

IRREGULÁRIS BESZÉD REGULÁRISSÁ ALAKÍTÁSA BESZÉDKÓDOLÁSON ALAPULÓ MÓDSZERREL

Csapó Tamás Gábor – Németh Géza

Bevezetés

Az emberi beszédben a zöngés hangok képzésekor a hangszalagok általában közel tökéletesen, kváziperiodikusan rezegnek, azaz az egyes zöngeperiódusok között csak kis változások figyelhetők meg. Ezt reguláris, modális, ideális zöngéképzésnek nevezzük. A természetes beszéd során azonban hosszabb-rövidebb időtartamra irregulárisra válhat a hangszalagok rezgése. Ez eltér a modális zöngéképzéstől, és ezért ezt irreguláris fonációnak, glottalizációnak, érdes zöngének vagy recsegő beszédnek nevezik (Gósy 2004: 30; Markó 2005: 61; Böhm 2009). A jelenség a zöngeperiódusok hosszának és/vagy amplitúdójának hirtelen megváltozásából adódik. Az irreguláris fonáció előfordul egészséges és patológus beszélők esetén is (Redi-Shattuck-Hufnagel 2001), általában szakaszhatárokon (pl. mondat végén) (Markó 2011) vagy magánhangzó-magánhangzó kapcsolatokban (Markó 2012a; Markó 2012b). Gyakran kíséri extrém alacsony alaphangfrekvencia és a glottális pulzusok hirtelen változása (Blomgren et al. 1998). Érzetileg recsegő, rekedtes, érdes jellegű beszédet jelent (Fónagy-Magdics 1967; Gósy 2004). Ha az irreguláris fonáció rövid ideig van jelen a beszédben (néhány periódus), akkor nem is nagyon érzékeljük, tehát azt lehet mondani, hogy az ilyen irreguláris zöngéképzés a legtöbb esetben nem zavarja a hallgatót.

A glottalizáció előfordulása függ a prozódiai szerkezettől (gyakran egybeesik prozódiai határokkal, például szünetekkel (Slifka 2007) és hangsúlyos szótagokkal (Dilley et al. 1996)), valamint információt hordoz a beszélő személyről, nyelvjárásáról, hangulatáról, érzelmi állapotáról és arról, hogy a hangszalagok egészségesek-e (Gobl-Chasaide 2003; Malyska 2008). A glottalizáció akár a beszédhangok 15%-ában is előfordulhat egy-egy beszélő esetén, így egyáltalán nem elhanyagolható jelenség (Böhm et al. 2009). A legtöbb beszédtechnológiai módszert idealizált beszéd feldolgozására készítették. Az irreguláris fonáció problémákat okozhat a beszédanalízis-módszerekben (pl. F0-mérés és spektrális analízis). A fentiek miatt az irreguláris zöngével képzett beszéd megfelelő modellezése hozzájárulhat a pontosabb beszédanalízis-módszerekhez és természetesebb hangzású beszéd szintetizátor rendszerek elkészítéséhez.

Ugyan már léteznek módszerek a glottalizáció jelenségének elemzésére (Markó 2005; Böhm 2009), detektálására és osztályozására (Böhm et al. 2009; Beke-Heltovics 2010; Kane et al. 2013), reguláris-irreguláris transz-

formációjára (Bóhm et al. 2008; Ruinskiy–Lavner 2008), valamint az irregularis zöngével kiegészített beszédszintézisre (Silén et al. 2009; Raitio et al. 2013; Csapó–Németh 2013), de az irregularis zöngéképzés gépi javításával, vagyis az irregularis–reguláris transzformációval eddig keveset foglalkoztak.

A szakirodalomban számos beszédkódoló módszerről olvashatunk, melyeknek eredeti célja a beszéd paramétereire bontása (kódolás, analízis lépés) azért, hogy a távközlési csatornán minél kisebb sávszélesség mellett lehessen átvinni a jelet (beszédet) (Németh–Olaszy 2010: 244). Az átvitel után, a vevő oldalon a paramétereket visszaalakítják beszédjellé (dekódolás, szintézis lépés). A kódolási technikákat három csoportba lehet osztani: hullámforma-kódolás, parametrikus vagy forráskódolás, valamint hibrid kódolás. A hullámforma-kódolás tetszőleges sávkorlátozott jel digitális tárolására alkalmas, és a jel redundanciájának csökkentésével törekszik az alakhűség megtartására. A parametrikus kódolás esetén forrásmodellt (beszédkeltési modellt) alkalmaznak, ami miatt ez a kódolási forma csak beszédjelre alkalmazható. A parametrikus kódolók családjába tartozik az LPC (*Linear Predictive Coding*) kódoló, valamint ennek továbbfejlesztett változatai, melyek az elsődleges cél mellett alkalmasak a beszédjel tulajdonságainak változtatására is (pl. F0-módosítás). A hibrid kódolás az előző kettő előnyeit ötvözi. A jelen kutatáshoz egy újonnan kifejlesztett parametrikus kódolót használunk fel (Csapó–Németh 2012).

