

DOKTORI ÉRTEKEZÉS

**Számítógép alapú kiejtésoktatás és beszédfejlesztés
műszaki tudományos alarendszerének
elméleti és gyakorlati kidolgozása**

Írta:

Vicsi Klára

Budapest
2003.

Tartalomjegyzék

Bevezetés, tézisek

1. A beszédtanítást segítő rendszerek történeti áttekintése	10
1.1. Analóg rendszerek	10
1.2. Részben digitális rendszerek	11
1.3. Mai számítógépes rendszerek	12
1.4. Magyarországon fejlesztett rendszerek	18
2. Az új többnyelvű, hallási és látási visszacsatolás együttesével kialakított beszédoktató és rehabilitációs rendszer alapkonceptiója – Beszédkorrektor	20
3. Multiszenzoros beszédterápiás környezet	24
3.1. Akusztikai előfeldolgozás	24
3.1.1. Spektrumelemzés	25
3.1.2. Alaphang meghatározás	28
3.2. Vizuális ábrázolás	29
3.2.1. Intenzitásváltozás képe	29
3.2.2. Hanglejtés képe	30
3.2.3. Hallási spektrum képe	31
3.2.4. Hallási spektrogram képe	33
3.2.5. Spektrális eltérés képe	36
3.2.6. Adatbázis rögzítő egység	37
3.2.7. Szegmentáló egység	37
3.2.7.1. Szegmentáló egység alapkonceptiója	37
3.2.7.2. Osztályozás neurális hálóval	39
3.2.7.3. A nyelv befolyása a szegmentálás pontosságára	43
3.3. Hangvisszajátszás	45
3.4. Kvalitatív kiértékelés	45
4. Adatbázisok létrehozása	48
5. Az adott nyelv fonémáit reprezentáló beszédhangok spektrális modelljei	50
5.1. Szubjektív lehallgatási kísérletek	51
5.2.. A helyes kiejtésű spektrális modellek	52

5.3. Az életkor szerepe a fonémák kiejtésének fejlődésében	56
6. A kiejtés jóságának automatikus megítélése	60
6.1. Spektrális távolságszámító eljárás kiválasztása	60
6.1.1. Módosított spektrális távolság	61
6.1.2. Módosított spektrális négyzetes távolság	62
6.2.3. A kész rendszer tesztelése	63
7. A kezelt beszédhangok akusztikai tulajdonságai és képi megjelenítésük	65
7.1. A sziszegő hangok akusztikai leírása	65
7.2. A magánhangzók akusztikai leírása	68
8. Beszédoktatási módszertan	73
8.1. A gyakorlatok sorrendjének jelentősége	74
8.2. Az előkészítés	74
8.3. A beszédhang kialakítása	75
8.4. Hanglejtés	77
8.5. Szabad gyakorlás	78
8.6. Hallásfejlesztés	78
8.7. Kapcsolat a betűkkel	80
9. Hatékonyságvizsgálat	81
Köszönetnyilvánítás	87
10. Irodalom	88
11. Mellékletek	
1.sz. Text and pictures of the reference database	
Report of the partners	
2.sz. Children speech databases for the statistical examination	Report of the partners
3. sz. A magyar nyelv beszédhangjainak hallási spektruma	
12. Demonstrációs CD-ROM a magyar, angol, svéd és szlovén beszédkorrektor programjaival	

Bevezetés

A számítástechnika és a digitális beszédfeldolgozás legújabb eredményei ma már lehetővé teszik olyan beszéd-oktató, -fejlesztő rendszerek összeállítását, amelyek a beszédtanításban és gyakorlásban valóban hatékony segítséget tudnak nyújtani. A technológia magában foglalja a beszédfelismerés, beszéd-szintézis, beszédelemzés, vizuális megjelenítés legújabb eszközszerét.

A Békésy György Akusztikai Kutatólaboratóriumban a nyolcvanas évek elejétől kezdve külső és belső munkatársakkal együttműködve intenzíven foglalkoztam a digitális beszédfeldolgozás, különösen a digitális hangelemzés különböző módszereinek kutatásával.

Ezt követően a gépi beszédfelismerés témakörén belül először izolált szavas (Vicsi és Berényi, 1989) majd akusztó-fonetikus szintű folyamatos beszédfelismerő (Vicsi és Víg, 1995) kutatását végeztem. Nemzetközi együttműködésben hozzájárultam az olvasott, tisztabeszéd adatbázisok elkészítési metodikájának kialakításához. Kidolgoztam a beszédkutatásokhoz, beszédfelismeréshez szükséges magyar beszédatadtbázisok létrehozásának technikáját. Vezetésem alatt készültek el azok a beszédatadtbázisok, amelyek a mai magyar fonetikai kutatások, valamint beszéd felismerési eljárások alapjául szolgálnak (Vicsi, 2003). Ezekkel és a hozzá kapcsolódó egyéb munkákkal (Vicsi és Víg, 1998) sikerült megteremteni azt a műszaki és tudományos háttérrel, amely lehetőséget adott a multiszenzoros beszédfejlesztő és oktatórendszer létrehozásához.

A feladat:

A kilencvenes évek elejétől kezdődött el egy többéves kutató-fejlesztő munka (Vicsi et al. 1993), (Vicsi, 1995), amelynek célja egy olyan többnyelvű multiszenzoros beszéd-oktató és -gyakorló rendszer elméleti alapjainak a kidolgozása majd gyakorlati megvalósítása volt, amely ép-halló beszédhibás, valamint hallássérült gyermekek és felnőttek helyes beszédének kialakításában nyújt segítséget, továbbá a nyelvtanulásban, és a kiejtés oktatásában is alkalmazható.

A rendszer a beszédképzéskor keletkező hang jellemző fizikai paramétereit a számítógép képernyőjén szemléletes módon jeleníti meg. Ezt a képet nevezem hangképnek. A hangképző szervek helyzetének megváltozásából adódó azonnali hangképváltozások követhetősége miatt módszeremmel lehetőséget adok a folyamatos beszéd lényegének vizuális követésére. A hibás képzéskor megjelenő hangkép eltér a helyes ejtést jellemző referencia hangképtől. Ez ad lehetőséget arra, hogy a rendszer hatékonyan rávezessen a helyes hangsor előállítására.

A képi megjelenítéssel egy időben az adott hang, szótag, szó vagy mondat is elhangzik, így biztosítva a több érzékszervi csatornán keresztüli (multiszenzoros) információáramlást (Vicsi, et al. 2001).

A kiejtés- és beszédfejlesztés e módszerével segítem a gyermekek artikulációs bázisának megteremtését, az adott nyelv beszédhangjai helyes kiejtésének kialakítását, rögzítését és a beszédprodukciónak automatizálását. A kutatások alapján megalapozott rendszer lehetőséget nyújt továbbá az alapvető, általános beszédjellemzők helyes kialakítására, gyakorlására is. Ilyen jellemzők a hangosság, a hangmagasság, a ritmus, a hanglejtés és a hangszín. Az általános mérő, feldolgozó és megjelenítő rendszer megalkotása mellett részletes oktatási tematikát dolgoztunk ki az ép-halló beszédhibás, valamint a nagyothalló gyermekek és felnőttek helyes beszédképzésének kialakítására.

A kutatások eredményeire épített rendszer két részből tevődik össze, az egyik rész egy mérő és feldolgozó rendszerből, amelyben az akusztikai előfeldolgozás, a vizuális ábrázolás és egyéb számítási algoritmusok nyelvfüggetlenek. A másik rész a kutatásom nyelvi oldalát megtestesítő fonetikai editor, amely lehetőséget ad különböző beszédoktatási feladatok (beszédjavítás, nyelvtanulás) ellátásához, különböző nyelvekre különböző típusú szöveg, hang és kép adatbázisok előállítására, szerkesztésére.

Az első, gyakorlatban megvalósított szoftver rendszert, (az úgynevezett Beszédkorrektor első változatát) magyar nyelvre készítettem el (Vicsi, et al., 1993).

A magyar kísérleti eredményekre támaszkodva általános tematikát dolgoztam ki, amellyel először német (Vicsi és Hacki, 1996),- majd az EU támogatásával (SPECO no. 977126) project vezetőként több nyelvre (angol, szlovén és svéd nyelvre) is megvalósítottam a beszédoktató programot, az adott nyelv sajátosságait figyelembe véve, külföldi kutatók bevonásával (K. Vicsi et al. 2001).

Minden nyelvre két különböző adatbázis készült el: egy helyes kiejtésű szótagokból, szavakból és kifejezésekből álló referencia gyermekbeszéd adatbázis, amelyet beépítettünk az oktató rendszerbe valamint egy általános gyermekbeszéd adatbázis, amelyben helyes és kevésbé helyes kiejtésű szavak és mondatok kerültek rögzítésre. Ez utóbbi adatbázis a statisztikai vizsgálatokhoz nyújtott alapot.

Hallássérült és épp-halló beszédhibás gyermekek egyéni foglalkozásain végzett széleskörű és részletes hatásvizsgálat azt mutatta, hogy ez az új módszer jelentősen hatékonyabb oktatási lehetőséget ad a pedagógusok kezébe, mint a hagyományos módszerek (Vicsi et al., 2000).

Bővíti a metodikai eszközrendszert, és reményeink szerint a speciális pedagógiában az újabb technikák, technológiák bátrabb felhasználásához is elvezet.

