

GYERMEKEK NEMÉNEK ÉS ÉLETKORÁNAK MEGHATÁROZÁSA A BESZÉDÜK ALAPJÁN

Tóth Andrea

Bevezetés

Az alkalmazott fonetika egy sokat kutatott területe a beszélői profilalkotás, melynek során a beszéd azon akusztikai jellemzőit keresik, amelyek képesek jellemezni a beszélő egy adott tulajdonságát, például életkorát, nemét, pszichikai és érzelmi állapotát vagy testalkatát (Lass–Davis 1976; Amir et al. 2012). Köznapi tapasztalataink is azt mutatják, jóllehet korlátozottan, de képesek vagyunk ezeknek a paramétereknek a becslésére. A beszéd tehát hordoz olyan egyedi akusztikai jellemzőket, amelyek utalnak a produktum létrehozójára (Gósy 2001).

Fiziológiailag az emberi hangképzés alapja a kilégzés okozta levegőáramlás, a hangszalagok fonációs beállítottsága és rezgése, illetve a toldalékcső rezonátortevékenysége (Gósy 2004). A beszédképzésben részt vevő szervek (tüdő, gége, toldalékcső) az életkor előrehaladtával változnak, ami a beszéd akusztikai lenyomatának változását eredményezi (Gósy 2004). Kisgyermekkorban a gége nagysága, a hangmagasság, a hangterjedelem és az átlagos beszédhangképzés tekintetében nincs különbség a nemek között (Balázs 1993).

A beszédképzés során a zöngé a vokális traktusban „formálódik tovább”, amely rezonátorüregként funkcionál. A rezonátorterek által felerősített felhangok a formánsok, amelyek meghatározzák a beszédhangok szerkezetét (Gósy 2004). A formánsértékek az artikulációs gesztusokkal, illetve a toldalékcső alakjával mutatnak összefüggést. Az első formáns az állkapocs nyitáshoz közelebb és a nyelvállás fokával, a második formáns a nyelv vízszintes mozgásával és az ajakműködéssel áll kapcsolatban (Kassai 2002). A vokális traktus növekedésének mértéke és tempója az anatómia és a fonetika sokat kutatott területe (l. például Vorperian et al. 2009). Az anatómiai változás fontos vetülete a beszéd akusztikumában végbemenő változás, amely például a nembeli eltérések kapcsán válik fontossá. Több kutatás tanúsága szerint (Lee et al. 1999; Vorperian et al. 2009) a pubertáskor előtt nincs különbség a vokális traktus hosszában a nemek között. Perry és munkatársai (2001) ezzel szemben azt találták, hogy a beszéd akusztikumában a nembeli eltérések már négyéves korban is kimutathatók, és ezek a testmérettel is kapcsolatban vannak. Az életkor előrehaladtával a formánsértékek csökkennek (Huber et al. 1999; Lee et al. 1999). A formánsok értékei a gyermekeknél a legmagasabbak és a felnőtt férfiaknál a legalacsonyabbak (Huber et al. 1999). A nemek

tekintetében a fiúk formánsértékei alacsonyabbak a velük azonos korú lányokénál (Lee et al. 1999).

Az akusztikai magánhangzótér a gyermekeknél a legnagyobb és a legmagasabb (Deme 2012). A magyar nyelvben a magánhangzókat az első két formánsuk értékével határozzák meg, míg az F_3 értéke a beszélőre jellemző paramétereket hordozhat (Gósy 2004).

Az életkor és a nem beszédalapú meghatározásában a legtöbb fonetikai kutatás az alaphangmagasság és a magánhangzóformánsok szerepét vizsgálja. Hasek és munkatársai (1980) 5 és 10 éves korú gyermekek beszédének vizsgálatakor azt találták, hogy 5–6 éves korban még nincs eltérés az f_0 -ban a nemek szerint, 7 és 10 éves kor között viszont a fiúk alaphangmagassága alacsonyabb volt a velük azonos korú lányokénál. Whiteside és Hodgson (1999) hasonló eredményre jutott 6 és 10 éves korú gyerekek beszédének vizsgálatakor: a lányoknál az alaphangmagasság csökkenése 6 és 10 éves kor között, míg a fiúk esetében 8 és 10 éves kor között jelentkezett. Bennett (1983, idézi Whiteside et al. 2002) longitudinális elemzésében 8 és 11 éves kor között vizsgálta az f_0 változását: nem talált eltérést a nemek között. Perry és munkatársai (2001) 4, 8, 12 és 16 évesek magánhangzóinak vizsgálatakor azt találták, hogy a nemeket jellemző alaphangmagasság-változás csak 12 éves korban jelenik meg a beszédben, ugyanakkor a magánhangzók formánsértékei már gyermekkorban is mutatják a nembeli eltéréseket, az F_3 értékei már 4 éves korban is. Busby és Plant (1995) kutatásának tanúsága szerint a nemek beszédalapú megkülönböztetését 4 és 12 éves kor között elsődlegesen a formánsok, 12 éves kor felett pedig az f_0 biztosítja. A nemek tekintetében az F_2 értékében általánosan, míg az F_1 értékében csak az alsó nyelvállású magánhangzóknál ([a], [æ]) találtak eltérést, ezért feltételezték, hogy ezek a magánhangzók érzékenyebbek a vokális traktus életkor- és nembeli változásaira, mivel képzésükkor a szájüreg nyílt, a szájnyílás pedig tág. Lee és munkatársai (1999) eredményei szerint a formánsfrekvenciák csökkenése a fiúk esetében 11–15 éves kor között, a lányoknál 7–12 éves kor között következik be, ami arra enged következtetni, hogy a vokális traktus növekedése a lányoknál előbb fejeződik be.

