

A MÁSSALHANGZÓ-KÖRNYEZET ÉS A ZÖNGEINDULÁS HATÁSA AZ ÉNEKELT MAGÁNHANGZÓK ÉSZLELÉSÉRE

Deme Andrea

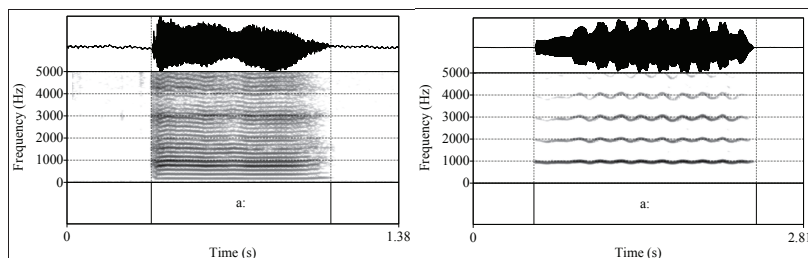
Bevezetés

Az énekelt magánhangzók kérdése a fonetikai szakirodalom egyik különleges témájának számít. Így van ez azért, mert az éneklés egy, a mindennapi beszédétől eltérő beszédképzési mód. Azonban ennél fontosabb (és kitüntetett helyzetét is indokolja) az a látszólagos ellenmondás, mely szerint bár az éneklés artikulációs és akusztikai sajátosságai miatt a magánhangzók ejtése jelentősen eltérhet a beszédben megszokottól, a hallgató mégis képes rá, hogy bizonyos mértékig, illetve bizonyos körülmények között az ejtési szándéknak megfelelően azonosítsa ezeket a vokálisokat.

A beszélt és énekelt magánhangzók közötti különbségek egyfelől az éneklés-alkotás sajátos (artikulációs és akusztikai) kivánalmainak, másfelől a magas alaphang okozta akusztikai változásoknak köszönhetőek.

Az operanéklésben elvárt, hogy az éneklés szintje a teljes alaphangtartományon homogén hangszínezetű és nagy intenzitású legyen. Ezt az éneklés – a magasabb alaphangokon – olyan artikulációs gesztusok segítségével éri el, melyek sok esetben nem kedveznek a magánhangzókra a mindennapi beszédben jellemző artikulációs konfigurációnak. A magánhangzók percepciója szempontjából is talán legfontosabb ilyen gesztus az állkapocs nyitásszögének az alaphang emelésével párhuzamos növelése. Ez a technika főként azon alaphangmagasságok esetében használatos, melyek esetében igaz az, hogy az alaphang (f_0) frekvenciaértéke nagyobb, mint az ejteni kívánt magánhangzó a beszédben megjelenő első formánsának (F_1) értéke. Ezekben a hangmagasságokon ugyanis az artikuláció megváltoztatása nélküli ejtés során az F_1 nem jelenhet meg a magánhangzó spektrumán, és az alaphang intenzitása is alacsonyabb, következésképpen az egész spektrum energiája is kisebb lenne. Az állkapocs nyitásszögének növelése azonban megemeli az F_1 értékét, így a fent említett helyzetben a magas frekvenciaértékű f_0 -nál alacsonyabb frekvenciaértékű F_1 feljebb (az f_0 értékére vagy akár afölé is) hangolható (Sundberg 1987; Garnier et al. 2010). Artikulációs tekintetben tehát az állkapocs nyitásszöge, míg akusztikai tekintetben az F_1 az alaphang függvényében változik – legalábbis azokban az esetekben, amikor az alaphang eléri a beszédben megjelenő F_1 értékét.

Am az énekelt magánhangzók akusztikai szerkezetét nemcsak az artikulációs változások, hanem a magas f_0 artikulációs mozgásoktól független akusztikai sajátosságai is befolyásolják. A mélyebb és magasabb alaphangon ejtett magánhangzók a magas f_0 -ra visszavezethető akusztikai különbségét szemlélteti az 1. ábra, melyen az [a:] hang keskeny sávú spektrogramja látható kétféle ejtésben: a beszédben ($f_0 \sim 190$ Hz) (bal oldal) és a B5 zenei hangon ($f_0 = 988$ Hz) énekelve (jobb oldal), egy szoprán énekesnő ejtésében.



1. ábra

Balra: az [a:] hang keskeny sávú spektrogramja a beszédben ($f_0 \sim 190$ Hz);
 jobbra: az [a:] hang keskenysávú spektrogramja a B5 énekelt
 alaphangmagasságon ($f_0 = 988$ Hz)

A keskeny sávú spektrogramon látható vízszintes csíkok a felharmonikusok, melyek frekvenciaértékei az alapfrekvencia egész számú többszörösei. A spektrogramok szemléltetik, hogy a magasabb alapfrekvenciájú hang esetében a felharmonikusok (az alaphanggal való összefüggésük következtében) jóval távolabb esnek egymástól, mint a mélyebb alaphangú vokálisoknál. A spektrografikus megjelenítésben a frekvenciakomponensek intenzitása is megjelenik, melyet a sötétedés mértéke mutat. Ez jól láthatóan eltér az ábrázolt beszédhangok különböző felharmonikusai esetében – annak megfelelően, hogy a felharmonikusok beleesnek-e valamely formáns (azaz a toldalékcső mint rezonátorüreg valamely sajátfrekvenciájának) sáv szélességébe. A spektrogram kellően mély alaphang és sűrű felhangszerkezet esetén (1. ábra, első kép) tehát a felhangok intenzitásának különbségein keresztül (sötétebb sávokként) „kirajzolja” a formánsokat (itt a beszéd esetében ezek 1 kHz, 1,5 kHz és 3 kHz környékén láthatók), míg erre a magas alaphangú hang felhangszerkezete (1. ábra, második kép) csak jóval kisebb mértékben képes. A magas f_0 következtében egymástól távolabb eső felharmonikusok ugyanis spektrálisan alul-mintavételezik (*undersampling*, l. Goldstein 1980: 235) a toldalékcső átviteli karakterisztikáját, azaz a toldalékcső artikulációs konfigurációját jellemző függvényt, illetve a magánhangzóspektrumot (l. pl. Goldstein 1980; Ryalls–Lieberman 1982). Ennek eredményeként a magas alaphangon produkált énekhang akusztikai kimenete sokkal kevesebb információt

közvetít az artikulációs konfigurációról, azaz az ejteni kívánt beszédhang képzésére berendezkedő beszédképző szervek aktuális állásáról, mint a mélyebb alaphangokon énekelt vagy beszélt vokálisoké.

