

Gyermekek nemének és életkorának meghatározása a beszédük alapján

Jelige: „virágcserep”

Gyermekek nemének és életkorának meghatározása a beszédük alapján

1. 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). Arra vonatkozóan, hogy melyek azok az akusztikai jegyek, amelyek meghatározóak az azonosításban, számos hazai és nemzetközi kutatás született.

A beszélő testalkatának becslése a hang alapján nehéz feladatnak bizonyul. Gonzalez (2003) kutatásában a magasság és a testsúly csak 14%-ban korrelált a becsült értékekkel. A válaszadók a beszédtypustól (pl. suttogott, normál ejtésű beszéd) függetlenül következetesek voltak döntésükben. Gonzalez az alaphangmagasság és a testalkat között nem talált összefüggést. Lass–Davis (1976) kutatásában a válaszadók a véletlennél biztosabban képesek voltak a magasság meghatározására, a testsúly esetében ez az eredmény csak a férfiakra vonatkozóan volt igazolható. A befolyásoló akusztikai jellemzők közül az alaphangmagasságot, az intonációt, a hangerőt és a levegős hangképzés fokát emelte ki. Gósy (2001) testalkat- és életkorbecsléssel foglalkozó tanulmányában a beszélők testsúlyát 38%-ban, termetét 56%-ban határozták meg helyesen a válaszadók. Gósy a beszédtempó, az alaphangmagasság, a frekvencia és az intenzitás állandóságának befolyásoló szerepét vizsgálta: a beszédtempó és a testalkat között nem talált összefüggést. A magasabb alaphangú nőket magasabbnak, fiatalabbnak és vékonyabbnak becsülték. A mély hangú nők és férfiak magasságának megítélése sikeresebbnek bizonyult. Gósy (2001) feltételezte, hogy az életkor becslése azért lehet sikeres a testalkatéval szemben, mivel az életkor konkrét, a termet és a súly viszont relatív értékek. Ennek következményeként feltételezte, hogy „az agy neurális spektogramjai őriznek adatokat az életkorról”. Evans et al. (2006) a vokális jellemzők és a testalkat közötti összefüggéseket kutatta. Férfihangok vizsgálata alapján bizonyította, hogy a nők a mély alaphangú férfiakat tartják vonzóknak, és ezekhez a hangokhoz idősebb, izmosabb testalkatot társítanak. Hasonló eredményre jutott kutatásában Gocsál–Huszár (2003) és Collins (2000) is, utóbbi eredményeit biológiai alapon magyarázta: az alaphangfrekvencia a tesztoszteron hormon szintjével hozható összefüggésbe, ezért a nők számára a mély alaphang a jó gének hordozóját „jelenti”. A formánsfrekvenciák és a testalkat között ugyanakkor nem

talált összefüggést Collins. A női hangokat tekintve a magasabb alaphangú nőket vonzóbbnak és fiatalabbnak ítélték meg a férfiak (Collins–Missing 2011).

A beszéd és a személyiség közötti összefüggések vizsgálatával foglalkozott Gocsál (2010), aki a beszéd akusztikai paramétereit vetette össze a beszélő személyiségvonásaival, a Big Five vonáselméletet használva (Christal–Tupes 1961, idézi Mirnics 2006). A kutatásában elemzett akusztikai paraméterek közül a szünettartás korrelációt mutatott a barátságosság vonásával. A beszélők megítélésére vonatkozóan a mély alaphangmagasság a női beszélőknél összefüggött a nyugodtság vonásával. A beszédtempó értékei szintén a női beszélőkről alkotott benyomások kialakításában játszottak szerepet. Krauss et al. (2002) kutatásában a résztvevőknek a beszédhangokat fényképekhez kellett társítaniuk. A kísérlet a véletlennél jobb eredményeket hozott. A szerző az alaphangmagasság és a beszéd folyamatosságának mutatóját nevezte meg befolyásoló tényezőként a személyiség és a beszéd kapcsolatában. A beszédhang létrehozójának emocionális állapotára is utal. Scherer (1995, idézi Gocsál 1998) kutatásában azt találta, hogy az öröm növeli az alaphangmagasság átlagát, variabilitását és energiáját, míg a szomorúság ezzel ellentétes hatást vált ki.

A beszélői profilalkotás másik kutatási területe a nem és az életkor becslésére irányul, amelyben a beszédhangot jellemző nem- és életkorspecifikus paramétereket keresik. Fiziológiailag az emberi hangképzés alapja a kilégzés okozta levegőáramlás, a hangszalagok fonációs beállítottsága és rezgése, illetve a toldalékcső rezonátortevékenysége (Gósy 2004). Az elsődleges hangképzési terület a gége, az itt keletkezett hang továbbhalad a gége feletti üregekbe, ahol rezgésformája módosul. A beszédképzésben résztvevő szervek (tüdő, gége, toldalékcső) az életkor előrehaladtával változnak, ami a beszéd akusztikai lenyomatának változását eredményezi (Gósy 2004). Gyermekkorban a gége nagysága, a hangmagasság, a hangterjedelem és az átlagos beszédhangképzés lényegében egyforma, a nemek tekintetében nincs különbség (Balázs 1993). A hangszalagok hossza a gyerekeknél kb. 1 cm (10 éves korig), a felnőtt nőknél 1,5–2 cm, a férfiaknál 1,9–2,9 cm (Gósy 2004). Ez a fizikai növekedés a zöngé alaphangjának értékét is befolyásolja, hiszen az alaphangmagasság a gége méreteitől és a hangszalagok hosszától függ (Kassai 2002; Smith–Patterson 2005). A felnőtt férfiak átlagos alaphangmagassága 80–140 Hz, a nők 160–260 Hz, míg a gyermekeké az 500Hz-et is elérheti (Kassai 2006). Anatómiailag 4 és 12 éves kor között a leglátványosabb változás, a 4 éves kori állapothoz viszonyítva a férfiak alaphangmagassága 21%-kal, a nők 20%-kal csökken (Smith–Patterson 2005).

Az életkori változások a rezonátorként funkcionáló artikulációs csatornát is érintik, amelynek hossza az újszülötteknél kb. 7–8 cm, a nőknél kb. 14 cm (térfogata kb. 130 cm³), a

férfiaknál kb. 17 cm, térfogata kb. 170 cm³ (Németh–Olaszy 2010). A vokális traktus növekedésének mértéke és tempója az anatómia és a fonetika sokat kutatott területe (l. pl. 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 et al. (2001) ezzel szemben azt találta, hogy a beszéd akusztikumában a nembeli eltérések már négyéves korban is kimutathatók.

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). Több kutatás is igazolta, hogy az életkor előrehaladtával a formánsértékek csökkennek (Huber et al. 1999; Lee et al. 1999). Huber et al. (1999) kutatásának tanúsága szerint a formánsok értékei a gyermekeknél a legmagasabbak és a felnőtt férfiaknál a legalacsonyabbak. A formánsértékek az artikulációs gesztusokkal mutatnak összefüggést: az első formáns az állkapocs nyitásszögével és a nyelvállás fokával, a második formáns a nyelv vízszintes mozgásával és az ajakműködéssel áll kapcsolatban (Kassai 2002). A nemek közti anatómiai és beszédakusztikai különbségek vizsgálata igazolta, hogy a fiúkra és a lányokra eltérő tendenciák jellemzőek növekedésük ütemében és a hormonális változásokban (Childers–Wu 1991b). Ennek következményeként a fiúk formánsértékei minden életkorban alacsonyabbak a velük egykorú lányokéinál (Lee et al. 1999).