Traunmüller és Bezooijen (1994) kutatásának tanúsága szerint 5-ről 11 éves korra 19%-ról duplájára nő a nem felismerésének aránya, és a normál és suttogott ejtésű beszédminták felismerési arányai között csak minimális a különbség. Az azonosításban a szerző szerint a formánsok és az f_0 mellett szupraszegmentumok (például beszédallam) is szerepet játszhatnak. Amir et al. (2012) 8 és 10 éves gyerekek beszédének elemzésekor azt találta, hogy a nemek felismerési aránya 70%-ról 80%-ra nő az életkor előrehaladtával. Az azonosítást segítő akusztikai jegyek közül a szupraglottális térhez köthető jellemzőket tartotta dominánsnak.

Lass és munkatársai (1976) kutatásukban suttogott, normál ejtésű és szűrt magánhangzók segítségével vizsgálták a nemek beszédalapú meghatározását

felnőtteknél. Azt találták, hogy a normál és a szűrt magánhangzók alapján több mint 90%-ban, míg a suttogott ejtésűek alapján 75%-ban különíthetők el a férfi és női beszélők. Ebből arra következtek, hogy az alaphangmagasság elsődleges a vokális traktus rezonáns tevékenységéhez képest. Schwartz (1968) kísérletében a válaszadók a véletlennél jobb arányban voltak képesek a nem megbecslésére zöngétlen frikatívák alapján. Az azonosításban a hangképző rendszer érése mellett szociolingvisztikai, tanult minták is szerepet játszanak (Whiteside–Hodgson 1999).

Az életkor becslésére vonatkozó kutatások változatos eredményeket hoztak: Amir és munkatársai (2012) kísérletükben az életkor becslése 37%-ban volt sikeres 8–18 évesek beszéde alapján. A felnőtt beszélőkkel készített kutatások a következő eredményeket hozták: Gocsál elemzésében (1998) következetes, de pontatlan becslésekről írt. Gósy (2001) kutatásában az életkor meghatározása 70%-ban, Ptacek és Sander (1966) kísérletében 78%-ban volt sikeres. Gorham-Rowan és Laures-Gore (2006) szerint a fiatal és idős beszélők elkülönítése az időskori beszéd sajátosságai miatt lehetséges (például az alaphangmagasság csökkenése, a hangtartás rövidülése, a hangerő csökkenése). A formánsfrekvenciák és az alaphangmagasság mellett vizsgálták még a beszédtempó (Gósy 2001) szerepét is az életkor meghatározásában.

A nem és az életkor becslésére vonatkozó kutatások módszertani, korpuszbeli és résztvevői oldalainak változatossága miatt az eredmények kevésbé összevethetők. A férfi és női beszédhangok eltérései részben magyarázhatók az anatómiai különbségekkel, ugyanakkor a paralingvisztikai jegyek, a tanult genderspecifikus viselkedésminták és azok akusztikai vetületei, például a nőkre jellemzőbb levegős hangképzés (Childers–Wu 1991), az artikuláció pontossága (Diehl et al. 1996) vagy a gége lezorításával elért mélyebb alaphang a fiatal fiúknál (Huszár 2009) sem hagyhatók figyelmen kívül. A testalkat szintén befolyásoló tényező lehet (Xue et al. 2010).

A beszélő nemének és korának meghatározása az automatikus osztályozó rendszerekben is fontos lépés, mivel a nemfüggő modellek működése pontosabb a nemfüggetlenségénél (Rao–Nagesh 2011). Az osztályozók felhasználása sokrétű, l. például multimédiás alkalmazásokban (Harb–Chen 2005) és az ember-gép interakcióban (Feld et al. 2010). A beszélő nemének és életkorának meghatározásakor elterjedt az alaphangmagasság (Hu et al. 2012) és a hallás átviteli karakterisztikáját modellező ún. MFC-együtthatók (Melfrekvenciás kepsztrális együtthatók) használata (Rao et al. 2011). A nem és az életkor gépi osztályozásával számos kutatás foglalkozott (l. például Rao et al. 2011; Li et al. 2013). Az alaphangmagasság bevonásával készült modellek a beszélő nemét közel 100%-ban képesek azonosítani (Hu et al. 2012).

Az anatómiai különbségek akusztikai vetületeiből látható, hogy az alaphangmagasság és a formánsfrekvenciák összefüggnek az életkori és a társadalmi nembeli sajátosságokkal a felnőttek esetében. Az óvodás (5 éves) és általános iskolás korú (10 éves) gyermekeknél a vokális traktus hosszában be-

következő változások 12 éves kor előtt nem magyarázzák a nemek akusztikailag lehetséges elkülönítését; Vorperian és társai (2009) kutatása a vokális traktus más (például orális és faringális) részeinek fejlődését tartja befolyásolónak az anatómiai változások közül.

Kutatásomban a magyar nyelvre vonatkozóan elsőként foglalkozom gyermekek nemének és életkorának beszédalapú becslésével. A felnőttkort célzó kutatások alapján feltételeztem, hogy 1. a gyermekek neme és életkora a beszéd alapján meghatározható; 2. a felismerés pontossága javul az életkor előrehaladtával.

