

KISISKOLÁSOK ALAPHANGMAGASSÁGÁNAK VARIABILITÁSA

Beke András – Horváth Viktória

Bevezetés

A beszéd kutatás egyik legtöbbet vizsgált területe kezdetek óta az alaphangmagasság jellemzőinek és változásainak leírása. A magyar beszédre vonatkozó kutatások még a közelmúltban is elsősorban felolvasott szövegen és mondatokon alapultak (vö. pl. Bolla 1992; Gósy–Terken 1994; Varga 1994, 2002; Olasz 2002). A kutatások egy része beszédtechnológiai felhasználással készült, a gépi beszéd előállításának egyik fő célja ugyanis a prozódia megfelelő megvalósítása (pl. Olasz 1995, 2002). A beszéd felismerésben szintén fontos szerepet játszik a prozódia (Szaszák 2008). A technikai fejlődésnek köszönhetően az utóbbi két évtizedben megindult a spontán beszéd-korpuszok kiépítése, így lehetőség nyílt a spontán beszéd alaphang-szerkezetének leírására. A gazdagréti kábeltelevízió válogatott adásairól Varga László készített intonációs átiratot (1988). Markó Alexandra saját korpuszán elemezte többek között az alaphangmagasság jellemzőit a szövegtípus függvényében (Markó 2005).

A BEA spontán beszéd-adatbázis (Gósy et al. 2012) lehetővé tette az alaphangszerkezet vizsgálatát jó minőségű, nagy mennyiségű anyagon. Beke a felolvasás és a spontán beszéd alaphangszerkezetét hasonlította össze, illetve a beszélő felismeréshez modellezte az alapfrekvencia-eloszlást (2008a, 2008b). Markó (2009) vizsgálata szerint a stigmatizált szökő dallamzár gyakrabban fordult elő spontán beszédben, mint felolvasásban. A fiatal és idős nők spontán beszédének és felolvasásának elemzése azt mutatta, hogy a beszéd mód nagyobb mértékben befolyásolta a vizsgált paramétereket, mint az életkor (Markó–Bóna 2012). A prozódiai szerkezet és a tagolás összefüggéseit is több kutatás elemezte (Gósy 2003; Markó 2010; Váradi 2013). Máty (2012) a fókusz jelölésének prozódiai eszközeit vizsgálta felolvasásban és spontán beszédben. Eredményei szerint a spontán beszédben a fókusz típusok megkülönböztetésében az alábbi paraméterek töltenek be elsődleges szerepet: F_0 -minimum és F_0 -maximum, az ezek közötti tartam és az F_0 -maximum pozíciója.

A szintaktika és a prozódia kapcsolatát vizsgálták olvasott és spontán beszédben magyar nyelvre (Szaszák–Beke 2012). Az eredmények azt mutatták, hogy a felolvasásban a prozódia alapján a szintaktika magasabb szintjei visszakövethetők, azonban a rendszer adaptálása spontán beszédre

kevésbé volt eredményes (Szaszák–Beke 2012). Vizsgálták továbbá, hogy a spontán beszédben prozódiai jellemzőkkel és nem-ellenőrzött módszerrel milyen pontossággal lehet automatikusan detektálni az intonációs frázisokat, illetve az azon belüli fonológiai frázisokat. A rendszer nagyon jó minőségben képes ezen frázisokat automatikusan címkézni (Beke et al. 2014).

Az alaphangszerkezetre vonatkozó kutatások nagyrészt a felnőttek beszédével foglalkoztak. A beszélő életkora alapvetően meghatározó az alaphangmagasság szempontjából. A gége működése hormonális befolyásoltság alatt áll. A csecsemők gégejének mérete – nemtől függetlenül – nagyjából egyforma, ezáltal a „kisgyermek hangmagassága, hangszíne, hangterjedelme és átlagos beszédhangfekvése lényegében egyforma” (Balázs 1993: 157). Ez a gyermekhang csak a pubertás idejében, hormonális hatásra változik meg; a mutálás eredményeként különül el a férfi és női hang. A nemzetközi kutatások különböző életkorokra teszik ennek idejét; de abban nagyjából megegyeznek az eredmények, hogy a folyamat a lányoknál és a fiúknál nem egy időben zajlik le. 4–6 éves korban még nincs eltérés az F_0 értékében a nemek szerint (Hasek 1980; Nygren et al. 2012); 7–8 éves korban a fiúknál azonban alaphangmagasság szignifikánsan csökken – a lányoknál már nem változik nagymértékben ebben az életkorban (Ferrand–Bloom 1996). Más kutatások szerint azonban a lányoknál éppen ebben az életkorban megy végbe az alaphangmagasság csökkenése, a fiúknál pedig 8–9 éves kor között zajlik (Hacki–Heitmüller 1999). Egy további vizsgálatban az alaphangmagasság csökkenése a lányoknál 6 és 10 éves kor között, a fiúknál 8 és 10 éves kor között jelentkezett (Whiteside–Hodgson 1999). Mások szerint a mutálás eredményeként bekövetkező alaphangmagasság-változás csak 11 éves korban (Lee et al. 1999), illetve 12 éves korban (Perry et al. 2001) mutatható ki, a női és férfi hang elkülönülése 15 éves korra fejeződik be. Ennek megfelelően a 6–10 éves fiúk és lányok átlagos alaphangmagassága nem mutat szignifikáns különbséget (Sorensen 1989). Guzman és munkatársai (2014) szintén azt találták, hogy a 7–10 éves lányok és fiúk beszédhangjának elkülönítésében az F_0 nem meghatározó tényező. (A nemzetközi szakirodalomban olvasható egymásnak ellentmondó tendenciák természetesen adódhatnak a vizsgálati személyek egyéni jellegzetességeiből.)