A fiziológiai változások akusztikai vetületeinek vizsgálatát célzó kutatások egyik fő iránya az alaphangmagasság elemzésével foglalkozik. A szakirodalom az eltérő metodológiai eljárások miatt nem egységes az alaphangmagasság-változás életkorra vonatkozó megítélésében. Hasek (1980) 5 és 10 éves korú gyermekek beszédének vizsgálatakor azt találta, hogy 5–6 éves korban még nincs eltérés az F0-ban a nemek szerint, 7 és 10 éves kor között viszont a fiúk alaphangmagassága alacsonyabb volt a velük azonos korú lányokénál. Whiteside–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 három éves periódusban vizsgálta az F0 változását 8 és 11 éves kor között: nem talált eltérést a nemek között. Perry et al. (2001) 4, 8, 12 és 16 évesek magánhangzóinak vizsgálatakor azt találta, hogy a nemeket jellemző alapfrekvencia-változás csak 12 éves korban jelenik meg a beszédben, ugyanakkor a magánhangzók formánsértékei már gyermekkorban is mutatják a nembeli eltéréseket. Busby és Plant (1995) kutatásának

tanúsága szerint a nemek észlelése 4 és 12 éves kor között elsődlegesen a formánsok, 12 éves kor felett pedig az F0 által biztosított. A döntő változást tehát a serdülőkor hozza el.

Az életkor és a nem beszéd alapú meghatározásában az alaphangmagasság mellett a legtöbb kutatás a formánsok szerepét vizsgálja (l. pl. Bennett 1981; Busby–Plant 1995;). A magánhangzók formánsértékei a nemek szerint már a gyermekkorban elkülönülnek, a lányok értékei jellemzően magasabbak (Bennett 1981; Busby–Plant 1995). Busby és Plant (1995) szerint 5 és 7 éves kor között a formánsértékek változása (csökkenése) nagyobb mértékű, mint 9 és 11 éves kor között. A nemek tekintetében az F2 értékében általánosan, míg az F1 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 et al. (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.

Smith–Patterson (2005) szerint a nem és az életkor beszédbeli eltéréseit a hangszalag és a vokális traktus különböző mérete okozza, ami az akusztikai kimenetben eltérésekhez vezet: egy magánhangzó esetében az alaphangmagasság és a formánsfrekvenciák változását (is) jelenti, hiszen a vokális traktus (méretétől és hosszától függően) más-más frekvenciákat erősít fel. Kutatásában a VTL (vocal tract length) és a GPR (glottal pulse rate) interakcióját elemezve határozta meg a jegyek nemre és életkorra gyakorolt hatását, izolált magánhangzók felhasználásával.

Lass et al. (1976) kutatásában suttogott, normál ejtésű és szűrt magánhangzók segítségével vizsgálta a nem beszéd-alapú meghatározását. Azt találta, 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 a következtetett, 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).

Traunmüller–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 F0 mellett szupraszegmentumok (pl. beszéddallam) is szerepet játszhatnak. Bennett–Weinberg (1971, idézi Perry et al. 2001) spontán beszéden végzett kutatásában az 5–

6 évesek nemét átlagosan 75%-ban azonosították helyesen a válaszadók. Amir et al. (2012) 8 és 10 éves gyerekek beszédének (felolvasott mondatok) 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 (pl. beszéddallam, hanglejtés) tartotta dominánsnak. Bachorowski–Owren (1999) felnőtt beszélőkkel végzett kutatásukban a VTL (vokális traktus hossza) és az alaphangmagasság szerepét emelték ki a nemek meghatározásában. Linville–Fisher (1985) kutatásában nők életkorát kellett megbecsülniük a résztvevőknek normál és suttogott ejtésű beszédminták alapján. A válaszadók átlagosan 75%-ban azonosították helyesen a beszélők életkorát. Linville–Fisher szerint ehhez normál ejtésű beszédben az F0, suttogott ejtésben pedig az F1 szolgáltatott domináns információkat.

Az életkor becslésére vonatkozó kutatások változatos eredményeket hoztak: Amir et al. (2012) kísérletében az életkor becslése 37%-ban volt sikeres 8–18 évesek beszéde alapján; Gocsál elemzésében (1998) következetes, de pontatlan becslésekről írt. Gósy (2001) kutatásában az életkor meghatározása 70%-ban, Ptacek–Sander (1966) kísérletében 78%-ban volt sikeres. A gyermek, felnőtt és időskori hangok felismerésében meghatározó akusztikai kulcs volt az alaphangmagasság (Gósy 2001). Gorham–Rowan–Laures–Gore (2006) szerint a fiatal és idős beszélők elkülönítése az időskori beszéd sajátosságai miatt lehetséges, ilyen például az alaphangmagasság csökkenése, a hangtartás rövidülése, a hangerő csökkenése, a beszédtempó lassabbá válása és a fokozott szünettartás. Ez az akusztikai lenyomat az időskorban végbemenő anatómiai változások következménye (Balázs 1993).

A formánsfrekvenciák és az alaphangmagasság mellett vizsgálták még a beszédtempó (Gósy 2001), a zöngékezdési idő (Bóna 2011) és más akusztikai jegyek szerepét is (l. Amir et al. 2012) az életkor meghatározásában.

A nem és életkor becslésére vonatkozó kutatások módszertani, korpuszbeli és résztvevői oldalainak változatossága miatt az eredmények kevéssé összevethetők. A férfi és női beszédhangok eltérései részben magyarázhatók az anatómiai különbségekkel, ugyanakkor a paralingvisztikai jegyek, a tanult „nem-specifikus” viselkedésminták és azok akusztikai vetületei (például a nőkre jellemzőbb levegős hangképzés) hatása sem hagyható figyelmen kívül elemzésük során (Childers–Wu 1991a). A testalkat szintén befolyásoló tényező lehet (Xue et al. 2010).

Az anatómiai különbségek akusztikai vetületeiből látható, hogy az alaphangmagasság és a formánsfrekvenciák fontosak az életkor és a nem meghatározásában a felnőttek esetében. Az óvodás és általános iskolás korú gyermekeknél anatómiailag a beszédképző szervek

nagyságában nincs különbség, ezért nem beszélhetünk ezek kapcsán akusztikai következményekről (Lee et al. 1999). A humán percepció számára a beszéd alapján a nemek elkülönítése mégis lehetséges (Perry et al. 2001). A jelen kutatásban arra keresem a választ, hogy melyek azok az akusztikai jellemzők, amelyek segítségével a beszélők neme és életkora elkülöníthető gyermekkorban. A felnőttkort célzó kutatások alapján feltételeztem, hogy 1. a gyermekek neme és életkora a beszéd alapján meghatározható; 2. a felismerés pontossága az életkor előrehaladtával javul; 3. az életkor és a nem meghatározásában az alaphangmagasság és a magánhangzó-formánsok fontos akusztikai kulcsként működnek.