A beszédoktató rendszer kutató munkája olyan eredményeket is adott, amelyek a beszédfeldolgozás más területein is hasznosíthatók. Ilyen például a nyelvfüggetlen automata szegmentáló rendszer kialakítása, vagy a fonémák spektrális statisztikus modelljeinek a létrehozása.

Hasznosítás

A kutatás alapján előállított rendszer neve Beszédkorrektor, a magyar nyelvű, gyermekek számára készült program a Varázsdoboz fantázianevet kapta a forgalmazótól. A Varázsdoboz Magyarországon számos alsó, középső, és felső szintű tanintézménynél, egyéni felhasználóknál használatban van (ELTE Bárczi Gusztáv Gyógypedagógiai Kar, Hallássérültek Pedagógiája Tanszék, Fonetikai és Logopédiai Tanszék, A siket gyermekek nevelését, oktatását ellátó Óvoda, Általános Iskola és Diákotthon, Dr. Török Béla Óvoda, Általános Iskola és Diákotthon, Nagyothallók Általános Iskolája, és megközelítőleg még 40 intézmény).

Németországban néhány intézményben a régi első verziót már mindennaposan használják (Straubingi Hallássérültek Intézete, Regensburg Klinika). Svéd, angol, szlovén nyelvterületen kipróbálása sikerrel befejeződött, az oktatásba való bevezetése folyamatban van.

Az európai „IST Prize” értékelő csoport a Beszédkorrektort a 2002. évben az "European IST Prize Nominee" díjjal tüntette ki.

Jelenleg készül a Beszédkorrektor új német nyelvű változata, valamint a francia beszédkiejtést oktató, gyakorló program, magyar anyanyelvű diákok számára.

Új tudományos eredmények tézises összefoglalása

1. Az általam kidolgozott multiszenzoros (hallási, látási) érzékeltetés jelentősen növeli a beszédoktatás hatékonyságát, azáltal hogy több érzékszervi csatornát, egymás hatását erősítve alkalmaz, és helyes módszertani lépésekben optimalizál.

Ezen többcsatornás érzékelést alkalmazó oktatási módszer esetében a látási információ kompenzálja a hiányzó hallási információt, miáltal **jól alkalmazható hallássérültek beszédfejlesztésére**. A vizualitás akkor is jelentős teljesítménynövelő tényező, ha a hallási csatornák épek, ezért a módszer bevezetése a nyelvtanulás kiejtésoktatásába igen ígéretes kutatási irány lehet a jövőben.

Beszédoktató rendszeremben a vizuális információ az akusztikai előfeldolgozással nyert fizikai paraméterek képi megjelenítése. Ezeket a megjelenített képeket nevezem hangképeknek. A hibás képzéskor megjelenő hangkép eltér a helyes ejtést jellemző hangképtől. Ez ad lehetőséget arra, hogy a rendszer hatékonyan rávezessen a helyes hangsor előállítására.

Kidolgoztam egy multiszenzoros **mérő, editáló és adatbázis kezelő rendszert, amely magában foglalja az egyes beszédhang paraméterek valósidejű mérési eljárását, a paraméterek vizuális megjelenítésének módszerét, vagyis a hangképek létrehozását** (3. fejezet). A megjelenített hangképek a következők: hallási spektrum (hallási szűrősávokban mért intenzitás dB-ben), a hallási spektrumon alapuló spektrogram, spektrális különbségek, dinamika, hanglejtés, ritmikai képletek hangképei.

Az elméleti alapokra épített Beszédkorrektor rendszer bevezetése olyan jelentős oktatáshatékonysági javuláshoz vezetett, ami egzakt számadatokkal jellemezhető és mérhető (9. fejezet).

2. A fizikai paraméterek képi megjelenítésében a hangképekben a lényeges illetve kevésbé lényeges részek rajzos módon magyarázhatóak(vizuális információ kiemelés), háttér képek alkalmazásával. A megvalósított Beszédkorrektor rendszerben ezek a háttér képek a helyes kiejtésű beszédhangok statisztikus spektrális adatainak rajzos megvalósításai (5. fejezet). Az oktatási gyakorlat azt mutatta, hogy ezek a háttér képek jól alátámasztják a spektrális információt, és segítenek a színeképek helyes értelmezésében (5. és 9. fejezet). A háttér képek alkalmazása a vizuális lényegkiemelés olyan új módszere, amelyet tudomásom szerint eddig még egyetlen vizuális visszacsatolásos rendszerben sem alkalmaztak.

A háttér képek a fenti statisztikus spektrális adatok megjelenítései. Ugyanezeket a statisztikus spektrális adatokat használom a kiejtés jóságának automatikus értékelésénél, mint referenciát. Így biztosítom a vizuális információ és az automatikus értékelés visszajelzéseinek szinkronizált hatását.

Beszéd adatbázis helyes kiejtésű beszédhang mintáin statisztikai elemzést végeztem. Ennek során meghatároztam a magyar fonémákat reprezentáló beszédhangok spektrális paramétereinek (hallási szűrősávokban mért intenzitás dB-ben) átlagát, átlagos eltéréseit magyar nyelvre, külön-külön férfiakra, nőkre és gyermekekre (5. fejezet). Ezek a kutatások általános jellegűek, nem kizárólag a beszédoktatáshoz kötődnek.

3. Automatikus kiejtés értékelő eljárást dolgoztam ki a paciens kiejtésének megítélésére, amely segíti a tanulást, alkalmat ad az önálló gyakorlásra, mivel azonnali visszajelzést ad a tanuló számára kiejtésének helyességéről (6.fejezet). A helyes kiejtést reprezentáló spektrális színeképi modell és a tanuló bemondási mintája közötti akusztikus hasonlóság mértékének megállapítására több módszert próbáltam ki, amelyek közül az euklideszi spektrális távolságszámítás adta a legjobb eredményt.

4. Megfelelően összeállított gyermek beszéd adatbázis szubjektív akusztikai vizsgálatával a gyermekek beszéd-fejlődése jól követhető (5.3. fejezet).

Meghatároztam az átlagosan beszélő egészséges 4-től 14 éves gyermekek beszédének korszerinti alakulását, szubjektív lehallgatási módszerrel. Megállapításaim az alábbiak:

- **a magánhangzók határozott, helyes kiejtése korábban fejlődik ki, mint a mássalhangzóké,**
- **a magánhangzók kialakulása 6 éves korra befejeződik.**
- **a mássalhangzók teljes kialakulása a 9-10 éves kor közé esik.**

E pszicho-akusztikai eredmény, amely a gyermekek beszédfejlődését mutatja, jó összhangban van a gyermekek artikulációs szerveinek teljes kiépülésével, továbbá magyarázatot ad arra a jelenségre, hogy a 10 év alatti gyermekek beszédhangjainak spektrális és idő adatai jóval nagyobb szórást mutatnak, mint a felnőtt férfiaké és a nőké, ami általános akusztikai jelenség. Ez a gyermekek beszédfejlődésével áll kapcsolatban.

Ezek a gyermeki beszédfejlődésre vonatkozó kutatási eredmények szintén általános jellegűek, nem kötődnek kizárólag a beszédoktatáshoz, a beszédkutatás egyéb területein is felhasználhatóak.

5. Érthetőségvizsgálati eljárást dolgoztam ki a beszéd-oktató programok hatékonyságának értékelésére (9. fejezet). Az eljárással kapott eredmény szerint a Beszédkorrektor hatékonysága minden csoportnál lényegesen jobb eredményt adott, mint a

kontrollcsoportban, ahol hagyományos beszédterápia folyt. A hatékonyság mértéke a sérülés típusától függ.

A részletes, széleskörű hatásvizsgálatot vezető szurdopedagógus szakember segítségével végeztem el, különböző beszédsérülések esetében a Dr. Török Béla Óvoda, Általános Iskola és Diákotthonban (Nagyothallók Általános Iskolája). Összesen 40, 6-8 éves gyermek szerepelt a vizsgálatban a következő csoportosítás szerint: normál halló beszédhibás, gyengén nagyothalló, közepesen hallássérült és súlyos hallássérült gyermekcsoport.

6. A magyar nyelvre kifejlesztett módszert angol, svéd, szlovén nyelvre adaptáltam, anyanyelvű fonetikus munkatársak együttműködésével, a nyelvi sajátosságok figyelembevételével (1 sz. melléklet).

7. Az általam kidolgozott, beszédhang szintű automatikus szegmentálási és címkézési eljárás közel nyelvfüggetlenül (több európai nyelven, a gyakorlat számára elfogadható pontossággal (81-90%)) működik (3.2.7. fejezet).

A referencia beszédminták vizuálisan megjelenített hallási spektrogramjába, valamint a gyermekbeszéd adatbázis hanganyagába bejelöltük a beszédhangok határait, valamint magukat a beszédhangokat is. E feladat megoldása szakembert igényel, és nagyon időigényes. Ezért volt szükség az automata szegmentálás bevezetésére.

Az automatikus szegmentálás alapja az, hogy a bemondott hangsorban az akusztikailag közel homogén részek akusztó-fonetikus osztályokba sorolhatók, a jelen megvalósításban 9 ilyen osztályt különböztettem meg.

Az osztályba sorolásra egy rejtett rétegű (visszafelé terjedő algoritmussal dolgozó) neurális hálót tanítottunk be. Amennyiben a bemondott hangsorban az akusztikailag kvázi homogén részek határai automatikusan bejelölhetők és a bemondás betűkkel való leírását megadtuk, akkor a betűk a hanganyagban automatikusan bejelölhetők.