Az énekhang fent említett tulajdonságaiból látszik, hogy az énekelt magánhangzók ejtése és akusztikai szerkezete szoros kapcsolatban áll az f_0 értékével. Ezzel összefüggésben pedig az a feltételezés is megalapozottnak tűnik, hogy az f_0 értéke az énekelt magánhangzók észlelésében is fontos szerepet játszik. Ennek a feltételezésnek megfelelően az énekelt magánhangzók észlelését vizsgáló szakirodalomban számos kutatás fókuszál az f_0 hatására. Ezek a kutatások kimutatták, hogy az azonosítás sikeressége az f_0 emelésével csökken (Scotto di Carlo–Germain 1985; Gottfried–Chew 1986; Hollien et al. 2000; Evgrafova–Evdokimova 2012; Deme 2012). Ugyanakkor az adatok arra is fényt derítenek, hogy a jól látható csökkenés ellenére bizonyos hangzók bizonyos körülmények között még magasabb alaphangon is felismerhetők maradnak, továbbá arra is, hogy az azonosítás sikerességének csökkenése sem minden esetben mutat szigorúan csökkenő tendenciát (Scotto di Carlo–Germain 1985; Hollien et al 2000; Deme 2012). Ennek ellenére kevés az olyan jellegű kutatás, amely a vokálisok percepció stabilitását támogató paramétereket, például a mássalhangzó-környezetben rejlő (dinamikus) akusztikai kulcsokat vizsgálná – a legtöbbször csak a magánhangzók ejtési sajátosságainak és az éneklés artikulációs kívánalmainak egybevetésével vagy eltéréseivel magyarázzák a felismerési arányok alakulását.

A beszédészlelés szakirodalmában széles körben elfogadott az a feltételezés, hogy könnyebb a magánhangzók azonosítása akkor, ha mássalhangzó-környezetben halljuk őket. Strange és Verbrugge (1976) híressé vált kísérletében /pVp/ szekvenciákban tesztelte a mássalhangzó-környezetben ejtett magánhangzók azonosítását, és azt a következtetést vonta le, hogy a hang-környezet pozitív hatása a mássalhangzó és magánhangzó határán álló koartikulációs hangátmenetekben kódolt információban rejlik. Bár a vizsgálat eredményei egyértelműen csak a beszélve ejtett magánhangzókra vonatkoznak, mégis előfordul, hogy a magas alaphangon énekelt beszédhangok észlelhetőségével kapcsolatban is említik a szerzők az észlelést valószínűsíthetően segítő mankók számbavételekor a mássalhangzó-környezetet (l. pl. Scotto di Carlo–Germain 1985; Hollien et al 2000). Teszik ezt annak ellenére, hogy a fentebb említett, az énekelt és beszélt ejtés közti jelentős artikulációs és akusztikai eltérések az éneklés kutatóinak körében köztudottak. Azon feltételezés tudományos igazolására, hogy a mássalhangzó-környezet megléte vagy hiánya valóban hatással van-e a magas alaphangon (szoprán éneklésben) ejtett vokálisok azonosítására, eddig összesen egyetlen kutatás született, egy további pedig mélyebb alaphangokon, egy férfi hangkategóriában (egy tenor énekes ejtésében) vizsgálta a kérdést.

Smith és Scott (1980) a magas alaphangon énekelt magánhangzók azonosítását /bVd/ hordozóhangsorban vizsgálta egy szoprán énekes ejtésében. Per-

cepciós adataik a szerzők értelmezése szerint az operaéneklésre nézve is megerősítik azt a beszédben tapasztalható tendenciát, miszerint a szomszédos mássalhangzók segítik a magánhangzók felismerését. Adataikat azonban nem szigorúan kontrollált körülmények között nyerték, hiszen a magánhangzót közrefogó mássalhangzó képzéshelye eltérő volt, így sem azt nem tudni, valóban mely mássalhangzó befolyását tesztelték, sem azt, hogy van-e különbség a magánhangzót megelőző és követő hangátmenetek hatása között. Az sem egyértelmű, hogy milyen módon kontrollálták a szerzők az énekes gégejének függőleges helyzetét, mely állításuk szerint szintén fontos befolyásoló tényezőnek bizonyult. Végül pedig a kapott eredmények interpretációjával kapcsolatban is kételyeink ébredhetnek, hiszen valóban egyértelmű tendenciákat csak az F5 zenei alaphangmagasság (azaz 698 Hz) fölött látunk.

Mélyebb alaphangokon (egy tenor énekprodukciónak) ejtett vokálisok azonosításával kapcsolatban Gottfried és Chew (1986) végzett (szintén /bVd/ hangsorokat tesztelő) vizsgálatot. Eredményeik (Smith és Scott 1980 eredményeihez hasonlóan) a mássalhangzó-környezet pozitív hatását mutatták. Ám ezzel a kutatással kapcsolatban is felmerülnek a már korábban is említett módszertani (ti. a hangkörnyezet kiválasztásával kapcsolatos) problémák, illetve további kérdések is, melyek az eredmények értelmezését nehezítik. Gottfried és Chew ugyanis az izolált magánhangzók felismerését a hordozóhangsorban ejtett magánhangzók tiszta fázisából kivágott 200 ms hosszúságú hangszeleten tesztelték, nem valóban izoláltan ejtett hangokon, így nem állítható, hogy ennek a kísérletnek az eredményei az izoláltan ejtett hangokra is közvetlenül vonatkoztathatóak volnának. Ezen kívül pedig nem tisztázott például az sem, hogy milyen ablakolással vágják ki a tesztelt hangszeleteket, amely paraméter szintén befolyásolhatja a kapott eredményeket.

Megjegyzendő, hogy az említetteken kívül még egy további olyan vizsgálat született, mely néhány, képzésmódban és képzéshelyben eltérő mássalhangzó alkotta fonetikai környezet hatását vizsgálta a magas alaphangon (női ejtésben) énekelt magánhangzók azonosításakor. Egy korábbi kutatásomban (2012) a nazálisok, illetve a zöngés és zöngétlen frikatívák között ejtett vokálisok azonosítását hasonlítottam össze, és nem találtam különbséget a mássalhangzók mentén (a kutatás CVC szekvenciákat vizsgált, melyekben az első és utolsó hang megegyezett, ebben a kutatásban azonban nem vettem össze az izoláltan és a CVC hangsorban ejtett hangzók felismerését).

Smith és Scott (1980) adataival azt is demonstrálta, hogy az izolált ejtésű magánhangzók sem válnak teljesen azonosíthatatlanná, még magasabb alaphangokon sem – ez pedig ellentmondani látszik az alaphanggal összefüggő artikulációs és akusztikai sajátosságok alapján esetlegesen elvárható tendenciáknak. Eredményeiket összehasonlítva Gottfried és Chew adataival pedig az is látszik, hogy ez az izolált ejtésben tapasztalt stabilitás nagyobb mértékű, mint ami a CVC hangsorból kivágott tiszta fázisok azonosításánál tapasztalható. A hangszeletek észlelésének sikeressége (Gottfried és Chew kísérleté-

ben) ugyanis még a mélyebb alaphangtartományban sem haladta meg a 30%-ot, míg az izoláltan ejtett vokálisokat (Smith és Scott kutatásában) mélyebb alaphangokon akár több mint 50%-ban is az ejtési szándéknak megfelelően azonosították a hallgatók (és drasztikus csökkenés csak az F5 zenei alaphangon következett be). A két kutatás összevetése nyomán tehát kézenfekvőnek látszik a feltetelezés, hogy az önmagukban ejtett magánhangzók valamelyest dinamikusabb szakasza, mégpedig a zöngé indulása is szolgáltathat olyan dinamikus akusztikai információt, melyre a hangazonosításkor észlelésünk támaszkodni tud. Ennek a vizsgálatára azonban ismereteink szerint még nem történt meg.

