

# A ZÖNGE PERIÓDUSIDEJÉNEK FUNKCIÓJA A HANGSZÍNEZETBEN

Nikléczy Péter

## Bevezetés

A beszédhangok képzésében jelentős szerepe van a hangszalagok által létrehozott, magas felhangtartalmú zöngeregzésnek, de a zöngeregzése önmagában nem tartalmazza az egyéni hangszínezet jellemzőit (Olaszy 1979). A hangszín kialakulása az illető személy artikulációs csatornájának, üregrendszerének alakjától és méretétől függ. Az általunk ismert személyeket az agyunkban elraktározott „hangképek” alapján ismerjük fel, amiben jelentős szerepe van az egyéni hangszínezetnek (Lavner et al. 1999). Kérdés az, hogy ezt az agyunkban tárolt egyéni hangszínt befolyásolja-e, ha a normál alaphangon beszélő személy hangját módosítjuk a zöngés hangok periódusidejének változtatásával. Vizsgálatunkkal arra próbáltunk választ kapni, hogy az alaphang mesterséges úton történő korrigálása az illető személy felismerését az észlelésben befolyásolja-e. Feltevésünk szerint, ha az adott hang kiejtésénél a személy artikulációs csatornája nem, csak az alaphangjának a frekvenciaértéke változik meg, akkor a megfelelő hangfekvésbe történő mesterséges transzformáció, lehetővé teheti a beszélő személy felismerését.

## A kísérlethez használt eszközök

A kísérleti hanganyagot Sound Blaster hangeditáló programmal készítettük el, a műszeres kiértékelést KAY CSL 4300B és PRAAT analízáló programokkal végeztük el. Az eredeti hanganyagot stúdióban rögzítettük Sony MZ-R900-as típusú minidisc hangfelvevővel.

## Anyag, a periódusok editálásának módszere, tesztelés

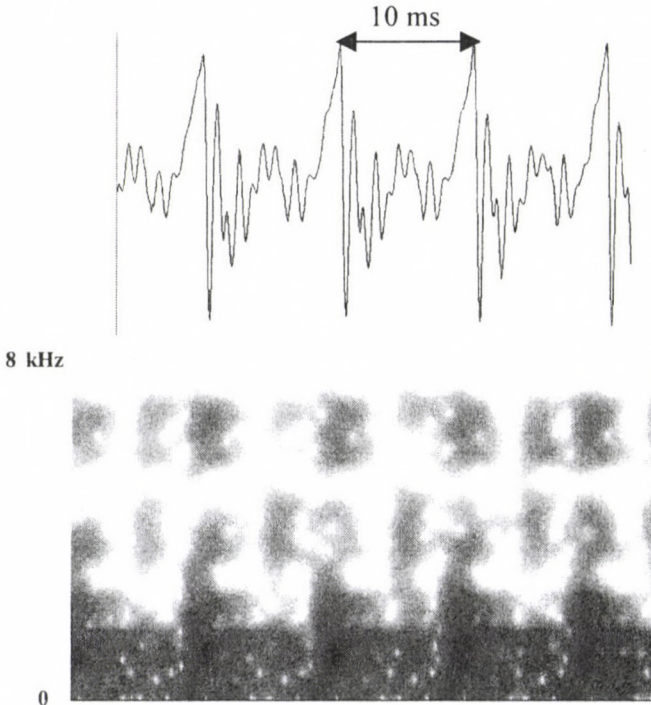
A kísérlethez használt hanganyag egy mondatrészlet volt, amelyet szándékosan úgy választottunk meg, hogy a percepció tesztelések során relatíve semleges szemantikai tartalmat képviseljen, továbbá hogy

rövidsége miatt a szupraszegmentális tényezők esetleges hatását a nullára redukáljuk. A hanganyag a következő volt: „*Lágyhang indításnál*”. A mondatrészletet egy 36 éves, magyar anyanyelvű, ép beszédű és hallású férfi mondta be. Kétféle bemondást rögzítettünk. Az első esetben normál ejtést kértünk, a második bemondáskor pedig mege-melt alaphangú artikulációt. A hanganyag időtartama mindkét esetben 1,7 s lett.

