

- Irreguláris zöngje jellemzői
(Gobl&Chasaide 2013, Markó 2011)
 - Extrém alacsony F0
 - Zöngjeperiódusok amplitúdójának hirtelen változása
 - Beszédtechnológiai problémák: F0 mérési hibák
- Létező módszerek
 - Reguláris-irreguláris osztályozás
(Böhm et al. 2009, Kane et al. 2013, ...)
 - Reguláris beszéd átalakítása irregulárisra
(Böhm et al. 2008)
 - Irreguláris zöngje beszéd szintézisben
(Raitio et al. 2013, Csapó&Németh 2013)

Bevezetés: beszédkódolás

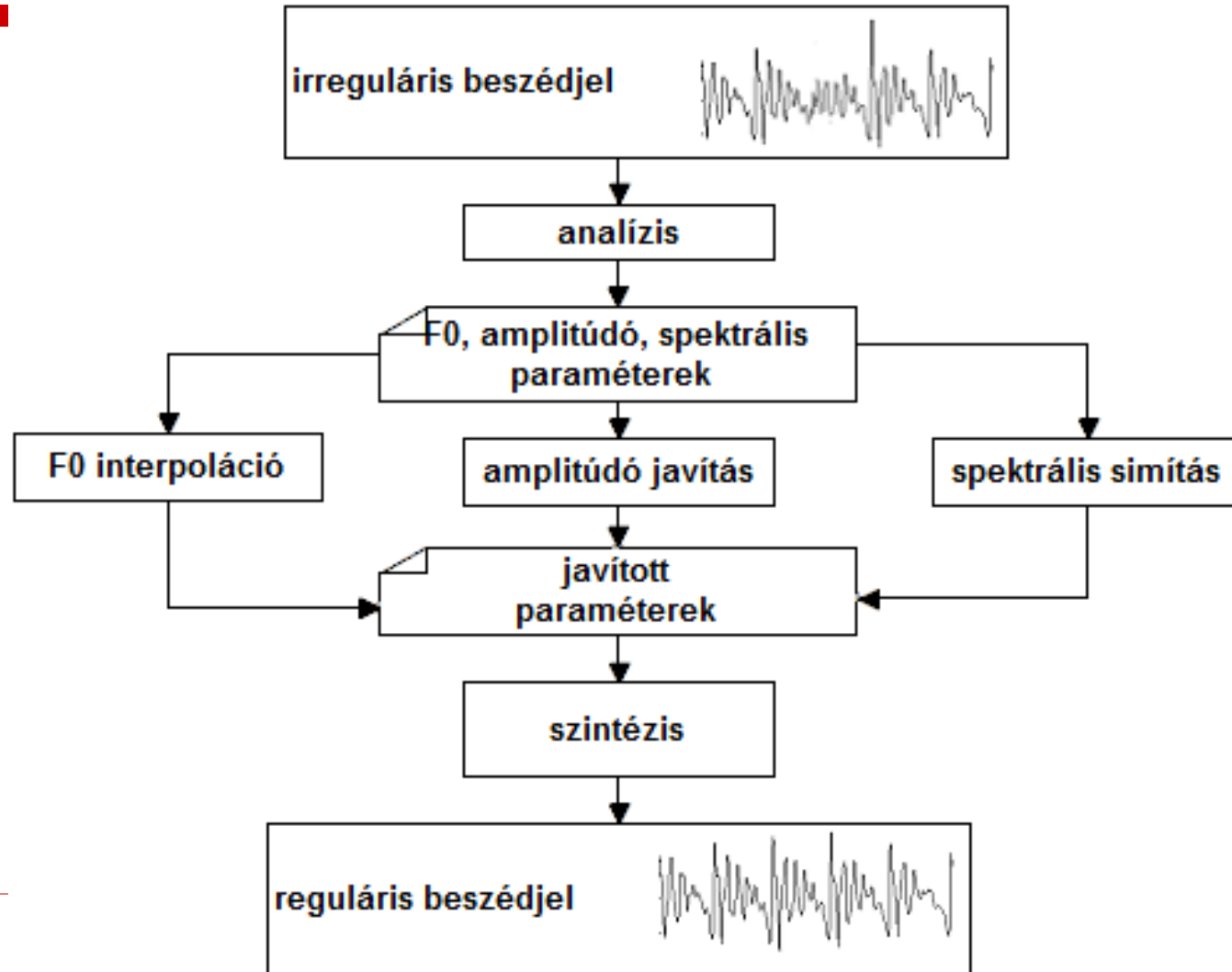
- Eredeti cél (Chu 2003, McCree&Barnwell 1995)
 - Beszéd átvitele távközlési csatornán
 - Minél kisebb sávszélesség
 - Beszédjel felbontása paraméterekre
- Felhasználás transzformációra
 - Paraméterekre bontás
 - Paraméterek módosítása
 - Beszédjel jellemzőinek változtatása (pl. F0 módosítás)

