
SZINTETIZÁLT [ɔ], [ɛ] ÉS [a:] HANGOK ÁLTAL KELTETT BENYOMÁSOK ÉS ÉLETKORBECSLÉSEK VIZSGÁLATA

Gocsál Ákos

Bevezetés

A beszéd bonyolult akusztikai struktúrái nem csak nyelvi üzenetet hordoznak. Laver (2003) szerint egy adott beszélő által egy adott helyzetben produkált hangzó három fő hatás együtteseként valósul meg. Egyrészt a hangzó egy meghatározott fonémát realizál. Másrészt paralingvisztikai attitűdkommunikációs funkciót is betölt. Harmadrészt pedig jelzéseket ad a beszélő testi, társadalmi vagy pszichológiai identitásáról. Ennek megfelelően a beszéd észlelése során sem kizárólag a nyelvi üzenetet dolgozzuk fel, hanem ezzel párhuzamosan a beszéd akusztikai struktúrái alapján a beszélő személyről benyomást alkotunk, és különféle tulajdonságaira következtetünk.

Azt, hogy a beszéd akusztikai szerkezetéből hogyan következtetünk a beszélő tulajdonságaira, a Brunswik-féle lencsemodell szemlélteti a legátfogóbban (Scherer 2003). Bár Scherer csak az érzelmek kommunikációjával összefüggésben ismertette az említett modellt, mégis könnyen értelmezhető bármely más emberi tulajdonságra vonatkoztatva. Röviden: a beszélő személy tulajdonságai befolyásolják beszédének akusztikai szerkezetét. Ezek az akusztikai paraméterek távoli felismerési kulcsokként (distal cues) funkcionálnak. A beszédészlelés során ezek a kulcsok leképeződnek, közeli kulcsokká (proximal cues) alakulnak, amelyek alapján a hallgató a beszélő személyhez valamilyen tulajdonságot rendel.

A modellből azonban nem az következik, hogy minden egyes emberi tulajdonságnak megvan a pontos akusztikai vetülete, és a hallgató is pontosan vissza tudná kódolni a beszélő adott tulajdonságát. A modell inkább ilyen jellegű hipotézisek felállítására alkalmas. Ugyanis a különféle emberi tulajdonságok más-más módon jelennek, jelenhetnek meg a beszéd akusztikai struktúráiban, másrészt a hallgató sem feltétlenül kódol vissza minden tulajdonságot helyesen. Egy kísérletben a beszélő nemének felismerése például – izolált magánhangzók alapján – az esetek 98,6%-ában volt sikeres (Eklund-Traunmüller 1997), de suttogott magánhangzók esetén is 91% volt. Más tulajdonságok megítélése azonban bizonytalanabb. Gósy (2001) testalkatbecslési kísérletének eredményei például azt mutatják, hogy a férfiak magasságát és súlyát pontosabban lehet megbecsülni hangjuk alapján, mint a nőkéét. A becslések pontossága függ a beszélő alaphangjától is.

DOI: 10.15775/Beszkut.2016.24.6

Jelen munkánkban a hallgatóban a beszélőről kialakuló benyomás két aspektusát vizsgáljuk. Egyrészt foglalkozunk azzal, hogy a kísérleti személyeknek lejátszott, a későbbiekben ismertető módon előállított magánhangzók kellemes vagy kellemetlen benyomást keltenek-e, másrészt megvizsgáljuk, hogy milyen becsléseket adnak az elképzelt beszélő életkorára.

A beszéd által keltett kellemes vagy kellemetlen benyomás kérdéskörével számos kutató foglalkozott. Coelho és munkatársai például azt találták, hogy a mélyebb alaphang, a változatosabb prozódia és a gyorsabb tempójú, de határozott szünetekkel megszakított beszéd kelt kellemesebb benyomást a hallgatókban (Coelho et al. 2008).

A beszélőt, hangjának észlelésekor, a hallgató saját szempontjából kedvező vagy kedvezőtlen tulajdonságokkal ruházza fel. Női hallgatóknál Bruckert és munkatársai (2006) megállapították, hogy a kellemesség benyomását izolált beszédhangok észlelésekor főleg azok intonációja határozza meg. Az emelkedő intonációval ejtett magánhangzók szignifikánsan kedvezőbb benyomást keltettek, különösen akkor, ha azokat mély hangú beszélő ejtette ki. Az életkor megállapításánál helyes becsléseket adtak, ehhez a formánsok értékéből és szóródásából kialakított formánskomponenst használták fel (Bruckert et al. 2006).

A mélyebb f_0 és formánsfrekvenciák férfi beszélők esetében vonzóbb beszélő benyomást keltették, ha azonban női beszélő produkált mélyebb f_0 -t és formánsfrekvenciákat, akkor ők kevésbé vonzóknak vélték a hallgatók (Pisanski–Rendall 2011). A szokásosnál mélyebb alaphang már nem eredményez vonzóbb benyomást, a hallgatók ekkor inkább a dominánsabbnak gondolják a beszélőt (Fraccaro et al. 2013). Egy másik kísérlet szerint pedig a kepsztrumon megfigyelhető csúcs kiemelkedése áll összefüggésben azzal, hogy mennyire tartja a hallgató vonzóknak a beszélőt (Balasubramaniam et al. 2012). Abend et al. (2015) eredményei pedig azt mutatják, hogy akinek a hangját vonzóknak tartja a hallgató, annak az arca is vonzó számára.

A beszéd által kiváltott kellemességérzet vizsgálata fontos lehet a beszédrehabilitáció szempontjából is. Eadie és Doyle (2004) a hang kellemessége és elfogadottsága között szoros összefüggést talált, ez viszont a beszédrehabilitáció számára is feladatokat jelenthet. Egy másik kutatás pedig azt mutatta ki, hogy a mélyebb alaphang a hallgatókban kellemesebb benyomást keltett, ezzel együtt pedig a beszélőt magasabb szocioökonómiai státuszúnak vélték a hallgatók (O'Connor et al. 2014).

A beszélő életkorának hangja alapján történő megbecslése meglehetősen bizonytalanságot mutat. Míg egyes esetekben pontosak a becslések, más beszélőknél a hallgatók következetesen idősebbnek vagy fiatalabbnak gondolják a beszélőt, mint a tényleges életkora (Gocsál 1998). Az életkorbecslések következetlenségét állapította meg Benjamin (1992). Természetes körülmények között felvett társalgásokat használt, és arra a következtetésre jutott, hogy a beszéd hangzásából nem lehet megbízhatóan megállapítani a beszélő élet-

korát. Az életkorbecslések elfogadhatóbb eredményt adnak, ha tágabb kategóriákat használunk (Cerrato et al. 2000; Schötz 2001). Érdekes eredmény ugyanakkor, hogy ha vizuális információ is rendelkezésre áll a beszélőről, akkor pontosabban becsülhető meg a beszélő életkora. Sőt ha a beszélő személyt bemutató videofelvételt szinkronizálták, és ugyanazon személy más-más hangon szólalt meg a kísérleti személyek számára, a beszéd eltérő hangzása sem befolyásolta szignifikánsan az életkorbecslést. Ha tehát a hallgató a beszélőt látja is beszéde közben, az életkorbecsléshez a kinézetét veszi alapul (Amilon et al. 2007).