Az így kapott két hanganyag idődiagramjába mesterséges úton beavatkoztunk, úgy, hogy a periódusokban manuálisan változtatásokat hajtottunk végre. A hangkorrekciót a zöngés periódusok módosításával értük el oly módon, hogy (i) az egyes periódusok lecsengő szakaszát nulla intenzitású szünettel toldottuk meg, illetőleg (ii) a megfelelő periódusidő eléréséhez a kívánt hosszúságú részt kivágtuk. Ezzel azt értük el, hogy a hang karakterét meghatározó alacsonyabb frekvenciájú, nagy energiájú komponenseket nem változtattuk meg, ugyanakkor alacsonyabb, illetve magasabb hangfekvést tudtunk létrehozni.

A vizsgálat leírása előtt ismertetjük a spektrografikus elemzés azon részét, aminek alapján a hangkorrekciós eredményeket kiértékeljük. Mivel a beszédhangok is tartalmazzák a hangot kiejtő személy hangszínezetét, a spektrogramon elvileg azokat az egyéni jellemzőket is láthatjuk, amelyek az adott személyre utalnak. Megállapíthatjuk a beszéd alaphangértékét, vagyis a másodpercenként ismétlődő periódusok számát, a formánsok elhelyezkedését, sáv szélességét és intenzitását. Ha a hangszíneként ezeket az értékeket olvassuk le, azt láthatjuk, hogy ugyanazon személy különböző időben kiejtett azonos hangjánál az adatok változnak attól függően, hogy a személy az adott időpillanatban milyen alaphangértékkel ejtette ki a hangot. Nézzük meg elsőnek azt, hogy a spektrogram hogyan ábrázolja az egyes periódusokat. A periódusok megegyeznek egy függőleges vonallal, amelyek a formánshelyeknél feketedést mutatnak, illetve a vonal után (széles sávú elemzés esetén) kevésbé intenzív köztes résszel. A függőleges vonalak – melyek a hangszalagok nyitódásával megegyező időpillanatot realizálják (1. ábra) – a periódus kezdetén jelennek meg, majd a periódus köztes részén elhalványodnak. Az ábra felső részén a hang oszcillografikus rezgésképe, az alsó részén a spektrogramja látható. Ha ennek a hangnak csak három teljes periódusát vizsgáljuk, láthatjuk, hogy a

széles sávú spektrogram vonalas szerkezete csak a periódusok első intenzív részének frekvenciatartományát mutatja. Ennek alapján feltételezhető, hogy a periódus idődiagramjának lecsengő részében történő beavatkozás a hangszínekben nem okoz jelentős változást. Ha egy hangsor zöngés hangjait periódusonként rövidítjük meg a kis energiájú rezgések elhagyásával, magasabb vagy nulla intenzitású rész beiktatásával, akkor mélyebb hangfekvést tudunk létrehozni. Feltételezésünk szerint ez a beavatkozás a felhangok elhelyezkedésében és a percepció megítélésben változást hoz létre, a formánshelyek elhelyezkedésében azonban nem okoz eltérést.



1. ábra  
Normál hangfekvésben ejtett [a:] hang  
oszcollogramja és spektrogramja

A meghallgatásos tesztelés célja az volt, hogy megtudjuk, a zöngéváltozás hatással van-e a beszélő személyének felismerésére. A percepciók kísérletben 4 fő vett részt, akik a hangsorokat bemondó személyt jól ismerték. A hangsorok meghallgatása során elsőként a megemelt hangú beszédet, majd a korrigált beszédet és végül a normál hangfekvésű beszédet hallgatták meg a tesztelő személyek. Minden elhangzó mondatrészlet után a résztvevőknek nyilatkozniuk kellett, hogy felismerik-e az elhangzott hangsor alapján a beszélő személyt.

### **A tesztanyag előállítása**

A tesztanyagokat háromféleképpen állítottuk elő, és a kapott hanganyagokat műszeres analízissal ellenőriztük.

