

Az automatikus irreguláriszöngé-detekció sikeressége az irregularitás mintázatának függvényében magyar (spontán és olvasott) beszédben



Markó Alexandra
Eötvös Loránd Tudományegyetem,
Fonetikai Tanszék
marko.alexandra@btk.elte.hu

Csapó Tamás Gábor
Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformaticai Tanszék
csapot@tmit.bme.hu

Bevezetés Irreguláris zöngé

- hangszalagok irreguláris rezgése
- elnevezés: irreguláris fonáció, glottalizáció, érdes zöngé, recsegő beszéd
- főleg magánhangzó-magánhangzó kapcsolatokban és szakaszhatárokon (pl. mondat vége) fordul elő a magyarban
- spontán beszédben akár a szótagok 47%-ában megvalósulhat
- fonetikai kutatás: irreguláris zöngé funkciói; más beszédparaméterekkel összefüggés
- beszédtechnológiai kutatás: irreguláris zöngé detekciója, transzformációja, szintézise;

Automatikus irreguláriszöngé-detekció

- beszédjelen mért akusztikai paraméterek alapján döntés hozatala
- [Kane2013]: lineáris predikció alapú maradékjelben méri a másodlagos csúcsok előfordulását
- [Drugman2014]: előbbi kiegészítése neurális háló alapú döntéssel + új paraméterek

A kutatás célja

- Mennyire illeszkedik a manuális és automatikus irreguláris zöngé címkézés?
- Milyen tényezők befolyásolják (rontják) az irreguláris zöngé detekciójának sikerességét a gépi és a humán annotációkban?

Módszerek

Hanganyag

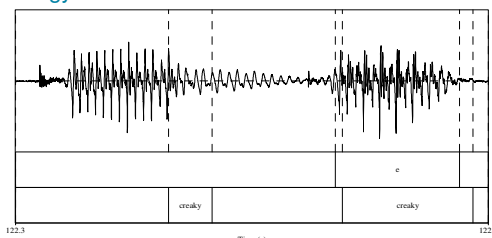
- BEA adatbázis: szövegfelolvasás és interjú; egy nő és egy férfi
 - 5 diszfóniás és 5 kontroll női beszélő mondatfelolvasásai
 - mondatfelolvasások, magánhangzó-magánhangzó kapcsolatok, 2 nő és 2 férfi
- irreguláris zöngével képzett szakaszok manuálisan annotálva

Irreguláriszöngé-detekció

- [Drugman2014] alapján automatikusan
- 10ms-onként irreguláris/reguláris döntés

Pontosság számítása

- referencia: az összes irreguláris zöngé címkével ellátott szakasz (a manuális és automatikus címkék uniója)
- pontosság: hangmintánként kiszámítottuk a referenciához képest a manuális és automatikus címkék számát
- az átfedő manuális / automatikus címkéket egyezőnek vettük

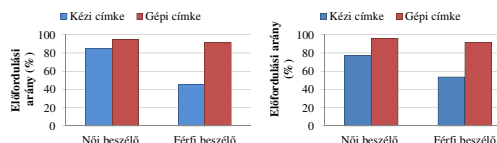


Példa manuális (felső) és automatikus (alsó) irreguláris zöngé címkére

Eredmények /1

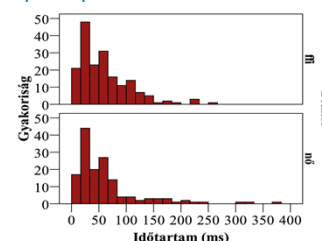
A: BEA adatbázis

- női beszélő: sokat glottalizál (felolvasás: 30%, spontán: 36%)
- férfi beszélő: mély alaphérfrekvencia, ami nehezebb teszi a zöngemínőség manuális címkézését
- cél: minden irreguláris szakasz címkézése



A címkézés pontossága a BEA adatbázis két beszélője esetében: balra az olvasott, jobbra a spontán hanganyag eredményei

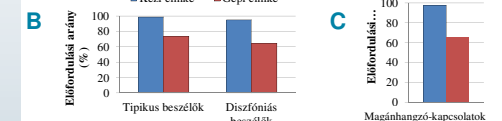
- olvasott és spontán beszéd esetén is pontosabb az automatikus címkézés
- fő ok: a manuális jelölésekre hatással van az irregularitás időtartama: mintegy 50 ms a humán percepció küszöbértéke



A kézi annotációból kimaradt gépi címkék gyakorisági eloszlása az irregularitás időtartama függvényében

- eltérés további oka: az automatikus annotálás a zöngés obstruensek zárszakaszában, illetve a nem teljes zárfelpattanás esetén is irregularitást jelöl -> rövid időtartamú címkék

Eredmények /2



A címkézés pontossága balra a diszfóniásbeszéd-korpuszban (magánhangzók jelölésében), jobbra a magánhangzó-kapcsolatok korpuszban

- cél: magánhangzók és magánhangzó-kapcsolatok irregularitásának címkézése
- mindkét korpuszban pontosabb a manuális címkézés
- fő ok: magánhangzós szókezdetek, magánhangzó-magánhangzó kapcsolatok, gégezárháng gyakran előfordul

Összefoglalás, következtetések

- az irregularitás általános célú annotációjára folyamatos szövegekben az automata algoritmus nagy hatásfokkal alkalmazható (kb. 90%-os pontosság)
- az automatikus annotáció bizonyos beszédhang-típusokat (pl. pergőhang, zöngés obstruensek) szükségtelenül is jelöl
- az automatikus detekció a gégezárhángot – az algoritmus sajátosságaiból adódóan – nem ismeri fel
- az általános irreguláriszöngé-címkézéshez 40-50 ms-os küszöbérték optimális

Főbb hivatkozások

Drugman, T., Kane, J., Gobl, C.: *Data-driven Detection and Analysis of the Patterns of Creaky Voice*. *Computer Speech and Language* 28, pp. 1233–1253, 2014.

Kane, J., Drugman, T., Gobl, C.: *Improved automatic detection of creak*. *Computer Speech and Language* 27, pp. 1028–1047, 2013.