A gyermekek spontán beszédének prozódiai szempontú vizsgálata a közelmúltban indult meg magyar nyelvűre (Markó et al. 2010; Deme 2012; Auszmann–Neuberger 2014; Tóth 2014). Ötéves gyermekeknél még nem volt különbség az F_0 értékében a nemek között (Tóth 2014). A 6–7 éves korosztályban a fiúk átlagos alaphangmagassága szignifikánsan magasabb volt a lányokénál (Deme 2012; Auszmann–Neuberger 2014). A 9 és 11 éves korosztályban nem volt szignifikáns különbség a fiúk és lányok alaphangmagasságának értékében, de a 13 éveseknél már a lányok alaphangja szignifikánsan magasabb volt (Auszmann–Neuberger 2014). Egy másik vizsgálati csoportban a 10 éves fiúk alaphangja szignifikánsan magasabb volt

a lányokénál (Tóth 2014). Valószínűsíthető, hogy 10 éves korra a lányok az alaphangmagasság-változás szakaszának végén, míg a fiúk a változás intenzívebb periódusában vannak. A változás tehát nem lineáris, és nemenként eltérő ütemben zajlik, ahogy ezt a nemzetközi kutatások is igazolták. A gyermekek fizikai változásai mellett feltételezhető egyéb tényezők – például szocializációs, tanult hatások – befolyása is (Deme 2012; Auszmann–Neuberger 2014; Tóth 2014).

A jelen kutatás célja kisiskolás gyermekek alaphangmagasságának vizsgálata spontán megnyilatkozásokban. A kutatás fő kérdése az volt, hogy az alaphangmagasság értékeiben kimutathatók-e életkori változások az iskolába lépéstől kezdve 3 éven keresztül. Ebben az életkori szakaszban indulhatnak be olyan biológiai folyamatok, amelyek jelentősen befolyásolhatják az alaphangmagasság jellemzőit. A szakirodalom ellentmondó eredményeket tartalmaz arra vonatkozóan, hogy például a nemek tekintetében ezek a paraméterek hogyan változnak a vizsgált életkorban.

Hipotéziseink szerint i) az alaphangmagasságban tendenciaszerű eltérést lehet adatolni az egyes életkori csoportok között, amely jellemzően a prozódiai eseményekben lesz kimutatható és nem az alaphangmagasság főbb statisztikai jellemzőiben, ii) feltételezzük, hogy a vizsgált életkori szakasz végén már statisztikailag is kimutatható különbségeket adathatunk a fiúk és a lányok F_0 -értékei között.

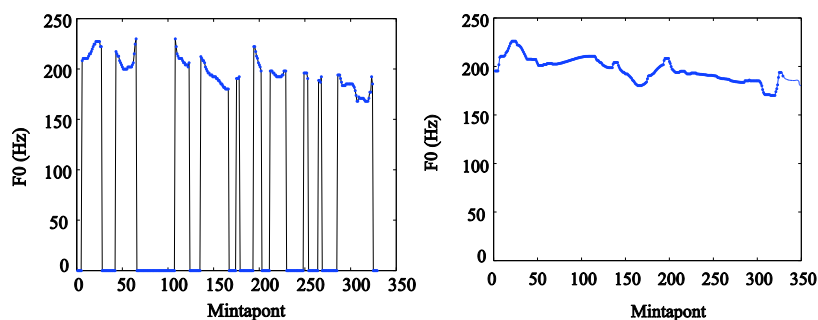
Kísérleti személyek, anyag és módszer

A kutatásban 7, 8 és 9 éves gyermekek vettek részt (korosztályonként 10 fő, 5 lány és 5 fiú minden csoportban). Mindannyian egynyelvűek, ép hallók, nem beszédhibásak és budapesti iskolába járnak. A gyermekekkel spontán narratívákat rögzítettünk. A protokoll minden esetben ugyanaz volt: a kísérletvezető előre megadta a témákat a gyermekeknek, ezt követően csak akkor szólalt meg, amikor segítő, a közlést továbbvivő kérdésre volt szükség. A narratívák témája a gyermekek családja, lakókörnyezete, hobbjai, iskolai elfoglaltságai voltak. A teljes korpusz időtartama 83 perc (gyermekenként 3–5 perc). A felvételek a megszokott iskolai környezetben, de csendes körülmények között készültek Sony ICD-SX700 típusú hangfelvevővel; a kísérletvezető minden esetben ugyanaz a személy volt. A gyermekek beszédének alaphangszerkezetét összevetettük felnőtt beszélőkével, ehhez a BEA-adatbázisból (Gósy et al. 2012) válogattunk narratívákat 10 felnőtt beszélőtől (5 férfi és 5 nő). Az adatközlők életkora 22–35 év. A korpusz időtartama 35 perc (adatközlőként 3–8 perc).

A korpuszt több szinten annotáltuk a Praat 5.3 programban (Boersma–Weenink 2013). Az annotálást követően kiválogattunk 1015 beszédszakaszt a következő kritériumok mentén: a megnyilatkozás nem tartalmaz zajos részt és irreguláris fonációt, kérdést.

Az alaphangmagasságot a MATLAB-ban írt YAAP ("Yet Another Algorithm for Pitch tracking.") szoftver segítségével nyertük ki (Zahorian–Hu 2008; vö. 1. ábra). A YAAP előnye, hogy robosztus mérési eljárást használ, amely alkalmassá teszi, hogy spontán beszédben mérjünk vele alaphangmagasságot. Mindemellett a tesztelési eredmények is azt mutatták, hogy a YAAP algoritmus szignifikánsan precízebben méri az F_0 -értékeket, mint a Praat (Boersma 1993), RAPT (Talkin 1995) vagy a YIN (Cheveigne–Kawahara 2002).

Az alaphangmagasságot a következő főbb paraméterezéssel végeztük: (25 ms-os Hamming típusú ablak 10 ms-os tolási értékkel, frekvencia küszöbök: [60–400]). Mivel az esetünkben folytonos F_0 görbére volt szükségünk, ezért a YAAP algoritmust úgy használtuk, hogy nem vettük igénybe a zöngétlen részek jelzését. Ebben az esetben az algoritmus kiegészítő funkciójával a zöngétlen részek által megszakított görbét inter- és extrapolációval, illetve mediánszűréssel lehet folytonossá tenni. A jelen kutatás során 5-pontos mediánszűrést végeztünk (1. ábra).