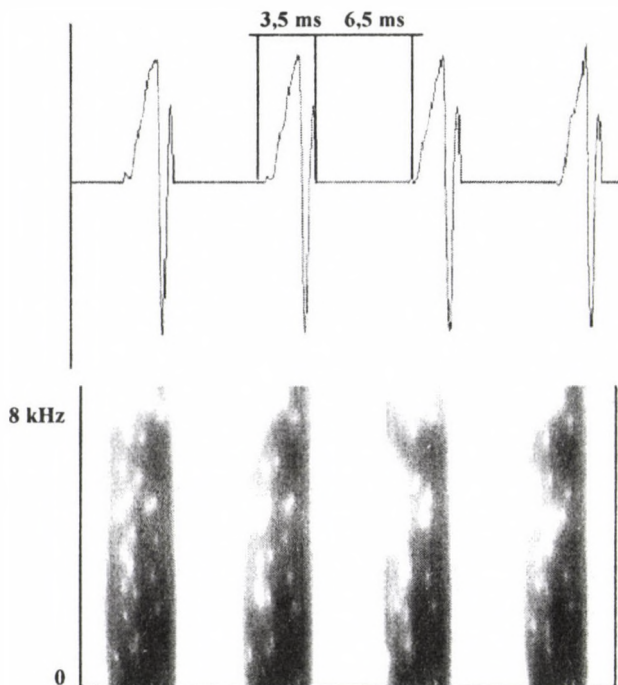
A hangedítálás formái a következők voltak:

1. Normál alaphangon ejtett férfi ( $F_0=100$  Hz) hangsor periódusaiból csak az első intenzív rezgést hagytuk meg, a többit nulla intenzitásúra változtattuk.
2. A normál ejtésű férfihang hangfekvését megemeltük oly módon, hogy a nulla intenzitású rész időtartamát csökkentettük. A csökkentés eredményeképpen háromszoros alaphangértéken szólalt meg a hangsor ( $F_0=300$  Hz).
3. A szándékosan torzított (háromszoros alaphangértékkel bemondott  $F_0=300$  Hz) hangsor időtartamát úgy módosítottuk, hogy a periódusok végére megfelelő időtartamú nulla intenzitású részt vágunk be. Ezzel az eredeti alaphangértéket állítottuk elő.

Az 1-es pont alapján elvégzett beavatkozásnál a meghagyott periódusrész időtartama 3,5 ms volt, a további rész 6,5 ms nulla intenzitású. A beavatkozás következtében a hangok időtartama és karaktere nem változott. A beszéd ugyan érdesebbé vált, de az egyén hangja felismerhető maradt. A 2. ábrán látható a periódusok belsejében bekövetkezett változás, ami a spektrogram vonalas részében nem okozott jelentősebb eltérést. Ezzel magyarázható, hogy a személy megítélésében a periódus 2/3-ának elhagyása nem okozott gondot.

A tesztelők a hangsor alapján az illető személyt 100%-os eredménnyel felismerték. Természetesen azt sem lehet figyelmen kívül

hagyni, hogy a beavatkozás a zár- és zörejhangokat nem érintette. Megítélésünk szerint ez azonban a kísérlet eredményét nem befolyásolja, mivel a többi pontban elvégzett tesztelés esetében sem változtattunk ezeken a hangokon.



2. ábra

A normál hangfekvésben ejtett [a:] hang oszcillogramja és spektrogramja periódusmódosítás után