Az automata szegmentálásra vonatkozó kutatások általános jellegűek, kizárólagosan nem a beszédoktatáshoz kötődnek, a beszédfeldolgozás egyéb területein is felhasználhatóak.

1. A beszédtanítást segítő rendszerek történeti áttekintése

Beszédterápiával, beszédkutatással foglalkozó szakembereket már sok – sok évtizeddel ezelőtt is foglalkoztatta, hogy hogyan lehet az adott technikai lehetőségeket a beszédterápiában felhasználni. A megvalósított eszközök bonyolultak, és nehezen használhatók voltak. Mindenesetre a használatuknál nyert tapasztalatok hozzájárultak a szakterület fejlődéséhez, ezért érdemes áttekinteni a különböző megoldásokat. A leírás nem tér ki minden részletre, hanem csak általános áttekintést kíván nyújtani.

1.1. Analóg rendszerek

A negyvenes évek első kísérletei után a hetvenes évek elején készültek olyan analóg berendezések, amelyek a beszédtanítást valóban segíteni tudták. Ezek az első jelentősebb fejlődési szakaszban készült berendezések két alapvetően különböző elméleti alapon nyugszanak. A vibrotaktilis készülékek a tapintási érzékeléssel kívánták a gyengén működő akusztikai érzékelést helyettesíteni, vagy kiegészíteni. Az ún. optovizuális berendezések a látási érzékelésre építettek. Mindkét módszerben a beszédhangot analóg módon dolgozták fel, ahogy azt az elektroakusztika és az elektroakusztikus technika az idő tájt lehetővé tette.

A jelfeldolgozás területének fő gyengéi közé tartoztak ezeknél a készülékeknél, hogy a folytonosan változó beszédhangok spektrális elemzése a célnak megfelelő felbontásban és sebességgel (az azonnali visszajelzés céljának érdekében) nem volt elvégezhető. A mért adatok csak mágnesszalagon voltak rögzíthetők. Továbbá a színek felbontása sem volt megfelelő.

A vibrotaktilis rendszerek a hallássérülteket segítő eszközök egy különálló csoportjának tekinthetők. Ezekre egy példa a FONATOR. A készülék elsősorban az energia nagyságának változásait mérte és részben képes volt a rezonáns és zörej jellegű hangok megkülönböztetésére, valamint ezek időbeli strukturálásának a közvetítésére is. Azt mindig figyelembe kellett venni az alkalmazás során, hogy taktilis úton jóval rosszabb a frekvenciafelbontás, mint a fülön keresztül. A fonátorral, illetve az újabb MINI-FONATOR-ral folytatott munkákról számos beszámoló készült (például Strauss, 1982).

Más rendszerek szűrősor segítségével, valamint a hozzácsatolt vibrátor-mátrixszal, leegyszerűsített spektrális információkat közvetítettek, melyekkel a hallási rendszer frekvencia megkülönböztetésének megfelelő hang-diszkriminációra nyílt lehetőség. Ilyen készülékekre példa a TACTILE ARTIFICIAL EAR és a különböző elektroakusztikus és mechanikus (bőrrel érintkező) rendszerek. Az előbbi készülék utódjaként mutatták be nemrégiben az alkaron viselt TAKTILE OHR TA4 rendszert. A legtöbb analóg feldolgozási móddal működő vibrotaktilis

készülék esetében vagy a beszéd prozódiai jegyeinek felismerhetőségét valósították meg, vagy a hangminőség megkülönböztetést (Neppert, 1989).

Hasonló helyzet figyelhető meg az optovizuális rendszereknél is, amelyek szintén ebből az időből származtak. Az úgynevezett hangsúlyjelzők, például az EMPHASIS INDICATOR, képek voltak az összenergia időbeli változásának megjelenítésére, de nem tudtak színeképeket ábrázolni. Éppen fordított volt a helyzet a VISIBLE SPEECH UNITS különböző típusainál, az artikulációs készülékeknél (Kiefer, 1981; Köster, 1981), és már az 1947-ből származó minden Visible Speech kísérlet klasszikusának tekinthető eljárásnál (Potter, 1947), amelyek kizárólag spektrális információkat közvetítettek.

A kiejtéstanítás során olyan módszertani megoldásokat kellett bevezetni, amelyek alkalmazkodtak a készülék nyújtotta lehetőségekhez, például alkalmazkodni kellett ahhoz, hogy többnyire csak a minta futólagos ábrázolására volt lehetőség. Az artikuláció által létrehozott hangábrázolás tehát nem marad fenn a tüzetes megfigyelés, a megbeszélés, vagy az utánzás céljára. Speciálisan az optovizuális készülékeknek egy másik gyengéje volt hogy többnyire csak egyszerű, egyszínű, vonalszerű képi megjelenítés volt rájuk jellemző az akusztikai-fonetikai regisztrálás során, ami csak kis mértékben ösztönző a tanulásra. Ezek az adottságok tehát nem kimondottan előnyösek abból a szempontból, hogy éppen az ilyen jellegű grafikus megjelenítések és az alapjukul szolgáló artikulációs folyamatok közötti viszonyokat kialakítsák, illetve őket megtanítsák. (Neppert, 1989). A tanításnál, a már felsorolt hátrányokon kívül, nagy gondot jelent ekkor még az elhangzott hanganyag tárolása vagy előhívása. Mindezen hátrányok az akkori technikai színvonalból adódtak.

1.2. Részben digitális rendszerek

A hetvenes évek második felétől kezdve olyan rendszereket mutattak be, amelyek új alkotóelemek beépítése révén (mint pl. IC-k, tárolóchipek és processzorok) sokoldalúbbak lettek és nagyobb teljesítményre voltak képesek. Az elméleti alapok viszont ezeknél a készülékeknél is a régié maradtak. Technikai értelemben ezek a készülékek, még részben analógok, de kisebb-nagyobb részt már a digitális jelfeldolgozást is alkalmazták. Legtöbbjük még ma is használatban van. A SPRACHSICHTGERÄT (a beszédet láthatóvá tevő készülék) alkalmas volt a spektrumok tetszőleges ideig történő tárolására és a képernyő felosztására, így a beszédminta és a tanulói próbálkozás egyszerre jelenhetnek meg. Egy beépített kazettás magnetofon lehetővé tette a hosszú távú tárolást, és ezzel a tanári beszédminta és a tanulói

próbálkozás megőrzését a magnószalagon. Ez mindenekelőtt a tanulási szakaszok és a tanulási eredmények dokumentálására szolgált.

Hasonló lehetőségeket kínál a VISI-PITCH. Ezt a készüléket remekül lehet kombinálni a személyi számítógéppel, aminek köszönhetően funkcionálisan még flexibilisebbé tehető.

A SPRACH-FARBILD-TRANSFORMATIONSGERÄT (SFT) ugyanazokat a pedagógiai fontos funkciókat mutatja, mint a SPRACHSICHTGERÄT, és az első olyan optovizuális rendszer, ami megkísérli, hogy együttesen közvetítse a prozódiai paramétereket és a hangminőséget meghatározó paramétereket. Továbbá a hangszíneképeket nem mint absztrakt spektrogramokat ábrázolja, hanem meghatározott frekvenciasávban megjelenő energia más - más színnel, valamint az intenzitással arányos nagysággal jelenik meg, úgy, hogy az hangminőségnek a képernyőn színminőségek felelnek meg. Ez a hang igen szemléletes képi megjelenítését adja (Esser, 1987). A tanár, vagy a helyesen beszélő szülő bmondhatja a tanuló előtt a gyakorlandó kifejezést, így a tanuló összehasonlíthatja őket egymással. Az eredeti verzió hátránya az volt, hogy használatához mindig kellett egy jól beszélő személy, mivel a helyes hangképek nem voltak tárolhatók.

Az SFT és a SPRACHSICHTGERÄT készülékek továbbfejlesztett verziójánál ma már lehetőség van a tanári beszédminta és a tanulói próbálkozás hosszú távú tárolására.

A SPEECH SPECTROGRAPH DISPLAY /SSD/ ezzel szemben azok közé a rendszerek közé tartozik, amelyek csak spektrális információkat közvetítenek. A beszédhangot a szonogramhoz hasonlóan elemzi és ábrázolja, vagyis a színek időbeli változását mutatja. A beszélt nyelv megértése szempontjából fontos időbeli viszonyokról; az amplitúdó és az alaphang változásokról szinte nincs információközvetítés.

A vibrotaktilis rendszereken belül meg kell említenünk az ehhez a fejlődési fokhoz tartozó mechanokután HÖRPROTHESE-t (hallási protézis), a hozzá tartozó PVDF-vibrátor-átalakítóval (Leysieffer, 1986). Ez a készülék a beszéd prozódikus és hangminőséget jellemző paramétereit kombináltan továbbítja. Különlegesen jó inger-lokalizációt alkalmaz, ami az újfajta piezoelektronikus műanyagból készült vibrátorgyűrűk segítségével történik.

1.3. A mai számítógépes rendszerek

Ahogy a számítástechnika és a beszédtechnológia fejlődik, a beszédoktató rendszerek megoldási lehetőségei nőnek. Az új kutatásokat a beszédfelismerés, a beszéd-szintézis, a beszédelemzés, a vizuális megjelenítés legújabb kutatási eredményei támogatják. (Vicsi,

2004). A beszédhibás hallássérült emberek beszédoktatásán kívül egy egészen új irányzat annak a vizsgálata, hogy hogyan lehetne hatékonyan az idegennyelv-oktatásban hasznosítani a számítógépes támogatás adta lehetőségeket (Computer Aided Language Learning CALL).