Anyag, módszer, kísérleti személyek

A jelen kutatásban 14 óvoda (életkoruk 5 év) és 14 általános iskolás (életkoruk 10 év) spontánbeszéd-anyagát használtam fel, amelyet óvodai és iskolai környezetben rögzítettem, egy Sony ICD-SX800 típusú digitális diktafonnal. Mindannyian ép halló, fővárosi, magyar anyanyelvű gyermekek voltak. A 10 éves korosztály kiválasztásának alapja az volt, hogy a serdülőkori (fiziológiai) változások ebben az életkorban kezdődnek el, és ezek a beszédprodukciónak is hatással vannak (Gósy 2004).

Minden gyermektől ugyanazon képsorozat leírását rögzítettem. Az így rögzített, átlagosan 1 perces hanganyagokból kiválasztottam egy-egy 7 s-os részletet, amely a percepció teszt anyagát képezte. A percepció tesztben 30 egyetemi hallgató (28 nő és 2 férfi) vett részt, átlagéletkoruk 23,5 év volt. A hanganyagot csoportosan, zárt térben, számítógépről lejátszva, aktív hangfalon keresztül hallgatták meg. A hanganyagokat minden esetben ugyanolyan sorrendben játszottam le. A kísérleti személyek feladata az volt, hogy egyszeri meghallgatás után határozzák meg a beszélő nemét és életkorát (óvoda vagy iskola).

A hanganyagot kézi címkézéssel, hangszinten annotáltam a Praat 5.3 szoftver segítségével (Boersma–Weenink 2013). A percepció tesztben használt 7-7 másodperces hanganyagok beszédhangot tartalmazó részein automatikusan, 0,01 másodpercenként kinyertem a beszélő alaphangmagasságát, amelyet 95%-on újraszűrtem – vagyis a sorba rendezett számok alsó és felső 2,5%-át figyelmen kívül hagytam –, az esetleges zajok kizárása végett. A magánhangzók formánsainak mérését a hanganyagban leggyakrabban előforduló hét magánhangzó esetében végeztem el, ezek az [ɔ], [a:], [ɛ], [e:], [i], [o], [u] voltak. A magánhangzók előfordulási arányai közel azonosak voltak. A formánsok mérésére egy erre a célra írt szkriptet használtam, amely a hangidőtartam 50%-án elhelyezett mérőpontban adta meg az F_1 , F_2 és F_3 értékét. A kinyert adatokat ebben az esetben is újraszűrtem, 95%-on.

A statisztikai elemzést az SPSS 13.0 program segítségével végeztem el. Az elemzett akusztikai jellemzők szerepét egytényezős varianciaanalízis (ANOVA) segítségével elemeztem, minden esetben 95%-os konfidenciaszinten.

A nem és az életkor felismerésének tesztelésére létrehoztam egy gépi osztályozót is. Az osztályozáshoz MFC-együtthatókat és GMM-SVM algoritmust használtam. Az osztályozót a Matlab szoftverben valósítottam meg.

A beszélőfelismerésben az MFCC-k kiszámítását a beszéd előfeldolgozása előzi meg, amelynek során előkiemelést hajtanak végre (a jel/zaj viszony javítására), illetve kiszűrik a beszédjelet nem tartalmazó részeket, ezt beszéddektálásnak nevezik (Beke 2012). Az előfeldolgozást a jellemzőkinyerés követi, ami a jelen kutatásban az MFC-együtthatók kiszámítását jelenti. Ennek részeként a beszédjelen gyors Fourier-transzformációt (FFT) hajtanak végre, ezt egy ún. ablakfüggvénnyel súlyozzák, így megkapják 10 ms-onként a beszédjel spektrumát. Második lépésként a kapott spektrumot szűrősoros elemzésnek vetik alá (az ún. Mel-skála alapján), amely a hallás kritikus sávjainak megfelelő érzékenységet követ (Beke–Szaszák 2009). A kapott jellemzők eloszlásából kevert Gauss-modellekkel (GMM) becsültem meg a mintaközepet (átlag) és szórását (Reynolds 1995).

A gépi osztályozásra SVM-et használtam. Az SVM bináris osztályozási problémák megoldásában használt matematikai konstrukció, amely szupportvektorok segítségével ún. döntési hipersíkot határoz meg, és így különíti el a két osztály mintáit (Beke 2011). Az osztályozás két részből áll: a tanítás során modelleket készítenek – a jelen kutatásban fiú-lány, illetve óvodás-iskolás modelleket –, majd a teszteléskor az osztályozó a bejövő beszédmintát összehasonlítja a tanító modellel, azaz mintaillesztést hajt végre.

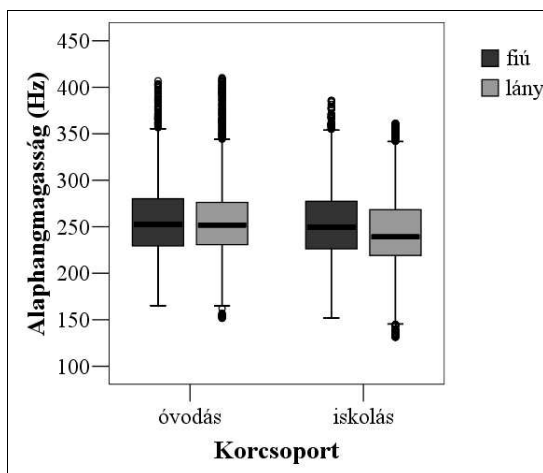
Eredmények

Az akusztikai elemzés eredményei

A gyermekek alaphangmagassága a nemek függvényében óvodáskorban nem mutatott eltérést [ANOVA: $F(1, 4210) = 0,617$; $p = 0,432$], hasonlóan Hasek és munkatársainak (1980) eredményeihez (1. ábra). Az ötéves lányok átlagos alaphangmagassága 261,6 Hz (szórás 49 Hz) volt, míg a fiúk átlagos alaphangja 260,5 Hz (szórás 42 Hz). Az eredmények közelítőleg megfelelnek az angol kutatásokban találtaknak (vö. Lee et al. 1999).