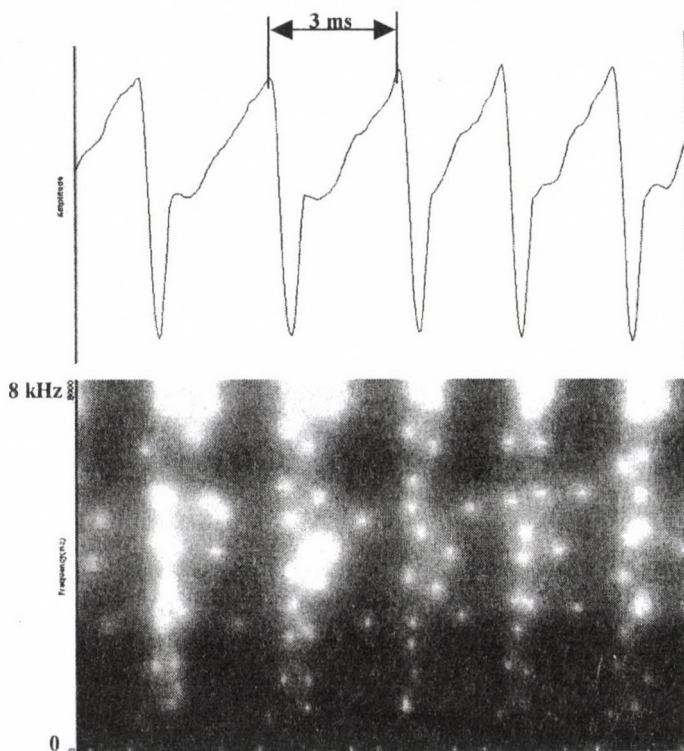
A 2. pontban leírt változtatást a periódusok nulla intenzitású részének elhagyásával hajtottuk végre úgy, hogy a megmaradt 3,5 ms időtartamú részek közvetlenül egymáshoz kapcsolódtak. Eredményképpen magas, közel 300 Hz-es ( $F_0 = 286$  Hz) hangfekvésű hangsort kaptunk. A periódusidő 3,5 ms volt. Mivel a zöngés hangok a normál

tempójú kiejtéshez képest egyharmadra rövidültek, a hangsor időtartamát korrigálni kellett az eredeti hangsor időtartamára. Ezt úgy oldottuk meg, hogy a hang minden periódusát kétszer megismételtük. Az így elkészített hangsor alapján a tesztelő személyek a beszélő személyt nem tudták azonosítani. Mivel a hangsorról készített spektrogramikus vizsgálat a formánsszerkezetben és a formánsintenzitás-szerkezetben lényeges változást nem mutatott ki, megítélésünk szerint csak a beszéd alaphangjának lényeges emelkedése tette felismerhetetlenné a beszélő személyt. Kijelenthetjük tehát, hogy a beszélő személy megszokott hangjától eltérő szokatlan hangfekvés az egyén hangszínezetének változását okozza a hallgató számára. A 3. ábrán a módosított, [a:] hang spektrogramját láthatjuk. A formánsszerkezetben döntő változás nem következett be, tehát a személy felismerését csak a csökkentett periódusidő befolyásolta.

A 2. számú kísérletről kiindulva, feltételezésünk szerint, ha a hangsor pozitív irányú korrekcióját hajtjuk végre, tehát a szándékosan magasan kiejtett hangsort redukáljuk le normál alaphangra, akkor az így kapott hangsorral már a személy egyéni hangszíne is reprodukálásra kerülhet. Ebből kiindulva végeztük el a 3. pontban jelzett kísérletet.

A 3-as számú hangkorrekciós vizsgálatnál az előbbiekkal azonos hangsort ejtett ki a beszélő személy lényegesen megemelt alaphangon. A normál hangfekvésénél kb. 3-szor magasabb alapfrekvencia-értékű beszédet tesztelve, a hallgatók számára a beszélő személy nem volt azonosítható. Az előbbi kísérlethez hasonlóan tehát az alaphang-változtatás a személy felismerhetőségét megghiúsította. A megemelt hangú beszédnél a mért periódusidő 3 ms volt, tehát a normál hangfekvés létrehozásához a periódusidőt 7 ms tartamú nulla intenzitású résszel kellett meghosszabbítani.