A beszélő nemének és korának meghatározása az automatikus osztályozó rendszerekben is fontos lépés, mivel a nem-függő modellek működése pontosabb a nem-függetleneknél (Rao et al. 2011). Az osztályozók felhasználása sokrétű, l. például multimédia alkalmazásokban (Harb–Chen 2005) és a humán-gép interfész interakcióban (Feld et al. 2010).

A beszélőfelismerő rendszerek létrehozásakor kiemelten fontos azoknak az akusztikai jellemzőknek a megtalálása, amelyek az adott beszélőre releváns információkat hordoznak (Beke 2012a). A beszélő nemének és életkorának meghatározásakor elterjedt az alaphangmagasság (Hu et al. 2012) és a hallás átviteli karakterisztikáját modellező ún. MFCC-együtthatók (Mel-frekvenciás kepsztrális együtthatók) használata a jellemzőkinyerésben (Rao et al. 2011). Az osztályozás során fontos a releváns, az osztályok elkülönítésében nagy szeparációs képességgel bíró jellemzők bevonása (Beke 2012b). Erre számos osztályozó algoritmus ismert, például GMM (kevert Gauss-modell), SVM (szupportvektorgép), ANN (neurális hálózat), HMM (rejtett Markov-modell), illetve ezek kombinációi (Németh–Olaszy 2012).

A nem és életkor gépi osztályozásával számos kutatás foglalkozott. Rao et al. (2011) MFCC-alapú GMM felismerője a felnőtt beszélők nemét 73%-ban osztályozta helyesen. Hasonló módszerrel Gaikwad et al. (2012) 90%-os teljesítményt ért el. Az alaphangmagasság bevonásával Harb–Chen (2005) az osztályozás pontosságát 98%-ig, Hu et al. (2012) 98,7%-ig tudta növelni. A jelen kutatásban használt SVM-GMM osztályozóhoz hasonló modellel Li et al. (2013) 76%-os felismerési eredményt ért el. Magyar beszéden végzett kutatásában Beke (2012a) MFCC-alapú GMM-UBM osztályozója a felnőtt beszélők nemét 90%-os eredménnyel tudta elkülöníteni.

Az életkor osztályozását célzó, kettőnél több életkori csoport bevonásával készült kutatások többsége a véletlen alatti eredményeket hozott (l. pl. Feld et al. 2010; Li et al. 2013). Shafran et al. (2003, idézi Schötz 2007) HMM-alapú osztályozója öt korcsoport

azonosításában 70%-os felismerési pontosságot ért el. A középkorú és idős beszélők osztályozása GMM és ANN-modellekkel 90%-os volt (Minematsu et al. 2002, idézi Schötz 2007; Müller et al. 2003, idézi Schötz 2007). Müller (2006) kutatásában további jellemzők (pl. jitter, shimmer, HNR, F0) és további korcsoportok bevonásával 63,5%-os felismerési eredményt ért el.

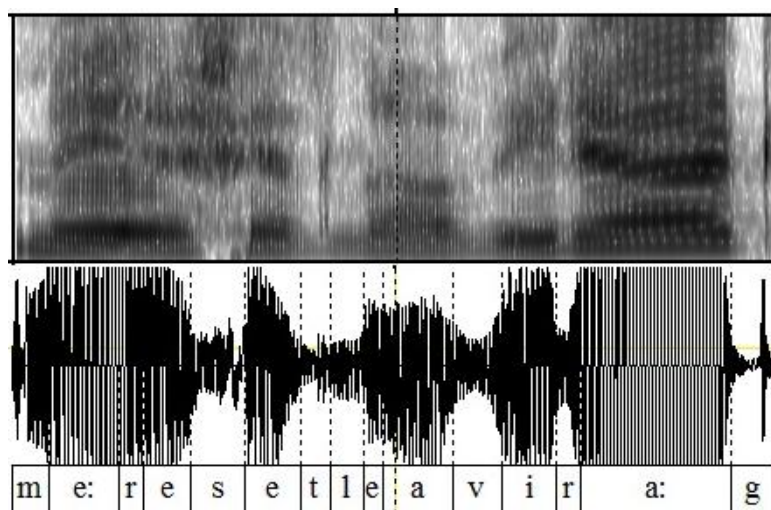
2. Anyag, módszer, kísérleti személyek

A jelen kutatásban 14 óvodás (életkoruk 5 év) és 14 általános iskolás (életkoruk 10 év) spontánbeszéd-anyagát használtam fel, amelyet óvodai és iskolai környezetben rögzítettem. 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 beszédprodukcióra is hatással vannak (Gósy 2004).

Minden gyermektől ugyanazon képsorozat leírását rögzítettem (l. melléklet). Az azonos feladat lehetővé tette a nembeli (pl. szókincsbeli) eltérések előfordulási valószínűségének csökkentését. Az így rögzített átlagosan 1 perces hanganyagokból kiválasztottam egy-egy 7 másodperces részletet, amely a kutatásban használt percepciós teszt anyagát képezte. Ez többségében az első kép leírását rögzítette. A percepciós tesztet 30 egyetemi hallgató töltötte ki (életkoruk 20–27 év, többségében nők). Feladatuk az volt, hogy egyszeri meghallgatás után határozzák meg a beszélő nemét és életkorát (óvodás vagy iskolás).

A hanganyag annotálását a Praat 5.3 szoftver (Boersma–Weenink 2012) segítségével készítettem el (l. ábra). A percepciós tesztben használt 7-7 másodperces részleteken akusztikai elemzést végeztem az alaphangmagasságra vonatkozóan, amelyet 95%-on újrászűrtem, az esetleges környezeti zajok kizárása miatt. A magánhangzók formánsainak mérését a hanganyagban leggyakrabban előforduló hét magánhangzó esetében végeztem el, ezek az [ɔ], [a:], [ɛ], [e:], [i], [o], [u] voltak. A formánsok mérésére az erre a célra írt szkriptet használtam, a kinyert adatokat ebben az esetben is újrászűrtem, 95%-on.

A statisztikai elemzést az SPSS 13. program segítségével végeztem el. Az elemzett akusztikai jellemzők szerepét egytényezős varianciaanalízis (ANOVA) segítségével elemeztem, minden esetben 95%-os konfidencia-intervallumban.



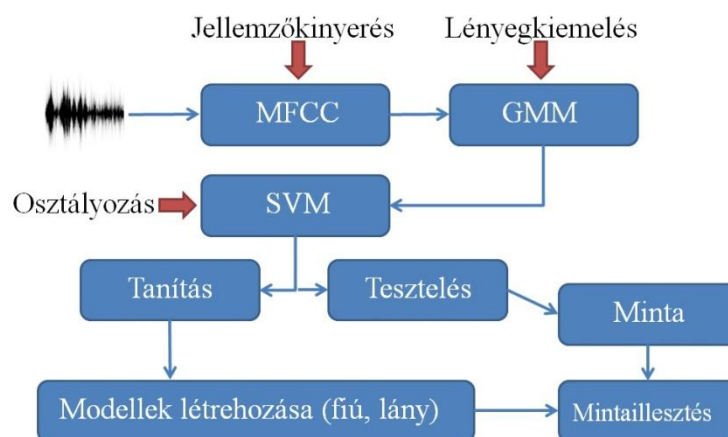
1. ábra

A kutatásban használt iskolás (fiú) beszélő hangszín- és rezgéseképe

Az nem- és életkor felismerésének tesztelésére létrehoztam egy gépi osztályozót is. Az osztályozáshoz MFCC-együtthatókat és GMM-SVM algoritmust használtam. Az osztályozót a MATLAB szoftverben valósítottam meg.