A tanulmányban bemutatunk egy új eljárást, amely irregularis beszéd regulárisá alakítására alkalmas egy jelfeldolgozási módszerrel, a beszédkódoló analízis-szintézis felbontáson alapulva. A módszer a forrásszűrő modell felhasználásával az eredeti glottalizált beszédet szétbontja gerjesztési és spektrális komponensekre (analízis), melyek a glottális forrást, illetve az artikulációs csatornát modellezik. Az analízis során kapott paraméterek (pl. F0) értékeit a glottalizált szakaszokon javítjuk, majd a visszaalakításkor előállítjuk a javított beszédet (szintézis). A tanulmány további részében percepció és akusztikai tesztekben vizsgáljuk a módszer eredményességét.

Anyag, módszer, kísérleti személyek

Az irregularis zöngéképzéssel kapcsolatos kísérleteket a PPBA adatbázisból kiválasztott négy magyar anyanyelvű beszélő felolvasott anyagán végeztük (Olaszy 2013). Három férfitől (FF1, FF3 és FF4) és egy nőtől (NO3) származó, professzionális körülmények között rögzített, 44,1 kHz-es mintavételezéssel 16 biten digitalizált hangfelvételt használtunk fel. Az adatbázis beszélőnként közel ugyanazt az 1940 mondatot tartalmazza, amely nagyjából 2 órányi tiszta hangfeltételt jelent. Az adatbázishoz szöveges címkézés, automatikus fonetikus átírat és ennek kézzel javított változata, valamint hang- és szóhatárjelölés is tartozik. A precíziós annotálás eredményeként az adatbázis megbízható, 99,9%-os: a hanghatárok pontossága 10 ms-on belül van és a fonetikai átírat pontosan megfelel a hanghullámnak.

Az irreguláris beszéd regulárisá alakításának módszere

A jelen tanulmányban bemutatott módszer a korábbi Csapó–Németh (2012) analízis-szintézis módszert (mely egy forrásszűrő szétválasztáson alapuló gerjesztési modell) egészíti ki egy olyan jelfeldolgozó eljárással, amely alkalmas a glottalizált beszéd modálissá alakítására, tehát az irreguláris zöngékepzés gépi javítására. Az új eljárás három fő lépésből áll: analízis, transzformáció és szintézis. Az analízis hasonlóan történik, mint a Csapó–Németh (2012) gerjesztési modellben, azaz az eredeti beszédjelből spektrális paramétereket, majd a lineáris predikció alapú inverz szűréssel származtatott maradékjelből gerjesztési paramétereket mérünk. A gerjesztési paraméterek modellezik a glottális forrást, a spektrális paraméterek pedig az artikulációs csatornát. Az analízis után a transzformáció lépésben a paramétereket módosítjuk, majd a gerjesztési modell szintézisével a paraméterekből visszaállítjuk a javított beszédjelet.

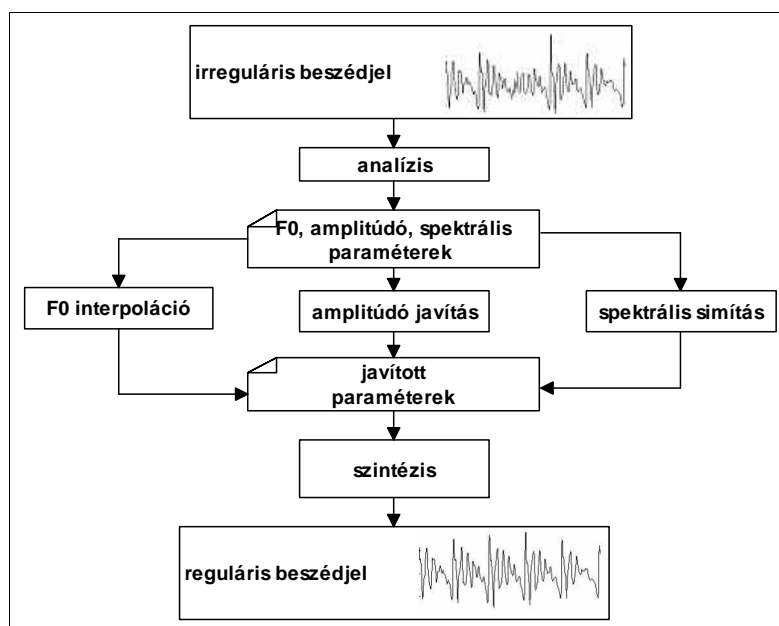
A transzformáció során az eredeti beszédből származtatott maradékjelnek csak azon szakaszait vizsgáljuk, amelyet irreguláris zöngé címkék jeleznek, a modális zöngés és zöngétlen maradékjel részeket változatlanul hagyjuk. A transzformációs eljárás működését az 1. ábra ismerteti.