Az, hogy pontosan milyen akusztikai kulcsokat használunk a beszélő életkorának megállapításához, még nem teljesen tisztázott, de bizonyos kutatási eredmények már rendelkezésre állnak. Tóth (2014) kísérletében gyermekkorú beszélők életkorát becsülték meg a kísérleti személyek. Meglehetősen jó arányban (90, illetve 86,4 %-ban) tudták helyesen megállapítani, hogy a beszélő óvodás vagy általános iskolás-e. A szerző az alaphangmagasság akusztikai kulcsként betöltött szerepét feltételezi, mivel az szignifikánsan különbözött a két korcsoportban. Ugyanakkor – legalábbis 9 és 11 éves gyerekek magánhangzóit összehasonlítva – szignifikáns különbség mutatkozott a magánhangzók formánsszerkezetében is, de a felnőttekéhez hasonló F_1 - és F_2 -értékek csak 11 éves kor fölött alakulnak ki (Auszmann 2015). A gyermekeknél a magánhangzó-háromszög területe nagyobb, mint korábbi kutatásokban szereplő felnőtteké (Deme 2012). A magánhangzók szerkezete azonban az életkor előrehaladtával tovább változik. Fiatalok és idősek magánhangzóinak akusztikai szerkezete között több szignifikáns eltérést is talált Bóna (2009). Az [ɔ] hang első és második formánsa is szignifikánsan magasabb volt a fiataloknál, mint az időseknél, az [a:] hang esetében a második formáns, az [o] hangnál viszont az első formáns volt magasabb. Bár igen jelentős egyéni különbségek is mutatkoztak, a 80 év fölötti beszélőknél jellemző volt a centralizáltabb ejtés, míg a fiataloknál jobban elkülönültek a magánhangzók. Így, mint távoli kulcs, feltételezhető a magánhangzók formánsainak szerepe az életkorbecslésben. Egyes akusztikai paramétereknek ugyanakkor torzító hatása lehet. Prakup (2012) kísérletében fiatalabbnak ítélték a kísérleti személyek azokat a beszélőket, akiknek a hangjában alacsonyabb volt a jitter értéke, azaz szabályosabb zöngével beszéltek. Ennek megfelelően a hallgatók az énekeseket szignifikánsan fiatalabbnak vélték valós életkoruknál. Az f_0 és a shimmer ugyanakkor nem állt kapcsolatban a becsült életkorral.

A fentiek alapján az alábbi kérdéseket fogalmazzuk meg. 1. A kellemes vagy kellemetlen benyomás kialakulásában van-e szerepe annak, hogy egy adott magánhangzót milyen formánsokkal realizált a beszélő? 2. Ugyanazon magánhangzó különböző formánsokkal realizált változatait hallva különböző életkorúnak gondolja-e a hallgató a beszélőt? Röviden: a hallgató a perceptuális ítéletének kialakításakor akusztikai kulcsként használja-e a magánhangzók formánsait? A kérdések megválaszolásához két kísérletet végeztünk.

1. kísérletünkhöz azt a hipotézist fogalmazzuk meg, hogy nincs különbség az adott beszédhangok különböző változatai által kiváltott kellemességérzetben.
2. kísérletünkhöz pedig azt a hipotézist állítottuk fel, hogy a magánhangzók különböző változatai alapján nem következhetnek különböző életkorú beszélőkre a hallgatók.

Anyag és módszer

A kísérletekhez szintetizált magánhangzókat állítottunk elő a Praat 5.3.82 programmal (Boersma–Weenik 2014), majd ezeket lejátszottuk kísérleti személyeknek. A szintézishez Gósy–Bóna (2014) által közölt formánsadatokat vettük alapul, ennek megfelelően csak az ott használt [ɔ], [a:] és [ɛ] hangokat használtuk. Első lépésben a program *pitch tier* utasításával létrehoztunk egy 110 Hz-es és egy 220 Hz-es, 1 másodperc hosszúságú alaphangot. Ezek segítségével az említett magánhangzók férfi, illetve női ejtészváltozatait hoztuk létre. A következő lépésben a *formant tier* utasítással beállítottuk a formánsokat. Csak az első három formánst használtuk. Minden más paramétert a program által felkínált alapértéken hagytunk. Minden egyes magánhangzót öt változatban állítottunk elő. Az első változat minden magánhangzó esetében a Gósy–Bóna (2014) kutatásában közölt átlagos formánsértékeket realizálta. Az így előállított magánhangzókat nevezzük a továbbiakban „középsőnek”, mivel az $F_1 \times F_2$ síkon a többi négy változat által kirajzolt négyszögnek megközelítőleg a közepén helyezkednek el (l. a 7. ábrát később). A többi négy változat – szándékunk szerint – ettől lényegesen eltérő színezetű hangzó volt. Ezeknek a „szélsőséges” ejtésű hangzóknak az előállítása a következőképp történt. Gósy–Bóna (2014) táblázatából alapul vettük az F_1 - és F_2 -értékek fiatalok, illetve idősek beszédén mért átlagértékeit. Az egyik szélső eset formánsainak a megállapításához megkerestük, hogy melyik átlagérték a nagyobb. Például a fiatal női beszélőknél az idézett szerzők az [ɔ] hang esetében átlagosan 670 és 1434 Hz-et, az időseknél pedig 616 és 1377 Hz-et kaptak a két első formánsra. Ezek közül kiválasztottuk a nagyobbat (670 és 1434 Hz), majd azért, hogy nagyobb különbséget idézzünk elő az egyes variánsok között, a közölt átlagos eltérés értékét is hozzáadtuk, így 81 és 136 Hz-et. A másik „szélsőséges” ejtésű változat esetében a kisebb átlagértékeket használtuk fel, amelyekből kivontuk az átlagos eltérés értékeit. Hasonló módon két köztes ejtésű hangzót is generáltunk, ezeknél alacsony F_1 -hez magas F_2 járult, és fordítva. Az $F_1 \times F_2$ síkon az így kapott négy pont az említett szerzők cikkében közölt 3. ábrán megközelítőleg a ponthalmaz sűrűbb részének a peremén helyezkednének el. A jelen tanulmányban az 1. és a 2. táblázat tartalmazza a kísérlethez használt F_1 - és F_2 -értékeket.