A jelen kutatás célja, hogy kibővítsé Smith és Scott (1980) korábbi kísérletét (Smith és Scott 1980, valamint Gottfried és Chew 1986 eredményeinek összehasonlítása nyomán) a magánhangzó-azonosítást befolyásoló változók szigorúbb kontrollálása mellett. Célunk nem csak a mássalhangzó-környezet, de a zöngéindulás hatásának tesztelése is. Hipotézisünk szerint a magánhangzók észlelése mássalhangzó-környezetben a legsikeresebb, melynél az izoláltan ejtett hangzók azonosítása valamivel kevésbé pontos, a legrosszabb azonosítási arányokat pedig a zöngéindulásra vonatkozó információt nem tartalmazó vokálisok esetében tapasztalhatunk. Ugyanakkor azt is feltételezzük, hogy a mássalhangzók és a zöngéindulás pozitív hatása az alaphang emelésével gyengül a magas alaphang okozta alul-mintavételezés, azaz a „rossz felbontás” miatt.

Anyag, módszer, kísérleti személyek

A kutatásban egy szoprán énekes énekprodukciónak rögzítettük. Az énekes a magyar nyelv három, akusztikai tekintetben egymástól legtávolabb eső magánhangzóját (/a: i: u:/) ejtette izoláltan („V” kondíció) és /bVb/ hangkörnyezetben („CVC” kondíció), hat énekelt alaphangmagasságon (F3 = 175 Hz, B3 = 247 Hz, F4 = 349 Hz, B4 = 494 Hz, F5 = 698 Hz, B5 = 988 Hz) és beszélve (melynek átlagos alaphangmagassága 191 Hz volt). Ezen kívül disztraktor hangsorokat is rögzítettünk ugyanazon énekessel. A disztraktorok további magyar magánhangzókat (ø ε e: o: ø: y:) tartalmaztak /bVb/ hangkörnyezetben és izolált ejtésben, beszélve, illetve a B4, F4, F5 zenei alaphangokon énekelve. Ezen stimulusok percepciósi eredményeit nem elemeztük. A kutatásban használt hordozóhangsor megválasztását a következők motiválták. A mássalhangzó-környezet szigorú kontrollálása érdekében olyan CVC hordozóhangsört használtunk, melyben a két mássalhangzó azonos, így az adataink egyértelműen az adott képzéshelyű mássalhangzóra vonatkozathatóak. Azonban ahhoz, hogy az eredményeinket a Smith és Scott (1980) kísérletében látottakkal is összehasonlíthatóvá tegyük, az általuk használt /bVd/ hangsor egyik mássalhangzóját kellett megtartanunk. Korábbi kutatási tapasztalataink szerint a CVC hangsor első mássalhangzója gondosabban artikulált az éneklésben, mint az utolsó, ezért a /bVd/ hangsor első mássalhangzójának a megtartása mellett döntöttünk. A magánhangzó zöngéindulásának hatását egy

harmadik kondíció előállításával teszteltük: egy hangeditáló program, a Wavesurfer (Sjölander–Beskow 2009) segítségével („fokozatos beúszás” hatás alkalmazásával) eltávolítottuk az izoláltan ejtett hangok elejét („Vágott” kondíció). A beúsztatást 30 ms-os időtartamon, logaritmikus függvény segítségével valósítottuk meg. A felvételeket csendesített szobában rögzítettük 44,1 kHz-es mintavételezéssel, 16 biten.

A hanganyagokat egy, a Praat programban (Boersma–Weenink 2013) írt és futtatott percepciós tesztben vizsgáltuk 22 ép hallású felnőtt részvételével. A teszt (3 kondíció \times 3 magánhangzó \times 7 alaphangmagasság =) 63 stimulust és 30 disztraktort tartalmazott. Minden tesztelő minden stimulust kétszer hallott, véletlenszerű sorrendben, így egy ülésben összesen 156 hangminta hangzott el. A teszt előtt a kísérleti személyek azt az utasítást kapták, hogy azonosítsák az önmagában vagy hangsorban elhangzó magánhangzót, és döntésüket a képernyőn megjelenő kilenc magánhangzó egyikének megjelölésével rögzítsék. A képernyőn a magyar hosszú (vagy hosszan ejthető) magánhangzókat (/v a: e: i: o: ø: u: y:/) jelenítettük meg ortografikus alakban. A tesztelők a stimulusokat binaurálisan, fejhallgatón keresztül hallották, és a számítógéphez csatlakoztatott egérrel kattinthatnak a kiválasztott magánhangzóra. A teszt átlagosan 10-15 percet vett igénybe, mivel azonban a stimulusok sorrendje minden tesztben eltérő volt, nem volt szükséges figyelembe vennünk a fáradás következtében esetlegesen felléphető tesztelést.

A válaszokat tévesztési mátrixokban összesítettük minden stimulus-válasz párt alaphangmagasságonként rendezve. Ezen a ponton le kell szögezünk, hogy a tanulmányban használt „helyes válasz” és „tévesztés” terminusok a (stimuluslistával irányított) ejtési szándék és a válasz közötti egyezésként vagy eltérésként értelmezendők.

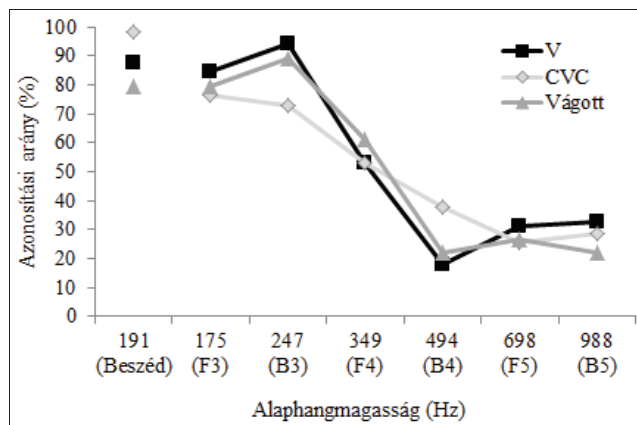
Az adatok statisztikai elemzését az R programmal végeztük (R Core Team 2013).

Eredmények

A szélsőséges teljesítményt nyújtó adatközlők kiszűréséhez minden adatközlőre „konzisztenciaarányt” számoltunk. Ez az érték azt a százalékos arányt mutatja meg, hogy az adott adatközlő az egyes stimulusokat tekintve mennyire volt következetes, azaz mindkét alkalommal ugyanúgy (bár nem feltétlenül helyesen) azonosította-e őket. Mivel minden adatközlő 63 stimulusra adott választ (kétszer), így a következetes válaszok számát (mely tehát maximum 63 lehetett) is erre a számra vetítettük. A konzisztenciaarány arra utal, hogy az adott adatközlőt mennyiben a hangok tényleges azonosítása, és mennyiben csak a találgatás vezette a válaszadáskor. Az első esetben ugyanis a két ismétlés megegyező azonosítása, a második esetben pedig a két ismétlés eltérő felismerése valószínűbb. Ugyanakkor mivel az f_0 emelésével a magánhangzók azonosíthatóságának csökkenését várjuk, ezért az alacsony értékek önmagukban nem az adatközlő megbízhatatlanságát jelzik. Ennek megfelelő-

en pusztán azt vizsgáltuk meg, hogy az értékek egymáshoz viszonyítva egyenletes eloszlásúak-e, azaz homogén csoportot képeznek-e az adatközlők. Mivel a χ^2 -próba tanúsága szerint nincs kilógó érték [$\chi^2(21) = 13,5221$; $p = 0,89$], és az adatok a Shapiro–Wilk-próba szerint normál eloszlásúak, így adatközlőink teljesítményét hasonlóknak ítéltük, és egyetlen adatközlő választait sem kellett kizárni az elemzésből.