1. ábra

Az F_0 reprezentálása utófeldolgozás előtt (balra) és után (jobbra)

Az egyes hangfelvételekben jelölt beszédszakaszok mindegyikében kiszámoltuk a fent bemutatott módon az F_0 kontúrt, amelyből statisztikai jellemzőket származtattunk. A statisztikai paraméterek közül a beszédszakasz normál eloszlását közelítő görbe középértékét, jelen esetben az átlagát, a szóródási mutatóját, az átlagos eltérést, a ferdeséget, a csúcosságot használtuk. A statisztikai vizsgálatok (varianciaanalízis, t-próba) az SPSS 20.0 programmal történtek.

Prozódiai események detektálása

A prozódiai esemény (PE) minden olyan az előzményekhez képesti változás, amely egy adaptív küszöbértéket meghalad. Mindezzel a folytonos jelet több kisebb egységre bontjuk úgy, hogy nem egy előre definiált, a

spontán beszédre sok esetben rosszul működő szabályrendszert alkalmazunk, hanem a percepciós mechanizmushoz inkább közelálló eseménydetektálást.

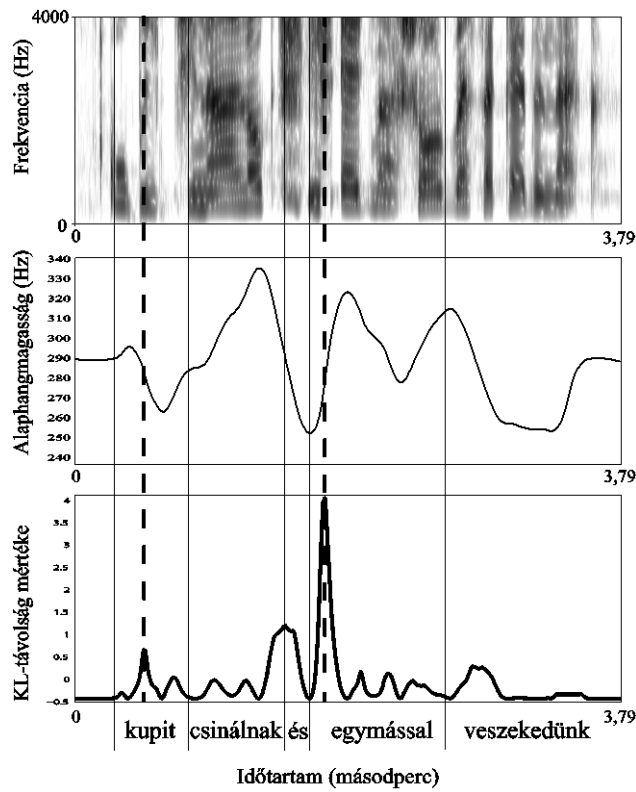
A Kullback–Leibler (*KL*) távolság az egyik leggyakrabban használt algoritmus arra, hogy hogyan mérhető a különbözőség két eloszlás között (Boite–Couvreur 1999). A *KL* távolságot számos területen alkalmazzák: beszélődetektálás, beszédfelismerés, beszélőfelismerés vagy beszéd-nem beszéd detektálás stb. Ezek mellett igen népszerű a *KL* távolság algoritmus használata szegmentálási problémák megoldására is, mint a beszéd vagy a zene szegmentálása. A jelen tanulmányban a *KL* távolságot a fonológiai frázisok határainak meghatározására alkalmazzuk. Siegler és munkatársai (1997) kimutatták, hogy a szimmetrikus Kullback–Leibler távolság egy olyan hatékony távolságmérő eljárás, amely könnyen mérhetővé teszi statisztikailag a különbözőség mértékének kifejezését két beszédjel között. A matematikai hátterét a következőkben ismertetjük: legyen *X* és *Y* két random eloszlás, *KL* pedig a különbözőség mértéke e két eloszlás között. A *KL*-ben az eloszlásokat Gauss-eloszlással modellezzük, így az egyes modelleket azok kovariancia mátrixával, és a középpérték vektorokkal írjuk le.

A Kullback–Leibler távolság ugyan nemnegatív, de nem valódi metrika, mivel nem szimmetrikus, azaz megkülönböztetheti a modellt és modellezett eloszlást. A *KL* aszimmetrikus távolságot szimmetrikussá lehet tenni a következő lépéssel:

$$KL2(X;Y)=KL(X;Y) + KL(Y;X)$$

Mint korábban írtuk, ha a két eloszlás Gauss-eloszlással közelíthető, akkor a szimmetrikussá tett formában is létezik *KL2* szimmetrikus *KL* távolság.

A jelen munka során a *KL2* távolságot a beszédjelben egymást követő részek között számoltuk, amely részek 4 keret hosszúságúnak felelnek meg, vagyis 40 ms időtartamúak (2. ábra). Az ablakhossz 1 keretnyi volt, ami 10 ms-os időtartam. Minden egymást követő beszéd-rész között számolt *KL2* érték egy folytonos görbét adott. A következő feladat az volt, hogy ebben a folytonos jelben megtaláljuk a csúcsokat, amelyek azt jelezték, ahol a két beszéd-szegmens között a legnagyobb eltérés jelentkezett. A magas *KL2* érték tehát azt feltételezi, hogy a két beszéd-szegmens között jelentős az eltérés, míg az alacsony *KL2* érték az azonosságot feltételezi. A csúcsdetektálás szempontjából igen fontos, hogy milyen ablakhosszban keressük az adott csúcsot. Ezért különböző ablakhosszokat alkalmaztunk, amelyeket 10 kerettől (100 ms) 40 keretig (400 ms) növeltünk a *KL2* folytonos görbén.