1. táblázat: A szintetizált hangok első és második formánsai (férfi)

Beszédhang	Változat	F ₁	F ₂
[ɔ]	középső	579	1251
	1	479	1666
	2	650	1666
	3	479	1049
	4	650	1049
[ɛ]	középső	553	1720
	1	450	1832
	2	663	1832
	3	450	1630
	4	663	1630
[a:]	középső	682	1478
	1	560	1642
	2	760	1642
	3	560	1355
	4	760	1355

2. táblázat: A szintetizált hangok első és második formánsai (nő)

Beszédhang	Változat	F ₁	F ₂
[ɔ]	középső	644	1406
	1	551	1563
	2	751	1563
	3	551	1241
	4	751	1291
[ɛ]	középső	646	1920
	1	541	2082
	2	763	2082
	3	541	1798
	4	763	1798
[a:]	középső	781	1782
	1	649	1938
	2	841	1938
	3	649	1438
	4	841	1438

Az így előállított magánhangzókat a kísérleti személyeknek véletlenszerű sorrendben játszottuk le. 1. kísérletünkhöz minden egyes magánhangzó elhangzása után egy ötfokú skálán kellett jelölniük, hogy kellemesnek vagy kellemetlennek érezték-e az adott hangzót (1 = kellemetlen, 5 = kellemes), il-

letve a 2. kísérlethez az elképzelt személy életkorát is meg kellett becsülniük. A kísérletben összesen 43 egyetemi hallgató vett részt (életkoruk: 18–27 év, 30 nő, 13 férfi).

A hangok lejátszása csoportosan, zavaró körülmények kizárásával, a PTE Művészeti Kar Zeneművészeti Intézetének előadótermében történt, professzionális hangszórók alkalmazásával. Minden kísérleti személy tisztán hallotta a lejátszott hangokat. Minden egyes magánhangzó lejátszása előtt a kutató bemondta az adott magánhangzó sorszámát, majd egy 1 másodperces, 440 Hz magasságú, figyelemfelhívó hangjelzés, illetve egy 2 másodperces szünet következett. A következő magánhangzó lejátszását a kísérletvezető akkor indította el, amikor minden kísérleti személy kitöltötte az előző magánhangzóra vonatkozó adatokat.

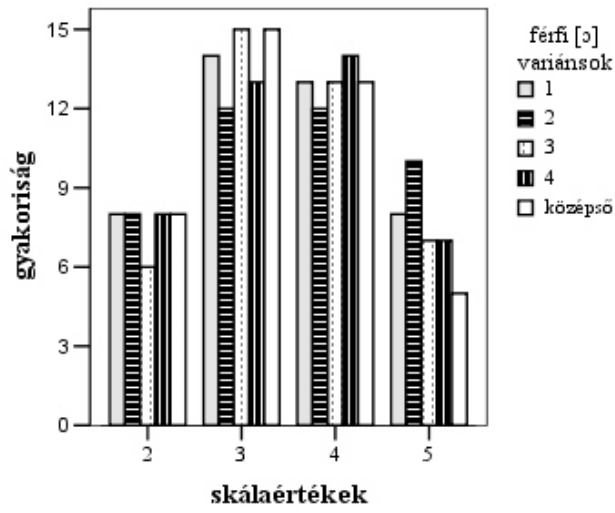
Az 1. kísérlet eredményei

A kutatás során a kísérleti személyek ötfokozatú skálán jelölték, hogy a hallott magánhangzót kellemesnek vagy kellemetlennek érezték. Az így kapott adatokat először gyakorisági grafikonokon ábrázoltuk. A grafikonokról az olvasható le, hogy a kísérleti személyek hány alkalommal karikázták be az 1-es, 2-es stb. skálaértéket. A skálaértékeknél öt oszlop látható, ezek az adott magánhangzó különböző variánsaira vonatkoznak.

Az eredményeket matematikai statisztikai módszerekkel elemeztük. Az egyes magánhangzók öt variánsára kapott skálaértékeket a Friedman-teszt segítségével hasonlítottuk össze. Nullhipotézisként azt állítottuk, hogy nincs eltérés az egyes magánhangzó-variánsok által keltett benyomások között. Amennyiben a Friedman-teszt szignifikáns eredményt ad, további, páronkénti összehasonlításokra van szükség Wilcoxon-próbák alkalmazásával.

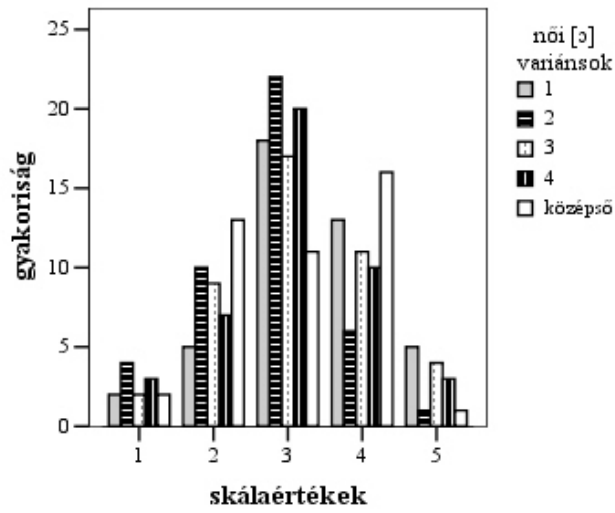
Az 1. ábrán láthatók a férfi ejtésű [ɔ] hang esetében kapott gyakorisági eloszlások. Az ábráról leolvasható, hogy jellemzően a 3-as és a 4-es skálaértékeket jelölték meg a kísérleti személyek. Szembeötlő eltéréseket nem látunk a grafikonon. A Friedman-teszt eredménye: $\chi^2(4) = 3,019$, $p = 0,555$, ami azt jelenti, hogy nullhipotézisünket nem vethetjük el, nincs szignifikáns különbség a hallgatókban kialakult benyomások között az egyes ejtészváltozatokat illetően.

A 2. ábra az [ɔ] hang női változatának lejátszásakor kapott skálaértékek eloszlását mutatja. Az előzőhöz képest csúcsosabb eloszlás arra utal, hogy többen voltak, akikben semleges érzetet keltettek az egyes [ɔ] hangok. Bár az átlagértékek alapján számolt, „középső” variáns esetében számszerűen több volt a 4-es érték, mint a többi variáns esetén, a számítás eredménye nem adott szignifikáns eltérést [$\chi^2(4) = 7,124$, $p = 0,129$]. Ez szintén arra utal, hogy a hallgatókban nem váltott ki szignifikánsan eltérő benyomást az [ɔ] magánhangzó öt, női ejtésű formánsértékeknek megfelelően generált változata.



1. ábra

A férfi ejtésű [ɔ] hangok esetében kapott skálaértékek eloszlása



2. ábra

A női ejtésű [ɔ] hangok esetében kapott skálaértékek eloszlása

A 3. ábrán látható a férfi ejtésű [ɛ] magánhangzó esetében kapott skálaértékek gyakorisági eloszlása. Az egyenletesnek tűnő eloszlást a Friedmann-teszt eredménye is megerősíti: $\chi^2(4) = 3,023$, $p = 0,554$, ami azt jelenti, hogy ezúttal sem adódott szignifikáns különbség a különböző variánsok által keltezt benyomásértékek között.