Az általános iskolások esetében a lányok átlagos alaphangmagassága 246,4 Hz (szórás 39 Hz), a fiúké 254,5 Hz (szórás 40 Hz) volt. A két csoport közötti különbség szignifikáns [ANOVA $F(1, 16289) = 161,471$; $p \leq 0,001$]. A kutatásban váratlan eredmény, hogy a fiúk alaphangmagassága a lányokénál magasabbnak bizonyult. Mindez ellentmondani látszik annak a szakirodalmi tapasztalatnak, amely 12 éves korra (serdülőkor) teszi az f_0 elkülönülésének idejét (l. Perry et al. 2001). Deme (2012) 6–7 évesek spontán beszédének vizsgálatakor – a jelen eredményekhez hasonlóan – azt találta, hogy a lányok átlagos alaphangmagassága alacsonyabb. A szerző szerint a lányoknál 6–7 éves korra tehető egy jelentős ugrás az f_0 csökkenésében, amely a fiúknál csak később jelentkezik. Ezzel magyarázható, hogy átmenetileg a lányoknál alacsonyabb alaphangmagasság mérhető (Hacki–Heithmüller 1999, idézi

Deme 2012). A jelen kutatás eredményei alapján ez a tendencia még 10 éves korban is megfigyelhető, bár a csökkenés mértéke kisebb, mint a megelőző életkorban. 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 (vö. Hasek et al. 1980). Ennek igazolására azonban további életkori csoportok vizsgálata szükséges. Ugyanakkor az alaphangmagasság eltéréseinek vizsgálatakor a szociokulturális (pszichológiai és társadalmi) vonatkozások sem hagyhatók figyelmen kívül, amelyek a beszédprodukción is befolyásolhatják (Kohn et al. 2012, idézi Deme 2012), illetve a testalkat is további befolyásoló tényező lehet (Perry et al. 2001).



1. ábra

Az alaphangmagasság eloszlása nemek és életkorok szerint

Deme (2012) hat fonéma realizációit vizsgálta 6–7 évesek spontán beszédében. Az F_1 esetében az [a:] és [o], az F_2 -nél az [ε], [i], az F_1 és F_2 -ben pedig az [e:] esetében talált eltérést a nemek között. A jelen kutatásban az óvodások esetében az F_1 -ben az [ɔ], [e:], az F_1 és F_2 -ben az [o], az F_3 -ban az [ɔ], [a:], [ε] hangok esetében találtam szignifikáns eltérést a nemek között. A statisztikai próba (ANOVA) eredményeit az 1. táblázatban foglaltam össze.

Az [ɔ], [a:], [ε] hangok esetében kimutatott különbségek hasonlóak a külföldi szakirodalomban foglaltakhoz (vö. Busby–Plant 1995). Anyagomban az [ɔ], [a:], [ε] hangok esetében az F_3 is szignifikáns különbséget mutatott a nemek között.

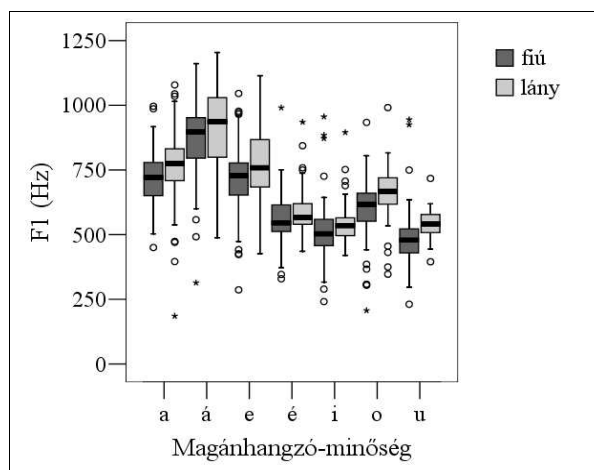
A magánhangzó-minőségek közötti eltérések utalhatnak hangképzésbeli különbségekre, amely fakadhat a toldalékcso méretbeli különbözőségeiből

(Lee et al. 1999). Ezt bizonyíthatja, hogy az eredmények szerint a legtöbb magánhangzó esetében a fiúk formánsértékei alacsonyabbak a lányokénál. A jelen kutatásban a hangképző szervek méretére vonatkozó adat nem áll rendelkezésemre, ezért csak a szakirodalmi tapasztalatoknak megfelelően tudom valószínűsíteni, hogy a fiúk testméretei már ebben az életkorban is nagyobbak a lányokénál.