A beszélőfelismerésben az MFCC kiszámítását a beszéd előfeldolgozása előzi meg, amelynek során előkiemelést hajtanak végre (a jel/zaj viszony javítására), illetve kiszűrik a beszédjelet nem tartalmazó részeket, ezt beszéd-detektálásnak nevezik (Beke 2012a). 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 tanítói modellel, azaz mintaillesztést hajt végre. A gépi osztályozás menetét a 2. ábra foglalja össze.



2. ábra

A gépi osztályozás menete

3. Eredmények

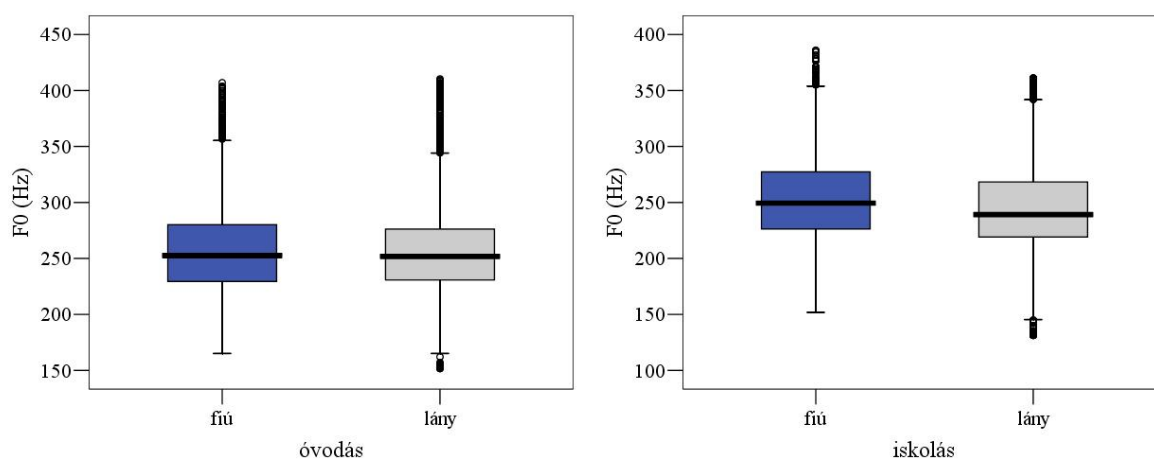
3.1. Az akusztikai elemzés eredményei

3.1.1 Az alaphangmagasság

A gyermekek alaphangmagassága a nemek függvényében óvodáskorban nem mutatott eltérést (ANOVA: $F(1, 4210) = 0,617$; $p = 0,432$), hasonlóan Hasek (1980) eredményeihez (3. 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 ($F(1, 16289) = 161,471$; $p < 0,005$). 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 F0 elkülönülésének idejét (vö. Perry et al. 2001). Deme (2012) 6–7 évesek spontán beszédének vizsgálatakor – a jelen eredményekhez hasonlóan – azt találta, hogy a lányok átlagos alaphangmagassága magasabb. A szerző szerint a lányoknál 6–7 éves korra tehető egy jelentős ugrás az F0 csökkenésében, amely a fiúknál csak később jelentkezik. Ezzel magyarázható, hogy átmenetileg a lányoknál alacsonyabb alaphangmagasság mérhető. 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, és nemenként eltérő ütemben zajlik (vö.

Hasek 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).



3. ábra

Az alaphangmagasság eloszlása nemek és életkorok szerint

3.1.2 A magánhangzóformánsok

A magánhangzók formánsértékei a testi növekedés során csökkennek, azonban az értékek alakulásában az életkor mellett a nemek közti különbségek is megmutatkoznak (Deme 2012). Ezek a nagyothalló és ép halló gyermekek esetében is fennállnak a formánsszerkezetben (Beke 2010). 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 F3 értéke a beszélőre jellemző paramétereket hordozhat (Gósy 2004).

Deme (2012) hat fonéma realizációit vizsgálta 6–7 évesek spontán beszédében. Az F1 esetében az [a:] és [o], az F2-nél az [ɛ], [i], az F1 és F2-ben pedig az [e:] esetében talált eltérést a nemek között. A jelen kutatásban az óvodások esetében az F1-ben az [ɔ], [e:], az F1 és F2-ben az [o], az F3-ban az [ɔ], [a:], [ɛ] hangok esetében találtam szignifikáns eltérést a nemek között (vö. Deme 2012). A statisztikai próba (ANOVA) eredményét az 1. táblázatban foglaltam össze. Az [ɔ], [a:], [ɛ] hangok esetében kimutatott különbségek hasonlóak a külföldi szakirodalomban foglaltaknak (vö. Busby–Plant 1995). Az [ɔ], [a:], [ɛ] hangok esetében az F3 is szignifikáns különbséget mutatott a nemek között. Perry et al. (2001) kutatásában azt

találta, hogy az F3 értéke az életkorra és a nemre is érzékeny bizonyos magánhangzók esetében. Ebből arra következtethetünk, hogy 5 éves korban az alsó nyelvállású hangzóknál az F3 hordozhat a nemek elkülönítésére alkalmas információkat.

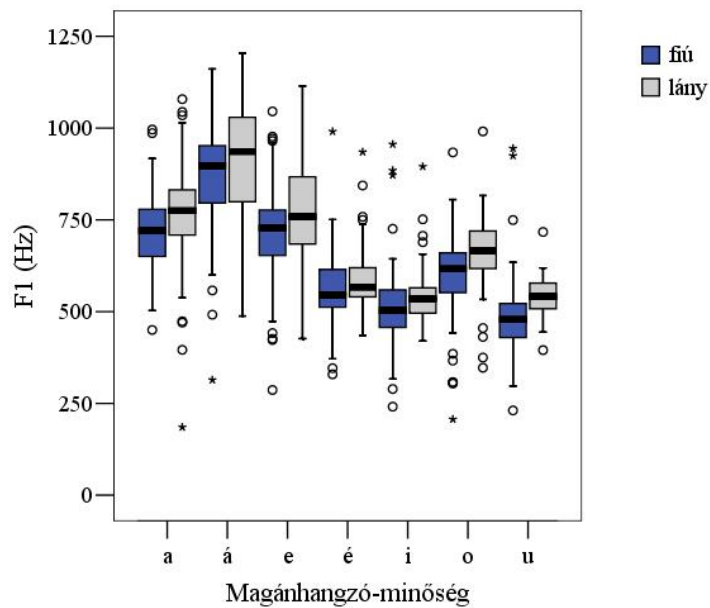
1. táblázat

A magánhangzók formánsértékeinek nemenkénti eltérése óvodáskorban (a szignifikáns eredményeket a szürke szín jelzi)