Az [ɛ] magánhangzó női variánsainál kapott gyakorisági eloszlásokat a 4. ábra mutatja. A Friedmann-teszt nem mutatott ki szignifikáns eltérést: $\chi^2(4) = 6,628$, $p = 0,157$ adódott a számítás során.

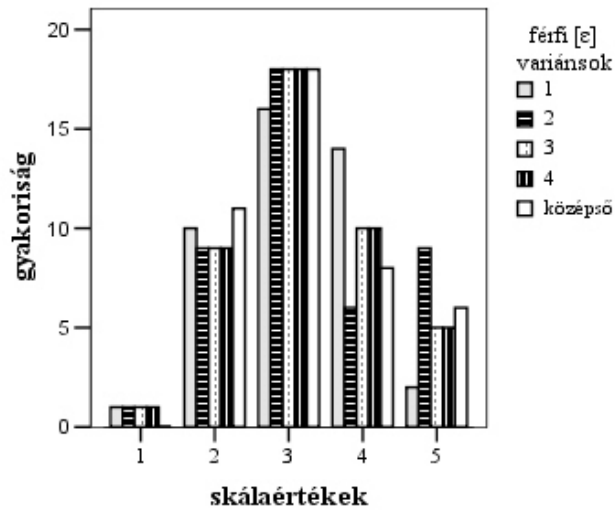
Az 5. és a 6. ábrán látható az [a:] magánhangzó férfi és női változatai esetében kapott skálaértékek eloszlása. A férfi ejtésű [a:] hang öt variánsa nem váltott ki szignifikánsan eltérő benyomásokat a hallgatók között, $\chi^2(4) = 7,915$, $p = 0,095$, adódott, ami a 0,05-ös határértéket még nem éri el, de megközelíti. Ennek oka az 5. ábráról leolvasható tendenciaszerű eltérés lehet. Megfigyelhető, hogy a 2. variánsnál a legtöbben a 2., a 3. variánsnál pedig a 4. skálaértéket jelölték. Hasonlóképpen nem szignifikáns, de ahhoz közeli értéket adott a Friedmann-teszt a női ejtésű [a:] hangok esetében is [$\chi^2(4) = 9,032$, $p = 0,06$], amit valószínűleg a 2. variáns esetében előforduló, többségében alacsony skálaérték okozhatott.

Összefoglalva: az eredmények azt mutatják, hogy a vizsgált szintetizált magánhangzók nem váltottak ki szignifikánsan eltérő kellemesség-benyomást a hallgatókban.

A kutatást megelőzően az volt a célunk, hogy ugyanazon magánhangzó-fonéma különböző, markánsan eltérő realizációi által előidézett benyomásokat vizsgáljuk. Ismeretes ugyanakkor, hogy az $F_1 \times F_2$ síkon az egyes magánhangzókhoz tartozó ponthalmazok gyakran átfedik egymást, így léteznek olyan magánhangzók, amelyek formánsszerkezete nagyon hasonló, annak ellenére, hogy elviekben különböző fonémák realizációiként jelennek meg a beszédben. A jelen kutatáshoz használt, különböző magánhangzókhoz tartozó pontok sem különülnek el határozottan egymástól az $F_1 \times F_2$ síkon.

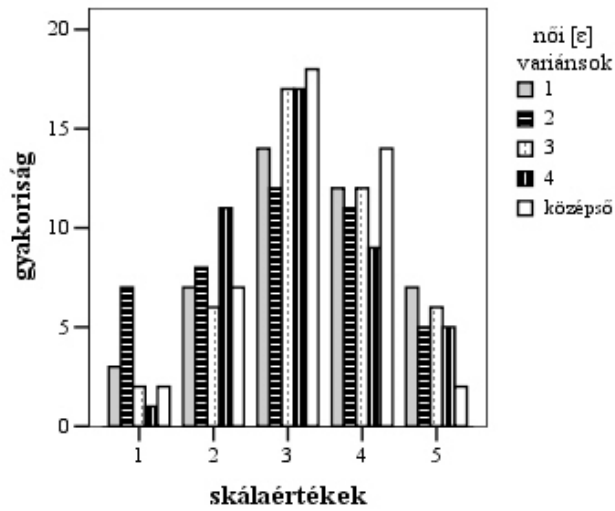
Felmerülhet tehát a kérdés, hogy a hasonló formánsszerkezetű magánhangzók esetében is fennáll-e a fentiekben kimutatott jelenség, azaz az, hogy nem idéznek elő különböző benyomásokat.

Az 1. és a 2. táblázat, illetve az 7. ábra alapján megállapítható, hogy a kísérletünkben használt, 5-5 magánhangzó-realizáció ponthalmazai is átfedik egymást az $F_1 \times F_2$ síkon. Megfigyelhető, hogy az egymást átfedő területeken több hangzó is meglehetősen közel helyezkedik el egy másik magánhangzó-fonéma valamelyik realizációjához. A férfi ejtést imitáló magánhangzók közül négy ilyen, egymáshoz közeli pár esetében kapott benyomásértékeket hasonlítottunk össze Wilcoxon-próbával.