Új irreguláris-reguláris transzformációs eljárás /1

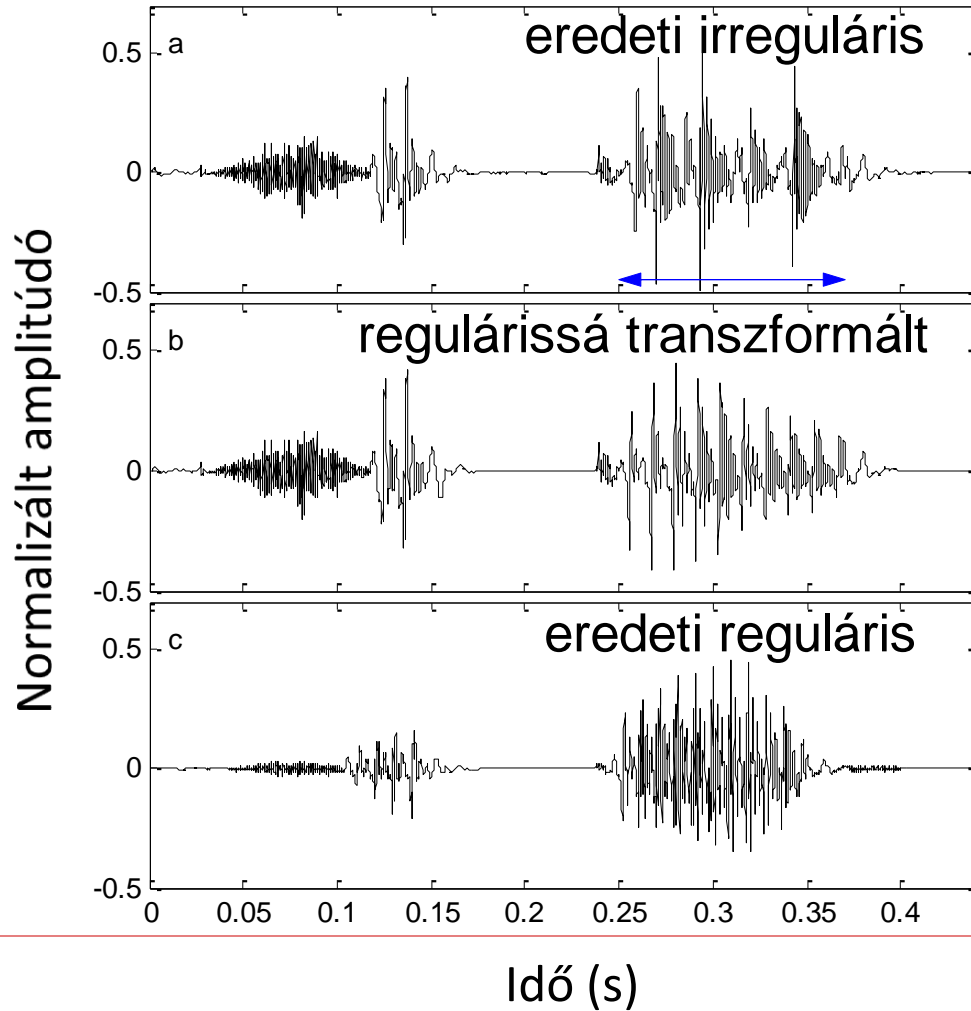
□ Módszer

- Forrás-szűrő szétválasztáson alapul
- Gerjesztés:
Lineáris Predikció alapú maradékjel
- Spektrális paraméter:
Mel-Általánosított Kepsztrum
- Korábbi analízis-szintézis jelfeldolgozás
módszer alapján (Csapó&Németh 2012)