2. ábra

A prozódiai egységek szegmentálásának folyamata

A csúcsetektálás szempontjából igen fontos feladat a küszöbérték megválasztása is, mivel ettől függ, hogy az adott $KL2$ értéket váltási pontnak fogadjuk el, vagy sem. Ennek megválasztására két adaptív küszöbölési technikát alkalmaztunk (thr_A és thr_B). Az első adaptív küszöbértéket úgy számoljuk, hogy az adott keretben található érték középértékét vesszük, majd megszorozzuk egy konstanssal:

$$thr_{\alpha} = \alpha \frac{1}{2N_1} \sum(F).$$

ahol F a jellemzővektor, N_1 az ablak hossza, és α a konstans.

A prozódiai események határainak detektálásához az adott értékek nagyobbak kell lennie, mint thr_β , amelyet a következőképpen számolhatunk:

$$thr_\beta = \sigma_F + \beta \frac{1}{2N_1} \sum (F)$$

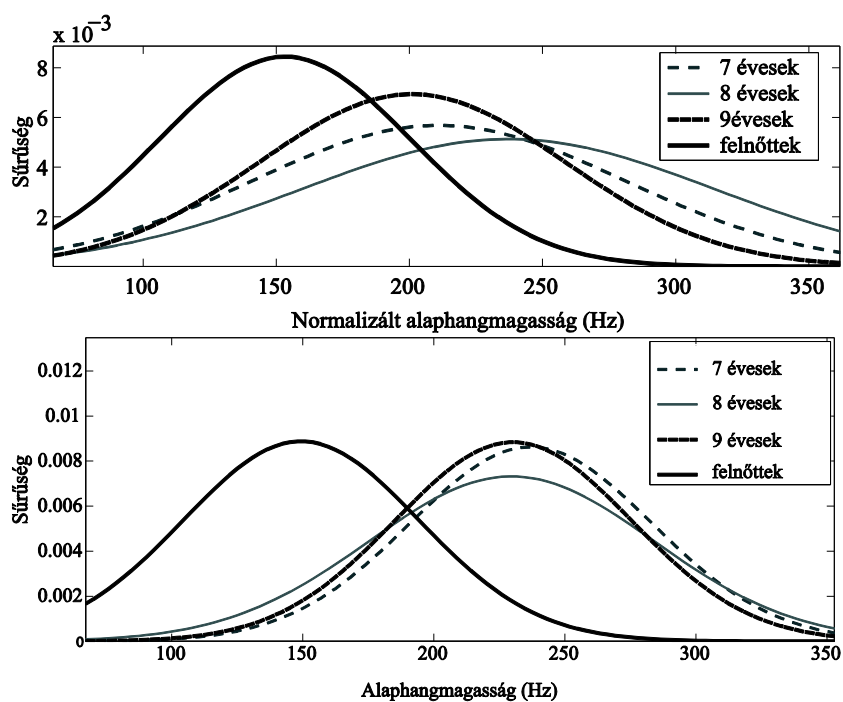
ahol σ_F az adott ablakhosszban lévő értékek átlagos eltérése, β az ablak hossza. Az első küszöbérték azt biztosítja, hogy az adott érték nagyobb, mint a környező területen számított értékek, amely egy rövid idejű ablakra számolandó. A második küszöbérték – amelyet egy hosszabb ablakra számolunk – kezeli az általános tendenciákat az ablakon kívüli adatok változásának figyelembe vételével.

Az ablakok méretét 3 és 4 másodpercre állítottuk, ez a küszöbölési technika használata biztosította, hogy a téves elfogadások száma csökkenjen, és csak a valóban magas $KL2$ értéket fogadja el, amelyek a prozódiai események határait jelentették.

Eredmények

A többszemponos varianciaanalízis szerint a beszélő neme meghatározza az alaphangmagasság átlagos értékét [$F(1, 10) = 6,446, p = 0,016, \eta^2 = 0,168$]; a 95% konfidenciaintervallum alsó értékét [$F(1, 10) = 6,614, p = 0,015, \eta^2 = 0,171$] és felső értékét [$F(1, 10) = 6,282, p = 0,017, \eta^2 = 0,164$]. A csoportszintű elemzéshez t-próbát használtunk. Az eredmények szerint a 7 éveseknél nincs szignifikáns különbség az átlagos alaphangmagasság értékében. A fiúk F_0 -átlaga 241 Hz (átl. elt.: 19 Hz), a lányoké valamivel alacsonyabb, 236 Hz (átl. elt.: 21 Hz). A 8 éves korosztályban sincs szignifikáns különbség a vizsgált csoportban a fiúk és lányok alaphangmagassága között. A fiúk F_0 -átlaga 232 Hz (átl. elt.: 22 Hz), a lányoké 235 Hz (átl. elt.: 24 Hz). A 9 éveseknél sem különbözik egymástól szignifikáns mértékben a lányok és fiúk alaphangmagassága, noha kicsit nagyobb eltérés volt adatolható a két csoport között, mint a fiatalabb korosztályokban. A fiúk F_0 -átlaga 224 Hz (átl. elt.: 8 Hz), a lányoké 239 Hz (átl. elt.: 16 Hz). Az elvártaknak megfelelően a felnőtt férfiak és nők alaphangmagassága szignifikánsan különbözik, az átlagok közötti eltérés 40 Hz ($t = -8,779, p < 0,001$). A férfiaknál mért átlagérték 133 Hz (átl. elt.: 6 Hz), a nőknél pedig 176 Hz (átl. elt.: 8 Hz). A kutatás fő szempontja a beszélő életkorának hatása volt az F_0 -ra, ezért az adatokat úgy normalizáltuk a későbbi elemzésekre, hogy eltűnjenek a beszélők neméből adódó, illetve az egyéni ejtésből fakadó különbségek. Az adatokat csoportszinten normalizáltuk két lépésben. A *z-normalizálás* után az akusztikai jellemzőt újra skáláztuk a teljes populációban mért akusztikai jellemző minimum és maximum értékére. Az így normalizált adatok voltak a korcsoportokra

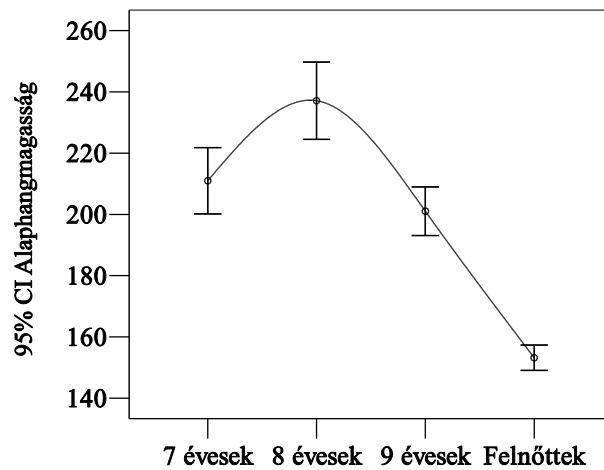
vonatkozó statisztikai elemzés bemenetei, az eljárás eredményét a 3. ábra mutatja.