A gerjesztési modell analízis lépésének eredményeként kapott paraméterek, melyeket 50 ms-os keretekre számolunk:

- **F0**: a keret alapfrekvenciája a Snack RAPT algoritmussal számítva (Talkin 1995),
- **gain**: a keret energiája átlagos négyzetes középként számítva,
- **MGC**: a keret spektrális paraméterei Mel-általánosított kepsztrum (Mel-Generalized Cepstrum) alapon (Tokuda et al. 1994).

Az analízis eredményeként kapott F0-értékeket interpoláljuk, míg a gain és MGC értékeket simítjuk az irreguláris szakaszokon. Az analízis során a glottalizáció megzavarhatja az F0-detekciót: a hirtelen alapfrekvencia- és amplitúdóváltozás miatt előfordulhat, hogy egy egyébként nyelvileg zöngésnek várt keretet (például magánhangzóban) zöngétlennek jelöl a detektor, vagy az eredeti érték felét méri. Ezeket a helyeket megjelöljük, majd a mért F0-menetet interpoláljuk (korrigáljuk az F0-értékeket) azokban a zöngés szakaszokban, ahol az algoritmus nem detektált zöngét. Ezen szakaszokat a beszédmintához tartozó hanghatárjelölés alapján határozzuk meg. Ha egy magánhangzóban egyáltalán nem mért F0-értéket a zöngedetektor, akkor a mondat dallamának megfelelő ereszkedő F0-menetet állítunk be. A kísérletek során minden F0-menetet kézzel ellenőriztünk és javítottunk, emiatt a módszer félautomatikus működésű. Az irreguláris fonáció kis perturbációkat okoz a keretenkénti gain és MGC értékekben az irreguláris zöngéperiódusok amplitúdójának hirtelen változása miatt. Emiatt 5 pontos simítást végeztünk ezeken a paramétereken, amely tapasztalataink szerint megfelelőnek bizonyult a perturbációk eltüntetésére. A szintézis további lépései megegyeznek a Csapó–

Németh (2012) gerjesztési modellben ismertetett lépésekkel, azaz a paramétereknek megfelelő maradékjel-elemeket keresünk a kódkönyvből a célköltség és összefűzési költség felhasználásával, majd ezeket átlapolt összedással összefűzzük. A zöngés és zöngétlen részeket egyesítve az energia megfelelő beállítása után spektrális szűréssel kapjuk meg a transzformáció kimeneti beszédjelét.



1. ábra

Az irreguláris–reguláris transzformáció működése. Az eljárás a Csapó–Németh (2012) maradékjel-kódkönyv alapú gerjesztési modellt alkalmazza

Percepció teszt

A transzformációs eljárás eredményességét internetes percepció (meghallgatásos) kísérlettel vizsgáltuk. A kísérletben a tesztelők az egyes hangminták meghallgatása után 1–5 skálás MOS (*Mean Opinion Score*) jellegű kérdésekre válaszoltak. A teszt elején a kísérleti személyek egy ismertetőt olvashattak a kísérlet témájáról és menetéről, majd néhány adat (nem, kor, eszköz, beszédtechnológiai ismeretek) megadását kértük tőlük.

Az irreguláris–reguláris transzformáció működését a PPBA adatbázis négy beszélőjének (3 férfi: FF1, FF3 és FF4 és egy nő: NO3) hanganyagán teszteltük (Olaszy 2013). Mind a négy beszélő gyakran használ irreguláris fonációt, elsősorban szakaszhatárokon (pl. a mondatok végén). Kiválasztottunk 4-4

szót, amelyek reguláris és irreguláris formában is előfordultak az adatbázisban. Ezután az irreguláris változatot transzformáltuk a fenti módszerrel. Bizonyos esetekben csak egy magánhangzó volt glottalizált, míg más mintákban a teljes zöngés szakaszt irreguláris módon ejtette a beszélő. Azokban az esetekben, amelyben az F0-interpoláció nem volt megfelelő, kézzel javítottuk az F0-menetet, hogy illeszkedjen a mondat ereszkedő dallammenetéhez. A szavak 3-3 változatát (eredeti irreguláris, regulárisra transzformált és eredeti reguláris) meghallgatásos tesztben hasonlítottuk össze. A 3. ábra egy példát mutat a teszt egyik szavának 3-3 változatára.

