

Egy egyszerű módszer modális beszéd glottalizálttá alakítására

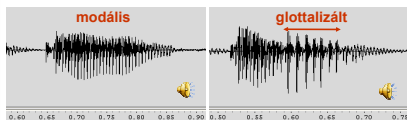
Bóhm Tamás, Németh Géza

Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Médiainformatikai Tanszék
1117 Budapest, Magyar Tudósok krt. 2. bohm@tmit.bme.hu



Glottalizáció

- A zöngékepzés során a hangszalagok nagyjából állandó időközönként összecsapódnak
- Modális zöngé:** a hangszalagregzés reguláris, kváziperiodikus
- Glottalizált zöngé:** a rezgés irreguláris, azaz nagy mértékű hirtelen változások láthatóak a pillanatnyi frekvenciában, amplitúdóban vagy mindkettőben



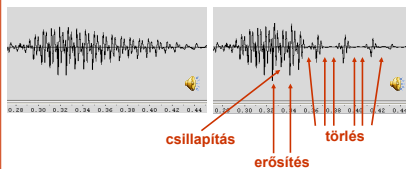
- Jellemző még az alacsony alapfrekvencia és a glottális impulzusok gyors lecsengése
- Érdes, rekedt hangként érzékeljük
- Gyakori jelenség mind egészséges, mind sérült gégefunkciójú beszélők esetén

Glottalizáció gépi előállítása

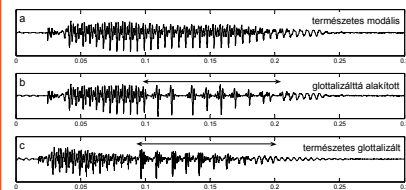
- Információt hordoz:
 - a prozódiai szerkezetéről
Amerikai angolban: prozódiai egység-határok, hangsúlyos szótagok, VV hiátus (Dilley és szerzőtársai, 1996)
 - a beszélő személyéről, nyelvjárásról (Henton, Bladon 1987)
 - a beszélő hangulatáról, érzelmi állapotáról (Gobl, Ni Chasaide 2003)
- Fontos megfelelően kezelni, ha természetes hangzású, érzelmi töltettel rendelkező és szemlélyre szabott beszéd-szintézis rendszerek szeretnénk építeni
- Korábbi eljárások:
 - formáns-szintézis:** több tucat paraméter kézi beállítása szükséges (Pierrehumbert, Frisch 1997)
 - jitter növelése:** csak áttételes hatás, nem természetes hangzás (Verma, Kumar 2005)
- Nem tudunk általánosan alkalmazható módszerről

Módszer

- Kézzel kiválasztott alapperiodusokat törünk, más periódusokat felerősítünk vagy csillapítunk
- Bemenet: beszédjel *pitchmark*okkal (glottális impulzusok időpontjai)
- Első és harmadik feldolgozási lépés PSOLA algoritmus szerint (Moulines, Charpentier 1990)



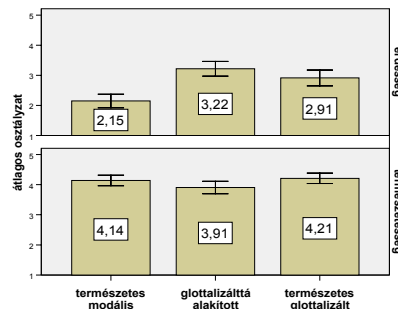
- Analízis:** a beszédjel szétbontása alapperiodusokra
 - Két periódus hosszú „aszimmetrikus” Hanning-ablak
- Manipuláció:** minden periódust beszorunk egy s faktoral
 - $s > 1$: erősítés $s < 1$: csillapítás
 - $s = 0$: törlés $s = 1$: nincs módosítás
 - Periódusok törlésével érdes hangzás érhető el
 - Fokozható több egymás utáni periódus törlésével és az amplitúdók perturbálásával
- Szintézis:** módosított beszédjel előállítása
 - Faktorokkal megszorított periódusok átfedve összeadása (overlap-and-add)
 - Ha minden $s = 1$, akkor visszakapjuk az eredeti jelet (kerekítési hibától eltekintve)



Értékelés

- Érdesesség és természetesség értékelése két külön meghallgatásos teszttel
- Felvételek:** 4 rövid szót 2 bemondó modálisan és glottalizált befejezéssel is felolvasott
- Manipuláció:** modális felvételeket glottalizálttá alakítottuk; természetes glottalizált felvételek mintaként szolgáltak; s faktorok iteratív finomítása
- 12 kísérleti személy (8 adatait elemeztük)

- Érdesesség teszt:** a kísérleti személyek a felvételeket (természetes modális, glottalizálttá alakított, természetes glottalizált) egyesével, véletlenszerű sorrendben hallották és 5 pontos skálán értékelték: 1: egyáltalán nem érdes 5: nagyon érdes
- Természetesség teszt:** érdesesség teszttel azonos 1: nagyon természetellenes 5: nagyon természetes



- Adatelemzés:** egyutas ANOVA, Tukey-féle post hoc próbák 5%-os szinten
- Manipuláció hatása:** szignifikáns növekedés az érdeségben; természetesség csökkenése nem szignifikáns

Összefoglalás

- Félaautomatikus eljárás modális felvételek glottalizálttá alakítására; meghallgatásos értékelés alapján jó minőség érhető el
- Egyszerű, kis számításigényű, könnyen integrálható PSOLA-t alkalmazó beszéd-szintézis-alkalmazás
- További feladatok:
 - Automatizálás
 - Glottalizált \rightarrow modális transzformáció

Köszönetnyilvánítás

- A szerzők szeretnék köszönetüket kifejezni Stefanie Shattuck-Hufnagelnek szakmai útmutatásáért és a Magyar-Amerikai Fulbright Bizottságnak a kísérleti személyek toborzásáért.
- A kutatást részben az NKFP 2. keretprogramja támogatta (szerződésszám: 2/034/2004).

Irodalom

- Dilley, L., Shattuck-Hufnagel, S., Ostendorf, M. (1996): Glottalization of word-initial vowels as a function of prosodic structure. *J. Phonetics* 24, 423–444
- Gobl, C., Ni Chasaide, A. (2003): The role of voice quality in communicating emotion, mood and attitude. *Sp. Comm.* 40, 189–212
- Henton, C.G., Bladon, A. (1987): Creak as a sociophonetic marker. In: Hyman, L.M., Li, C.N. (eds.): *Language, speech and mind*. Routledge, London, 3–29
- Moulines, E., Charpentier, F. (1990): Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Sp. Comm.* 9, 453–467
- Pierrehumbert, J.B., Frisch, S. (1997): Synthesizing Allophonic Glottalization. In: van Santen, J.P.H. et al. (eds.): *Progress in Speech Synthesis*. Springer, New York, 9–26
- Verma, A., Kumar, A. (2005): Introducing roughness in individuality transformation through jitter modeling and modification. *Proc. ICASSP 2005*, 5–8