3. ábra

Az alaphangmagasság (fent) és a normalizált alaphangmagasság (lent) értékei az egyes korcsoportokban

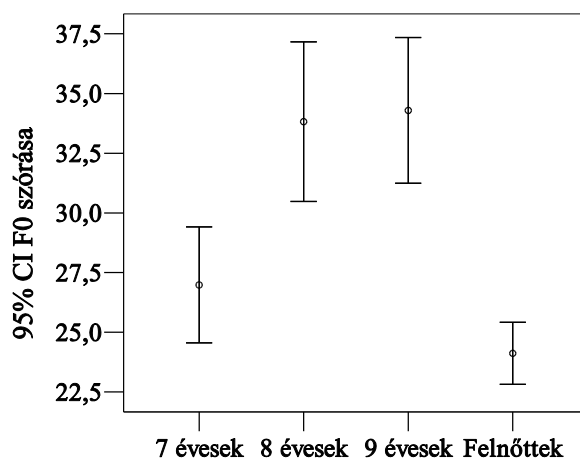
Az egyes beszélőkre meghatározott átlagos alaphangmagasság normalizált értéke a 7 éveseknél 210 Hz (átl. elt.: 70 Hz), a 8 éveseknél 237 Hz (átl. elt.: 77 Hz), a 9 éveseknél 201 Hz (átl. elt.: 57 Hz). A felnőttek csoportjában 153 Hz volt az F_0 átlagos értéke (átl. elt.: 47 Hz). A többszempontos varianciaanalízis szerint a kor hatással van a beszélő alaphangmagasságára [$F(3, 1015) = 104,502, p < 0,001, \eta^2 = 0,236$]. Szignifikáns különbség volt igazolható a felnőttek és mindhárom kisiskolás csoport között (minden esetben $p < 0,001$, vö. 4. ábra). A 8 évesek átlagos alaphangmagassága szignifikánsan magasabb, mint a 7 éveseké és a 9 éveseké (mindkét esetben: $p < 0,001$).



4. ábra

Az F_0 középértéke az életkor függvényében

Az F_0 szórása szignifikáns eltérést mutat a beszélő életkorának függvényében [$F(3, 1015) = 22,602, p < 0,001, \eta^2 = 0,063$]. A szórás értéke a 7 éveseknél 26 Hz (átl. elt: 15 Hz), a 8 éveseknél 33 Hz (átl. elt: 20 Hz), a 9 éveseknél 34 Hz (átl. elt: 22 Hz), a felnőtteknél pedig 24 Hz (átl. elt: 20 Hz). A 8 és 9 éveseknél meghatározott szórásértékek szignifikánsan nagyobbak a felnőtteknél (mindkét esetben $p < 0,001$). A szórás értékében a 8 és 9 évesek kivételével szignifikáns a különbség a kisiskolás csoport között (minden esetben $p < 0,001$). Az F_0 variabilitása tehát változik az életkorral, és felnőttkorban jóval kevésbé szórnak az értékek, mint a gyermekeknél (5. ábra).



5. ábra

Az F_0 szórása az egyes korosztályokban

Az F_0 eloszlásra illesztett normál eloszlás ferdesége a 7 éveseknél 0,1 (átl. elt: 1,1), a 8 éveseknél $-0,4$ (átl. elt: 1,3), a 9 éveseknél $-0,03$ (átl. elt: 2,06), a felnőtteknél pedig $-0,1$ (átl. elt: 1,5). Az egyes életkori csoportoknál meghatározott F_0 eloszlása balra ferde, amely arra utal, hogy az átlagos alaphangmagasság értékéhez képes az alsóbb frekvenciaértékek súlya nagyobb. Az egyes korosztályok között nem volt szignifikáns eltérés.

Az F_0 eloszlásra illesztett normál eloszlás csúcossága a 7 éveseknél 3,6 (átl. elt: 2,03), a 8 éveseknél 5,2 (átl. elt: 3,4), a 9 éveseknél 8,4 (átl. elt: 7,3), a felnőtteknél 5,7 (átl. elt: 4,2). A különbség statisztikailag szignifikáns az egyes korosztályok között [$F(3, 1015) = 33,865, p < 0,001, \eta^2 = 0,091$]. Mindhárom kisiskolás csoport értéke szignifikánsan különbözik a felnőttektől (minden esetben $p < 0,001$). A 8 évesek ebben a jellemzőben is különböznek a 7 és 9 évesektől (mindkét esetben: $p < 0,001$).

Elemeztük a hangköz értékeit is, vagyis az F_0 maximumának és minimumának hányadosát. A hangköz értéke 7 éveseknél 1,43 (átl. elt: 0,33), a 8 éveseknél 1,5 (átl. elt: 0,37), a 9 éveseknél 1,48 (átl. elt: 0,4), a felnőtteknél 1,44 (átl. elt: 0,22). Ebben a paraméterben tehát nincs különbség a beszélő életkorának függvényében.