1. táblázat: A magánhangzók formánsértékeinek nemenkénti eltérése óvodáskorban (a *p* értéke; a szignifikáns eredményeket a szürke szín jelzi)

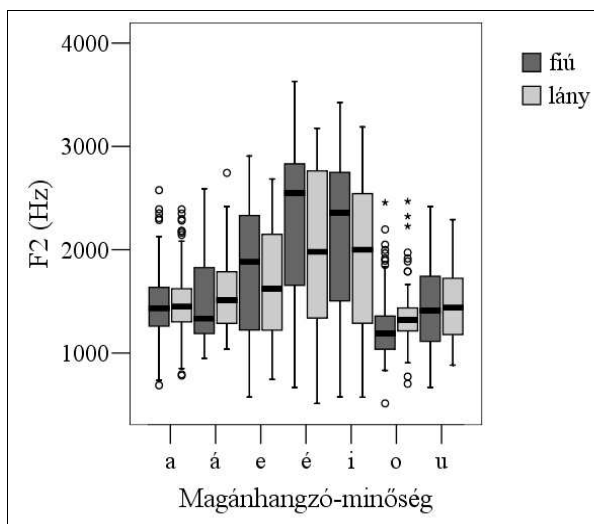
	[ɔ]	[a:]	[ɛ]	[e:]	[i]	[o]	[u]
F ₁	< 0,001	0,170	< 0,001	0,076	0,039	< 0,001	0,072
F ₂	0,694	0,333	0,064	0,095	0,111	0,015	0,458
F ₃	0,001	0,050	< 0,001	0,301	0,014	0,347	0,828

A 2. ábrán látható, hogy az [ɔ], [a:], [ɛ] hangok az F₁ mentén nagy szórással realizálódtak, ami a nyelvvállásfok artikulációjának variabilitására utal. Ez a változatos megvalósulás az F₂-paraméter esetében nem jelentkezett, egyenletesebbek voltak az értékek (3. ábra). Az adatok közötti átfedések a nemek szerint elkülönített csoportokban mind az F₁, mind az F₂ esetében nagyok voltak, a hangzók minőségének ejtésében ezért valószínűsíthetően kicsi volt a különbség. Az [ɔ], [a:], [ɛ] hangok ejtése a fiúknál zártabb, az [ɛ], [e:], [i] velárisabb volt, hasonlóan Deme (2012) eredményeihez.



2. ábra

A magánhangzók első formánsainak értékei nemek szerint óvodáskorban



3. ábra

A magánhangzók második formánsainak értékei nemek szerint óvodáskorban

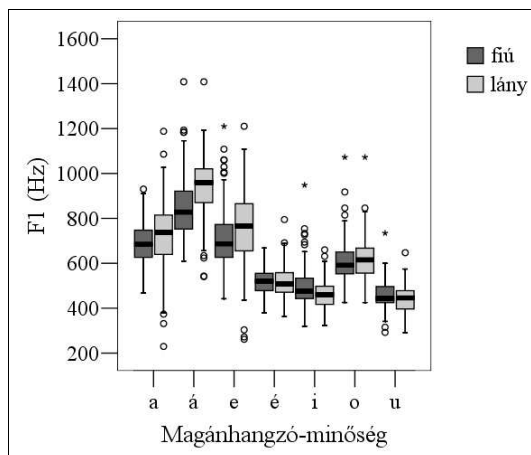
A 10 éves korú gyermekeknel a nemek szerint az F_1 paramétere az alsó és legelső nyelvállású hangzók ([ɔ], [a:], [ɛ]) és az [i] esetében mutatott szignifikáns eltérést, az F_2 az [a:], [ɛ] és [o] hang esetében, az F_3 pedig az [e:] és [i] hangoknál. A statisztikai próba (ANOVA) eredményeit a 2. táblázatban foglaltam össze.

2. táblázat: A magánhangzók formánsértékeinek nemenkénti eltérése iskoláskorban (a p értéke; a szignifikáns eredményeket a szürke szín jelzi)

	[ɔ]	[a:]	[ɛ]	[e:]	[i]	[o]	[u]
F_1	0,001	< 0,001	< 0,001	0,724	0,001	0,164	0,179
F_2	0,249	< 0,001	0,004	0,198	0,894	0,022	0,880
F_3	0,290	0,031	0,809	0,006	0,002	0,837	0,537

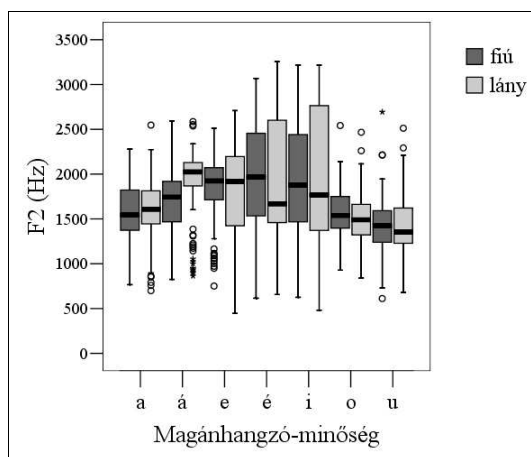
Az eredmények alapján feltételezhető, hogy az alsó és legelső nyelvállású hangok érzékenyebbek a fizikai növekedéssel bekövetkező változásokra (például toldalékcso méretének változása), hiszen az artikulációt tekintve ezeknél a hangoknál a legnyíltabb a szájüreg és a legtágabb a szájnyílás. Az [ɔ], [a:], [ɛ] hangok F_1 -értékei az óvodásokéhoz hasonlóan nagy szóródást mutatnak, ami változatos artikulációs megvalósulásukra utal (4. ábra). A fiúk ejtése az [ɛ] hangok esetében még zártabb, mint a lányoké. Az F_2 paraméter tekintetében az értékek szóródása kisebb, ami összevetve az 5 éves kori érté-

ekkel, az artikuláció stabilizálódására is utalhat. A variabilitás az F_2 esetében az [ɛ], [e:], [i] hangoknál a legnagyobb (5. ábra).