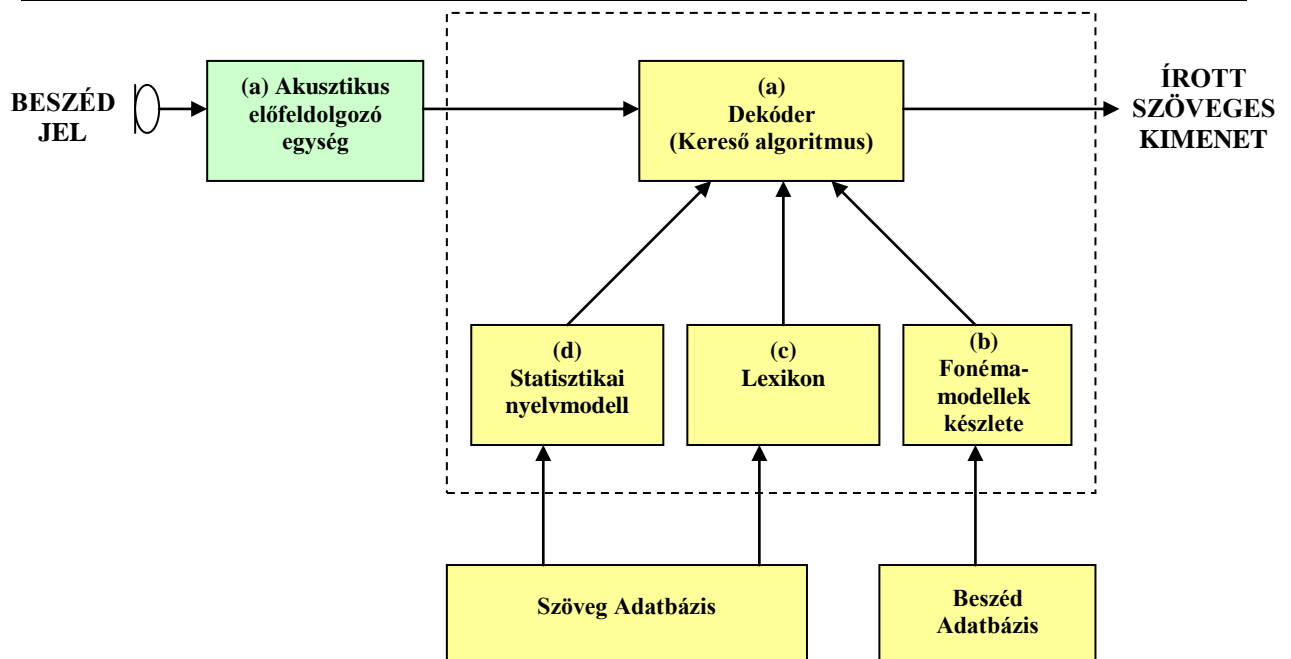
Nagyon fontos azonban, hogy a gyors technikai fejlődés, fejlesztés mellett figyelembe vegyük a fonetikai, fonológiai oktatási szempontokat, a beszédfejlesztés különböző lépéseit, a felhasználók károsodás mértéke vagy az életkor szerinti szellemi képességeit. Sajnos az utóbbi években kialakított beszédoktató rendszerek legtöbbje csak átveszi változtatás nélkül a legújabb beszédtechnológiai eljárásokat és nem alakítja azokat a speciális alkalmazáshoz, feladathoz. Erre példa a gépi beszédfelismerők széleskörű alkalmazása a kiejtés helyességének megítélésére.

Rendszerint fonéma alapú, Rejtett Markov Modell alapján működő beszédfelismerőket használnak (Jelinek, 1996), amelyek amint azt az 1. ábrán bemutatjuk, lényegében 5 alapegységből épülnek fel:

- (a) Akusztikus előfeldolgozó egység, amely a szinképelemzést és egyéb paraméterek kiemelését végzi;
- (b) Fonémamodellek készlete, amelyet nagymennyiségű beszéd adattal tanítanak be; E beszédanyagban vegyesen fordul elő jó és kevésbé, jó, esetleg hibás kiejtés is. Így készítik fel a felismerőt a kevésbé korrekt kiejtések felismerésére is.
- (c) Lexikon, a fonémaláncok szavakká történő átalakítására;
- (d) Statisztikai nyelvi modell, amely a lehetséges szavak sorrendbeli előfordulási valószínűségére épül;
- (e) Dekóder, amely egy kereső algoritmus a kimondott kifejezés és a szóba jöhető szóláncok legjobb illesztésére.

A feladatuk éppen az, hogy olyankor is felismerjék az adott kifejezést, ha annak kiejtése nem igazán jó, például pongyola, renyhe ejtés, hadarás, vagy esetleges hangkihagyás esetében is.

A beszédfelismerők használatán alapuló automatikus visszajelzés hatékonyságát illetően a tanárok véleménye sem igazán pozitív (Wallace, 1998). Tapasztalataik szerint vagy nem megfelelőek ezek az automatikus ítéletek az automatikus beszédfelismerési technológia alapján, vagy pedig nem elég érzékenyek az apróbb különbségek észrevételéhez, ami félrevezeti a tanulókat. Ezáltal rosszabb eredményeket érhetnek el, mint az automatikus visszacsatolás használata nélkül.



1. ábra

A beszédfelismerő egység szerkezeti felépítése

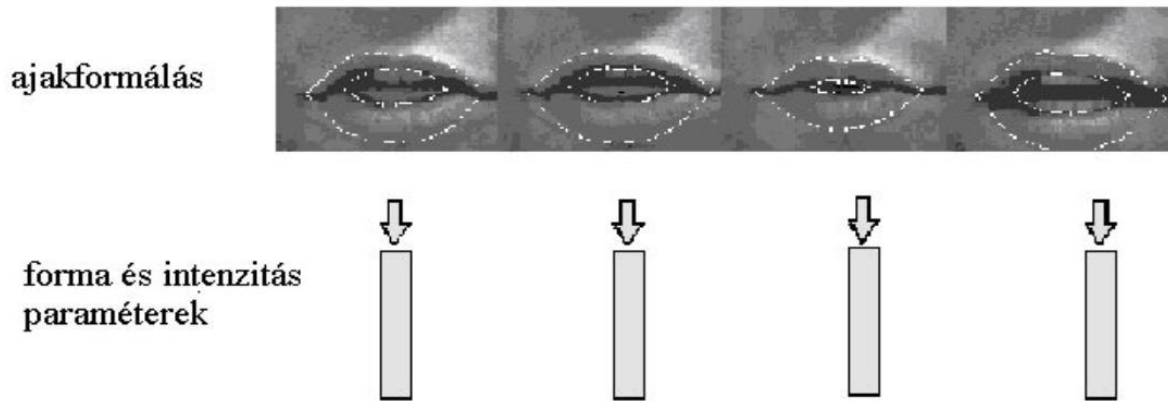
Az 1. ábrán bemutatott beszédfelismerőt először is nem beszédoktatási célra fejlesztették

Ilyen automatikus visszajelzésen alapuló kiértékeléses eljárással dolgozik az „ISTRA” és „ISLE” (Interactive Spoken Language Education) nyelvoktató rendszer. E program is a fonéma alapú Rejtett Markov Modelleket alkalmazó beszédfelismerési technológiát használja fel a kiejtés megítélésére. Hasonlóan működnek továbbá a „TALK TO ME”, „TELL ME MORE” nyelvoktató programok. (<http://www.auralog.com/us/schools.html>); (Narusa, 1999, Kawai et al. 1998).

E programok használhatóságát segítené, ha valamilyen más, például vizuális visszacsatolást is alkalmaznának. Néhány program hullámforma megjelenítést ugyan használ, és a fonetikus, akusztikus szakember el is igazodik a hullámformán, de egy gyermek biztosan nem. Ugyanakkor a látott időfüggvény nem mindig egyértelmű. A „TALK TO ME” Magyarországon is ismert nyelvoktató program például használja ezt a hullámforma megjelenítést. A programba való beillesztésének az oka az lehet, hogy mérése és ábrázolása egyszerű, nem szükséges külön beszédtechnológiai szakértő bevonása a fejlesztésbe.

A kialakított rendszerek egyik csoportjánál magát az artikulációt mutatják be a tanulóknak a beszéd közbeni artikulációs mozgás grafikai megjelenítésével. Közvetlenül a beszédképző szervek pontos beállítására teszik a hangsúlyt. Ez az ún. „folyamat orientált” megközelítés. Az előállított szintetikus arc artikulál a beszédhanggal szinkronban. Ilyenkor a beszéd vizuális gépi előállításának technikáját alkalmazzák. Az artikuláció azonos idejű modellezése ma még technológiailag nehéz feladat (Hardcastle, et al., 1999., Gibbon, et al., 1998., PCLX, England)

A paraméterekkel vezérelt vizuális beszédzintézis az arc 3D-s poligonális modelljén alapszik (Massaro, 1998., Cole, et al. 1998.) Ezek a rendszerek a beszéd vizuális képsorozatának jellemzéséhez, megjelenítéséhez nyomon követik és meghatározzák a beszélő szájmozgását, amint az a 2. ábrán látható.