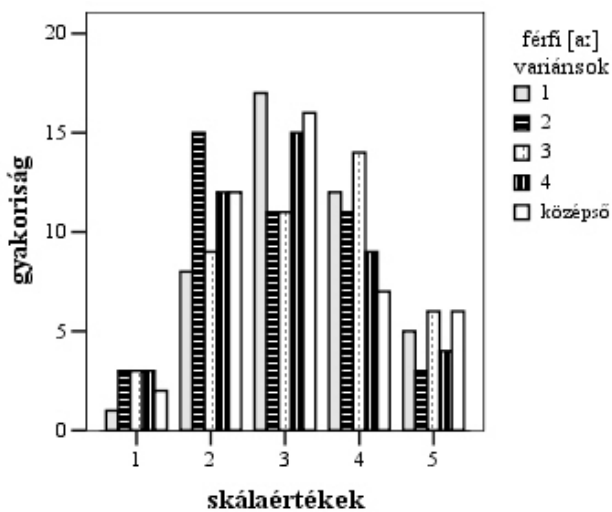
3. ábra

A férfi ejtésű [ɛ] hangok esetében kapott skálaértékek eloszlása



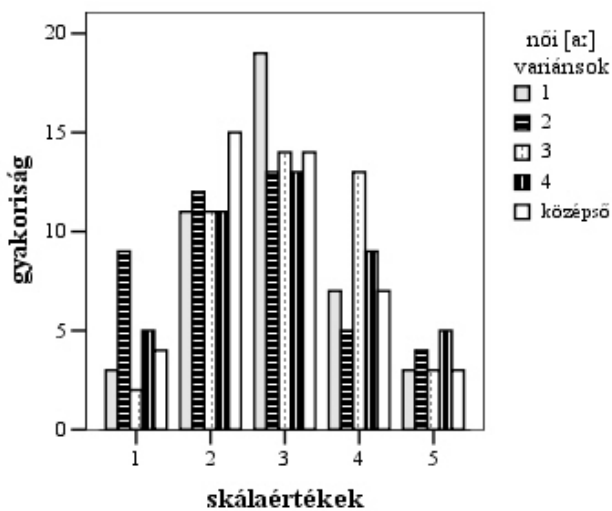
4. ábra

A női ejtésű [ɛ] hangok esetében kapott skálaértékek eloszlása



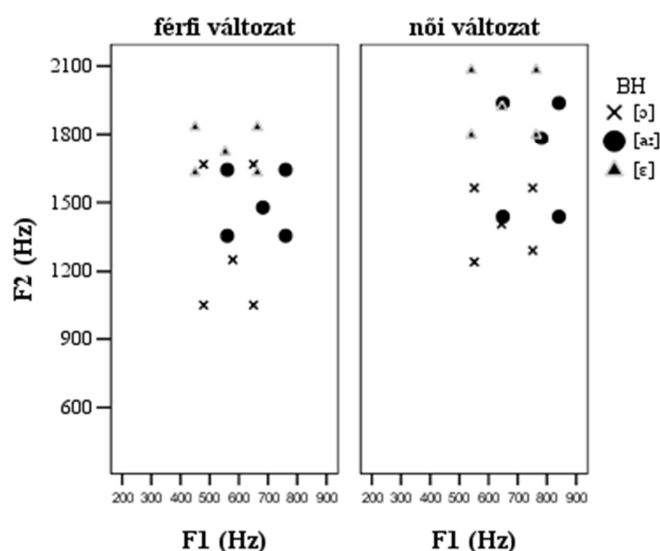
5. ábra

A férfi ejtésű [a:] hangok esetében kapott skálaértékek eloszlása



6. ábra

A női ejtésű [a:] hangok esetében kapott skálaértékek eloszlása



7. ábra

A szintetizált magánhangzók elhelyezkedése az $F_1 \times F_2$ síkon

Az [ɛ] hang 3. variánsa és az [ɔ] hang 1. variánsának összehasonlítására a Wilcoxon-próba nem mutatott szignifikáns eltérést ($Z = -0,758$, $p = 0,449$). Az [ɛ] hang átlagértékek alapján generált változata és az [a:] hang 1. variánsának formánsai szintén közel állnak egymáshoz. Ebben az esetben sem mutatott a Wilcoxon-próba szignifikáns különbséget ($Z = -0,421$, $p = 0,674$). A harmadik összehasonlított hangzópár az [ɛ] hang 4. variánsa és az [ɔ] hang 2. variánsa. A Wilcoxon-próba eredménye ebben az esetben sem szignifikáns ($Z = -1,839$, $p = 0,066$), de a p 0,05-ös szignifikanciaszintet megközelítő eredménye tendenciaszerűen az [ɔ] hang kellemesebb hangzására utal. Végül a férfi változatban szintetizált hangzók közül az [a:] hang 3. variánsát és az [ɔ] hang átlagértékek alapján előállított változatát hasonlítottuk össze. A Wilcoxon-próba azt az eredményt adta, hogy a negatív rangok összege megegyezett a pozitív rangok összegével, így $Z = 0,000$, $p = 1,000$, azaz a nullhipotézis nem vethető el. Ez azt jelenti, hogy a két hangzó esetében rendkívül hasonló benyomásértékeket rögzítettek a kísérleti személyek.

A női ejtésű magánhangzók közül három olyan párt találtunk, amelyek formánsszerkezete hasonló. Az első összehasonlítást az [ɛ] hang átlagértékek alapján előállított változata, illetve az [a:] hang 1. variánsa között végeztük. A Wilcoxon-próba eredménye: $Z = -1,799$, $p = 0,072$. A kapott eredmény nem mutat szignifikáns különbséget, de tendenciaszerű eltérést mutat az [ɛ] hang átlagértékek alapján előállított változata javára. A második összehason-

lítást az [a:] hang átlagértékek szerinti változata és az [ɛ] hang 4. variánsa között végeztük. A Wilcoxon-próba $Z = -2,278$, $p = 0,023$ eredményt adott, ami szignifikáns különbségre utal. A hallgatók az [ɛ] hangot szignifikánsan kellemesebbnek ítélték, mint az [a:] hozzá hasonló változatát. A harmadik összehasonlítást az [ɔ] hang átlagértékek alapján előállított változata, illetve az [a:] hang 3. variánsa között végeztük. A Wilcoxon-próba eredménye: $Z = -0,504$, $p = 0,614$, ami nem utal szignifikáns különbségre.

A 2. kísérlet eredményei

2. kísérletünkben arra a kérdésre kerestük a választ, hogy a szintetizált magánhangzók észlelése során a hallgatókban kialakul-e bármiféle benyomás az elképzelt beszélő életkorával kapcsolatban, illetve hipotézisünknek megfelelően, van-e bármilyen eltérés az adott magánhangzók különböző variánsai által kiváltott életkorbecslésekben. Kísérletünkben a korábban írottakhoz hasonlóan tágabb kategóriákat használtunk, azaz a kísérleti személyeknek a „fiatal”, „középkorú”, illetve „idős” kategóriák valamelyikébe kellett besorolniuk az elképzelt beszélőt.

A kutatás során nyert adatokat keresztábrák segítségével, χ^2 -próbákkal elemeztük, melyek sorait a vizsgált magánhangzó-variánsok, oszlopait pedig az életkori kategóriák alkották. A keresztábra egyes celláiba gyakorisági értékek kerültek, tehát az, hogy hány alkalommal fordult elő, hogy az adott hangot fiatal, középkorú vagy idős beszélőnek tulajdonították a kísérleti személyek. Az adatok kiértékelése során azt vizsgáltuk meg, hogy az egyes cellákban az elméletileg elvárt gyakorisághoz képest kisebb vagy nagyobb számban fordultak-e elő az egyes életkori kategóriákhoz kapcsolódó, ténylegesen kapott becslési adatok. Azt, hogy a keresztábra valamely cellájában túl kevészer vagy túl sokszor szerepelt valamely érték, a korrigált reziduumok (adjusted residuals) elemzésével állapítottuk meg. Abban az esetben, ha a korrigált reziduum értéke -2 és 2 közé esik, a kísérlet során kapott gyakorisági adatok nem különböznek szignifikánsan az elméletileg elvárt értéktől. Ha a korrigált reziduum értéke -2 -nél kisebb, akkor alulreprezentált, ha 2 -nél nagyobb, akkor felülreprezentált a gyakorisági érték, azaz az elvárthoz képest túl kevészer, vagy túl sokszor fordult elő. A 3. táblázat illusztrálja, hogy milyen elrendezés szerint értelmezhetők a kapott adatok.