4. ábra

A magánhangzók első formánsainak értékei nemek szerint iskoláskorban



5. ábra

A magánhangzók második formánsainak értékei nemek szerint iskoláskorban

A percepció teszt eredményei

A kutatásban az óvodások nemét átlagosan 69%-ban, az általános iskolás gyerekeket átlagosan 85,3%-ban osztályozták helyesen a válaszadók (6. áb-

ra). Az 5 évesek esetében a fiúk és lányok felismerési arányának átlaga között nem volt különbség: a lányokat 69%-ban, a fiúkat 69,5%-ban azonosították helyesen, azonban a lányok felismerése a fiúkénál nagyobb szóródást mutatott (a szórás a lányoknál 9,2%, a fiúknál 5,7%), ami a bizonytalanabb észlelést jelzi.

Az iskolások esetében szintén a fiúk felismerése volt pontosabb (95,7%), szemben a lányokéval (75,3%). A lányok azonosítása ebben az életkorban is nagyfokú bizonytalanságot mutatott, a szórás értéke 9,7% volt, szemben a fiúk 1,5%-os értékével.

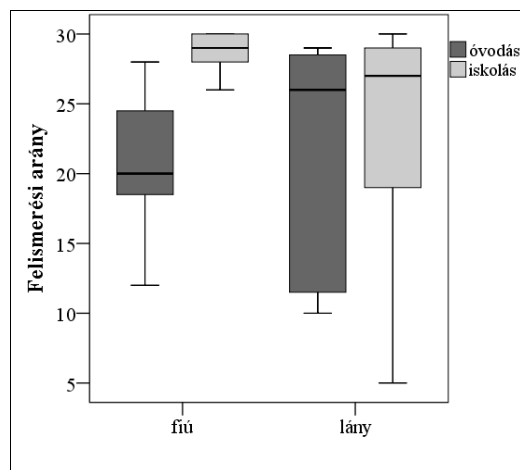
Általánosan elmondható, hogy az életkor előrehaladtával javult a felismerés pontossága a nemekre vonatkozóan. A fiúk felismerése mindkét korcsoportban sikeresebbnek bizonyult.

Elemeztem a felismerési eredményeket az életkor tekintetében is. Az óvodásokat átlagosan 90%-ban, az általános iskolásokat átlagosan 86,4%-ban, tehát közel azonos arányban ismerték fel a percepció tesztet kitöltők. Az észlelésben az életkorra vonatkozólag fontos akusztikai kulcs lehetett az alaphangmagasság, amely az óvodások esetében szignifikánsan magasabb volt, mint az általános iskolásoknál. Mivel a kutatásban csak két korcsoport szerepelt, ezekből az eredményekből messzemenő következtetés nem vonható le az életkorbecslésre vonatkozóan. Tendenciaként azonban megfigyelhető, hogy az óvodáskorú gyermekek életkorának felismerése a humán percepcióban relatíve biztosnak mondható.

A gépi osztályozás eredményei

A kutatásban létrehozott MFCC-alapú SVM-GMM osztályozó az óvodások nemét 48%-ban, az általános iskolásokat 73%-ban osztályozta helyesen. Mivel az óvodások esetében az osztályozás a véletlen határán mozgott, az eredményeket tovább nem elemzem. Az iskolások csoportján belül a lányok osztályozása volt sikeresebb, a gépi címkézés pontossága 85%, a fiúk esetében ez az érték 61,9% volt. A gépi osztályozó eredményei elmaradtak a humán percepciótól. Az a tendencia azonban itt is megfigyelhető, hogy az életkor előrehaladtával javul az osztályozás pontossága. A gépi osztályozás eredménye az életkor tekintetében 47%-os volt, hasonlóan több külföldi kutatáséhoz (Feld et al. 2010; Li et al. 2013).

Fontos hangsúlyozni, hogy a gépi osztályozás eredménye a humán percepcióval csak fenntartásokkal hasonlítható össze, mivel nincsenek bizonyos ismereteink arról, hogy melyek azok a konkrét akusztikai jellemzők, amelyek szerepet kapnak a nem és az életkor azonosításában. A gépi osztályozásban számos jellemző bevonásával próbálják meg modellezni az emberi hallás – ezáltal észlelés – folyamatát (l. például Müller 2006), azonban ezek a jellemzők csak az adott tanítói bázist veszik (és vehetik) alapul, ami az emberi tapasztaláshoz, a szocializációs folyamat során tanult mintákhoz képest relatíve kis „korpuszt” jelent. Ennek is köszönhető, hogy a gépi osztályozás eredményei elmaradnak a humán észleléstől.



6. ábra

Felismerési eredmények a nemek és az életkor függvényében

Következtetések

A jelen kutatásban 5 és 10 éves gyermekek nemének és életkorának becslését vizsgáltam a spontán beszédük alapján. Kutatásom a magyar nyelvben elsőként foglalkozik a gyermekek nemének és életkorának becslésével, illetve a percepció és akusztikum ilyen irányú kapcsolatával.

Feltételeztem, hogy a gyermekek nemének és életkorának azonosítása lehetséges a beszéd alapján. Első hipotézisem igazolódott, az óvodások csoportjában a lányokat 69%-ban, a fiúkat 69,5%-ban azonosították helyesen a válaszadók. Az iskoláskorú gyerekeknél a pontosabb felismerésben az alaphangmagasság és egyes hangoknál a magánhangzóformánsok fontos akusztikai kulcsként működhetnek a nemek felismerésében. Az óvodások esetében a magánhangzóformánsok mellett más akusztikai jegyek (például szupraszegmentális térhez köthető jellemzők) befolyásolhatták az észlelést. A kutatás felismerési eredményei hasonlatosak a szakirodalomban foglaltakhoz (vö. Amir et al. 2012). Az a tapasztalat, miszerint a fiatalabb életkorú beszélőket inkább lánynak feltételezik, a jelen kutatásban nem volt igazolható (vö. Traummüller–Bezooijen 1994).