Az internetes meghallgatásos tesztben összesen 48 szót kellett értékelni (4 beszélő * 4 szó * 3 változat) természetesség és érdekesség szerint. A teszt megkezdése előtt a kísérleti személyeket megkértük, hogy hallgassanak meg néhány glottalizált beszédmintát, hogy egyértelműsítsük az „érdes” kifejezés jelentését. A tesztelők minden minta meghallgatása után két MOS-jellegű kérdésre válaszoltak: 1) „Kérlek, jelöld be, hogy mennyire érzed érdesnek (glottalizáltak) a hallott hangot!” („1 - egyáltalán nem érdes” ... „5 - nagyon érdes”); 2) „Kérlek, jelöld be, hogy mennyire érzed természetesnek a hallott hangot!” („1 - nagyon természetellenes” ... „5 - teljesen természetes”). A szavakat a tesztelők véletlen sorrendben hallgatták meg.

A tesztet összesen 9 magyar anyanyelvű tesztelő végezte el (mind a 9 férfi volt). Közülük hárman beszédkutatók voltak, míg a többiek egyetemi hallgatók. Az átlagos életkor 23,67 év volt (szórás: 3,20 év). Heten fejhallgatóval végezték a kísérletet, ketten hangszórón hallgatták a mintákat. Átlagosan 6,92 percig tartott a teszt kitöltése (szórás: 1,39 perc).

Akusztikai teszt

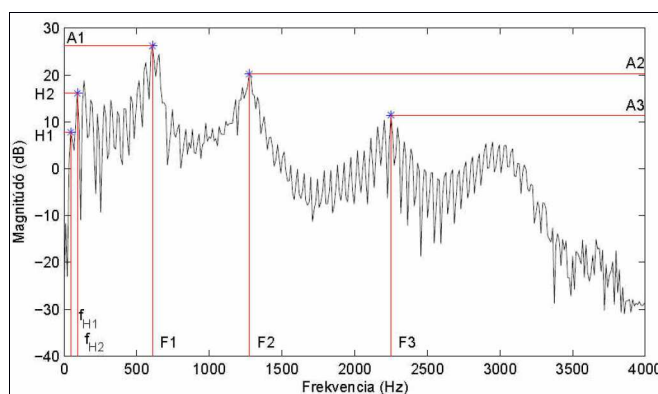
A meghallgatásos teszthez kiválasztott beszédmintákon akusztikai elemzést is végeztünk. A zöngeminőségnek számos akusztikai megfelelője van, melyeket a szakirodalomban következetesen használnak (Blomgren et al. 1998). Ez alapján megvizsgáltuk az eredeti irreguláris, regulárisra transzformált és eredeti reguláris beszédmintákat néhány akusztikai jegy szempontjából.

A szakirodalomból kiválasztottunk három olyan akusztikai jegyet, amelyeket korábban irreguláris és reguláris beszéd megkülönböztetésére használtak (Klatt–Klatt 1990; Böhm et al. 2008). Ezek alapján irreguláris zöngéképzés esetén a hangrés nyitott idejének aránya, vagyis a nyitott hányad (*open quotient*, OQ) alacsonyabb, mint reguláris zöngében. Az első formáns sávszélessége (*first formant bandwidth*, B1) nagyobb a glottalizáció során a gégeben előforduló nagyobb mértékű akusztikai veszteség miatt. Irreguláris zöngéképzés során a hangszalagok záródása összefüggéstelenebb, azaz a spektrális lejtés (*spectral tilt*, TL) meredekebb, mint reguláris beszédben. A transzformáció hatását az OQ, B1, TL akusztikai jellemzőkre mérésekkel vizsgáltuk. A méréseket spektrális tartományban végeztük, mivel így könnyebb a paraméterek számítása (Böhm 2009: 53). Holmberg és munkatársai (1995)

kimutatták, hogy az OQ arányos az első és második harmonikus dB-ben mért különbségével (H1–H2), B1 fordítottan arányos H1 és az első formáns amplitúdójának különbségével (H1–A1), míg a TL korrelál H1 és a harmadik formáns amplitúdójának különbségével (H1–A3).

A H1, H2 és A3 értékeket a formánsok hatása befolyásolhatja, ezért az Iseli–Alwan (2004) által javasolt kompenzációt alkalmaztuk a VoiceSauce program implementációjában. Ez alapján a H1 és H2 értékét az első és második formáns frekvenciája szerint korrigáltuk (H1* és H2*), valamint az A3 értékét az első három formáns frekvenciája szerint kompenzáltuk (A3*). Végül kiszámítottuk az amplitúdó különbségeket: H1*–H2*, H1*–A1 és H1*–A3*.