A kísérleti személyek közül a férfi ejtésű [ɔ] hangokat 1-6 fő fiatal beszélőnek, 15-27 fő középkorúnak, 14-25 fő pedig idősnek tulajdonította. A χ^2 -próbát ebben a formában nem lehetett elvégezni, mert a keresztábrában több esetben is 5-nél kisebb várt érték szerepelt. Emiatt azt az áthidaló megoldást alkalmaztuk, hogy a „fiatal” és a „középkorú” kategóriát összevontan kezeltük, így a továbbiakban az „idős” és a „nem idős” kategóriákat használtuk. Az így kiszámolt χ^2 -próba nem adott szignifikáns eredményt: $\chi^2(4) = 8,749$, $p = 0,068$, Cramer's $V = 0,202$, de a p megközelítette a 0,05-ös küszöbértéket. A korrigált reziduumok elemzése az öt hangzóból egy esetben adott olyan ér-

téket, amely -2 -nél kisebb vagy 2 -nél nagyobb. Az [ɔ] hang 4. variánsát 29 esetben nem idősnek, 14 esetben idős beszélőnek tulajdonították a kísérleti személyek, itt $-2,7$, illetve $2,7$ adódott a korrigált reziduumok értékére. Az 1., 2. és 3. változatú a hangokat többségében idős beszélőnek tulajdonították, de a korrigált reziduumok értéke -2 és 2 közötti értékeket vett fel. Az átlagértékek alapján generált, középső [ɔ] hangot pedig 21 fő tulajdonította idős, illetve 22 fő nem idős beszélőnek, emiatt 0 körüli érték adódott.

3. táblázat: A férfi ejtésű [ɔ] hangok gyakorisági eloszlásai az életkori kategóriákkal

[ɔ] variánsai		Idős	Nem idős
középső	Gyakoriság	21	22
	Várt gyakoriság	22	21
	Korrigált reziduum	$-0,3$	$0,3$
1. variáns	Gyakoriság	25	18
	Várt gyakoriság	22	21
	Korrigált reziduum	1	-1
2. variáns	Gyakoriság	24	19
	Várt gyakoriság	22	21
	Korrigált reziduum	$0,7$	$-0,7$
3. variáns	Gyakoriság	26	17
	Várt gyakoriság	22	21
	Korrigált reziduum	$1,4$	$-1,4$
4. variáns	Gyakoriság	14	29
	Várt gyakoriság	22	21
	Korrigált reziduum	$-2,7$	$2,7$

A női ejtésű [ɔ] hang esetében mindegyik variánsnál többségében voltak, akik fiatal beszélőt képzeltek el (23 – 29 fő), kevesebben gondolták a beszélőt középkorúnak (12 – 19 fő), és még kevesebben (0 – 3 fő) idősnek. A χ^2 -próba elvégzése során kiderült, hogy öt cellában 5-nél kisebb várt érték adódott, ezért az „idős” és a „középkorú” kategóriák összevonásával „fiatal” és „nem fiatal” kategóriákat használtunk. Az így elvégzett számítás eredménye: $\chi^2(4) = 2,102$, $p = 0,717$, Cramer's $V = 0,099$, ami arra utal, hogy a különböző ejtésű a hangok alapján szerzett benyomás a beszélő életkoráról (fiatal-nem fiatal) nem különbözött szignifikánsan. A korrigált reziduumok értéke is minden esetben -2 és 2 között maradt. Noha a számítás nem adott szignifikáns eredményt, megjegyzendő, hogy minden esetben többen voltak, akik az [ɔ] hang alapján a beszélőt fiatalnak vélték, mint azok, akik középkorúnak, illetőleg nem fiatalnak.

A férfi ejtésű [ɛ] hang esetében összesítve a következő eredmény adódott. Fiatalnak 1 – 4 fő, középkorúnak 13 – 23 fő, idősnek pedig 17 – 29 fő vélte az

elképzelt beszélőt. A keresztábra ismét tartalmazott olyan cellákat, amelyekben 5-nél kisebb várt értékek jelentek meg, emiatt a „fiatal” és a „középkorú” kategóriák összevonásával a vizsgálatot „idős” és „nem idős” kategóriákra végeztük el. A számítás eredményeként $\chi^2(4) = 11461$, $p = 0,022$ szinten szignifikáns különbség adódott, Cramer's $V = 0,231$ értékkel. A korrigált reziduomok értéke az [ε] hang 1. variánsa esetében esett kívül a [-2, 2] intervallumon, értéke 2,5 illetve -2,5 volt, ekkor 29 fő ítélte a beszélőt idősnek, 14 fő pedig nem idősnek.

A számítást elvégezve az [ε] magánhangzó női ejtésű változataira, a következő eredményt kaptuk. Az öt ejtésváltozat alapján a válaszadók túlnyomó része (24–28 fő) az elképzelt beszélőt fiatalnak gondolta. Középkorúnak 13–18 fő vélte, míg idősnek 0–7 fő. Az egyes változatok között a χ^2 -próba nem adott szignifikáns különbséget, de több cellában is 5-nél kisebb szám fordult elő, így a korábbiakhoz hasonló módon az „idős” és a „középkorú” kategóriákat együttesen kezeltük „nem fiatal” kategóriaként. Az így elvégzett számítás eredményei az alábbiak: $\chi^2(4) = 1,654$, $p = 0,799$ Cramer's $V = 0,088$, $p = 0,799$, ami azt jelenti, hogy a különböző ejtésű a hangok alapján szerzett benyomás a beszélő életkoráról (fiatal vs. nem fiatal) ebben az esetben sem különbözött szignifikánsan. Ezt az adatot megerősítik a korrigált reziduomok értékei is, ugyanis minden esetben 0 körüli értéket vettek fel.

Végül az [a:] hangon elvégezve a számításokat az alábbi eredményre jutottunk. A férfij ejtést imitáló [a:] hangok alapján az elképzelt beszélőt kevesen gondolták fiatalnak (1–6 fő), többen vélték középkorúnak (15–31 fő), illetve idősnek (11–16 fő). Tekintettel arra, hogy itt is több olyan cellát találtunk, amelyben 5-nél kisebb volt a várt érték, a „fiatal” és a „középkorú” kategóriát összevontan kezelve, a számítást „idős” és „nem idős” kategóriákra végeztük el. A számítás eredménye: $\chi^2(4) = 13,216$, $p = 0,01$ szinten szignifikáns, Cramer's $V = 0,248$ értékkel. A korrigált reziduomok egy esetben adtak a [-2, 2] intervallumon kívül eső értéket, azaz -3,4-et az „idős” kategóriára, illetve 3,4-et a „nem idős”-re, mégpedig a 4. variáns esetében. Ekkor a kísérleti személyek közül lényegesen többen vélték az elképzelt beszélőt nem idősnek (32-en), mint idősnek (11-en).