Új irreguláris-reguláris transzformációs eljárás /2



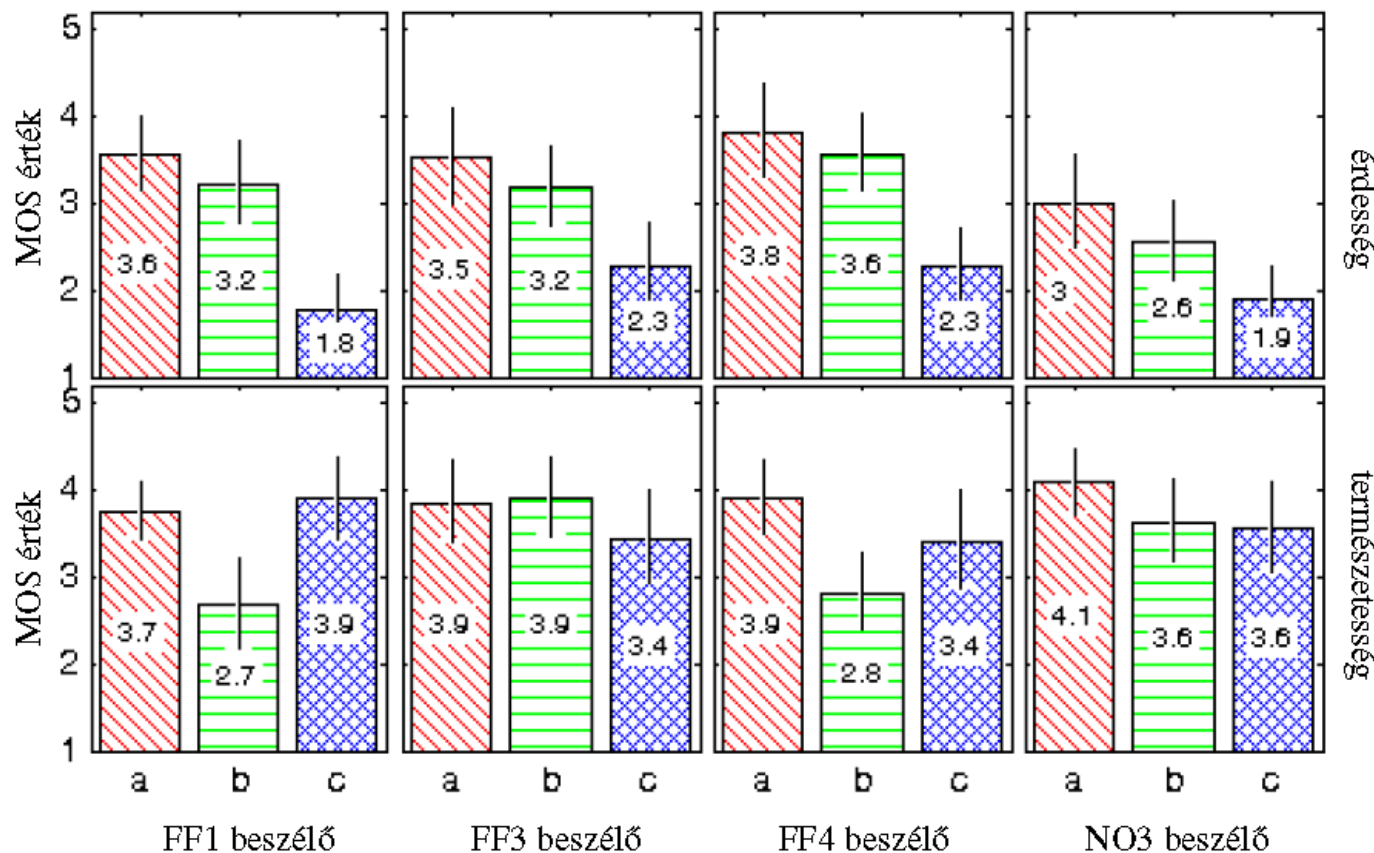
Hangminták



Percepciós teszt /1

- 4 beszélő (3 férfi, 1 nő), 4-4 szó
- Eredeti irreguláris, regulárissá transzformált, eredeti reguláris változat
- Kérdések
 - Mennyire érzed érdesnek (glottalizáltnak) a hallott hangot? 1: egyáltalán nem érdes ... 5: nagyon érdes
 - Mennyire érzed természetesnek a hallott hangot? 1: nagyon természetellenes ... 5: teljesen természetes
- 9 tesztelő (9 férfi; átlag 24 év)