	[ɔ]	[a:]	[ɛ]	[e:]	[i]	[o]	[u]
F1	p < 0,05	p = 0,170	p < 0,05	p = 0,076	p = 0,039	p < 0,05	p = 0,072
F2	p = 0,694	p = 0,333	p = 0,064	p = 0,095	p = 0,111	p < 0,05	p = 0,458
F3	p < 0,05	p < 0,05	p < 0,05	p = 0,301	p < 0,05	p = 0,347	p = 0,828

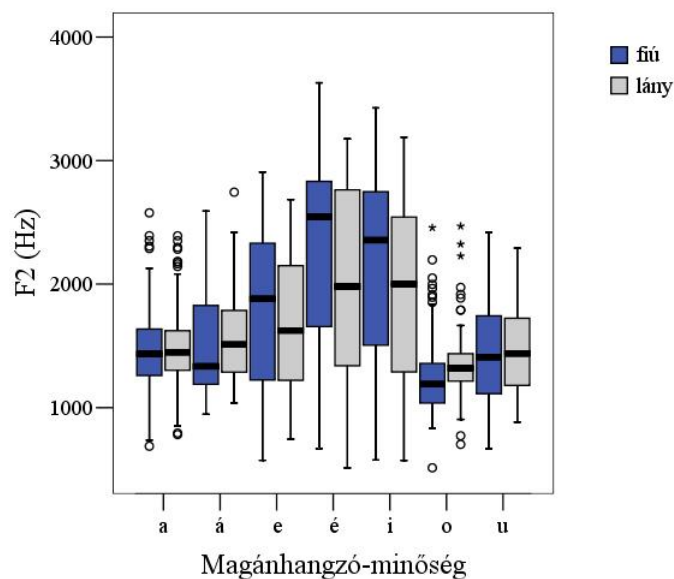
A magánhangzó-minőségek közötti eltérések utalhatnak hangképzésbeli különbségekre, amely fakadhat a toldalékcso méretbeli különbözőségeiből (Lee et al. 1999). Ezt bizonyíthatja, hogy az eredmények szerint a legtöbb magánhangzó esetében a fiúk formánsértékei alacsonyabbak a lányokénál. A jelen kutatásban a hangképző szervek méretére vonatkozó adat nem állt rendelkezésemre, ezért csak a szakirodalmi tapasztalatoknak megfelelően tudom valószínűsíteni, hogy a fiúk testméretei már ebben az életkorban is nagyobbak a lányokénál (vö. Perry et al. 2001). Ugyanakkor a hangképzésbeli – artikulációs – különbségek esetében a szociális tényezők befolyása sem hagyható figyelmen kívül.

A 4. és 5. ábrán látható, hogy [ɔ], [a:], [ɛ] hangok az F1 mentén nagy szóródással realizálódtak, ami a nyelvállásfok artikulációjának variabilitására utal. Ez a változatos megvalósulás az F2-paraméter esetében nem jelentkezett, egyenletesebbek voltak az értékek. Az adatok közötti átfedések mind az F1, mind az F2 esetében nagyok voltak, ami arra utal, hogy a hangzók minőségének ejtésében valószínűsíthetően kicsi volt a különbség. Az [ɔ], [a:], [ɛ] hangok ejtése a fiúknál zártabb, az [ɛ], [e:], [i] velárisabb volt, hasonlóan Deme (2012) eredményeihez.



4. ábra

A magánhangzók első formánsainak értékei nemek szerint óvodáskorban



5. ábra

A magánhangzók második formánsainak értékei nemek szerint óvodáskorban

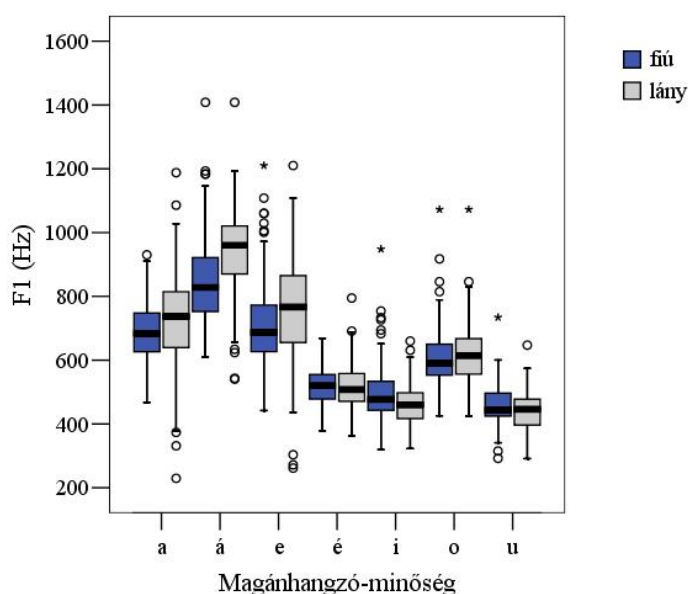
A 10 éves korú gyermekeknél a nemek szerint az F1-paramétere az alsó és legalsó nyelvállású hangzók ([ɔ], [a:], [ɛ]) és az [i] esetében mutatott szignifikáns eltérést, az F2 az [a:], [ɛ] és [o] hang esetében, az F3 pedig az [e:] és [i] hangoknál. A statisztikai próba (ANOVA) eredményeit a 2. táblázatban foglaltam össze.

2. táblázat

A magánhangzók formánsértékeinek nemenkénti eltérése iskoláskorban (a szignifikáns eredményeket a szürke szín jelzi)

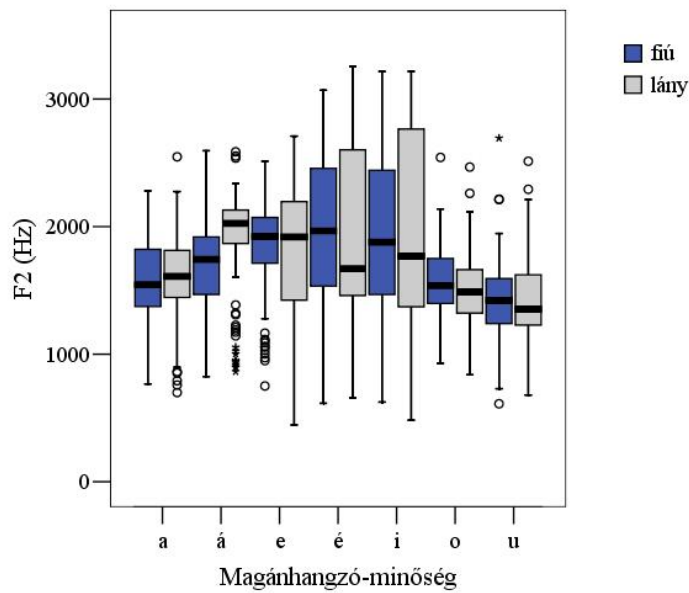
	[ɔ]	[a:]	[ɛ]	[e:]	[i]	[o]	[u]
F1	p < 0,05	p < 0,05	p < 0,05	p = 0,724	p < 0,05	p = 0,164	p = 0,179
F2	p = 0,249	p < 0,05	p < 0,05	p = 0,198	p = 0,894	p < 0,05	p = 0,880
F3	p = 0,290	p = 0,31	p = 0,809	p < 0,05	p < 0,05	p = 0,837	p = 0,537