A 2. ábra a három magánhangzó helyes azonosításának százalékos arányát (y tengely) mutatja az alaphangmagasság függvényében (x tengely), összesítve. Az adatok megjelenítése azt sugallja, hogy a vokálisok azonosítása csökken az f_0 emelésével. A kéttényezős ANOVA szerint, melyben az f_0 és a kondíció hatását teszteltük, ez a benyomás helyes, és az f_0 paraméter valóban hatással van az azonosítási arányok alakulására [$F(1) = 33,33$, $p < 0,001$]. A lehetséges összes pár összehasonlítása (Bonferroni-korrekcióval) azt is felfedte, hogy az egymást alaphangmagasság szerint követő énekelt stimulusok között egészen B4-ig (494 Hz) mutatkozik páronként szignifikáns eltérés ($p < 0,02$), mely alaphangtól felfelé azonban a különbségek már nem ilyen jelentősek. A Pearson-féle korrelációs tesztek azt mutatták, hogy az f_0 és az azonosítási arány kapcsolatát viszonylag erős negatív korreláció jellemzi. Amennyiben a három magánhangzóra kapott adatokat egy csoportban vizsgáljuk, erős összefüggést találunk: $r = -0,631$, $p < 0,001$. Ha pedig szétválasztjuk az adatokat a három kondíció mentén, némileg eltérő erősségű, de minden esetben szignifikáns összefüggéseket látunk: V: $r = -0,590$, $p = 0,005$; CVC: $r = -0,609$, $p = 0,003$; vágott: $r = -0,703$, $p < 0,001$.



2. ábra

A három vizsgált magánhangzó azonosításának összesített százalékos aránya az alaphang függvényében

A kéttényezős ANOVA azonban nem igazolta, hogy a kondíció is hatással lenne az észlelési eredményekre (sem önmagában, sem pedig az f_0 változóval kombinálva). A 2. ábrát szemlélve az is kitűnik, hogy a kondíció és az észlelési arányok közötti összefüggés nem is konzisztens, azaz nemcsak az állítható, hogy nem bizonyított a magánhangzók jobb azonosítása mássalhangzó-környezetben, hanem az is, hogy néha (például a B3 zenei alaphangon) kifejezetten rosszabbnak bizonyul a CVC szekvenciában elhangzó vokálisok felismerése. A különbségek statisztikailag ugyan nem igazolhatóak, általánosságban azonban mégis úgy látszik, nagyobb az eltérés a V és CVC, illetve a Vágott és a CVC minták felismerési aránya között, mint a V és Vágott minták esetében.

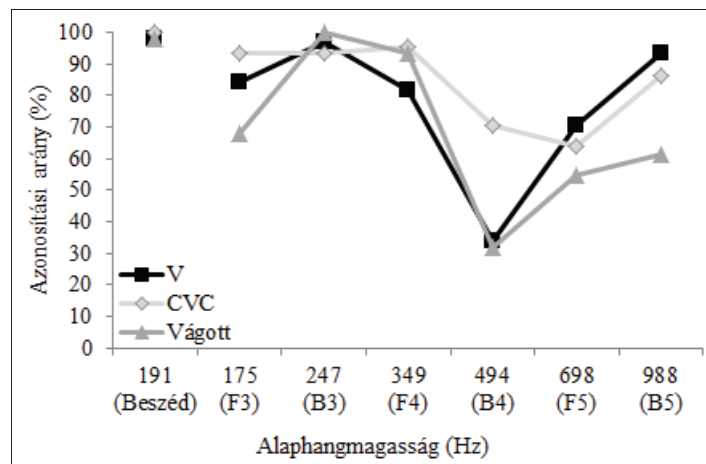
Bár az éneklésben nem igazolódott a fonetikai kontextus (a mássalhangzó-környezet és a zöngéindulás) hatása, a beszédben az elvárt tendenciákat látjuk: a legmagasabb azonosítási arányt a CVC környezet mutatja, melyet az izolált ejtésű, és végül pedig a vágott stimulusok követnek.

Ugyan a fonetikai kontextus hatása a magánhangzó-azonosításra nem igazolódott, az f_0 , a kondíció és a magánhangzó-minőség változókkal lefuttatott varianciaanalízis szerint nemcsak az f_0 , de a magánhangzó-minőség [$F(2) = 22,673$; $p < 0,001$] is, illetve e két faktor kombinációja is [$F(2) = 7,419$; $p = 0,002$] befolyásolja a hangzóazonosítási értékeket. Ezért az azonosítási arányok változását a magánhangzó-minőségek mentén is elemeztük.

A magánhangzó-minőségek közti összes lehetséges páros szembenállás összevetése szerint Bonferroni-korrekció után is szignifikáns különbség mutatkozik az /i:/-/a:/ és /u:/-/a:/ párok, tehát a nyílt és zárt képzésű magánhangzók között ($p < 0,018$), de nincs különbség az állkapocsnyitásszögben egyező ejtésű, zárt vokális, azaz /i:/ és /u:/ között. Ezt az elkülönülést rajzolja ki az f_0 és az azonosítási arány közti korreláció is, ha azt a magánhangzók csoportjain belül vizsgáljuk, ugyanis a két változó csak az /i:/ és /u:/ esetében mutat erős, szignifikáns összefüggést [/i:/: $r = -0,781$ $p < 0,001$; /u:/: $r = -0,900$; $p < 0,001$], míg az /a:/ esetében csak gyenge nem szignifikáns összefüggés látszik ($r = -0,384$; $p = 0,086$). A statisztikai elemzések, továbbá a 3., 4. és 5. ábra tanúsága szerint tehát a zárt képzésű vokálisok azonosításának sikeressége csökken az alaphang emelésével, míg a nyílt képzésű beszédhanghoz tartozó százalékok az f_0 értékétől megközelítőleg függetlenül alakulnak.