2.a. ábra

A vizuális beszédjellemezők (forma és intenzitás) meghatározása

A meghatározott beszédjellemezők leírják az ajkak mozgását, a száj nyitott területét. A szájnyitás fő módusai inkább a beszélőre jellemzőek, mint a beszédinformációra. Bár a kisebb módusok jellemzőek a beszédre, és leírják a fogak és a nyelv láthatóságát is. Néhány rendszerben ezeket a modell paramétereit és időbeli függésüket Rejtett Markov Modellel jellemzik.



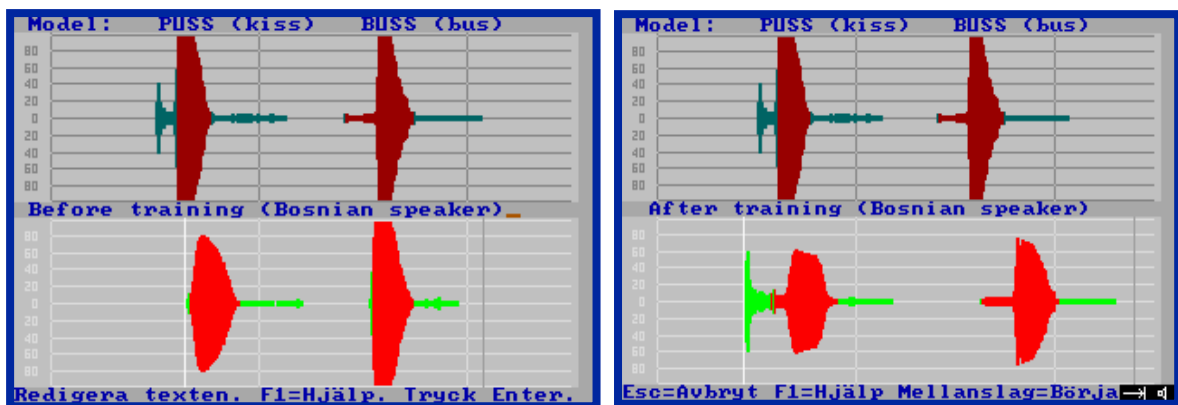
2.b. ábra

Massaro artikulációs mozgást modellező szintetikus arca, a természetes módon nem látható artikulációs szervek kiemelésével

Egy kedves beszélő fej van beépítve Massaro és munkatársai (Massaro, 1998) által fejlesztett oktató rendszerbe, melynek neve „CSLU Speech Toolkit”. Kutatói segédeszköz, melynek honlapja a következő címen található: <http://cslu.cse.ogi.edu/toolkit>). Ezek artikulációs mozgást modellező rendszerek szintetikus arcok, mesterségesen előállított beszéddel.

Ezek az oktató rendszerek arra a vizuális visszacsatolásra építenek, ami a természetes beszédkommunikációban is jelen van. Hallássérült gyermekek esetén problémát jelent, hogy a belső hangképző szervek pozícióját nem lehet látni. A fejlesztők igyekeznek a modellekben a rejtett beszédszerveket is láthatóvá tenni, de egyelőre ezek a rajzok még elég riasztóak. Egyet közülük a 2.b. ábrán mutatunk be.

A beszédoktató rendszerek másik csoportjánál különböző akusztikai paramétereket jelenítenek meg. Tehát nem az artikulációs szervek beállítását hangsúlyozzák, hanem azt az akusztikai produktumot, amit a beszélő előállít, és felhasznál a kommunikációban. Ez az ún. „**produktum orientált**” megközelítés. Rögzítik a beszédjelet, visszajátsszák, közben

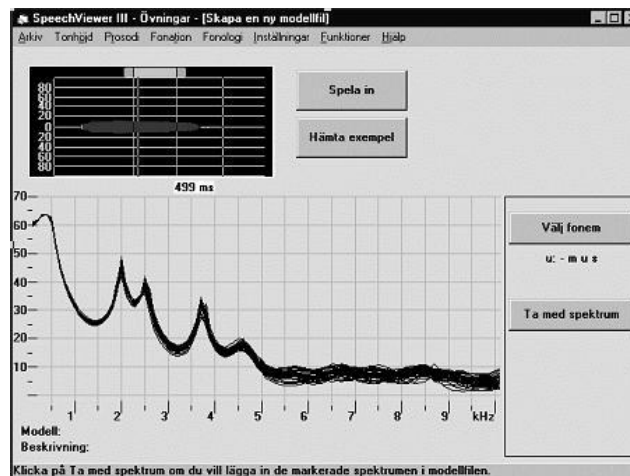


3. ábra

Hullámforma megjelenítés zárhangok gyakorlására. A gyakorlás előtti, és a gyakorlás utáni állapotokat lehet összevetni. (IBM Speech Viewer)

megjelenítik a képernyőn valamely formában. Ezeknek az eszközöknek, vagy módszereknek a sikeressége azon múlik, hogy milyen paramétereket mérnek és milyen megoldást vezetnek be a mért paraméterek vizuális visszacsatolására. Ebbe a csoportba tartozik az IBM ”Speech Viewer” programja, amely kifejezetten hallássérült gyermekek számára készült. A megjelenített képek vagy ábrák, kétfélek. Az ábrák nehezen feldolgozhatóak amint az a 3, 4. ábrákon látható. Ezeket az ábrákat a fonetikus szakértő tudja értelmezni, de a gyermek nem.

Az 5. a),b) ábrán látható képi megjelenítés viszont valójában inkább egy játék. Például az alma leesik a fáról, ha a kiejtés helyes.



4. ábra.
Az 'u' hang spektrális megjelenítése
(IBM Speech Viewer)



a



b

5. ábra
Az 'u' hang helyes (a), és helytelen kiejtésben (b) (IBM Speech Viewer)

A program használatakor problémát jelent, hogy a gyermek nem tudja felmérni, hogy milyen a jó kiejtés, és azt sem, hogy az ő kiejtése milyen messze van az optimálistól, mit kell csinálnia, hogy az adott hang képzése jó legyen. Egy oktatásra jól használható rendszernél a megjelenített hangkép érdekes, de ugyanakkor fonetikailag helyes kell, hogy legyen. Meg kell adni az oktatáshoz azt az információt, hogy az egyik kiejtés miért helyes, a másik miért helytelen. Rá kell vezetni a pácienszt a helyes hangképzés és artikuláció kialakítására.

1.4. Magyarországon fejlesztett rendszerek

A magyar beszédtanítás terén az első, világszerte elismert próbálkozás Kempelen Farkas (1734-1804), (magyar nyelvű fordítás 1989.) nevéhez fűződik, aki a hangképző szervek működésének megfigyelése alapján beszélőgépet szerkesztett. Tanulmánya, amely a beszédfiziológiai tételeit taglalja, a beszédszervek tevékenységére vonatkozóan sok máig is érvényes megfigyelést tartalmaz. Kempelen beszélőgépe, amely az emberi hangképző szervek felépítésének és fizikai, akusztikai tulajdonságainak megfigyelésén alapult, a mesterséges, gépi beszéd létrehozásának korszakát nyitotta meg, kutatási eredményei hosszú időn keresztül alapul szolgáltak az általános fonetikai vizsgálódásokhoz. Ő már a 18. század végén alkalmazta a 20. század 50-es éveiben kidolgozott, szintézissel történő analízisnek nevezett módszert (Kassai, 1998, p. 22., Nikléczy, P., Olaszy, G., 2002).

Az 1950-ben Groningenben rendezett Siketnémaügyi Nemzetközi Kongresszuson terjesztették elő az oszcillográf alkalmazását siketek hangos beszéd tanításában. Az alacsonyfrekvenciás, hosszú utánvilágítású oszcilloszkóp kb. 12 cm átmérőjű kerek képernyőjén rövid ideig megmaradt a folyamatosan ejthető a magánhangzók jellemző hangnyomás - időfüggvény görbéje, ez a megjelenítés a beszédritmus és a beszédtempó vizuális ábrázolását is szolgálta (Farkas Miklós, 2002).

Több éves tudományos kutatómunka eredményeként az előzőekben tárgyalt tapasztalatokból kiindulva Gáspár Árpád Gyógypedagógiai Tanárképző Főiskola Hallássérültek Pedagógiája Tanszékének tanára összeállította a Chordoscop műszercsoportot. Lényegében hangfrekvenciás erősítőtől, alacsony frekvenciás oszcilloszkópból, száj- és gégemikrofonból, valamint fejhallgatóból áll. A Chordoscop műszercsoport a rezgéseket kristálmikrofon segítségével feszültséggé alakítja át, majd megfelelő erősítés után egy alacsony frekvenciás hosszú utánvilágítású oszcilloszkópra, illetve süllyesztett képernyőre viszi fel, ahol a rezgés kép látható (vizuális) formába jelenik meg az idő függvényében. A tanár által adott példa, mint norma, a tanuló számára leutáNZandó feladatként jelentkezik. A rezgés kép fotórekorder segítségével filmre rögzíthető.

A szájmikrofon alkalmazása lehetővé teszi, hogy tanár és tanuló a megfelelő levegőáramoltatást regisztrálja. A száj- illetve a gégemikrofon alkalmazása lehetővé teszi, hogy tanár és tanuló a hangszalagokkal való bánni tudást, továbbá a tiszta, csengő zöngét egyidejű kontrollal produkálja (Gáspár, 1971).

Hégely Gábor, a hallássérültek szegedi intézetének igazgatója, jelenleg is próbálkozik műszeres segédeszköz kifejlesztésével a Szegedi József Attila Tudományegyetemmel együttműködve.