Második hipotézisem – amely szerint a felismerési pontosság javul az életkor előrehaladtával –, szintén igazolódott, hiszen az iskolás csoportban a lányokat már 75,3%-ban, a fiúkat pedig 95,7%-ban azonosították helyesen a kísérletben részt vevők. Az életkor felismerése mind az óvodások (90%), mind az iskolások (86,4%) esetében sikeresnek bizonyult.

Az akusztikai elemzésből kiderült, hogy az óvodásoknál az alaphangmagasságban a nemek szerint nem adatolható eltérés. A magánhangzó-formánsok elemzésékor az F_1 és F_2 értékei csak bizonyos magánhangzók esetében hordoz-

tak különbséget a nemtől függően. Az alsó és a legalsó nyelvállású hangok esetében az F_3 értékében eltérést találtam a két csoport között. Az F_3 paraméter ezért a jelen kutatás szerint az alsó nyelvállású hangoknál hordozhat a beszélő nemére vonatkozó információt. Az F_1 -értékek magas szóródásából arra következtettem, hogy az [ɔ], [a:], [ɛ] hangokat mind az óvodások, mind az iskolások esetében változatos artikulációs megvalósulás jellemezte. Az alsó és legalsó nyelvállású hangok genderspecifikussága az iskolások csoportjában is megmutatkozott, vagyis ezek érzékenyebbek lehetnek a vokális traktus fizikai növekedés során bekövetkező méretbeli változásaira, mivel képzésüket tekintve a legnyíltabb hangzók.

Az iskolások esetében a nemek között az alaphangmagasságban adatoltam eltérést. A kutatásban a fiúk alaphangja magasabb volt a lányokénál, hasonló jelenséget Deme (2012) is megfigyelt. Ugyanakkor több kutatásban találták azt, hogy a lányoknál 6–10, a fiúknál 8–10 éves kor között megfigyelhető az alaphangmagasság csökkenésének tendenciája (Whiteside–Hodgson 1999). A jelen kutatás eredményeiből az derül ki, hogy ez a folyamat még 10 éves korban is tart. Ezért vannak olyan időpillanatok a növekedésben, amikor ez a nemenként eltérő ütemű változás a lányoknál – időlegesen – alacsonyabb alaphangmagasságot eredményez (Deme 2012). A kérdés pontosabb vizsgálatához azonban további kutatások szükségesek. A növekedésbeli, fizikai változások mellett természetesen feltételezhető egyéb tényezők – például szocializációs, tanult hatások – befolyása is.

A nem- és életkorbecslés a beszélőfelismerő rendszerek fejlesztésében is fontos lépés. A kutatásban megalkotott gépi osztályozó eredményei elmaradtak a humán észleléstől. A korpusz és az életkori csoportok (gyerekek) sajátosságai miatt az eredmények a külföldi kutatásokkal csak fenntartásokkal vehetők össze (vö. Li et al. 2013). A gépi osztályozás az iskolások csoportjában a lányok esetében sikeresebb volt, ami a jelen kutatásban a humán percepcióra kapott eredményekkel ellentétes, azonban a külföldi szakirodalmi tapasztalatoknak nem mond ellent (vö. Amir et al. 2012). Bár a humán percepció számára az óvodás és iskolás életkor elkülönítése nem okozott problémát, a gyenge gépi osztályozási eredmény magyarázható az egymáshoz közeli életkorokkal. Több kutatásban találták azt, hogy az életkort célzó gépi felismerés az egymástól távolabbi életkorok esetében sikeresebb (Müller 2006).

Az észlelés és felismerés folyamatában természetesen további jellemzők is szerepet kaphatnak, például a testalkat, illetve szocializációs, kulturális hatások. A kérdés további vizsgálata újabb akusztikai jellemzők bevonásával képzelhető el, ilyenek lehetnek például a beszédhang-minőség jellemzői (jitter, shimmer, időtartam stb.), illetve a szupraszegmentális jegyek (például beszédtempó, beszédhang, hangsúly) is. A jelenlegi eredmények a beszélői profilalkotás területén szolgálhatnak adalékokkal, hozzájárulnak a nemek közötti különbségek beszédbeli megjelenésének megismeréséhez, ezáltal a beszéd- és beszélőfelismerő rendszerek fejlesztéséhez.