A paramétereket a meghallgatásos teszt mintáin mértük (eredeti irregularis, reguláris transzformált és eredeti reguláris változatok). A hangfájlokat először 3,8 kHz-es aluláteresztő szűrésen engedték át, majd újramintavételeztük 8 kHz-en, ami biztosította, hogy a spektrumban csak a 3,8 kHz alatti tartomány látható. Ezután az eredeti irregularis változatokból kiválasztottunk egy glottalizált magánhangzót, majd ennek 3-3 pontját jelöltük ki (nagyjából egyenletesen elosztva és a zöngéhatároknak megfelelően). A Wavesurfer programot használva 512 pontos FFT analízist végeztünk a Hanning-ablakozott jelen, majd vizuálisan leolvastuk a H1, H2 amplitúdókat és ezek frekvenciáit, az F1, F2, F3, valamint A1 és A3 értékeket. Az irregularis változatokban gyakran erős alharmonikusok jelentek meg; itt a H1 és H2 értékét a spektrális csúcsok közül az első kettőben mértük. A formánsok értékeit a legnagyobb spektrális csúcs frekvenciájaként és amplitúdójaként mértük. A mérés módszere a 2. ábrán látható.



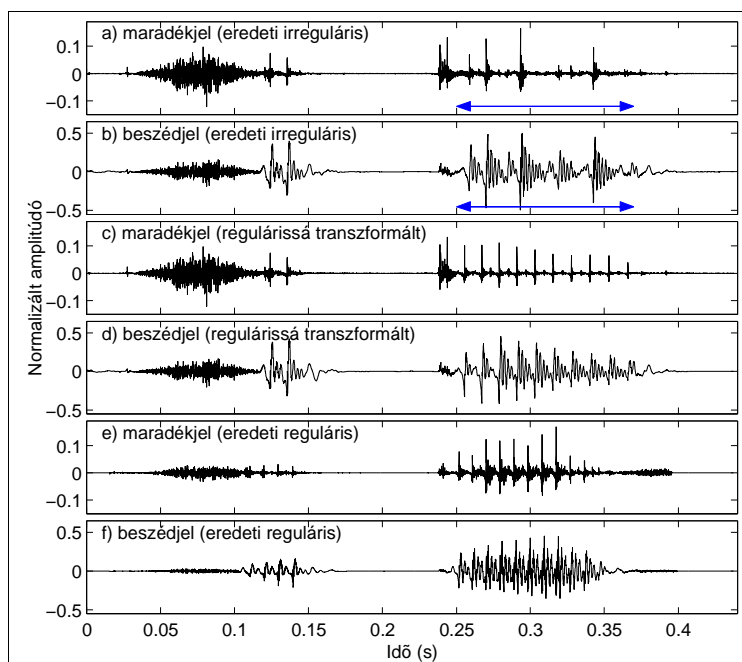
2. ábra

Az első két harmonikus (H1, H2) és az első három formáns (F1, F2 és F3) frekvenciájának és amplitúdójának (A1, A2 és A3) mérése az FFT spektrum alapján

A kutatás során a transzformációs eljárásokat a Matlab szoftverrel készítettük; a statisztikai tesztek az SPSS programban végeztük. A statisztikai elemzések során párosított mintás t -próbát, nem paraméteres összetartozó mintás Wilcoxon-tesztet és Tukey-HSD post-hoc teszttel kiegészített egytényezős ANOVA elemzést alkalmaztunk a Matlab és SPSS programokkal. Az elemzések során kétoldalas $p < 0,05$ szignifikanciaszint alatt (95% konfidenciaszint felett) vetjük el a nullhipotézist.

Eredmények

A 3. ábra az irreguláris–reguláris transzformáció eredményére mutat egy példát. Az ábrán észrevehető, hogy a „regulárisra transzformált” (c és d) és az „eredeti reguláris” (e és f) változatoknak hasonló zöngperiódusai vannak, míg az „eredeti irreguláris” (a és b) jel ettől lényegesen eltérő, és periódusonkénti amplitúdíngadozást tartalmaz. A c–e, illetve d–f ábrák közötti különbségek azért fordulnak elő, mert ez a szó két különböző realizációja, így kis eltérések láthatóak az egyes beszédhangok időtartamában és átlagos amplitúdójában.