Elemeztük a megnyilatkozások F_0 -kontúrját, ezt követően *k-közép* algoritmussal (a minták közötti hasonlóságot korrelációval mérve) a kontúrokat csoportokra bontottuk. Ezzel az eljárással megkaptuk a tipikus dallamenteket. A klaszterek számát (jelen esetben a tipikus kontúrmenetek) automatikusan választottuk meg a *silhouette* analízissel (Rousseeuw 1987). A *silhouette* eljárással a *k-közép* csoportba való sorolásának erősségét lehet

vizsgálni, illetve, hogy az egyes elemek a csoportban milyen erős tagsággal rendelkeznek. A *silhouette* analízis az alábbiak szerint kerül levezetésre:

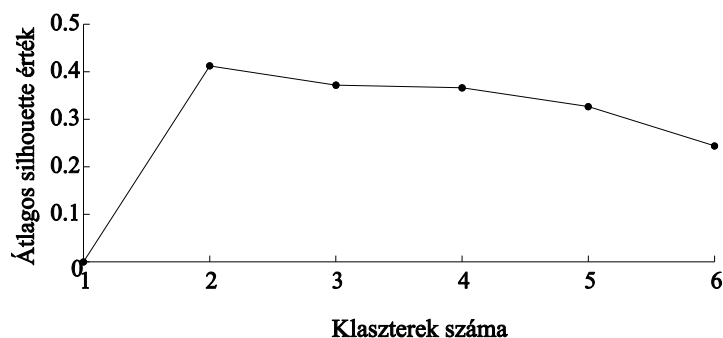
$$SW_i = (b_i - a_i) / \max(a_i - b_i)$$

ahol az átlagos távolság az *i*-edik pont és az *i*-edik pont klaszterének többi egyede között, míg *b* az átlagos távolság az *i*-edik pont és a többi klaszter egyedei között, így minden esetben: $-1 < SW_i < 1$. Az SW_i az alábbiak szerint értelmezhető (1. táblázat):

1. táblázat: A *silhouette* analízis értékei és a hozzájuk tartozó kategóriák

A <i>silhouette</i> elemzés kiértékelés	
SW_i értéke	Kapcsolat erőssége
0,71 – 1,00	erős
0,51 – 0,70	mérsékelt
0,26 – 0,50	gyenge
$\leq 0,25$	nincs

A csoportok számának megállapítására a *silhouette* eljárással kapott SW_i értékek középértékét vettük, így jellemezve a klaszterezés eredményességét (6. ábra). Az ábrán látható, hogy a *silhouette* átlagos értéke akkor volt a legmagasabb, ha csoportok számát kettőre választottuk.

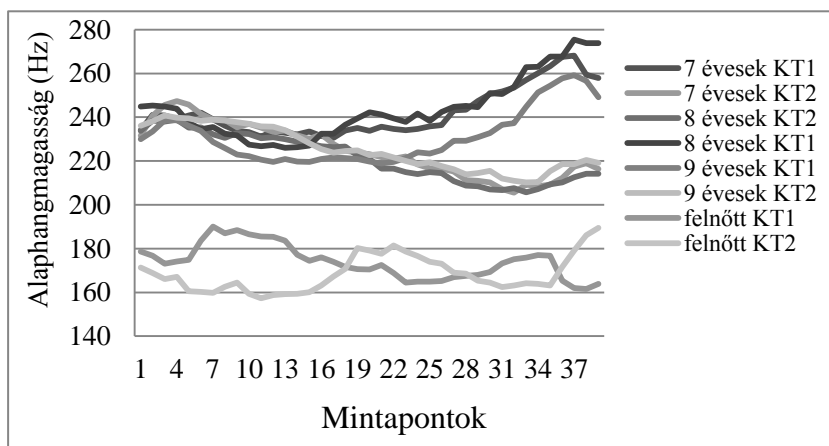


6. ábra

A *silhouette* érték a klaszterek számának függvényében

A tipikus menetek két csoportba voltak sorolhatóak. Az első típus esetén főként emelkedő tendenciát mutat az F_0 , míg a kettes kontúr típus inkább eső tendenciát. A gyermekek esetében ez a két kontúr típus jól elkülöníthető, erősebb eséssel vagy emelkedéssel valósul meg, amely főként a kontúr vége

felé adatható. A felnőttek kontúrja a gyermekekéhez képest inkább lebegő típusú, és a különbség inkább a kontúr elején látható (vö. 7. ábra).



7. ábra
Az F₀ fő kontúrjai az egyes korosztályokban

A kontúrtípusok előfordulási arányát a korpuszban az 2. táblázat mutatja.

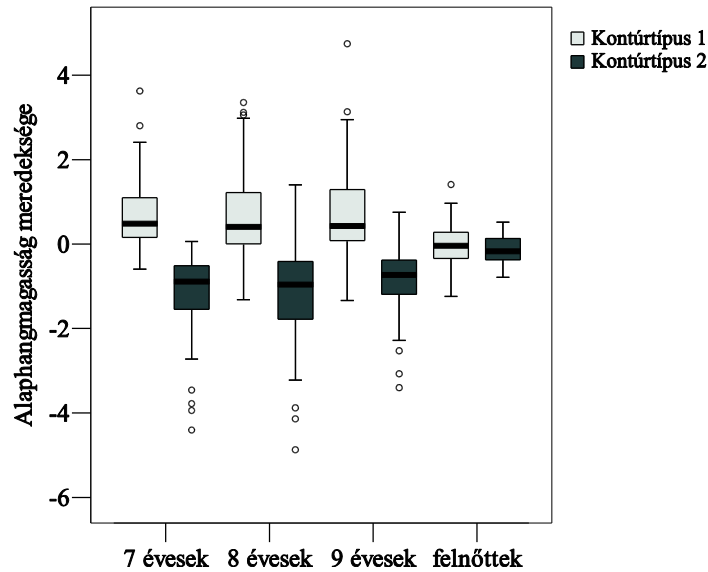
2. táblázat: Az F₀-kontúrtípusok előfordulása az egyes korcsoportokban

Az F ₀ -kontúrtípusok előfordulása (%)		
Korosztály	F ₀ -kontúr	
	1. típus	2. típus
7 évesek	59	41
8 évesek	74	26
9 évesek	68	32
Felnőttek	64	36

Meghatároztuk az egyes típusok meredekségét; az adatokra statisztikai vizsgálatokat végeztünk. Az 1. kontúrtípus meredekségét meghatározza a beszélő életkora [$F(3, 281) = 7,766, p < 0,001, \eta^2 = 0,077$], statisztikailag szignifikáns különbséget ugyanakkor csak a felnőttek és a kisiskolás csoportok között adatoltunk (minden esetben $p < 0,001$). Az első kontúrtípus meredeksége a 7 éveseknél 0,71 (átl. elt: 0,83), a 8 éveseknél 0,76 (átl. elt: 1,08), a 9 éveseknél 0,83 (átl. elt: 1,17), a felnőtteknél pedig -0,4 (átl. elt: 0,55).