Percepció teszt /2



a: eredeti
irreguláris

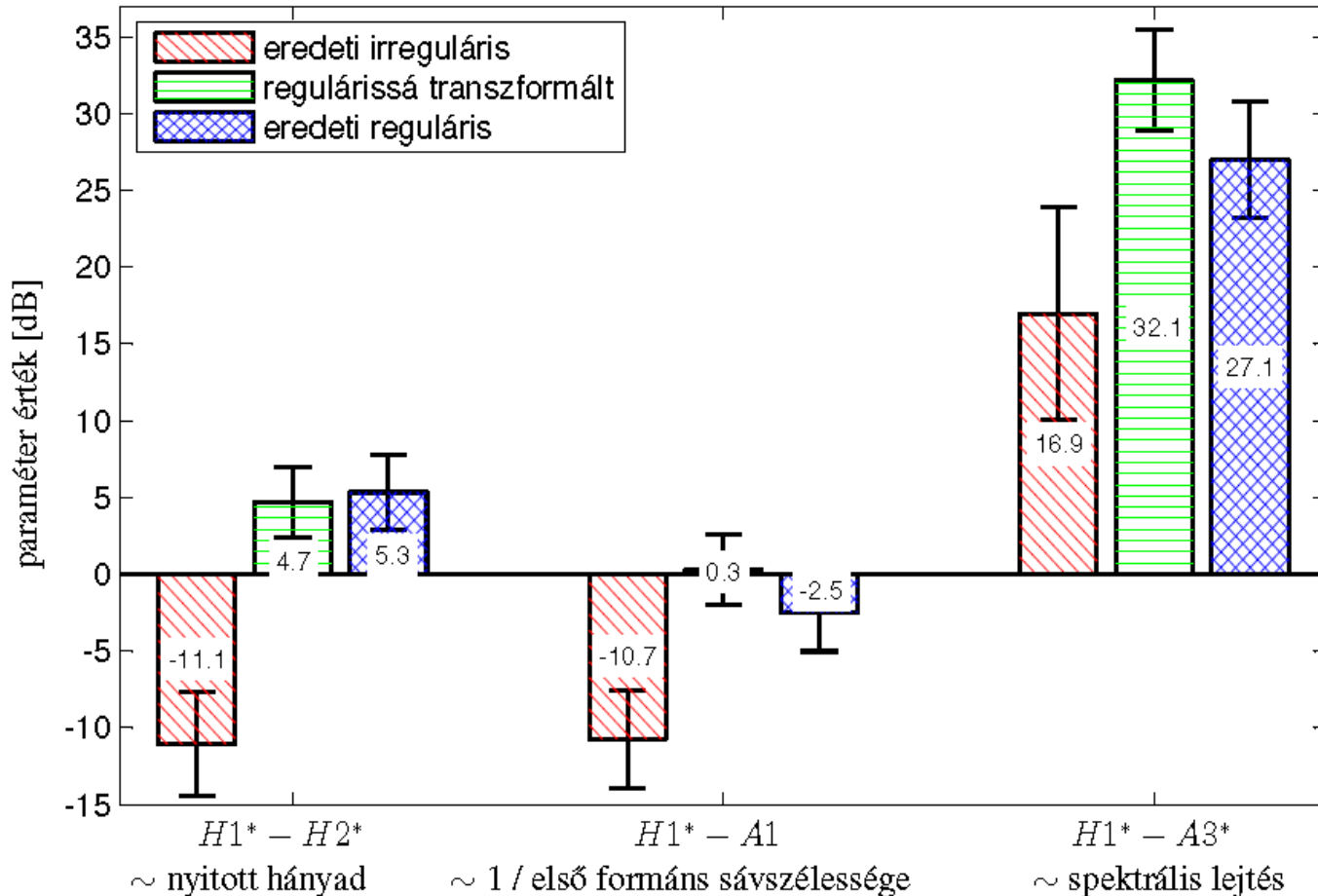
b: regulárissá
transzformált

c: eredeti
reguláris

Akusztikai teszt /1

- 3 releváns akusztikai paraméter
irreguláris és reguláris beszéd
megkülönböztetésére
(Klatt&Klatt 1990, Böhm et al. 2008)
 - Nyitott hányad
 - Első formáns sávzélessége
 - Spektrális lejtés

Akusztikai teszt /2



Csapó, Németh: Irreguláris beszéd regulárisra alakítása
beszédkódoláson alapuló módszerrel

Összefoglalás

- ❑ Beszédkódoláson alapuló módszer irreguláris-reguláris átalakításra
- ❑ Percepciós teszt: érdeesség csökkent, de a természetesség is
- ❑ Akusztikai teszt: reguláris beszéd irányába változás

További kutatás

- Beszédatbázis automatikus javítása, beszédszintézis célok
- Rekedtes, patológikus hangok szebbé tétele
- Más zöngeminőségek transzformációja (pl. levegős beszéd)
- Jobb F0 interpoláció, gépiesség csökkentése

Köszönöm a figyelmet!

Csapó Tamás Gábor, Németh Géza
{csapot,nemeth}@tmit.bme.hu

Hivatkozások

- Bóhm, T., Audibert, N., Shattuck-Hufnagel, S., Németh, G., & Aubergé, V. (2008). Transforming modal voice into irregular voice by amplitude scaling of individual glottal cycles. In *Acoustics'08* (pp. 6141–6146). Paris, France.