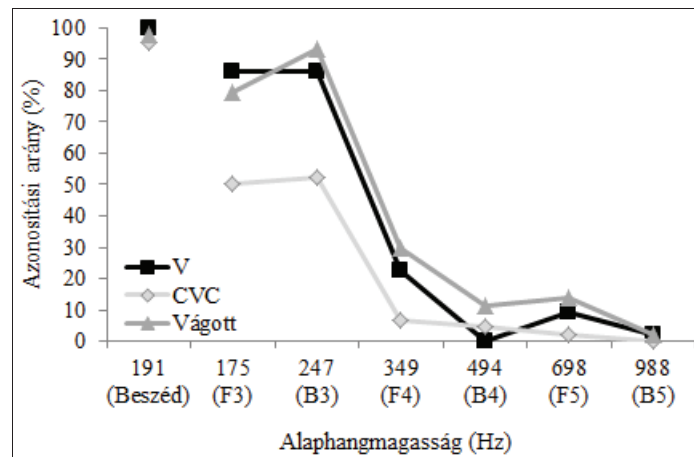
Az /a:/ azonosítási aránya B4 (494 Hz) alatt viszonylag magas minden kondícióban (3. ábra). B4-nél azonban hirtelen esést látunk. Ezen az alaphangon 46%-ban tévesztettek a tesztelők, és minden esetben a magasabb első formánssal ejtett (zártabb képzésű) /v/-t jelölték meg válaszként. A B4 fölött újabb emelkedés következik, majd az azonosítási arányok folyamatosan emelkednek a legmagasabb alaphangig, ahol izolált ejtésben 93%, CVC környezetben 86%, míg eltávolított zöngéindulással 61% a helyes válaszok százalékos aránya. F5-nél ismét az /v/ hang dominálja a tévesztéseket (az összes válasz 27%-a), míg B5-nél az elől képzett /ε/ (az összes válasz 8%-a). Mind-

két hangzó magasabb első formánssal realizálódik beszédben, mint az ejteni kívánt hang.



3. ábra

Az /a:/ azonosításának százalékos aránya az alaphang függvényében

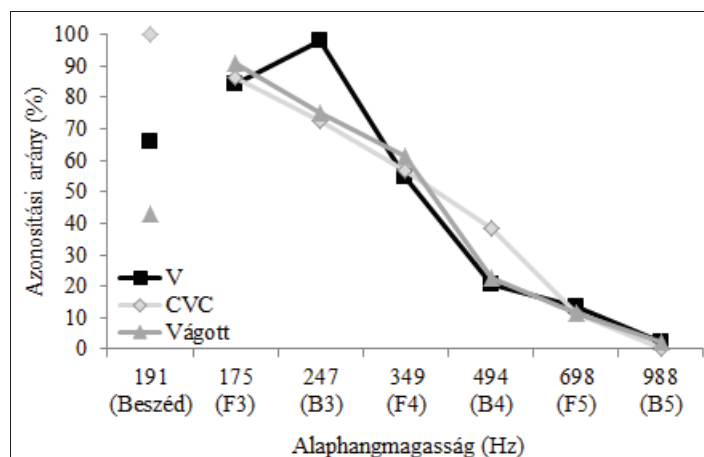


4. ábra

Az /i:/ azonosításának százalékos aránya az alaphang függvényében

Az /i:/ azonosításának százaléka az F4 alaphangnál mutatnak hirtelen csökkenést (4. ábra), itt és a magasabb alaphangok esetében is a hibák legnagyobb része /e/-re történő tévesztés (F4: 80%, B4: 94%, F5: 67% az összes

válaszból). A B5, azaz a legmagasabb alaphang esetében azonban az /i:/-t is leggyakrabban /a:/-ként ismerik fel a hallgatók (az összes válasz 80%-ában), a helyes azonosítás aránya 2% körüli (minden kondícióban).



5. ábra

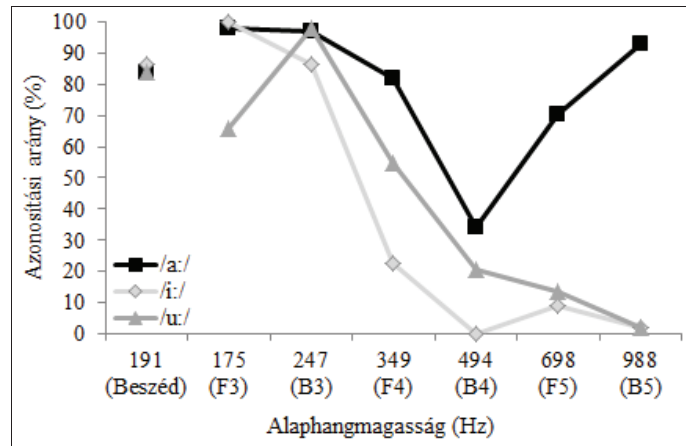
Az /u:/ azonosításának százalékos aránya az alaphang függvényében

Az /u:/ azonosítását leíró görbék (5. ábra) a 2. ábrán bemutatott átlageredményekhez hasonlóan alakulnak: az f_0 emelésével fokozatosan csökkenő százalékarányokat látunk, míg a V és CVC kondíciók között nem következetesen alakuló kapcsolatok mutatkoznak. A hibákat tekintve a legtöbb alaphangmagasságon leggyakrabban előforduló hangzó az /o:/ (F3 és B3: 13%, F4: 42%, B4: 62%, F5: 25% az összes válaszból), ugyanakkor a legmagasabb alaphangon (B5) itt is az /a:/ uralja a tévesztések hierarchiáját (az összes válasz 86%-ában), a vokális azonosítása pedig itt is maximum 2%-ban bizonyult az ejtési szándékkal megegyezőnek.

Az /u:/ beszélt ejtésben bemutatott realizációi (az /i:/-vel és /a:/-val szemben) a hipotézisben megfogalmazottaknak megfelelő tendenciát mutatják: a legjobb arányban a CVC hangsorbéli megvalósulást azonosították a hallgatók, melyet a V és végül a Vágott kondíciók követnek (5. ábra). Azonban ez az /u:/ esetében is csak a beszélt ejtésre korlátozódó mintázat (a többi hang esetében pedig sem a beszédben, sem másutt nem látjuk ezt a tendenciát).

Összehasonlítottuk az összes lehetséges hangzószembenállást a három kondíció alkotta csoporton belül is. Ezek a tesztek megmutatták az egyetlen olyan tendenciát, mely a jelen adatokban a kondícióval, és azon belül is a mássalhangzó-környezet meglétével is összefüggésbe hozhatónak bizonyult. Egyik csoporton belül sem találtuk ugyanis eltérőnek a különböző állkapocs-

nyitásszöggel képzett hangzókat, kivétel a CVC csoportot. Itt ugyanis az eltérő állkapocsnyitással képzett hangok azonosítási arányát eltérő mértékben látszott befolyásolni az f_0 értéke – bár csak a nyelv vízszintes mozgásának tekintetében jobban eltérő pár, az /i:/ és az /a:/ esetében találtunk a Bonferroni-korrekció után is ($p < 0,02$) szignifikáns eltérést (a nyelv vízszintes mozgásának tekintetében egymáshoz hasonlóbb /u:/ és /a:/ nem tért el). Az összehasonlításokat a 6., 7. és 8. ábrán szemléltetjük.