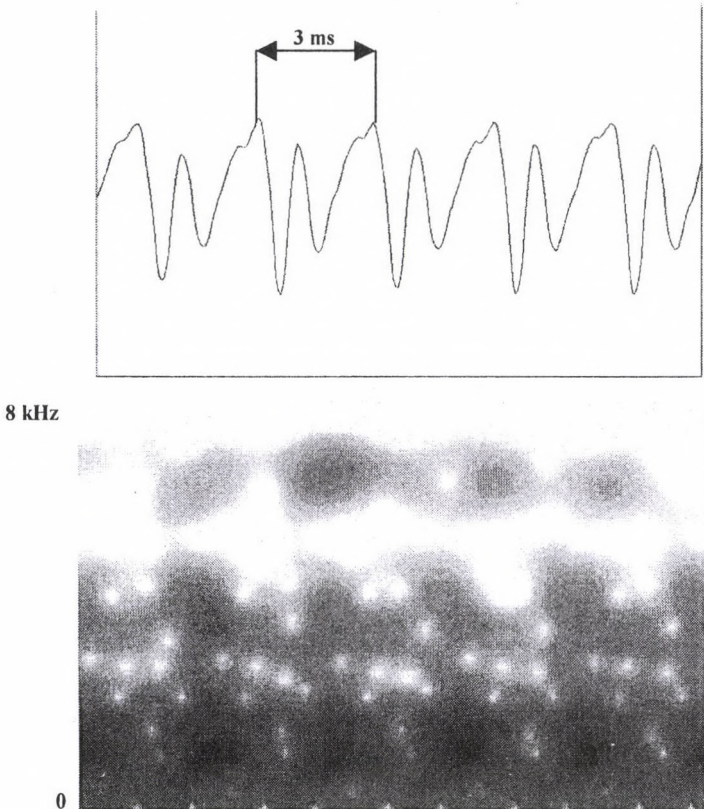
A beavatkozás következtében létrejött beszéd így alaphangértékben megegyezett a személy eredeti F<sub>0</sub>-jával, de időtartama a periódusok számának háromszoros értéke miatt hosszabb lett. A hangsor időtartamának korrekcióját így itt is el kellett végezni, de természetesen negatív irányban. Például a bemutatott hang normál kiejtésben 17 periódust tartalmazott, míg a megemelt változat 54-et. Ez esetben tehát



3. ábra

Az időben korrigált hang oszcillogramja és spektrogramja  
8 kHz-es tartományban  
(háromszoros alaphangérték)

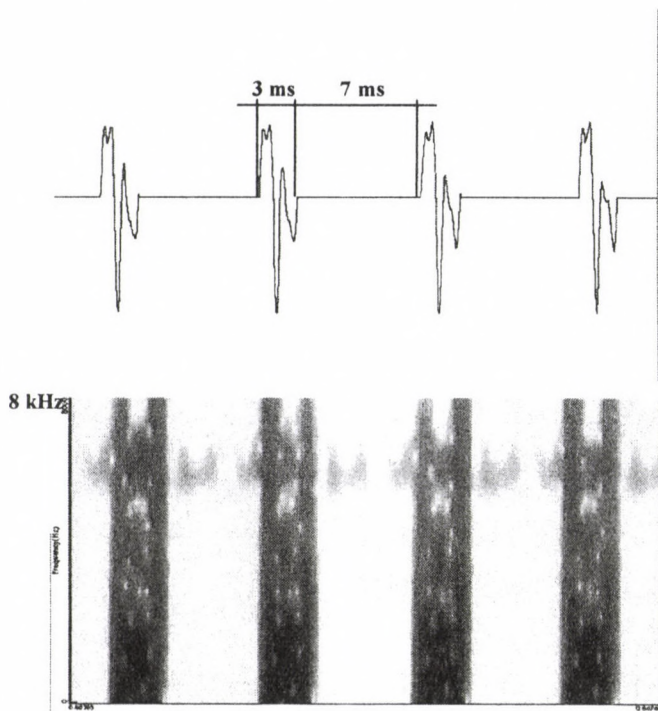
minden második és harmadik periódust kivágtuk. A zöngés hangok időtartama így megegyezett az eredeti, normál alaphangon kiejtett hangsorral. Az elvégzett teszt eredménye alapján a tesztelők 70%-os valószínűséggel azonosították a személyt, illetve úgy ítélték meg, hogy a hallott hanganyag alapján nem zárják ki az azonosságot.