A női beszélőt imitáló [a:] hangok esetében az alábbi eredmények adódtak. A kísérleti személyek az 5 variánst jellemzően fiatal beszélőnek tulajdonították (25–29 alkalommal), középkorúnak 7–13 alkalommal, idősnek pedig 3–7 alkalommal. A keresztábrában több esetben is előfordult az 5-nél kisebb várt érték, így az „idős” és „középkorú” kategóriákat összevontan, „nem fiatal” kategóriaként vettük figyelembe a további számítások során. Az így kapott eredmény: $\chi^2(4) = 1,281$, $p = 0,865$, Cramer's $V = 0,077$, ami egyértelműen az eltérés hiányát mutatja. Ezt a korrigált reziduomok is megerősítik, minden esetben 0 körüli értéket kaptunk. Ezek az eredmények arra utalnak, hogy az egyes variánsok között nincs szignifikáns különbség az elképzelt be-

szelő életkorát illetően. Ugyanakkor a válaszadók többsége – legalább 25 fő – minden hangzó esetén fiatalnak gondolta az elképzelt női beszélőt.

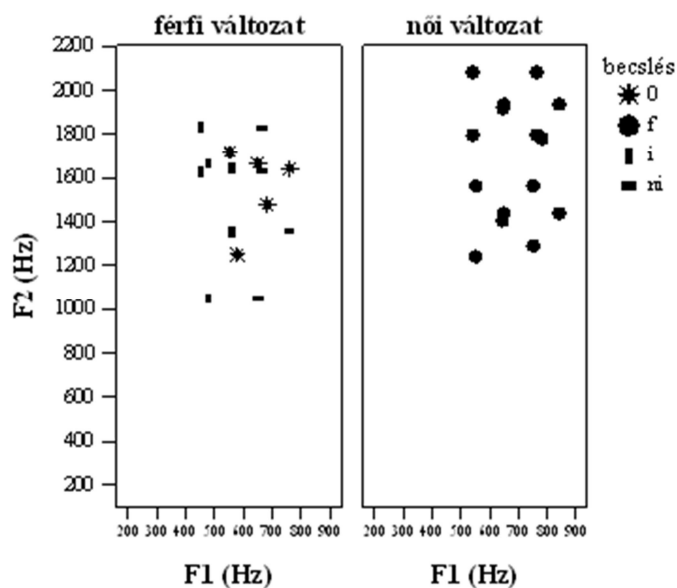
Következtetések

Első kísérletünk eredményei azt mutatják, hogy ugyanazon hangzók különböző formánsszerkezetekkel előállított változatai nem váltottak ki szignifikánsan eltérő benyomásértékeket, sem férfi, sem női ejtősváltozatban. A különböző magánhangzókat realizáló, de hasonló formánsszerkezetű magánhangzók összehasonlítása egy esetben adott szignifikáns különbséget, a női ejtősváltozatban előállított [ɛ] hangot kellemesebbnek ítélték a hallgatók, mint a hozzá hasonló [a:] hangot. Ez az egyetlen szignifikáns különbség azonban nem elegendő ahhoz, hogy általánosítva, valamilyen mögöttes folyamatra tudjunk következtetni. A szignifikáns eltérés hiánya inkább arra enged következtetni, hogy önmagukban a formánsok helye nem elegendő ahhoz, hogy szignifikánsan befolyásolja a kellemesség érzetét. Ezek az eredmények 1. hipotézisünket megerősítik, fenntartva az egy kivételt. Jelen kísérletünkben más paramétereken nem változtattunk, így azt a további következtetést fogalmazhatjuk meg, hogy ha a formánsoknak van szerepe a benyomás kialakításában, akkor az feltételezhetően csak más beszédakusztikai paraméterekkel való interakcióikkal lehetséges. További feladat lehet annak vizsgálata, hogy ugyanezek a magánhangzók más alaphangokkal vagy különféle sávzsélességekkel előállított változatai a közölt eredményektől eltérő benyomásokat váltanak-e ki a hallgatókban.

Második kísérletünkben a kísérleti személyek az elképzelt beszélő életkorát becsülték meg három kategória alapján. A férfi ejtésben létrehozott 15 beszédhangot átlagosan 21,5 esetben tulajdonították idős, és szintén 21,5 esetben nem idős beszélőnek. Ez a látszólagos kiegyenlítettség azonban jelentős eltéréseket takar, ugyanis a három magánhangzó egy-egy változata szignifikánsan más életkori kategóriába került, mint a többi. Ebből a három eltérésből azonban nehéz általánosítható következtetést levonni. Ugyanakkor tendenciaszerű eltérések megfigyelhetők, ha a korrigált reziduumok vizsgálatát a $[-1, 1]$ intervallum alkalmazásával végezzük. Ha megfigyeljük, hogy mely hangzóváltozatoknál esik a korrigált reziduum értéke ezen intervallumon kívül, illetve melyik életkori kategória az, amelyet a vártnál többen választottak. A 8. ábrán ennek megfelelően ábrázoltuk az egyes hangzókat az $F_1 \times F_2$ síkon, ez esetben eltekintve attól, hogy ténylegesen milyen hangzót hallottak a kísérleti személyek.

A férfi ejtésben létrehozott hangzóknál látható az a tendencia, hogy az F_1 alacsony értékeinél vélték többségében idősnek a beszélőt, míg a „nem idős” kategóriába tartozó hangokra jellemző volt, hogy magasabb F_1 -értékekkel rendelkeztek. Az F_2 értékeinél viszont ilyen jellegű eltéréseket nem figyelhetünk meg. Ebből azt a továbbiakban vizsgálandó hipotézist fogalmazhatjuk meg, hogy az F_1 értékének szerepe lehet az életkori kategória megítélésében.

A magánhangzók női változatai esetében a 15 hangot átlagosan 25,9 esetben tulajdonították fiatalnak, 17,1 esetben pedig nem fiatalnak. Az egyes hangok esetében valamennyi eltérést tapasztaltunk, de a korrigált reziduumok egyik esetben sem lépték túl az 1-et.



8. ábra

Az életkorbecslések tendenciaszerű megoszlása
 (*: kiegyenlített idős-nem idős arány, ●: többségében fiatal,
 ■ : többségében idős, ■: többségében nem idős)

Második hipotézisünket részben igazoltnak tekintjük azzal a kitételrel, hogy a női változatban előállított hangzók nem váltottak ki szignifikáns különbségeket az életkori kategóriák megállapításakor, ugyanakkor a férfi változatban előállított hangzók néhány esetben szignifikáns eltérést eredményeztek, illetve nem szignifikáns, de tendenciaszerű különbségeket is megfigyeltünk.