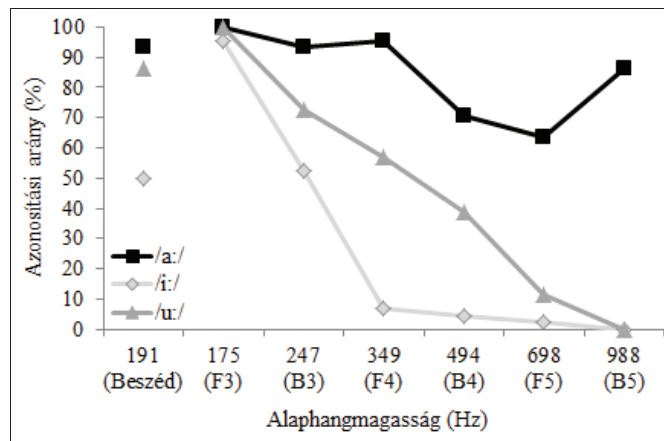
6. ábra

A magánhangzó-azonosítás aránya izolált ejtésben (V kondíció)

Az adatokból jól látható, hogy a különbségek megjelenése (illetve meg nem jelenése a többi kondícióban) annak köszönhető, hogy a CVC kontextusban tapasztalható leginkább az /a:/ percepciós stabilitása, ugyanakkor a zárt /i:/ is itt mutatja a legalacsonyabb felismerési arányokat. Míg a V és Vágot kondíciókban csak B4 fölött, addig a CVC kondíciókban már a jóval alacsonyabb F3 fölött is divergáló tendenciákat látunk a három magánhangzóra.

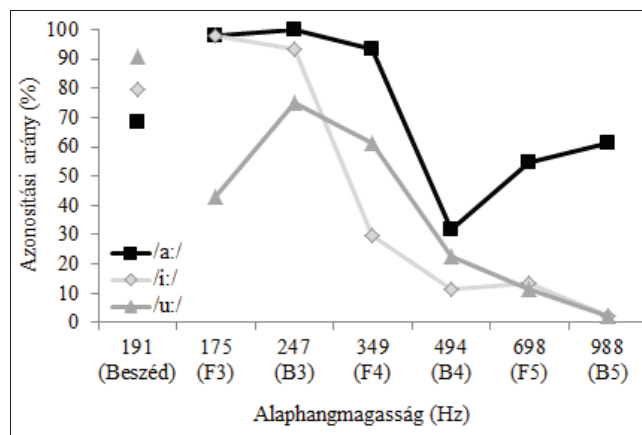
Következtetések

A jelen kutatásban a mássalhangzó-környezetnek és a zöngéindulásnak az énekelt magánhangzók percepciójára kifejtett hatását vizsgáltuk. A két, a témában folytatott korábbi kutatás eredményeire (Smith–Scott 1980 és Gottfried–Chew 1986), valamint a beszédpercepció szakirodalmára (Strange–Verbugge 1976) alapozva azt feltételeztük, hogy a formánsátmenetekben, illetve a zöngéindulásban kódolt dinamikus akusztikai információ segíti a magánhangzók azonosítását – bár ez a hatás az alaphang emelésével (a toldalékcso átviteli karakterisztikájának spektrális alul-mintavételezettségéből fakadóan) csökken.



7. ábra

A magánhangzó-azonosítás aránya mássalhangzó-környezetben (CVC kondíció)



8. ábra

A magánhangzó-azonosítás aránya a zöngéindulás nélkül bemutatott magánhangzók esetében (Vágott kondíció)

A hipotézist egy szoprán énekes ejtésében, három magánhangzó (/a: i: u:/), három kondíció (/bVb/ hangkörnyezet, izolált ejtés, eltávolított zöngéindulás), továbbá 6 énekelt alaphangmagasság (az F3 és B5 zenei hangok között) és a beszélt ejtés segítségével teszteltük egy percepció vizsgálatban, melyben 22 kísérleti személy vett részt.

A kutatás legfontosabb eredménye, hogy sem a mássalhangzó-környezet, sem az izolált ejtésű magánhangzó zöngéindításának pozitív hatását nem sikerült igazolnunk: a kondíciók páronként inkonzisztens eltéréseket mutattak. Az eltérések következetlensége azt jelenti, hogy nemcsak a CVC kontextus, illetve a zöngéindulás pozitív hatását nem sikerült demonstrálnunk, de éppen az elvárásokkal ellenkező példákat is találtunk (a CVC esetében általánosan a B3 alaphangon, a zöngéindulás esetében pedig a B3 alaphangon az /a:/ hangra). Továbbá mivel a V és Vágott kondíciók gyakorlatilag nem mutattak különülést, ezért azt is feltételezhetjük, hogy a magánhangzó és a mássalhangzók közti formánsátmenetek feltehetőleg nagyobb hatással bírnak az azonosításra, mint a zöngé indulása. Az eredményeket két, egymást kiegészítő tényező kiemelésével magyarázhatjuk: a magas alaphangon ejtett (zöngés) hangok spektrális alul-mintavételezettségével és az operaéneklésnek a mássalhangzók képzését érintő sajátosságaival.

Amint azt a bevezetőben bemutattuk, a magánhangzók spektruma a magasabb énekelt alaphangokon a felhangok nagy távolsága miatt alul-mintavételezetté válik. Azonban ez az alul-mintavételezettség nemcsak a vokálisok tiszta fázisát jellemzi, hanem minden olyan beszédhangot és hangrészletet is, melyek artikulációjának akusztikai lenyomatát a hangszalagműködés következtében létrejövő zöngé hozza létre. Ennek megfelelően a zöngés mássalhangzók és a bármely két szonoráns között realizálódó formánsátmenetek is alul-mintavételezetté válnak. Feltételezhetjük tehát, hogy a kutatás eredményei azt demonstrálták, hogy amiként a magánhangzók tiszta fázisának formánsait, úgy a hangkapcsolatokat átkötő formánsátmeneteket sem kódolja megfelelő pontossággal a magas alaphangú zöngé az éneklés akusztikai lenyomatában, így a formánsmozgásokban rejlő akusztikai kulcsok az éneklés-kor a percepció számára nem elérhetőek.

A másik, az előzőt is kiegészítő magyarázat az énekelt mássalhangzók ejtésén alapszik. Az operaéneklés (vagy másként a *bel canto* technika) egyik alapvetése a magánhangzók időtartamának maximalizálása a mássalhangzók időtartamának rovására. Ennek a mássalhangzó-redukciónak a módját és jellegét azonban, a szerző ismeretei szerint, kísérletes úton eddig még csak egyetlen alkalommal vizsgálták. Deme, Grácsi és Jankovics (2013) az énekelt obstruensek zöngéességének alakulására fókuszáló kutatásában a /b/ zöngés explozív rövidülését mutatta be szókezdő és szózáró helyzetben, illetve a szókezdő, szó belseji és szózáró explozívak és frikatívák zöngéességének megváltozását is leírják. Ezen túlmenően azonban a szerzők beszámolnak arról is, hogy a mássalhangzók címkézésekor gyakori problémát jelentett az explozívak felpattanásainak elmaradása (vagy legalábbis detektálhatatlansága a spektrogramon és az oscillogramon), illetve a réshangok turbulens zörej-összetevőinek alacsony intenzitása. Az obstruensek pontatlan képzése és egyes akusztikai összetevőinek csekély intenzitása vagy hiánya azonban nemcsak magukra a mássalhangzókra, illetve azok azonosíthatóságára van

hatással, hanem a koartikulációs formánsátmenetekre és azok akusztikai lenyomatára is: ha például a magánhangzót megelőző explozív felpattanása elmarad, a formánsátmenetek sem jelennek meg vagy nem lesznek tisztán ki-vehetőek. Ez pedig nyilvánvaló módon a formánsátmenetek percepció el-érhetőségét is akadályozni fogja.