Az óvodások formánsértékeivel összehasonlítva a legjelentősebb változás az alsó és legalsó nyelvállású hangok esetében jelent meg: az [a:] és [ɛ] hangok a nem szerint elkülöníthető váltak, hiszen az beszédhangot definiáló F1 és F2 értékében is szignifikáns különbséget adatoltam (vö. 6. és 7. ábra). Az eredmények alapján feltételezhető, hogy az alsó és legalsó nyelvállású hangok érzékenyebbek a fizikai növekedéssel bekövetkező változásokra (pl. toldalékcso méretének változása), hiszen az artikulációt tekintve ezeknél a hangoknál a legnyíltabb szájüreg és a legtágabb szájnyílás. Az [ɔ], [a:], [ɛ] hangok F1-értékei az óvodásokéhoz hasonlóan nagy szóródást mutatnak, ami változatos artikulációs megvalósulásukra utal. A fiúk ejtése az [ɛ] hangok esetében még zártabb, mint a lányoké. Az F2-paramétert tekintetében az értékek eloszlása egyenletesebb, amely összevetve az 5 éves kori értékekkel, az artikuláció stabilizálódására is utalhat. A variabilitás az F2 esetében az [ɛ], [e:], [i] hangoknál a legnagyobb.



6. ábra

A magánhangzók első formánsainak értékei nemek szerint iskoláskorban



7. ábra

A magánhangzók második formánsainak értékei nemek szerint iskoláskorban

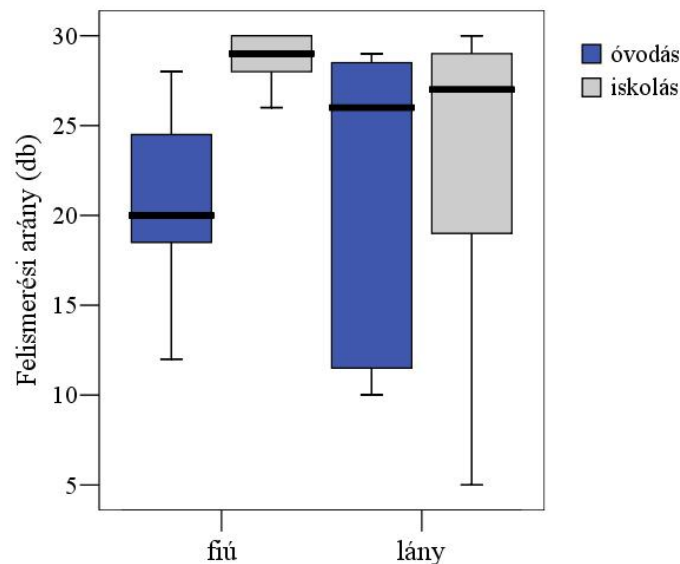
3.2 A percepció teszt eredményei

A jelen kutatásban az óvodások nemét átlagosan 69%-ban, az általános iskolás gyerekekét átlagosan 85,3%-ban osztályozták helyesen a válaszadók. 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, vagyis a két csoport elkülönítése között látszólag nincs különbség. A dobozdiagramon azonban látható, hogy a lányok felismerése a fiúkénál nagyobb szóródást mutat (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%). Ahogy a diagramon is látható, a lányok osztályozása ebben az életkorban is nagyfokú bizonytalanságot mutatott, a szórás értéke 9,7 volt, szemben a fiúk 1,5-es értékével.

Általánosan elmondható, hogy az életkor előrehaladtával javult a felismerés pontossága. Az iskoláskorú gyerekeknél a pontosabb felismerésben az alaphangmagasság és egyes hangoknál a magánhangzó-formánsok fontos akusztikai kulcsként működhetett a nemek felismerésében. Az óvodások esetében a magánhangzó-formánsok mellett más akusztikai jegyek (pl. szupraszegmentális térhez köthető jellemzők) lehettek befolyásolóak az észlelésben. A fiúk felismerése mindkét korcsoportban sikeresebbnek bizonyult. 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. Traunmüller–Bezooijen 1994). A percepció teszt eredményeit a 8. diagram mutatja.



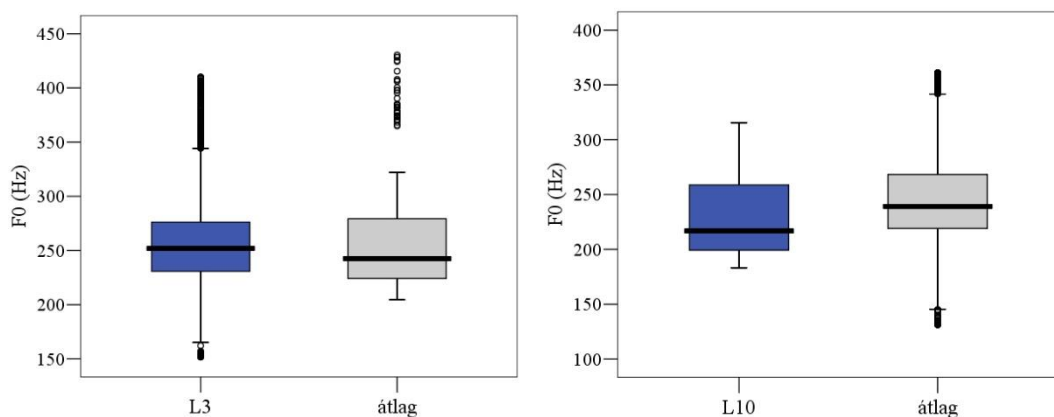
8. ábra

Felismerési eredmények a nemek és az életkor függvényében

Elemeztem a felismerési eredményeket az életkor tekintetében. 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 életkor tekintetében elsődleges 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 – ahogy azt több kutatás tapasztalatai is megerősítették, l. például Ptacek–Sander (1966) –, hogy az óvodáskorú gyermekek életkorának felismerése a humán percepcióban relatíve biztosnak mondható.

A humán percepcióban volt olyan beszélő, akinek nemét csak a válaszadók elenyésző része tudta helyesen meghatározni, ezért megvizsgáltam a leggyengébb felismerési arányt elért beszélők eseteit. Az óvodások csoportjában ez egy 33%-os felismerési arányt jelentett, a beszélő lány (továbbiakban L3) volt. L3 alaphangmagasságát (264,1 Hz) összehasonlítottam az óvodás lányoknál mért átlagos alaphangmagassággal (261,6 Hz), azonban nem találtam szignifikáns eltérést (ANOVA: $F(1, 2239) = 0,731$; $p = 0,393$). Az észlelésben valószínűsíthetően más akusztikai jellemzők (pl. beszédhang, hangsúly) játszhattak

szerepet. Az általános iskolások csoportjában a leggyengébb felismerési arányú beszélő szintén egy lány volt (L10), akinek nemét a válaszadók mindössze 17%-ban osztályozták helyesen. Megvizsgáltam L10 alaphangmagasságát az átlag függvényében, és szignifikáns eltérést találtam (ANOVA: $F(1, 10141) = 54,813$; $p < 0,005$). Az F0 L10 (230,4 Hz) esetében alacsonyabb volt az átlagosnál (246,4 Hz), amely az észlelésben a fiúként való azonosítást eredményezhette. Bár a kutatásban mért adatok nem támasztják alá a fiúk alaphangjának mélyebb voltát ebben az életkorban, a humán percepcióban meglévő előzetes tapasztalatok a mélyebb alaphanghoz férfi beszélőt társítanak (Lee et al. 1999). A leggyengébb felismeréseket a 9. ábra mutatja.