Kutatásunk során korlátozó tényező volt, hogy a kísérleti személyek neme közötti eloszlása nem volt egyenletes. További feladat lehet több férfi hallgató bevonása a kísérletbe, és a számításokat a hallgatók neme szerint elkülönítve is elvégezni. Másrészt a számításokat nemparaméteres eljárásokkal végeztük, mivel nominális és ordinális változókat használtunk. Más megközelítésmódban, például a tényleges formánsértékek használatával érzékenyebb, paraméteres statisztikai módszerek is alkalmazhatók lehetnek. Ezekről a korlátozó tényezőktől eltekintve azonban megállapíthatjuk, hogy koherens hallgatói be-

nyomásokat találtunk. Elvértve voltak olyan hallgatók, akik a férfiakat fiatalnak vélték volna. Ez azt jelenti, hogy ezek a szintetizált hangok nem tartalmaztak olyan akusztikai kulcsokat, amelyek a hallgatókban fiatal beszélő benyomását keltették volna, ugyanakkor a formánszerkezet bizonyos mértékig differenciáló tényezőként működött az életkori kategória megállapítása során. A női változatban előállított hangok viszont a kísérleti személyek többségében fiatal beszélő benyomását keltették, a formánszerkezet pedig nem differenciálta szignifikánsan az életkoritéseket, így feltételezhető, hogy más akusztikai paraméterek, vagy az itt vizsgált formánsok más tulajdonságai vagy kombinációi játszanak szerepet a női beszélők életkorának megítélésében.

Irodalom

- Abend, Peter – Pflüger, Lena S. – Koppensteiner, Markus – Coquelle, Michael – Grammer, Karl 2015. The sound of female shape: a redundant signal of vocal and facial attractiveness. *Evolution and Human Behavior* 36. 174–183.
- Amilon, Kajsa – van de Weijer, Joost – Schötz, Susanne 2007. The impact of visual and auditory cues in age estimation. In Müller, Christian (ed.): *Speaker Classification. II. Lecture Notes in Computer Science 4441*. Springer-Verlag, Berlin-Heidelberg. 10–21.
- Auszmann Anita 2015. Magánhangzók akusztikai szerkezete 9 és 11 éves iskolások spontán beszédében. *Beszédkutató 2015*. 164–175.
- Balasubramaniam, Radish Kumar – Bhat, Jayashree S. – Srivastava, Manav – Edose, Aimy 2012. Cepstral analysis of sexually appealing voice. *Journal of Voice* 26/4. 412–415.
- Benjamin, Gail R. 1992. Perceptions of speaker's age in natural conversation in Japanese and English. *Language Sciences* 14/1–2. 77–84.
- Boersma, Paul – Weenik, David 2014. *Praat: doing phonetics by computer* [computer program]. Version 5.3.82. <http://www.praat.org/> (A letöltés ideje: 2014. június 12.)
- Bóna Judit 2009. Az idős életkor tükröződése a magánhangzók ejtésében. *Beszédkutató 2009*. 76–87.
- Bruckert, Laetitia – Liénard, Jean-Sylvain – Lacroix, André – Kreutzer, Michel – Leboucher, Gérard 2006. Women use voice parameters to assess men's characteristics. In: *Proceedings of the Royal Society* 273. 83–89. <http://rspb.royalsocietypublishing.org/content/royprsb/273/1582/83.full.pdf> (A letöltés ideje: 2015. november 12.)
- Cerrato, Loredana – Falcone, Mauro – Paolini, Andrea 2000. Subjective age estimation of telephonic voices. *Speech Communication* 31. 107–112.
- Coelho, Luis – Braga, Daniela – Mateo, Carmen 2008. Voice pleasantness: on the improvement of TTS voice quality. In: *Proceedings of the V Jordanas en Tecnologia del Habla. 12-14 Nov 2008. Universidad del Pais Vasco, Bilbao*, 211–214. <http://lorien.die.upm.es/~lapiz/rth/JORNADAS/V/pdfs/actas.pdf> (A letöltés ideje: 2015. november 28.)
- Deme Andrea 2012. Óvodások magánhangzóinak akusztikai jellemzői. In Markó Alexandra (szerk.): *Beszédtudomány*. ELTE Bölcsészettudományi Kar – MTA Nyelvtudományi Intézet, Budapest. 77–99.

- Fraccaro, Paul J. – O’Connor, Jillian J. M. – Re, Daniel E. – Jones, Benedict C. – DeBruine, Lisa M. – Freinberg, David R. 2013. Faking it: deliberately altered voice pitch and vocal attractiveness. *Animal Behavior* 85. 127–136.
- Gocsál Ákos 1998. Életkorbecslés a beszélő hangja alapján. *Beszéd kutatás ’98*. 122–134.
- Gósy Mária 2001. A testalkat és az életkor becslése a beszéd alapján. *Magyar Nyelvőr* 125. 137–148.
- Gósy Mária – Bóna Judit 2014. Magánhangzók ejtése fiatalok és idősek spontán beszédében. *Magyar Nyelv* 110. 129–143.
- Eadie, Tanya L. – Doyle, Philip C. 2004. Scaling of voice pleasantness and acceptability in tracheoesophageal speakers. *Journal of Voice* 19/3. 373–383.
- Eklund, Ingegerd – Traunmüller, Hartmut 1997. Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica* 54/1. 1–21.
- Laver, John 2003. Three semiotic layers of spoken communication. *Journal of Phonetics* 31. 413–415.
- O’Connor, Jullian J. M. – Fraccaro, Paul J. – Pisanski, Katarzyna – Tigue, Cara C. – O’Donnell, Timothy, J. – Freinberg, David R. 2014. Social dialect and men’s voice pitch influence women’s mate preferences. *Evolution and Human Behavior* 35. 368–375.
- Pisanski, Katarzyna – Rendall, Drew 2011. The prioritization of voice fundamental frequency or formants in listeners’ assessments of speaker size, masculinity, and attractiveness. *Journal of the Acoustic Society of America* 129/4. 2201–2212.
- Prakup, Barbara 2012. Acoustic measures of the voices of older singers and non-singers. *Journal of Voice* 26/3. 341–350
- Scherer, Klaus 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40. 227–256.
- Schötz, Susanne 2001. A perceptual study of speaker age. *Lund University Working Papers* 49. 136–139. <http://journals.lub.lu.se/index.php/LWPL/article/viewFile/2387/1962> (A letöltés ideje: 2015. november 22.)
- Tóth Andrea 2014. Gyermekek nemének és életkorának meghatározása a beszédük alapján. *Beszéd kutatás 2014*. 98–111.

Pleasantness and age judgments evoked by synthesized [ɔ], [ɛ] and [a:] sounds

The purpose of this study was to determine if synthesized Hungarian [ɔ], [ɛ] and [a:] vowels of different formant structures evoke different pleasantness and age judgments in listeners. 43 undergraduate students were randomly played five versions of the three vowels, both male and female variants. With one exception, no significant differences were found in pleasantness judgements. Female vowels evoked no significant differences in age judgements, but were generally perceived as ‘young’. Few significant differences were found with male vowels but no general conclusion could be drawn. However, there was a non-significant tendency that male vowels with lower F₁ were more likely to be perceived as ‘old’.