Irodalom

- Amir, Ofer – Engel, Merav – Shabtai, Esther – Amir, Noam 2012. Identification of children's gender and age by listeners. *Journal of Voice* 26/3. 313–321.
- Balázs Boglárka 1993. Az időskori hangképzés jellemzői. *Beszédkutatás* 1993. 156–165.
- Beke András 2011. Szókezdetek automatikus osztályozása spontán beszédben. *Magyar Nyelvőr* 135/2. 226–241.
- Beke András 2012. Beszélőfelismerés kevert-Gauss modellekkel. In Markó Alexandra (szerk.): *Beszédtudomány. Az anyanyelv-elsajátítástól a zöngékezdési időig*. ELTE BTK–MTA, Budapest. 335–353.
- Beke András – Szaszák György 2009. A svávariációk automatikus felismerése magyar nyelvű spontán beszédben. *Beszédkutatás* 2009. 148–169.
- Boersma, Paul – Weenink, David 2013. *Praat: doing phonetics by computer [Computer program]. Version 5.3*. <http://www.praat.org>. (Letöltés ideje: 2013. júl. 9.)
- Busby, Peter A. – Plant, Geoff 1995. Formant frequency values of vowels produced by preadolescent boys and girls. *Journal of the Acoustical Society of America* 97/4. 2603–2606.
- Childers, Donald G. – Wu, Ke 1991. Gender recognition from speech. Part I: Coarse analysis. *Journal of the Acoustical Society of America* 90/4. 1828–1840.
- Deme Andrea 2012. Óvodások magánhangzóinak akusztikai jellemzői. In, Markó Alexandra szerk.: *Beszédtudomány. Az anyanyelv-elsajátítástól a zöngékezdési időig*. ELTE BTK–MTA Nyelvtudományi Intézet, Budapest. 77–99.
- Diehl, Randy L. – Lindblom, Björn – Hoemeke, Kathryn A. – Fahey, Richard P. 1996. On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics* 24. 187–208.
- Feld, Michael – Burkhardt, Felix – Müller, Christian 2010. Automatic speaker age and gender recognition in the car for tailoring dialog and mobile services. In: *Proceedings of Interspeech 2010*. Chiba. 2834–2837.
- Gocsál Ákos 1998. Életkorbecslés a beszélő hangja alapján. *Beszédkutatás* 1998. 122–134.
- Gorham-Rowan, Mary M. – Laures-Gore, Jacqueline 2006. Acoustic-perceptual correlates of voice quality in elderly man and women. *Journal of Communication Disorders* 39. 171–184.
- Gósy Mária 2001. A testalkat és az életkor becslése a beszéd alapján. *Magyar Nyelvőr* 125. 137–148.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Harb, Hadi – Chen, Liming 2005. Voice-based gender identification in multimedia applications. *Journal of Intelligent Information Systems - Special issue: Intelligent multimedia applications* 24/2. 179–198.
- Hasek, Carol S. – Singh, Sadanand – Murry, Thomas 1980. Acoustic attributes of preadolescent voices. *Journal of the Acoustical Society of America* 68. 1262–1265.
- Hu, Yakun – Wu, Dapeng – Nucci, Antonio 2012. Pitch-based gender identification with two-stage classification. *Security and Communication Networks* 5/2. 211–225.
- Huber, Jessica L. – Stathopoulos, Elaine T. – Curione, Gina M. – Ash, Theresa A. – Johnson, Kenneth 1999. Formants of children, women, and men: the effects of vocal intensity variation. *Journal of the Acoustical Society of America* 106/3. 1532–1542.
- Huszár Ágnes 2009. *Bevezetés a gendernyelvészetbe*. Tinta Könyvkiadó, Budapest.
- Kassai Ilona 2002. *Fonetika*. Nemzeti Tankönyvkiadó, Budapest.

- Lass, Norman J. – Davis, Margaret 1976. An investigation of speaker height and weight identification. *Journal of the Acoustical Society of America* 60/3. 700–703.
- Lass, Norman J. – Hughes, Karen R. – Bowyer, Melanie D. – Waters, Lucille T. – Bourne, Victoria T. 1976. Speaker sex identification from voiced, whispered, and filtered isolated vowels. *Journal of the Acoustical Society of America* 59/3. 675–678.
- 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.
- Li, Ming – Han, Kyu J. – Narayanan, Shrikanth 2013. Automatic speaker age and gender recognition using acoustic and prosodic level information fusion. *Computer Speech and Language* 27/1. 151–167.
- Müller, Christian 2006. Automatic recognition of speakers' age and gender on the basis of empirical studies. In *Proceedings of Interspeech 2006*. Pittsburgh. 17–21.
- Perry, Theodore L. – Ohde, Ralph N. – Ashmead, Daniel H. 2001. The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America* 109/6. 2988–2998.
- Ptacek, P. H. – Sander, E. K. 1966. Age recognition from voice. *Journal of Speech and Hearing Research* 9. 273–277.
- Rao, Rajeshwara R. – Nagesh, A. 2011. Source feature based gender identification system using GMM. *International Journal on Computer Science and Engineering* 3/2. 586–593.
- Reynolds, Douglas A. 1995. Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication* 17. 91–108.
- Schwartz, Martin F. 1968. Identification of speaker sex from isolated voiceless fricatives. *Journal of the Acoustical Society of America* 43/5. 1178–1179.
- Traunmüller, Hartmut – van Bezooijen, Renée 1994. The auditory perception of children's age and sex. In: *Proceedings ICSLIP*. 1171–1174. <http://www2.ling.su.se/staff/hartmut/barn.pdf> (Letöltés ideje: 2013.01.03.)
- Vorperian, Hourii K. – Wang, Shubing – Chung, Moo K. – Schimek, E. Michael – Durtschi, Reid B. – Kent, Ray D. 2009. Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of the Acoustical Society of America* 125/3. 1666–1678.
- 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.
- Whiteside, Sandra P. – Hodgson, Carolyn – Tapster, C. 2002. Vocal characteristics in pre-adolescent and adolescent children: A longitudinal study. *Logopedics Phoniatrics Vocology* 27. 12–20.
- Xue, Steve A. – Wing Chi Cheng, Regine–Manwa Ng, Lawrence 2010. Vocal tract dimensional development of adolescents: An acoustic reflection study. *International Journal of Pediatric Otorhinolaryngology* 74. 907–912.

Köszönöm Beke Andrásnak a gépi osztályozásban nyújtott segítségét.