Mindezek mellett azonban arról sem szabad megfeledkeznünk, hogy ma-gasabb alaphangok esetében a vokálisok artikulációjának jelentős megvál-tozását, elsősorban nagyobb mértékű állkapocsnyitást kell feltételeznünk. Ha tehát egy megváltozott ejtésű vokális (egy adott alaphangon) percepciósan már nem az ejtési szándéknak megfelelő, hanem egy másik magánhangzó-kategóriába kerül, a koartikulációban rejlő és esetlegesen akár el is érhető in-formáció sem az ejteni kívánt, hanem „csak” erre az utóbbi, az ejteni tudott hangzóra utalhat (azaz valamivel következetesebbé teheti a „tévesztési” ten-denciákat).

Természetesen az adatok kimerítő értelmezéséhez szükséges a különböző képzéshelyű mássalhangzók (tehát nemcsak a bilabiális, de az alveoláris és veláris felpattanók) vizsgálata is, melyek képzéshelyük (illetve eltérő frek-veciaszerkezetük) következtében eltérő formánsátmenetekkel realizálódnak a CVC szekvenciákban. Ezt, tehát a képzéshely okozta különbségeket követ-kező kutatásainkban tervezzük vizsgálni.

A jelen kutatás eredményei ellentmondanak a két korábbi vizsgálatban (Smith–Scott 1980; Gottfried–Chew 1986) olvashatóknak (ahogyan a be-szédpercepció szakirodalma, vö. Strange–Vergbrugge 1976, alapján megfo-galmazható feltételezéseknek is). Kérdés, hogy mivel magyarázható ez az el-lentmondás. Amint arra már a bevezetőben is utaltunk, Smith és Scott ered-ményei az adatokat befolyásoló változók nem szigorú kontrollálása következ-tében nehezen interpretálhatók – ezzel szemben a jelen vizsgálatban mind a felvételek, mind pedig a tesztelés alaposan kontrollált és felügyelt volt. Ezen túlmenően Smith és Scott mindössze 10 kísérleti személy anyagát használta fel – ehhez képest a jelen kutatásban több mint kétszeres adatmennyiséggel, 22 adatközlő válaszaival dolgoztunk. Harmadrészt feltehető az is, hogy a két kísérletben használt eltérő magánhangzók és azok képzésének (a centralizá-ció mértékének) különbségei is befolyásolták az adatokat. Smith és Scott ugyanis az /i ɪ e æ/ vokálisokat használta, egy feltehetőleg ausztrál akcentusú beszélő ejtésében. Végezetül pedig felmerülhet az a gyanú is, hogy a két kutatásban kapott eredmények a kísérletben részt vevő énekesek egyéni ének-technikai sajátosságaiból fakadnak. A szakirodalom alapján azonban jelenleg megválaszolhatatlan az a kérdés, hogy az egyes (profí) operaénekesek ren-delkezhetnek-e olyan mértékben eltérő egyéni artikulációs jellemzőkkel (mind a magán-, mind a mássalhangzók ejtését tekintve), melyek magyaráz-hatnák a korábbi elemzések és a jelen kutatás eredményei közötti jelentős elté-réseket, így ennek a feltételezésnek a vizsgálata további kutatásokat igényel. Az mindenesetre biztosnak látszik, hogy a kutatások eredményei közötti kü-

lönbség nem a jelen kutatás módszertani felépítésének esetleges problémáiból vagy a jelen adatok kétes hitelességéből fakad.

Bár a mássalhangzó-környezet és a magánhangzók zöngéindulása nem befolyásolta a magánhangzók azonosítását, az f_0 és a magánhangzó-minőség, továbbá a kettő interakciója statisztikailag is igazolható hatást gyakorolt a felismerési arányokra. Ebből következően az elemzés során a magánhangzó-minőségekkel és az alaphangmagassággal összefüggő tendenciákat is részletebben elemeztük. E két változó mentén megközelítőleg a korábban leírtakkal egybevágó tendenciákat találtunk (vö. Scotto di Carlo–Germain 1985; Gottfried–Chew 1986; Sundberg 1994; Deme 2012):

1. Általánosságban elmondható, hogy a magánhangzó-azonosítás feladata egyre nehezebbé vált (azaz a helyes azonosítások aránya csökkent) az alaphang emelésével. A magánhangzó-minőségek mentén azonban eltéréseket tapasztaltunk (lásd 2. és 3. pont).

2. A zárt képzésű /i:/ és /u:/ azonosításának aránya határozottan csökkent az f_0 növelésével párhuzamosan. A beszédben mindkét hangzó első formánsa alacsony értéken jelenik meg, melyet tehát már a közepesen magas alaphangok esetében is elér, tehát befolyásol(hat) az f_0 . Mindkét vokálist elsősorban (egy fokkal) nyíltabb, magasabb első formánssal rendelkező hangzókra tévesztették a hallgatók. Az /i:/-re és /u:/-ra kapott percepciós adatok tehát megfelelnek az artikulációs és akusztikai jellemzők alapján feltételezhetőeknek, melyek szerint az f_0 emelése az éneklésben nagyobb állkapocsnyitással, így az első formáns emelésével jár (az eredeti hangzó artikulációs jellemzőitől független módon), a hangzókat pedig ennek megfelelően nyíltabbként is észleljük.

3. A legnyíltabb magyar magánhangzó, az /a:/ azonosíthatósága (a zárt hangzókkal szemben) még magasabb alaphangok esetén is stabilabbnak bizonyult. Ez nem meglepő, hiszen az /a:/ ejtésének kedvez az éneklés egyik artikulációs szükségszerűsége, az állkapocsnyitás-növelés, mely az intenzív és homogén hangszínezetet fenntartó alaphangemelés eszköze magas alaphangon. Az akusztikum tekintetében ezt úgy lehet megfogalmazni, hogy az /a:/ produkcióját kevésbé befolyásolja az f_0 , hiszen az alaphang értéke az /a:/ magas első formánsát csak kifejezetten magas alaphangmagasságon éri el. Így tehát azt várhatjuk, hogy az /a:/ észlelésére is csak a kifejezetten magas alaphangú ejtés van hatással. A korábbi percepciós adatok egy része igazolja is ezt a teoretikus alapokon nyugvó elvárást (Scotto di Carlo–Germain 1985; Gottfried–Chew 1986; Sundberg 1994 egyes eredményei), ugyanakkor más kutatásokban az /a:/ percepciós instabilitásáról is beszámolnak (l. Deme 2012 összefoglaló munkáját). Az utóbb említett vizsgálathoz hasonlóan a jelen kutatásban is találkozunk az instabilitás példáival, a B4 és F5 hangmagasságok esetében ugyanis az /a:/ alacsonyabb azonosítási arányt mutatott, a tévesztések hierarchiáját legnagyobb részben az alacsonyabb első formánsú /ɒ/ uralta. Ezek a tendenciák azonban csak első ránézésre tűnnek váratlan eredménynek. A beszédhangok percepcióját vizsgáló szakirodalom ugyanis leírja, hogy a

vokálisok zártságának percepciója nem az első formáns abszolút értékében, hanem sokkal inkább annak az f_0 -hoz viszonyított tonotopikus elhelyezkedésében van kódolva: az első formáns és az f_0 frekvenciaértékének közelítése a zártabb ejtés, míg az értékek távolodása a nyitottabb ejtés érzetét kelti (Traunmüller 1981). Mindezek alapján tehát azt feltételezzük, hogy a jelen (és a korábbi) kutatás által bemutatott zártabbra tévesztések és a nyílt ejtésű hangzó perceptuális bizonytalansága az f_0 és az első formáns változó (csökkenő) viszonyából, és az ehhez kapcsolódó észlelési sajátosságból fakad.