4. ábra

A magas alaphangon ejtett beszéd  
( $F_0=300$  Hz) oszcillogramja és spektrogramja

A tesztelési eredményeknek a 6., 7. és 8. ábrán látható spektrografikus képek felelnek meg. Bemutatjuk a természetes hangfekvéssel bementett hangsort (6. ábra), a magas alaphangon ejtett hangsort (7. ábra) és a megemelt hangon ejtett hangorból redukált természetes hangfekvésű (8. ábra) hangsort 0-8 kHz-es tartományban.

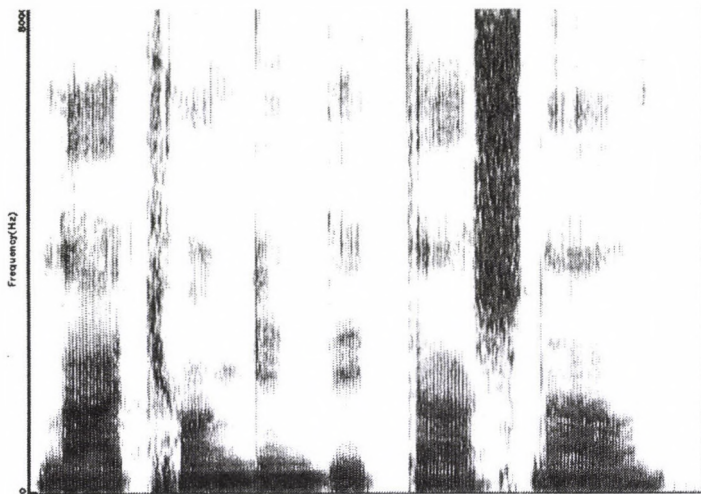


5. ábra

A magas alaphangú beszéd oszcillogramja és spektrogramja transzformálás után

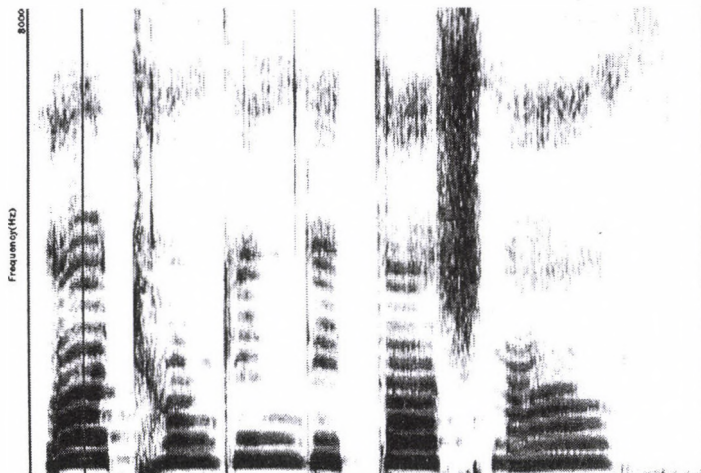
Jól látható az ábrákon, hogy a természetes alaphangú és a korrigált alaphangú hangsor spektrogramja nagyon hasonló, főbb akusztikai kulcsait tekintve megegyeznek, viszont a megemelt hangú hangsor hangszíneképe eltérő.

A hangsor [a:] hangjának három különböző alaphangértékű variánsának átlag-energiáspektruma látható a 9. ábrán. Jól kivehető, hogy a magas hangfekvéssel ejtett [a:] hang energiáspektruma (nyíllal jelölt) eltér a másik két görbétől, ami a hangon belüli energiaviszonyok különbségére utal.