9. ábra

A leggyengébb felismerési arányú beszélők (balra az óvodás, jobbra az iskolás)
alaphangmagassága az átlaghoz viszonyítva

3.3 A gépi osztályozás eredményei

A kutatásban létrehozott MFCC-alapú SVM-GMM osztályozó az óvodások nemét 48%-ban, az általános iskolásokét 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 korpusz és az életkori csoportok (gyerekek) sajátosságai miatt az eredmények a külföldi kutatásokkal csak fenntartásokkal vethetők össze (vö. Li et al. 2013). A gépi osztályozás az iskolások csoportjában a lányok esetében sikeresebb volt, ami a jelen kutatásban kapott humán percepció eredményeivel

ellentétes, azonban a külföldi szakirodalmi tapasztalatoknak nem mond ellent (vö. Amir et al. 2012).

A gépi osztályozás eredménye az életkor tekintetében 47%-os volt, hasonlóan több külföldi kutatásához (Feld et al. 2010; Li et al. 2013). 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 is azt találták, hogy az életkort célzó gépi felismerés az egymástól távolabbi életkorok esetében sikeresebb (Müller 2006).

Fontos kihangsúlyozni, hogy a gépi osztályozás eredménye a humán percepcióval csak fenntartásokkal hasonlítható össze, mivel nincsenek biztos ismereteink arról, hogy melyek azok a konkrét akusztikai jellemzők, amelyek szerepet kapnak a nem és életkor azonosításában. A gépi osztályozásban számos jellemző bevonásával próbálják meg modellezni az emberi hallás – ezáltal észlelés – folyamatát (l. pl. 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.

4. Összegzés, következtetések

A jelen kutatásban 5 és 10 éves gyermekek nemének és életkorának becslését vizsgáltam a spontán beszédük alapján. A nem és életkorbecslés a beszélői profilalkotás egyik kutatási területe, amely feltételezi, hogy a beszéd jellemzi a produktumának létrehozóját, vagyis vannak olyan akusztikai paraméterek, amelyek utalnak a beszélő bizonyos tulajdonságára, pl. testalkatára, személyiségére, nemére, életkorára, stb. (Lass–Davis 1976). A magyar nyelvben eddig kis számú kutatás foglalkozott a gyermekek nemének- és életkorának becslésével (l. pl. Gósy 1984, idézi Deme 2012; Deme 2012), illetve a percepció és akusztikum ilyen irányú kapcsolatával.

Célom annak vizsgálata volt, hogy elemezzem, lehetséges-e az óvodáskorú és általános iskolás gyermekek nemének és életkorának elkülönítése, pusztán a beszédük alapján. A nem- és életkorbecslés a tárgyalt életkorban (gyermekkor) azért releváns kérdés, mert több szakirodalom is megerősítette, hogy a gyermekek beszédképző szerveik (toldalékcső, gége) méreteiben nem különböznek a nemek tekintetében egymástól, ebből pedig azt lehetne valószínűsíteni, hogy beszédük akusztikai paraméterei sem térnek el. Perry et al. (2001) kutatásában azt találta, hogy a formánsfrekvenciák már négyéves korban is

mutatják a nemek közötti különbségeket. Ugyanakkor fontos hangsúlyozni, hogy a nemzetközi szakirodalom nem egységes e kérdés megítélésében (vö. Lee et al. 1999).

A jelen kutatásban feltételeztem – köznapai tapasztalataink és külföldi szakirodalmi kutatások eredményei alapján (l. pl. Perry et al. 2001) –, 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 osztályozták helyesen a válaszadók. Második hipotézisem – miszerint 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 osztályozták helyesen a kísérletben résztvevők. Az életkor felismerése mind az óvodások (90%), mind az iskolások (86,4%) esetében sikeresnek bizonyult.

Harmadik hipotézisem az életkor és nem felismerését biztosító akusztikai jellemzőkre vonatkozott. Feltételeztem, hogy az alaphangmagasság és a magánhangzó-formánsok fontos akusztikai kulcsként működnek e jellemzők meghatározásában, hiszen a fizikai növekedés hatására a toldalékcső és a hangszalagok méretében és formájában bekövetkező változások ezeket az akusztikai paramétereket is érintik. Mindemellett több külföldi kutatásban is azt találták, hogy a nemek és az életkor elkülönítésében az F0 és a formánsok fontos funkcióval bírnak (Busby–Plant 1995; Perry et al. 2001). Az akusztikai elemzéséből kiderült, hogy az óvodásoknál az alaphangmagasságban a nemek szerint nem adatolható eltérés, ezért a felismerésben valószínűsíthetően egyéb (pl. a szupraszegmentális térhez köthető jegyek) játszhatnak döntő szerepet. A magánhangzó-formánsok elemzésékor az F1 és F2 értékei csak bizonyos magánhangzók esetében hordoztak különbséget a nemtől függően. Az alsó és a legalsó nyelvállású hangok esetében az F3 értékében eltérést találtam a két csoport között. Az F3-paraméter ezért a jelen kutatás szerint hordozhat a beszélő nemére vonatkozó információt. Az F1-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 „nem-specifikussága” az iskolások csoportjában is megmutatkozott, vagyis ezek érzékenyebbek lehetnek a vokális traktus fizikai növekedés során bekövetkező méretbeli változásokra, mivel képzésüket tekintve a legnyíltabb hangzók.

Az iskolások esetében az alaphangmagasság tekintetében a nemek között adatoltam eltérést. Váratlan eredmény volt a kutatásban, hogy a fiúk alaphangja magasabb volt a lányokénál, hasonló jelenséget Deme (2012) is megfigyelt. Ez ellentmondani látszott annak a „tapasztalatnak”, miszerint a fiúk alaphangmagassága minden életkorban alacsonyabb a velük egykorú lányokénál (Lee et al. 1999). 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 megfigyelhető, ezért vannak olyan időpillanatok a növekedésben, amikor ez a nemenként eltérő ütemű változás a lányoknál – időlegesen – alacsonyabb alaphangmagasságot eredményez. A változás – összevetve Deme (2012) eredményeivel, aki 6–7 évesek spontán beszédét vizsgálta –, 10 éves korra a végéhez közeledhet. Erre a két csoport közötti kis F0-frekvenciaértékek különbsége utalhat. A kérdés pontosabb vizsgálatához azonban további kutatások szükségesek. A növekedésbeli, fizikai változások mellett természetesen feltételezhető egyéb tényezők – pl. szociális, tanult hatások – befolyása is.

Összefoglalásként elmondható, hogy az elemzett akusztikai paraméterek közül a nemek felismerésében ötéves korban az alaphangmagasság nem domináns akusztikai kulcs, a magánhangzó-formánsok (bizonyos hangoknál) viszont befolyással bírhatnak. Az iskolások tekintetében az F0 már domináns jeggyé válik, illetve a formánsértékek is, főként az alsó és legalsó nyelvállású magánhangzók esetében. Az életkorra vonatkozólag az F0 alapvető fontosságú az észlelésben.