2. Az új többnyelvű, hallási és látási visszacsatolás együttesével kialakított beszédoktató és rehabilitációs rendszer alapkonceptiója - Beszédkorrektor

Cél:

Céлом egy olyan multiszenzoros beszédoktató módszer kidolgozása volt, amelynél a hallási visszacsatoláson kívül, a vizuális visszacsatolás segíti a páciens a beszédterápiában. A vizuális visszacsatolás alapján a beszéd akusztikai tulajdonságainak képi megjelenítésével találják meg a páciensek a hangképző szerveik helyes beállítását.

Állítás:

Állítom, hogy a beszéd megfelelő akusztikai paramétereinek vizuális megjelenítése hangképek formájában, valamint a vizuális információk kiemelése a hangképekből háttér képek segítségével olyan oktatórendszer létrehozását teszi lehetővé, amelynek segítségével az ép-halló beszédhibás és a nagyothalló gyermekek és felnőttek is intenzívebb, hatékonyabb terápiában részesülhetnek, mintha a kezelés csak a korábbi hagyományos módszerrel történik. Az eredményességhez hozzájárul a megfelelő oktatási tematika, és a helyes kiejtések tára, a referencia adatbázis.

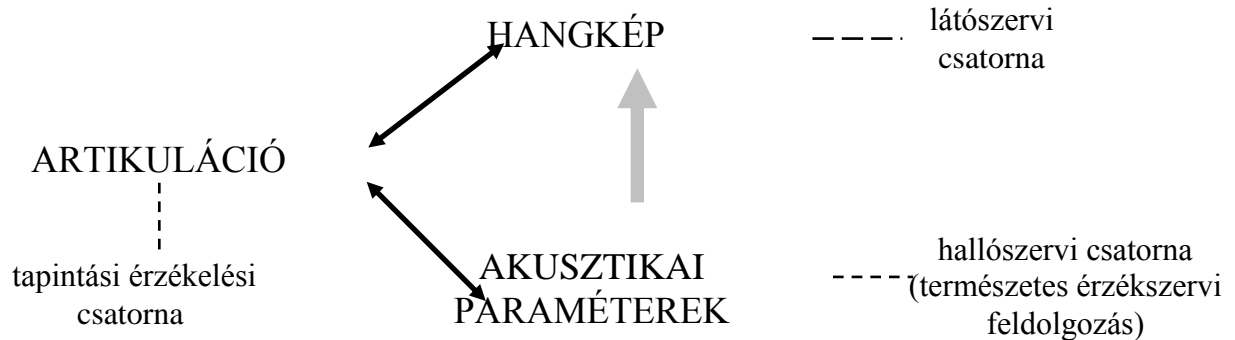
Leírás:

A hallási és látási visszacsatoláson alapuló környezetben a hangsúly a látási csatorna használatán, a beszéd akusztikai tulajdonságainak vizuális megformálásán van.

A hagyományos beszédterápia közvetlenül a beszédképző szervek pontos beállítására összpontosít. Az artikuláció azonos idejű számítógépes modellezésének felhasználása a kiejtésoktatásban szintén hasonló elvi alapokon nyugszik. Hátrányait, már az 1.3. fejezetben leírtam. Ezzel ellentétben én nem az artikulációs szervek beállítását hangsúlyozom, hanem az akusztikai produktumot. A természetes beszédtanulás során a gyermekek szintén az utóbbi megközelítést használják, vagyis az előállított beszéd akusztikai tulajdonságai alapján, saját akusztikai visszacsatolásuk segítségével találják meg a hangképző szerveik optimális beállítását.

Minden beszédterápia lényeges része a „modelltanulás”, amelynél a referenciabeszélő (például a terapeuta) beszéde utánzásának van fontos szerepe. Nemcsak a jó beszédhang funkcionális képzésmódjának utánzásáról van szó, hanem hallószervi, akusztikus utánzásról is. Az utánzáshoz használt természetes érzékszervi csatorna elsősorban a hallás. A hatás sokkal intenzívebbé tehető a látószervi csatorna bekapcsolásával. Professor Campbell a természetes, ember – ember közötti kommunikációban végzett erre vonatkozóan részletes vizsgálatokat (Campbell, 1998).

Módszeremben a visszacsatolást mind a két csatornán biztosítani kívánom. A látószervi csatornán keresztül úgy, hogy az elhangzott beszédet vizuálisan, hangképek formájában megjelenítem.



6. ábra

A beszédoktatásba bevont érzékszervi csatornák

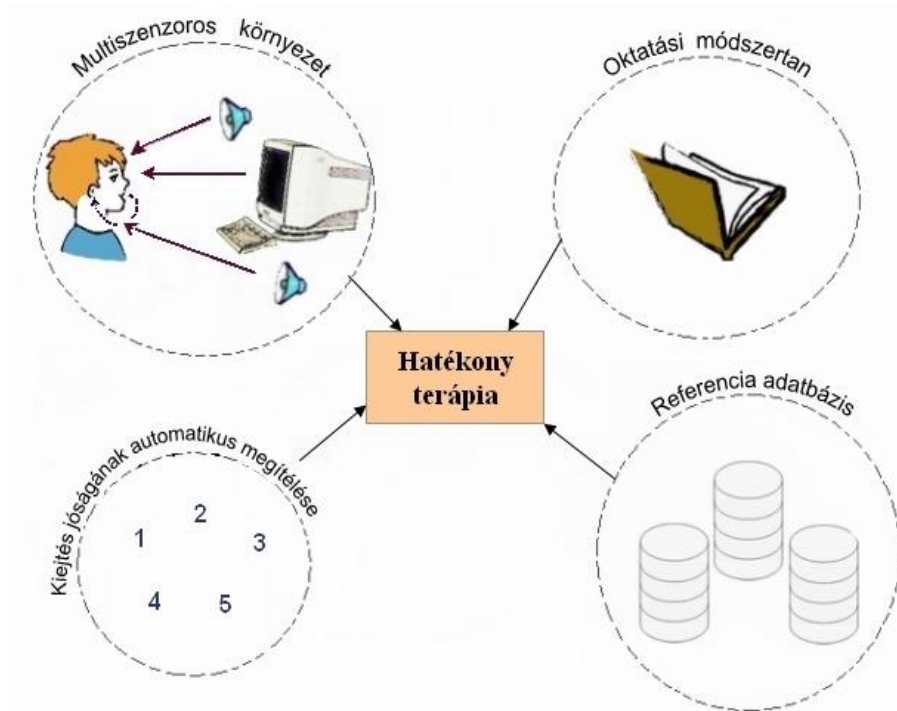
A látószervi visszacsatolásra az ad lehetőséget, hogy olyan akusztikai előfeldolgozást, és az előfeldolgozással nyert paraméterek olyan képi megjelenítését alkalmazom, amelynél az artikuláció és a megjelenített beszédparaméterek, azaz a hangképek között egyértelmű kapcsolat van (6. ábra).

Az általam kidolgozott módszer a 7. ábrán bemutatott 4 különböző építőelem harmonikus, együttes alkalmazására épül. Ezek a következők:

1. A hallási és látási visszacsatoláson alapuló beszédterápiás környezet.
2. Részletes oktatási módszertan az új környezetre építve.
3. Megfelelő mennyiségű beszédmintát tartalmazó referencia adatbázis, a gyakorlóanyag összeállításához, a fiziológiás szenzomotoros csatolások kiépítésére.
4. Automatikus értékelés a páciens kiejtési jóságának megítélésére.

Ezek az összehangoltan tervezett és felhasznált elemek kölcsönösen felerősítik egymás hatását a beszédterápiában, ahogy azt a Beszédkorrektor rendszer gyakorlati alkalmazásakor nyert tapasztalatok mutatják (Öster, A., 1999, 2002.), (Váry, 2001).

Ilyen komplex beszédoktató rendszer létrehozásáról a mai napig nincs tudomásom. A továbbiakban ismertetem a 4 építőelemet.



7. ábra

A Beszédkorrektor rendszer építő elemei

Multiszenzoros beszédterápiás környezet:

Rendszeremben az akusztikus paramétereket a periférikus hallásmodell (Zwicker, 1982., Zwicker and Terhardt, 1980.) alapján mérem. Ezeknek a paramétereknek újfajta képi megjelenítése, mint vizuális információ, igen alkalmas a beszéd egyes jellemzőinek helyes beállítására. A spektrális vizuális információt spektrumok és spektrogramok adják, amelyek a modell kimeneti számértékeit ábrázolják. A multiszenzoros környezet részletes leírását, az akusztikai előfeldolgozás, a vizuális ábrázolás módszerét, egyben az 1. tézis gyakorlati megvalósítását a 3. fejezetben részletezem.

A fizikai paraméterek egyszerű grafikai megjelenítésének értelmezése gyermekek számára nehéz, ezért azt háttér képek alkalmazásával segítem. A háttér képek a beszédhang minták spektrális statisztikus modelljeit jelenítik meg. A statisztikus modellek leírását a 5. fejezetben adom meg, amely a 2. tézis, gyakorlati megvalósítása. Ezek a spektrális modellek ugyanazok, mint amelyeket a kiejtés jóságának automatikus értékelésénél használok, mint referenciát. Így biztosítottam a vizuális információ és az automatikus értékelés visszajelzéseinek szinkronizált hatását.

Automatikus értékelés

Az automatikus visszacsatolás a kiejtés jóságának a program által végzett automatikus értékelését nyújtja, amelynek részletes leírását a 6. fejezetben adom meg. Osztályzatokkal, jutalmakkal értékeljük a paciens által kimondottakat.