A magánhangzó-azonosítási arányok kondíciók szerinti összehasonlításakor a mássalhangzó-környezettel összefüggésbe hozható egyetlen tendenciát is megtaláltuk: a magánhangzók ugyanis a CVC kondícióban mutatták a legszéttartóbb tendenciákat. Az /a:/ esetében itt tapasztaltuk a legmagasabb, míg az /i:/ esetében a legalacsonyabb arányokat, a két, az állkapocsnyitás szögében különböző hangzó észlelési tendenciái ebben a kondícióban tértek el a legnagyobb mértékben. Ez az eredmény egybecseng a korábbi megjegyzésünkkel, mely a magánhangzók kategorikus észlelésére hivatkozik, és paradoxnak tűnő módon, bizonyos tekintetben a hipotézisünket is alátámasztja. Az /a:/ esetében tapasztalható, a többi környezettel szemben nagyobb mértékűnek tűnő stabilitás könnyen belátható módon utalhat a mássalhangzó-környezet percepciót segítő hatására. Ugyanakkor az a tény, hogy az /i:/ már alacsonyabb alaphangon is rosszabb felismerési arányt mutat CVC kondícióban, szintén lehet éppen annak a jele, hogy a mássalhangzók segítik a hangazonosítást. Ha ugyanis az őt érintő artikulációs változások akusztikai következményeinek folyamánaképpen az /i:/ már alacsonyabb alaphangokon is egy másik percepciók kategóriába kerül, az ott tapasztalt alacsony azonosítási arányok esetleg éppen azt mutatják, hogy a hallgatók ezt a másik kategóriát ismerik fel, ráadásul jó arányban.

A produkció módja két okból is az észlelést befolyásoló tényezőnek tűnik. Egyfelől az énekelt mintákkal szemben a beszédben ejtett hangzók több esetben is (az összesített eredményekben, illetve az /u:/ esetében is) mutatták a feltételezett, de az éneklésre nem igazolódott tendenciát: a mássalhangzó-környezet és a zöngéindulás pozitív hatását. Másfelől az /a:/ és /i:/ esetében a beszélve ejtett realizációk mutatták a legmagasabb azonosítási arányokat (alaphangmagasságtól függetlenül) még azokhoz az énekelt alaphangmagasságokhoz képest is, melyek a beszéd átlagos f_0 -értékéhez közeliak voltak.

A kutatás eredményeit látva megállapíthatjuk, hogy a mássalhangzó-környezet az énekelt magánhangzók azonosítására gyakorolt hatásának vizsgálatához a jövőben nemcsak a képzéshely hatásából fakadó percepciók különbségek vizsgálata, de a mássalhangzók realizációjának és esetleges redukciójának kutatása is szükséges. Addig ugyanis nincs okunk feltételezni, hogy a mássalhangzó, illetve a koartikulációs formánsátmenetek bármilyen befolyással bírhatnak a percepciók működésekre az énekelt magánhangzók esetében, míg az sem tisztázott, hogy maga a konzonáns és a formánsátmenetek

egyáltalán megjelennek-e, és elérhetőek lesznek-e az akusztikumban, illetve a percepció számára.

Irodalom

- Boersma, Paul – Weenink, David 2013. *Praat: doing phonetics by computer [Computer program]. Version 5.3.* <http://www.praat.org>. (Letöltés ideje: 2013. júl. 9.)
- Deme, Andrea 2012. On the Hungarian sung vowels. *The Phonetician* 105–106. 73–87.
- Deme, Andrea – Grácz, Tekla Ételka – Jankovics, Julianna 2013. Obstruent voicing in singing. (Előadás) *15th International Summer School of Psycholinguistics*, 2013. május 26–30; Balatonalmádi, Hungary.
- Evgrafova, Karina – Evdokimova, Vera 2012. Perception of Russian vowels in singing. *Baltic HLT Frontiers in Artificial Intelligence and Applications*. IOS Press 247. 42–49.
- Garnier, Maëva – Henrich, Natalie – Smith, John – Wolfe, Joe 2010. Vocal tract adjustments in the high soprano range. *Journal of the Acoustical Society of America* 127/6. 3771–3780.
- Goldstein, Ursula 1980. *An articulatory model for the vocal tracts of growing children*. D.Sc. thesis. Massachusetts Institute of Technology, Cambridge, MA.
- Gottfried, Terry L. – Chew, Stephen L. 1986. Intelligibility of vowels sung by a countertenor. *Journal of the Acoustical Society of America* 79/1. 124–30.
- Hollien, Harry – Mendes-Scwartz, Ana P. – Nielsen, Kenneth 2000. Perceptual confusions of high-pitched sung vowels. *Journal of Voice* 14/2. 287–298.
- R Core Team 2013. *R: A Language and Environment for Statistical Computing*. <http://www.R-project.org>.
- Ryalls, John. H. – Lieberman, Philip 1982. Fundamental frequency and vowel perception. *Journal of the Acoustical Society of America* 72/5. 1631–1634.
- Scotto di Carlo, Nicole – Germain, Aline 1985. A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica* 42/2. 188–97.
- Sjölander, Kåre – Beskov, Jonas 2009. *Wavesurfer*. <http://www.speech.kth.se/wavesurfer>.
- Smith, Lloyd A. – Scott, Brian L. 1980. Increasing the intelligibility of sung vowels. *Journal of the Acoustical Society of America* 67/5. 1795–1797.
- Strange, Winifred – Verbrugge, Robert R. 1976. Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America* 60/1. 213–224.
- Sundberg, Johan 1987. *The science of the singing voice*. Northern Illinois University Press, Illinois.
- Sundberg, Johan 1994. Perceptual aspects of singing. *Journal of Voice* 8/2. 106–122.
- Traunmüller, Hartmut 1981. Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America* 69/5. 1465–1475.

A kutatás alatt a szerző a Balassi Intézet Campus Hungary programjának ösztöndíjasa volt. A felvételek a Kempelen Farkas Beszédkutató Laboratóriumban készültek (MTA Nyelvtudományi Intézet, Budapest), az adatokat részben a Department of Speech, Music and Hearing (KTH Royal Institute of Technology, Stockholm) intézményben elemeztük, professzor Johan Sundberg (KTH Royal Institute of Technology) szíves segítségével.