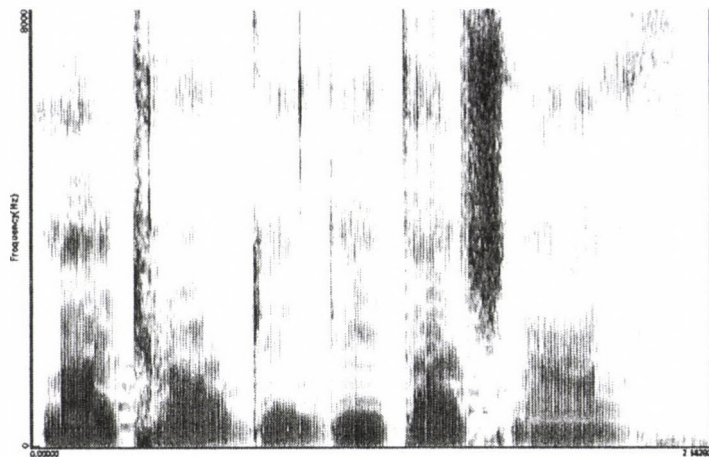
6. ábra

A normál alaphangon kiejtett hangsor spektrogramja 8 kHz-es tartományban



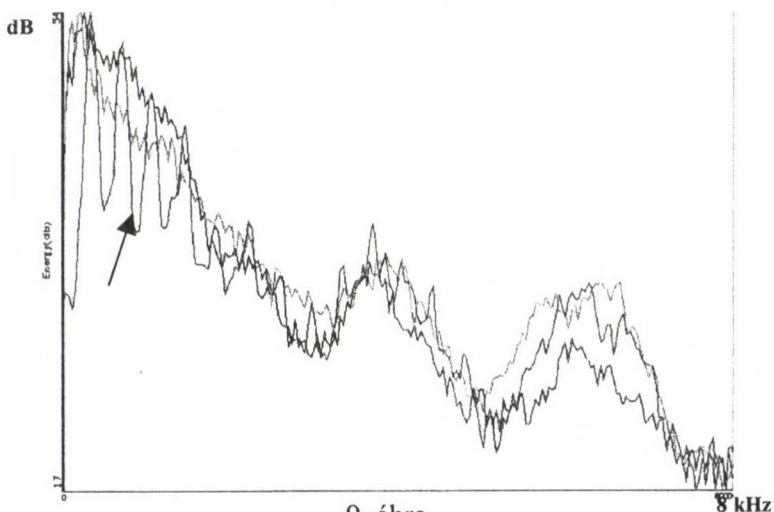
7. ábra

A magas hangfekvésben ejtett hangsor spektrogramja



8. ábra

A magas alaphanggal ejtett normál hangfekvésre korrigált hangsor spektrogramja



9. ábra

Természetes hangfekvésben ejtett, magas hangfekvésű és korrigált [a:] hang átlag-energiáspektruma

### **Következtetések**

A kísérlet eredményei alapján kimondhatjuk, hogy míg a szándékosan torzított (megemelt alaphangú) beszéd a beszélő személy felismerését megnehezítette, illetve sok esetben lehetetlenné tette, addig a korrekciós beavatkozás a személy kilétének megítélését 70%-os felismerési eredményre módosította. Az egyéni hangszínezet megítélésében tehát nagy jelentősége van a személyre jellemző hangfekvésnek, és kisebb a jelentősége az artikulációs csatorna hangszínezetet befolyásoló tényezőjének.

A kísérletnek az ad gyakorlati jelentőséget, hogy a személy felismerését sok esetben megnehezíti vagy lehetetlenné teszi a szándékosan mélyített vagy magasított beszédhang (Zetterholm 2003). Ha ez az eltérés az egyénre jellemző hangfekvéshez képest nagy (esetleg 2-3-szoros), a személy kilétének meghatározását sem percepciósi meghallgatással, sem spektrografikus elemzéssel nem lehet elvégezni. Ezekben az esetekben jelen kísérlet eredményei szerint (főleg a kriminálisztikai gyakorlatban) hatásosan alkalmazható a fent ismertetett eljárás. Meg kell azonban jegyeznünk, hogy a kísérletet stúdióban készített hangfelvételekkel végeztük el, így az szűkített spektrumú hangfelvételek esetében (pl. telefonvonalról rögzített hangfelvétel) csak korlátozottan alkalmazható.

### **Irodalom**

- Lavner, Y. – Rosenhouse, J. – Gath, I. (1999): The prototype model in speaker verification. In: Proceedings of Eurospeech '99. Budapest, 771-774.
- Olaszy Gábor (1979): A zöngé szerepe az egyéni hangszínezet kialakításában. Magyar Fonetikai Füzetek 4: 137-145.
- Zetterholm, E. (2003): Voice imitation. Lund University. Lund.