- Bóhm, T., Both, Z., & Németh, G. (2009). Automatic Classification of Regular vs. Irregular Phonation Types. In *NOLISP* (pp. 43–50). Vic, Spain.
- Chu, W. C. (2003). *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders*. Hoboken, New Jersey: John Wiley & Sons.
- Csapó, T. G., & Németh, G. (2012). A novel codebook-based excitation model for use in speech synthesis. In *IEEE CogInfoCom* (pp. 661–665). Kosice, Slovakia.
- Csapó, T. G., & Németh, G. (2013). Modeling irregular voice in statistical parametric speech synthesis with residual codebook based excitation. *IEEE Journal on Selected Topics in Signal Processing*, accepted.
- Drugman, T., Kane, J., Raitio, T., & Gobl, C. (2013). Prediction of Creaky Voice from Contextual Factors. In *Proc. ICASSP* (pp. 7967–7971). Vancouver, Canada.
- Kane, J., Drugman, T., & Gobl, C. (2013). Improved automatic detection of creak. *Computer Speech & Language*, 27(4), 1028–1047.
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, 87(2), 820–857.
- McCree, A. V., & Barnwell, T. P. (1995). A mixed excitation LPC vocoder model for low bit rate speech coding. *IEEE Transactions on Speech and Audio Processing*, 3(4), 242–250.
- Raitio, T., Kane, J., Drugman, T., & Gobl, C. (2013). HMM-based synthesis of creaky voice. In *Proc. Interspeech* (pp. 2316–2320).