A 2. kontúrtípus meredekségét szintén meghatározza a beszélő életkora [$F(3, 315) = 7,701, p < 0,001, \eta^2 = 0,083$], statisztikailag szignifikáns különbséget ennél a típusnál is csak a felnőttek és a kisiskolás csoportok között adatoltunk (minden esetben $p < 0,001$). A második kontúrtípus meredeksége a 7 éveseknél $-1,10$ (átl. elt: $0,88$), a 8 éveseknél $-1,17$ (átl. elt: $1,05$), a 9 éveseknél $-0,93$ (átl. elt: $0,89$), a felnőtteknél pedig $-0,14$ (átl. elt: $0,33$).

Elemeztük azt is, hogy az egyes korosztályokban az első vagy a második kontúrtípus valósul-e meg nagyobb meredekséggel (8. ábra). A kontúr típusa meghatározza a F_0 változás meredekségét [$F(1, 596) = 255,692, p < 0,001, \eta^2 = 0,303$]. A gyermekek esetében minden korcsoportban szignifikáns a különbség a két kontúrtípus között ($p < 0,05$). A felnőtteknél ez az eltérés azonban matematika nem igazolható.



8. ábra
Az F_0 fő kontúrjai az egyes korosztályokban

Következtetések

A jelen kutatásban 7, 8 és 9 éves kisiskolás gyermek alaphangmagasságát vizsgáltuk, felnőttekével összehasonlítva. Az eredmények azt mutatták, hogy kisiskolás korban a jelen vizsgálati csoportban még nem lehetett szignifikáns különbséget kimutatni a lányok és a fiúk alaphangmagassága között. A

tendencia azt mutatta, hogy a 7 éveseknél a lányok átlagos alaphangmagassága kicsit alacsonyabb a fiúkénál, a 8 éveseknél gyakorlatilag nincs különbség a két csoport között; a 9 éveseknél pedig már a lányok átlagos F_0 értéke magasabb a fiúkénál.

A kutatás fő kérdése az volt, hogy miként változik az alaphangmagasság paramétereinek értéke a beszélő életkorának függvényében. A felnőttek átlagos alaphangmagassága természetesen szignifikánsan alacsonyabb volt a gyermekekéhez képest; de a kisiskolás csoportok F_0 értékeiben is találtunk különbséget. A 8 évesek átlagos alaphangmagassága szignifikánsan magasabb volt a jelen vizsgálati csoportban, mint a náluk egy évvel fiatalabb, illetve egy évvel idősebb gyermekeké. Az F_0 értékek szórása a felnőtteknél szignifikánsan kisebb, mint a gyermekekénél, mivel a gyermek beszédében a zöngé modulálása kevésbé stabil a felnőttekéhez képest. Az F_0 kontúrok közül minden korosztályban az emelkedő jellegű gyakoribb, mint az eső jellegű. Mindkét típusnak meredeksége szignifikánsan nagyobb a gyermekekénél, mint a felnőttek beszédében, tehát az alaphangmagasság nagyobb változásokat mutat egy megnyilatkozáson belül a gyermekek esetében. A gyermekekénél jellemzőbb az emelkedés a kontúr vége felé. Ez a jellegzetes dallamemelés hozzájárul a gyermekek beszédéről kialakult percepciók benyomáshoz.

Irodalom

- Auszmann, Anita – Neuberger, Tilda 2014. Age- and gender-related differences in formant structure during the stabilization process of vowels. In: *Proceedings of the Olinco 2014*. (megjelenés alatt).
- Balázs Boglárka 1993. Az időskori hangképzés jellemzői. *Beszédkutató 1993*. 156–165.
- Beke András 2008a. A felolvasás és a spontán beszéd alaphangszerkezetének vizsgálata. *Beszédkutató 2008*. 93–107.
- Beke András 2008b. Az alaphangmagasság-eloszlás modellezése a beszélőfelismeréshez. *Alkalmazott Nyelvtudomány* 8/1–2. 121–133.
- Beke, András – Szaszák, György – Váradi, Viola 2013. Automatic phrase segmentation and clustering in spontaneous speech. In: *Proceedings of IEEE 4th International Conference on Cognitive Infocommunications, CogInfoCom 2013*. Budapest. 452–462.
- Boersma, Paul 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In: *Proceedings of the Institute of Phonetic Sciences 17*. University of Amsterdam. 97–110.
- Boersma, Paul – Weenink, David 2013. Praat: doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/>
- Boite, Jean-Marc – Couvreur, Laurent 1999. Speaker tracking in broadcast audio material in the frame work of the THISL project. In: *Proceeding of the workshop on accessing information in spoken audio (ESCAETRW99)*, 84–89.