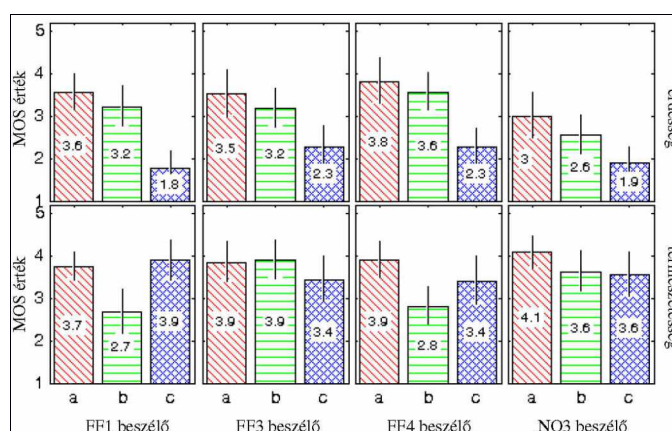
3. ábra

A kiejtett és transzformált „cipő” szó beszédhullámformái és maradékjelei FF3 beszélőtől. (Nyíl jelöli az irreguláris zöngét.)

Percepció teszt

A tesztelők értékelését a 4. ábra mutatja, melyet nem paraméteres összetartozó mintás Wilcoxon-tesztel hasonlítottunk össze. Az „érdesség” kérdésnél az „1” érték jelöli az egyáltalán nem érdekes, míg az „5” a nagyon érdekes válaszokat. A „természetesség” kérdésnél az „1” érték jelöli a nagyon természetellenes, míg az „5” a teljesen természetes válaszokat. Az elemzések szerint az eredeti irreguláris és reguláris transzformált minták nem különböznek szignifikánsan érdekes szempontjából ($p = 0,13$), amennyiben az összes beszélőt egybe vesszük. Az ábra MOS-értékeit megvizsgálva azt vehetjük észre, hogy az eredeti irreguláris minták érdekesége csökkent, de a statisztikai próba ezt nem mutatta ki. Összességében a módszer szignifikáns mértékben csökkentette az eredeti minták természetességét ($p < 0,05$). A természetesség megőrzésében nagy beszélőnkénti különbség látható: az FF3 és NO3 beszélők esetében nem csökkent szignifikánsan a minták természetessége, míg az FF1 és FF4 beszélő esetén igen.

A meghallgatásos teszt eredménye az, hogy az irreguláris–reguláris transzformációs módszer kis mértékben csökkentette a beszédminták érdekeségét, és a négyből két beszélő esetén ezt a természetesség csökkentése nélkül tudta megtenni.



4. ábra

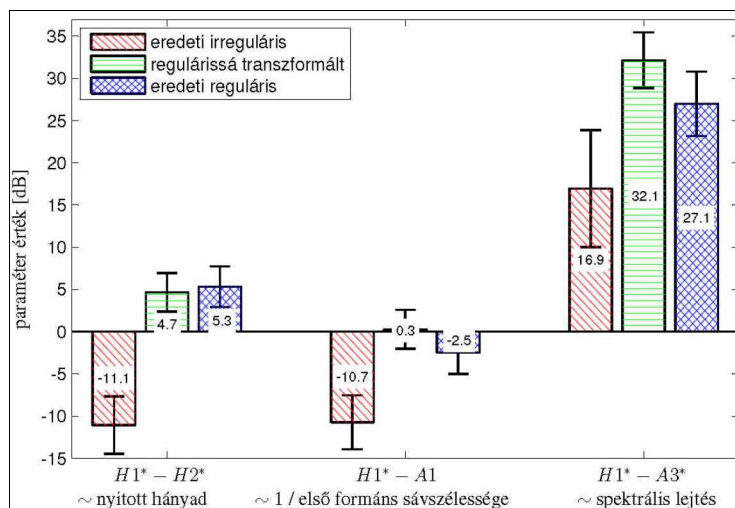
Az irreguláris–reguláris transzformációval módosított szavak percepció elemzésének eredménye: a) eredeti irreguláris b) reguláris transzformált c) eredeti reguláris. A függőleges fekete vonalak a 95%-os konfidenciaintervallumot jelölik

Akusztikai teszt

A három mért akusztikai paramétert ($H1^*–H2^*$: nyitott hányad, $H1^*–A1$: első formáns sáv szélessége, $H1^*–A3^*$: spektrális lejtés) a három beszédmin-

tatípuson az 5. ábra mutatja be. ANOVA-elemzést végeztünk, mely szerint a beszédminták típusának szignifikáns hatása volt mindhárom paraméterre ($F(2,141)=44,16; 17,26; 9,36$; sorban; $p < 0,001$). Tukey-HSD post-hoc tesztel hasonlítottuk össze a beszédmintatípusok átlagos paramétereinek értékét. Ez alapján megállapítottuk, hogy a $H1^*-H2^*$ megközelítőleg azonos az eredeti reguláris és a transzformált beszédrészekre ($p = 0,938$, n.sz. különbség), míg szignifikánsan különböző az eredeti irreguláris mintákhoz képest ($p < 0,001$). A nyitott hányad szempontjából a transzformált változatok tehát közel vannak a modális beszédhez. Az irreguláris zöngével képzett szavak $H1^*-A1$ és $H1^*-A3^*$ különbségei szintén szignifikánsan különbözőek az eredeti reguláris és a transzformált változatokhoz képest ($p < 0,001$ és $p < 0,05$), de az eredeti reguláris és a transzformált változatokban közel megegyeznek ($p = 0,336$ és $p = 0,321$, n.sz. különbség). Eszerint a transzformált minták közel vannak az eredeti modális felvételekhez B1 és TL tekintetében is. A spektrális lejtés szempontjából viszont a transzformált minták értékei jóval magasabbak a természetes mintákénál, ami magyarázhatja, hogy miért érezhették a meghallgatásos teszt résztvevői a minták egy részét kevésbé természetesnek.