Gyakorlás során a kiejtés jóságának megítélése és annak visszacsatolása megkönnyíti a tanulást, lehetőséget ad önálló gyakorlásra. Az automatikus visszajelzés spektrális távolságszámításon alapul. Azt mérjük, hogy a paciens által bemondott minta spektruma milyen messze van a statisztikus spektrális modellektől (3. tézis kifejtése, és bizonyítása a 6. fejezetben olvasható).

Oktatási módszertan:

Részletes oktatási módszertant dolgoztam ki a hang kialakításától, a helyes hangszín beállításán át a folyamatos szöveg érthető, automatizált, spontán képzéséig. Az egyes oktatási lépések, gyakorlatok kialakításakor megszerkesztettem a referencia minták szöveganyagát (Referenciaadatbázis). A vezérfonal a hagyományos oktatáshoz igazodik, az aktuális megoldások viszont az új technológiai háttérhez, az új beszédterápiás környezethez (ld. 8. fejezet).

Referencia beszéd adatbázis

A beszédterápia az új hangképzési folyamat kialakítására, valamint a rossz, régi, már automatizált és stabilizált hangképzés leépítésére, törlésére irányul. Ez egy tanulási folyamat, amely főleg szenzomotoros hatáskörben érvényesül. E tanulás során állandó gyakorlás és ismétlés révén a fizioológias szenzomotoros csatolásoknak kell létrejönniük, kialakulniuk, majd stabilizálódniuk. Ez a tanulási mód a gyakorlás állandó irányítását, ismétlését és korrekcióját igényli. A beszédterápiát tartalmilag az új beszédminták megtanulásának folyamata jellemzi, ami sok-sok gyakorlással járó feladat. Tanuláskor ehhez sok mintát kell a paciensnek hallani, látni, kiejteni. A tanuláskor hallott és látott mintákat a referencia adatbázis tartalmazza (ld. 5. fejezet).

Bizonyítás

A tézisekben megfogalmazott állításaim helyességét, különböző, az adott témához illeszkedő módszerrel bizonyítom. Mivel egy multiszenzoros beszéd-oktató rendszer műszaki tudományos alaprendszerének kidolgozásáról van szó, egy – egy elem vizsgálata nem ad igazi eredményt, hiszen a lényege a multiszenzoros hatásnak, az, hogy együttesen alkalmazva őket, egymás hatását fölerősítik. Éppen ezért a teljes komplex rendszer elméleti alapjainak helyességét, a gyakorlatban megvalósított, úgy nevezett Beszédkorrektor rendszer részletes hatásvizsgálatának eredményeivel fogom bizonyítani, amelyeket részletesen bemutatok és elemzek a 9. fejezetben.

3. Multiszenzoros beszédterápiás környezet

Ebben a fejezetben tárgyalom a 1.tézisben foglaltak konkrét megvalósításait, az egyes elemek kidolgozását, egységes rendszerbe illesztését. Vagyis megvalósítottam egy multiszenzoros érzékeltetésen (hallási, látási) alapuló olyan rendszert, amelyben a vizuális információ az akusztikai előfeldolgozással nyert fizikai paraméterek képi megjelenítése.

Kutatási módszer

A mérési, megjelenítési eljárások kidolgozásánál, az oktatási rendszer megalkotásánál, végig általános módszer volt az állandó visszacsatolás a paciensek és a terapeuták részéről. Használva az éppen kidolgozott módszert - akár hangképi kialakítást, akár új módszertani megoldást, - azokat a gyermekekkel kipróbáltuk és a terapeuták véleményét mindig kikértem. Észrevételeiket jegyzőkönyveztem, és a további kutatásnál hasznosítottam. Véleményem szerint ez a módszer segített ahhoz, hogy sikeres megoldásokat tudtam találni.

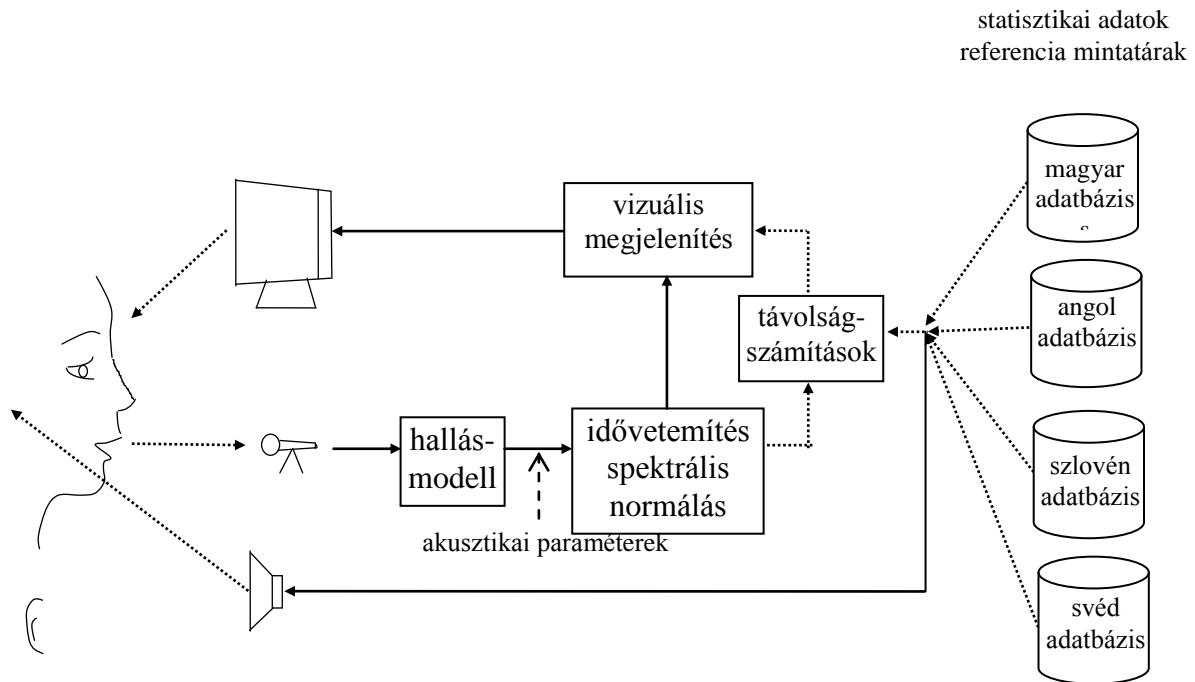
Kezdetben a laboratóriumban folytatott rendszeres oktatás adta a kutatáshoz a megfelelő alapot. Később OTKA program keretében lehetőségem nyílt arra, hogy a teszt programokat 4 iskolában telepíthessem, és az ott dolgozó terapeuták véleményét rendszeresen írásban kikértem.

A többnyelvű fejlesztés esetén szintén jellemző volt, hogy a témavezetők szoros kapcsolatot tartottak fenn a rendszert használó terapeutákkal, és előre meghatározott a kutatás adott fázisára vonatkozó tesztlapok kitöltésére kértük fel őket.

A kutatás, majd a fejlesztés során előfordult, hogy más kísérleti módszert alkalmaztam. Ezeket az eseteket a dolgozatban, az adott helyen részletesen kifejtem. Ilyen például a nyelvfüggetlen automata beszédhang szegmentálási módszer kidolgozása a 3.2.7. fejezetben, ahol a szegmentálás pontosságát például a betanítóanyag mennyiségének és milyenségének a függvényében vizsgáltam.

3.1. Akusztikai előfeldolgozás

A teljes, többnyelvű oktató és gyakorló rendszer, a Beszédkorrektor blokkvázlata a 8. ábrán látható. Az akusztikai előfeldolgozás, a vetemítő, távolságszámító algoritmusok és a képi megjelenítés nyelvfüggetlenek. A referencia hang és képszótárak, valamint az adott nyelvre vonatkozó nagy adatbázisból nyert statisztikai adatok nyelvfüggőek, azaz minden nyelvre külön kell őket létrehozni.