Az észlelés és felismerés folyamatában természetesen további jellemzők is szerepet kaphatnak, pl. a testalkat, illetve szociális, 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 (pl. beszédtempó, beszéd dallam, hangsúly) is. A testalkatra vonatkozó információk bevonása (testi és artikulációs jellemzők hatásának vizsgálata) szintén javíthatja a következtetéseket.

A nem- és életkorbecslés a beszélőfelismerő rendszerek fejlesztésében is fontos lépés. A kutatásban megalkotott gépi osztályozó eredményei elmaradtak a humán észleléstől. Az osztályozás hatásfoka más akusztikai jegyek bevonásával javítható. Ehhez azonban a beszéd további akusztikai-fonetikai elemzésére van szükség. A jelenlegi eredmények a beszélői profilalkotás területén szolgálhatnak adalékokkal, hozzájárulnak a nemek közötti különbségek beszédbeli megjelenésének megismeréséhez, ezáltal a beszéd- és beszélőfelismerő rendszerek fejlesztéséhez.

5. 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.

Bachorowski Jo-Anne–Owren, Michael J. 1999. Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America* 106/2. 1054–1063.

Balázs Boglárka 1993. Az időskori hangképzés jellemzői. *Beszédkutatás* 1993. 156–165.

Beke András 2010. Magánhangzó-minőségek nagyothalló gyermekek spontán beszédében. *Beszédkutatás* 2010. 223–242.

Beke András 2011. Szókezdetek automatikus osztályozása spontán beszédben. *Magyar Nyelvőr* 135/2. 226–241.

Beke András 2012a. Beszélőfelismerés kevert-Gauss modellekkel. In, Markó Alexandra szerk.: *Beszédtudomány: az anyanyelv-elsajátítástól a zöngékezdesi időig*. ELTE–MTA, Budapest. 335–353.

Beke András 2012b. Az egyszerre beszélések automatikus osztályozása spontán magyar társalgásokban. In, Bárdosi Vilmos főszerk.: *Tanulmányok. Nyelvtudományi Doktori Iskola. Asteriskos 1*. ELTE BTK, Budapest. 39–53.

Beke András–Szaszák György 2009. A svávariációk automatikus felismerése magyar nyelvű spontán beszédben. *Beszédkutatás* 2009. 148–169.

Bennett, Suzanne 1981. Vowel formant frequency characteristics of preadolescent males and females. *Journal of the Acoustical Society of America* 69/1. 231–238.

Boersma, Paul–Weenink, David 2012. Praat: doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/> (A letöltés ideje: 2013. 01. 04.)

Bóna Judit 2011. A [p, t, k] mássalhangzók zöngékezdesi ideje idősek és fiatalok spontán beszédében és felolvasásában. *Beszédkutatás* 2011. 61–72.

Busby, P. A.–Plant, L. 1995. Formant frequency values of vowels produced by preadolescent boys and girls. *Journal of the Acoustical Society of America* 97/4. 2603–2606.

Childers, D. G.–Wu, Ke 1991a. Gender recognition from speech. Part I: Coarse analysis. *Journal of the Acoustical Society of America* 90/4. 1828–1840.

Childers, D. G.–Wu, Ke 1991b. Gender recognition from speech. Part II: Fine analysis. *Journal of the Acoustical Society of America* 90/4. 1841–1856.

Collins, Sarah A. 2000. Men's voices and women's choices. *Animal Behaviour* 60. 773–780.

Collins, Sarah A.–Missing, Caroline 2011. Vocal and visual attractiveness are related in women. *Animal Behaviour* 65. 997–1004.

Deme Andrea 2012. Óvodások magánhangzóinak akusztikai jellemzői. In, Markó Alexandra szerk.: *Beszédtudomány: az anyanyelv-elsajátítástól a zöngékezdesi időig*. ELTE–MTA, Budapest. 77–99.

Evans, Sarah–Neave, Nick–Wakelin, Delia 2006. Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice. *Biological Psychology* 72. 160–163.

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.

Gaikwad, Santosh–Gavali, Bharti–Mehrotra S. C. 2012. Gender identification using SVM with combination of MFCC. *Advances in Computational Research* 4/1. 69–73.

Gocsál Ákos 1998. Életkorbecslés a beszélő hangja alapján. *Beszéd kutatás 1998*. 122–134.

Gocsál Ákos–Huszár Ágnes 2003. Csábító hangok. *Beszéd kutatás 2003*. 9–18.

Gocsál Ákos 2010. A beszédakusztikai paraméterek és a beszélő személyiségjegyei közötti összefüggések vizsgálata. PhD tézisleírás.
<http://doktori.btk.elte.hu/lingv/gocsalakos/tezis.pdf> (A letöltés ideje: 2013. 01. 03.)

Gonzalez, Julio 2003. Estimation of speakers' weight and height from speech: a re-analysis of data from multiple studies by Lass and colleagues. *Perceptual and Motor Skills* 96. 294–304.

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, T. 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.

Kassai Ilona 2002. *Fonetika*. Nemzeti Tankönyvkiadó, Budapest.

Kassai Ilona 2006. Fonetika. In, főszerk.: Kiefer Ferenc, *Magyar nyelv*. Akadémiai Kiadó, Budapest. 789–835.

Krauss, Robert M.–Freyberg, Robin–Morsella, Ezequiel 2002. Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology* 38. 618–625.

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.

Linville, S. E.–Fisher, H. 1985. Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females. *Journal of the Acoustical Society of America* 78. 40–48.

Mirnic Zsuzsanna 2006. *A személyiség építőkövei*. Bölcsész Konzorcium, Budapest.

Müller, Chirstian 2006. Automatic recognition of speakers' age and gender on the basis of empirical studies. In, *Proceedings of Interspeech 2006*, Pittsburgh. 17–21.

Németh Géza–Olaszy Gábor 2010. *A magyar beszéd: beszéd kutatás, beszédtechnológia, beszédinformációs rendszerek*. Akadémiai Kiadó, Budapest.

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.

Ptacek, P. H.–Sander, E. K. 1966. Age recognition from voice. *Journal of Speech and Hearing Research* 9. 273–277.

Rao, Rajeshwara R.–Nagesh 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.

Schötz, Susanne 2007. Acoustic analysis of adult speaker age. *Lecture Notes in Computer Science*. 88–107.

Schwartz, Martin F. 1968. Identification of speaker sex from isolated voiceless fricatives. *Journal of the Acoustical Society of America* 43/5. 1178–1179.

Smith, David R.–Patterson, Roy D. 2005. The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America* 118/5. 3177–3186.

Traunmüller, Hartmut–Bezooijen, Renée van 1994. The auditory perception of children's age and sex. In, *Proceedings ICSLIP*. 1171–1174.
<http://www2.ling.su.se/staff/hartmut/barn.pdf> (A letöltés ideje: 2013. 01. 03.)

Vorperian, Houri 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.

6. Melléklet

A képleíráshoz felhasznált képanyag