A transzformációs eljárás a vizsgált akusztikai jegyek (nyitott hányad, első formáns sávszélessége és spektrális lejtés) szempontjából tehát a reguláris beszédre jellemző értékek irányába módosítja az irreguláris beszédjelet.



5. ábra

Az irreguláris–reguláris transzformációval módosított szavak akusztikai elemzésének eredménye. A függőleges fekete vonalak a 95%-os konfidenciaintervallumot jelölik

Összefoglalás

A cikkben ismertettünk egy transzformációs eljárást, amely irregularis–regularis beszéd átalakítására alkalmas egy beszédkódoláson alapuló jelfeldolgozó módszerrel. A transzformáció eredményét percepciós kísérletben és akusztikai elemzés során vizsgáltuk. Ezek alapján az irregularisból regularissá alakított beszéd kevésbé érdes, mint az eredeti minták; valamint három akusztikai jegy szempontjából közel van az eredeti regularis beszédhez.

A beszédmintákat utólagosan megvizsgálva a két beszélőnél jelentkező természetesség csökkenését valószínűleg az alkalmazott Csapó–Németh (2012) gerjesztési modell okozta: a maradékjel-kódkönyvek felépítése és a transzformáció során az összefüzési költség számításának változtatásával várhatóan jobb eredmény érhető el. A szintézis lépésben az elemkiválasztásnál előfordult, hogy az összefüzési költség miatt ugyanazon maradékjel-periódus ismételten megjelent egymás után többször, így robotos, gépies hangzást eredményezett. A személyfüggés oka az lehet, hogy a glottalizáció különböző megjelenési formái közül a transzformációs algoritmus valószínűleg nem minden esetben tudja megtartani az eredeti beszéd természetességét.

A percepciós teszt eredményeit az is befolyásolhatta, hogy bizonyos tesztlők ismerték a beszélők hangjait, míg mások nem. Böhm–Ujváry (2008) kimutatta, hogy az irregularis fonáció gyakorisága jellemző az egyes beszélőkre. A jövőben tervezzük új percepciós teszt végrehajtását hosszabb beszédmintákkal, nagyobb tesztelői létszámmal, valamint különválasztva a beszélőket ismerő és nem ismerő tesztelőket. Az akusztikai elemzés megbízhatóságát növelheti az eredeti regularis és regularissá transzformált minták között más egyezőségi mutató (pl. korrelációs együttható, átlagos négyzetes hiba) használata.

A kutatásban bemutatott glottalizációjavító eljárást ki lehet terjeszteni hosszabb beszédszakaszokra is, amivel rekedtes, patológikus hangokat várhatóan szebbé, kellemesebbé lehet tenni beszédkozmetikai eljárással (pl. színészek, bemondók hangja). Az irregularis–regularis átalakító eljárás automatikussá kiegészített változatával beszédadatbázisokból el lehetne tüntetni az irregularis zöngéjű szakaszokat, ezáltal ideálisabbá téve a beszédet a további feldolgozás és felhasználás céljából.

Irodalom

- Beke András – Heltovics Éva 2010. A glottalizált magánhangzók automatikus osztályozása spontán magyar beszédben. *Beszédkutatás* 2010. 253–263.
- Blomgren, Michael – Chen, Yang – Ng, Manwa L. – Gilbert, Harvey R. 1998. Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *The Journal of the Acoustical Society of America* 103(5). 2649–2658.
- Böhm, Tamás – Audibert, Nicolas – Shattuck-Hufnagel, Stefanie – Németh, Géza – Aubergé, Véronique 2008. Transforming modal voice into irregular voice by