- Bolla Kálmán 1992. 1992. *Szupraszegmentális elemzések*. Egyetemi Fonetikai Füzetek 7. ELTE Fonetikai Tanszék, Budapest.
- de Cheveigne, Alain – Kawahara, Hideki. 2002. YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustic Society of America* 111. 1917–1930.
- Deme Andrea 2012. Óvodások magánhangzójának akusztikai jellemzői. In Markó Alexandra (szerk.): *Beszéstudomány: az anyanyelv-elsajátítástól a zöngelkedési időig*. ELTE–MTA, Budapest. 77–99.
- Ferrand, Carole T. – Bloom, Ronald R. 1996. Gender differences in children's intonational patterns. *Journal of Voice* 10 (3.) 284–291.
- Gósy Mária 2003. Virtuális mondatok a spontán beszédben. *Beszédkutatás* 2003. 19–43.
- Gósy Mária – Terken, Jacques 1994. Question marking in Hungarian: timing and height of pitch peaks. *Journal of Phonetics* 22. 269–281.
- Gósy Mária – Gyarmathy Dorottya – Horváth Viktória – Grácsi Tekla Etelka – Beke András – Neuberger Tilda – Nikléczy Péter 2012. BEA: Beszélt nyelvi adatbázis. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest, 9–24.
- Guzman, Marco 2014. Acoustic markers to differentiate gender in prepubescent children's speaking and singing voice. *International Journal of Pediatric Otorhinolaryngology* 78. 1592–1598.
- Hacki, Tamas – Heitmüller, S. 1999. Development of the child's voice: premutation, mutation. *International Journal of Pediatric Otorhinolaryngology* 49. 141–144.
- Hasek, Carol S. – Singh, Sadanand – Murry, T. 1980. Acoustic attributes of preadolescent voices. *Journal of the Acoustical Society of America* 68. 1262–1265.
- Lee, Sungbok – Potamianos, Alexandros – Narayanan, Shrikanth 1999. Acoustics of children's speech: developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America* 105/3. 1455–1468.
- Mády Katalin 2012. A fókusz prozódiai jelölése felolvasásban és spontán beszédben. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 91–107.
- Markó Alexandra 2005. *A spontán beszéd néhány szupraszegmentális jellegzetessége. Monologikus és dialogikus szövegek összevetése, valamint a hümmögés vizsgálata*. Phd-disszertáció. ELTE. Budapest.
- Markó Alexandra 2009. Stigmatizált hanglejtésforma a spontán beszédben. *Beszédkutatás* 2009. 88–106.
- Markó Alexandra 2010. A prozódia szerepe a spontán beszéd tagolásában. *Beszédkutatás* 2010. 82–99.
- Markó Alexandra – Grácsi Tekla Etelka – Imre Angéla 2010. A diskurzusjelölők használatának fejlődése: a hümmögés formai és funkcionális sajátosságai különböző életkorokban. In Navracsics Judit (szerk.): *Nyelv, beszéd, írás. Pszicholingvisztikai tanulmányok*. Tinta Könyvkiadó, Budapest. 82–92.
- Markó Alexandra – Bóna Judit 2012. Eltérő beszédmódok intonációs sajátosságai fiatal és idős korban. In Balázs Géza – Veszelszki Ágnes (szerk.): *Nyelv és kultúra – kulturális nyelvészet*. Magyar Szemiotikai Társaság, Budapest. 253–258.

- Nygren, Mariana – Tyboni, Mikaela – Lindström, Frederic – McAllister, Anita – van Doorn, Jan 2012. Gender differences in children's voice use in a day care environment. *Journal of Voice* 26 (6). 817–e15-8.
- Olaszy Gábor 1995. A kérés, a figyelmeztetés, a felszólítás és a kérdés prozódiaja a kijelentő mondat tükrében. *Beszédkutató* 1995. 46–61.
- Olaszy Gábor 2002. The most important prosody patterns of Hungarian. *Acta Linguistica Hungarica* 49. (3–4). 277–306.
- Perry, Theodore L. – Ohde, Ralph N. – Ashmead, Danieal H. 2001. The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America* 109/6. 2988–2998.
- Rousseeuw Peter J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics* 20. 53–65.
- Siegler, Matthew A. – Jain, Uday – Raj, Bhiksha – Stern, Richard M. 1997. Automatic segmentation, classification, and clustering of broadcast news audio. In: *Proceeding of the DARPA speech recognition workshop*, Chantilly, Virginia, February, 97–99.
- Sorensen, David N. 1989. A fundamental frequency investigation of children ages 6–10 years old. *Journal of Communication Disorders* 22 (2). 115–123.
- Szaszák György 2008. *A szupraszegmentális jellemzők szerepe és felhasználása a beszédfelismerésben*. PhD-disszertáció. Budapesti Műszaki és Gazdaságtudományi Egyetem.
- Szaszák, György – Beke, András 2012. Exploiting Prosody for Syntactic Analysis in Automatic Speech Understanding. *Journal of Language Modelling* 0(1).143–172.
- Talkin, David 1995. A robust algorithm for pitch tracking RAPT. In Kleijn, W. Bastiaan – Paliwal, Kuldip K. (eds.): *Speech Coding and Synthesis*. Elsevier Science, New York. 495–518.
- Tóth Andrea 2014. Gyermekek nemének és életkorának meghatározása a beszédük alapján. *Beszédkutató* 2014. 98–111
- Váradai Viola 2013. *A spontán beszéd szegmentálása produkciós és percepciósszempontról*. PhD-disszertáció. ELTE, Budapest.
- Varga László 1988. A gazdagréti kábeltelevízió műsorából válogatott anyag intonációs átirata. In Kontra Miklós (szerk.): *Beszélt nyelvi tanulmányok*. *Linguistica A/1*. MTA Nyelvtudományi Intézet, Budapest. 5–21.
- Varga László 1994. A hanglejtés. In Kiefer Ferenc (szerk.): *Strukturális Magyar Nyelvtan 2. Fonológia*. Akadémiai Kiadó, Budapest. 468–549.
- Varga László 2002. *Intonation and stress. Evidence from Hungarian*. Palgrave Macmillan, Houndmills, Basingstoke.
- Whiteside, Sandra P. – Hodgson, Carolyn 1999. Acoustic characteristics in 6–10-year-old children's voices: some preliminary findings. *Logopedics Phoniatrics Vocology* 24. 6–13.
- Zahorian, Stephen A. – Hu, Honbing 2008. A spectral/temporal method for robust fundamental frequency tracking. *Journal of the Acoustic Society of America* 123 (6). 4559–4571.

A tanulmány a Bolyai János Kutatási Ösztöndíj és az OTKA 108762 sz. pályázat támogatásával készült.