- amplitude scaling of individual glottal cycles. In *Acoustics '08*. Paris, France, 6141–6146.
- Bóhm, Tamás 2009. *Analysis and modeling of speech produced with irregular phonation*. PhD-disszertáció, BME TMIT, Budapest.
- Bóhm, Tamás – Both, Zoltán – Németh, Géza 2009. Automatic Classification of Regular vs. Irregular Phonation Types. In *NOLISP*. Vic, Spain, 43–50.
- Bóhm Tamás – Ujváry István 2008. Az irreguláris fonáció mint egyéni hangjellemző a magyar beszédben. *Beszédkutatás* 2008. 108–120.
- Csapó, Tamás Gábor – Németh, Géza 2012. A novel codebook-based excitation model for use in speech synthesis. In *CogInfoCom 2012*. Kosice, Slovakia, 661–665.
- Csapó, Tamás Gábor – Németh, Géza 2013. A novel irregular voice model for HMM-based speech synthesis. In *Proc. ISCA SSW8*. Barcelona, Spain, 229–234.
- Dilley, Laura – Shattuck-Hufnagel, Stefanie – Ostendorf, Mari 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24(4). 423–444.
- Fónagy Iván – Magdics Klára 1967. *A magyar beszéd dallama*. Akadémiai Kiadó, Budapest.
- Gobl, Christer – Ní Chasaide, Ailbhe 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication* 40(1-2). 189–212.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Holmberg, Eva B. – Hillman, Robert E. – Perkell, Joseph S. – Guiod, Peter C. – Goldman, Susan L. 1995. Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research* 38. (6). 1212–1223.
- Iseli, Markus – Alwan, Abeer 2004. An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. In *Proc. ICASSP*. Montreal, Quebec, Canada, 669–672.
- Kane, John – Drugman, Thomas – Gobl, Christer 2013. Improved automatic detection of creak. *Computer Speech & Language* 27. (4). 1028–1047.
- Klatt, Dennis H. – Klatt, Laura 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.
- Malyska, Nicolas 2008. *Analysis of nonmodal glottal event patterns with application to automatic speaker recognition*. PhD-disszertáció, MIT, Massachusetts, USA.
- Markó Alexandra 2005. *A spontán beszéd néhány szupraszegmentális jellegzetessége. Monologikus és dialogikus szövegek összevetése, valamint a hümmögés vizsgálata*. PhD-disszertáció, ELTE, Budapest.
- Markó Alexandra 2011. A glottalizáció határjelző szerepe a felolvasásban. *Beszédkutatás* 2011. 31–45.
- Markó Alexandra 2012a. Az irreguláris zöngé szerepe a magánhangzók határának jelölésében V(#)V kapcsolatokban. *Beszédkutatás* 2012. 5–29.
- Markó, Alexandra 2012b. Boundary marking in Hungarian V(#)V clusters with special regard to the role of irregular phonation. *The Phonetician* 105-106. 7–26.
- Németh Géza – Olasz Gábor (szerk.) 2010. *A MAGYAR BESZÉD; Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*. Akadémiai Kiadó, Budapest.

- Olaszy Gábor 2013. Precíziós, párhuzamos magyar beszédatadabázis fejlesztése és szolgáltatásai. *Beszédkutatás 2013*. 261–270.
- Raitio, Tuomo – Kane, John – Drugman, Thomas – Gobl, Christer 2013. HMM-based synthesis of creaky voice. In *Proc. Interspeech 2013*. Lyon, France, 2316–2320.
- Redi, Laura – Shattuck-Hufnagel, Stefanie 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29. (4). 407–429.
- Ruinskiy, Dima – Lavner, Yizhar 2008. Stochastic models of pitch jitter and amplitude shimmer for voice modification. In *IEEE 25th Convention of Electrical and Electronics Engineers in Israel*. Eilat, Israel, 489–493.
- Silén, Hanna – Helander, Elina – Nurminen, Jani – Gabbouj, Moncef 2009. Parameterization of vocal fry in HMM-based speech synthesis. In *Proc. Interspeech 2009*. Brighton, UK, 1775–1778.
- Slifka, Janet 2007. Irregular phonation and its preferred role as a cue to silence in phonological systems. In *ICPhS 2007*. Saarbrücken, Germany, 229–232.
- Talkin, David 1995. A Robust Algorithm for Pitch Tracking (RAPT). In Kleijn, W. B. – Paliwal, K. K. (szerk.) *Speech Coding and Synthesis*. Elsevier. 495–518.
- Tokuda, Keiichi – Kobayashi, Takao – Masuko, Takashi – Imai, Satoshi 1994. Mel-generalized cepstral analysis - a unified approach to speech spectral estimation. In *Proc. ICSLP 1994*. Yokohama, Japan, 1043–1046.

A szerzők ezúton fejezik ki köszönetüket Olaszy Gábornak a kézirattal kapcsolatos hasznos észrevételeiért. Emellett a szerzők köszönetet mondanak a percepció teszt résztvevőinek. A kutatást a PaeLife (Grant No. AAL-08-1-2011-0001) és az EIT-KIC (EITKIC_12-1-2012-001) projekt támogatta.