8. ábra

A Beszédkorrektor mérő és feldolgozó rendszer blokkvázlata

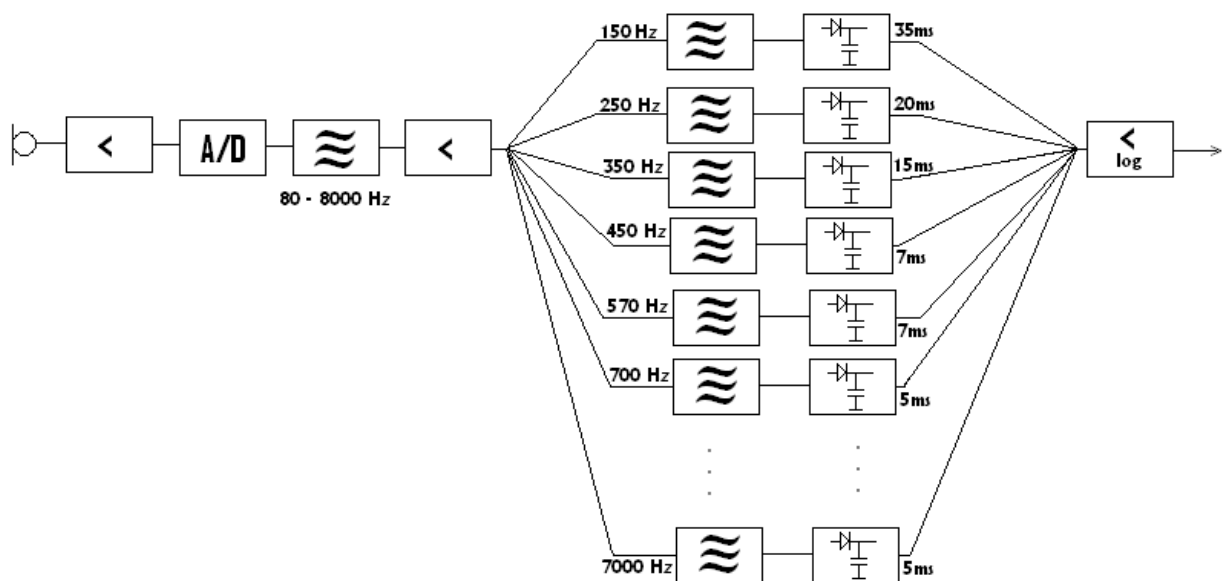
3.1.1. Spektrum elemzés

Az akusztikai feldolgozásnál a lényeges kérdés az, hogy melyik előfeldolgozási eljárás a legmegfelelőbb arra, hogy vizuálisan megjelenítve alkalmas legyen a helyes kiejtés szemléltetésére, vagyis a kép figyelésekor az agyi feldolgozás alapján a hangképzésre engedjen következtetni. Erre a legalkalmasabbnak tartottam egy egyszerűsített hallásmodell alkalmazását, amely az emberi hallórendszer periférikus szintű feldolgozási mechanizmusát modellezi (Zwicker, 1982., Zwicker and Terhardt, 1980.). A frekvenciaelemzést hallási szűrősávokban (Bark-sáv) végzi.

A modellbe beépített adatok átlagos beszédhang intenzitásra és átlagos beszédtempóra érvényesek. A szűrők erősen aszimmetrikusak (kis frekvenciák irányában enyhe meredekségűek, nagy frekvenciáknál viszont élesen levágnak, a hallási lefedési görbék fordítottjaként). A beszéd átlagos intenzitás szintjén (60-70 dB) ez a meredekség kis frekvenciák irányában 25 dB/Bark, nagyobb frekvenciák irányában 10 dB/Bark. Spektrumelemzési eljárásunk lényegesen eltér a hagyományos szonográf elemzési módszerétől. Lényeges különbség a frekvencia- és időfelbontásban van. Eltér továbbá a széles körben elterjedt Fourier-elemzésektől (FFT, DFT, Cepstrum (Gordos, Takács, 1983)) is.

Méréskor a hallási sávokat megvalósító digitális szűrősort használjuk a 100 Hz és 7700 Hz közötti szélső frekvenciasávban. Ebbe a frekvenciatartományba 20 hallási elemzési sáv esik, az 1. táblázat szerint. A szűrőkből kijövő jelet egyenirányítás, rövid idejű átlagolás után logaritmizáltam (10. ábra). A rövididejű átlagolás időállandója 1 kHz-ig a szűrő középfrekvenciájának megfelelő periódusidő ötszöröse ($5 \times T_0$), 1kHz felett állandó 5 msec.

Az e modell kimenő paramétereire alapított vizuális ábrázolás adja meg a tanulónak azt a lehetőséget, hogy a beszédet a vizuális látvány alapján tovább feldolgozza, és a feldolgozás eredményeként a szükséges korrekciót megtegye.



9. ábra
A modell működési blokkvázlata

A szűrőparaméterek matematikai leírását (Zwicker, 1982) a következő összefüggések adják meg, ahol z = a Bark-érzeti hangmagasság (Bark), és f = a futó frekvencia (kHz):

$$z = 13 \cdot \arctg(0,76f) + 3,5 \cdot \arctg(f / 7,5)^2 \quad (1)$$

a hallási sáv szélessége pedig:

$$\Delta f_{kr} = 25 + 75 \cdot [1 + 1,4f^2]^{0,69} \quad (2)$$

Általános beszédhangerőnél a szűrők átviteli jelleggörbéje logaritmikus számértékekben kifejezve:

$$10 \log F(x) = 15,8 + 7,5(x + 0,5) - 17,5 \cdot [1 + (x + 0,5^2)]^{1/2} \quad (3)$$

ahol $F(x)$ maximuma $x=0$ értéknél van.

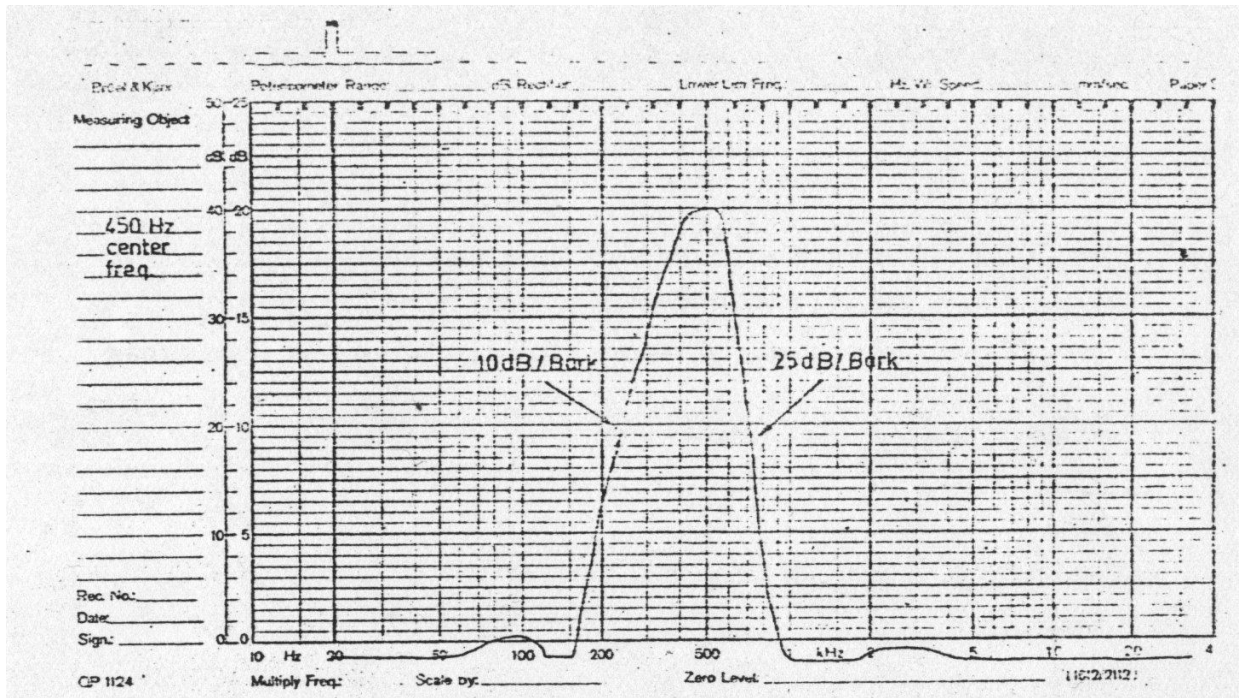
E jelleggörbét IIF digitális szűrőkkel közelítettük, és a közelítés pontossága az átviteli tartományban ± 1 dB, a zárótartományban, 38 dB-es pontokon kívül ± 3 dB, sávközépi erősítés 0dB ± 1 dB.

Mintavétel: 22 025 kHz, 16 bit. Dinamikatartomány 50 dB.

Táblázat 1.

A hallási sáv szélességű szűrősor adatai

A szűrő sorszáma	Középfrekvencia	Alsó vágási frekvencia	Felső vágási frekvencia	Sáv szélesség
Bark	Hz	Hz	Hz	Hz
1	150	100	200	100
2	250	200	300	100
3	350	300	400	100
4	450	400	510	110
5	570	510	630	120
6	700	630	770	140
7	840	770	920	150
8	1000	920	1080	160
9	1170	1080	1270	190
10	1370	1270	1480	210
11	1600	1480	1720	240
12	1850	1720	2000	280
13	2150	2000	2320	320
14	2500	2320	2700	380
15	2900	2700	3150	450
16	3400	3150	3700	550
17	4000	3700	4400	700
18	4800	4400	5300	900
19	5800	5300	6400	1100
20	7000	6400	7700	1300



10. ábra

A megvalósított 450 Hz középfrekvenciájú szűrő kimenet frekvencia függése

3.1.2. Alaphang meghatározás

A rövid idejű **Átlagos Magnitúdó Különbségfüggvény (AMDF)** módszert használtam (Gordos, 1983). Az adott feladatra az alaphang meghatározási pontossága ezzel az egyszerű algoritmussal is nagyon jó. **Minden pontjának számolásához az n-edik** sorszámával végződő N darab beszédmintát használunk fel.

$$D_n(k) = \frac{1}{N} \sum_{m=n-N+1}^n |s(m) - s(m-k)| \quad (4)$$

ahol $s(m)$: a beszédjel m . időpillanathoz tartozó mintája.

k : az AMDF változója

ahol

$N \cdot T = 20$ msec,

T : a mintavételi idő.

A függvénynek a periódusidőnek megfelelő k_m értéknél éles minimuma van. Az alapperiódus így $T_0 = k_m \cdot T$.

3.2. Vizuális ábrázolás

Amint azt az előző fejezetben is említettem az akusztikai előfeldolgozást a periférikus hallást modellező egység végzi, amelynek kimenetén megjelenő paraméterek képét a páciens dolgozza fel magas szintű információfeldolgozási képessége alapján. A hangkép megformálásánál az volt a cél, hogy olyan információt adjunk a hangról a páciensnek, amely rávezeti őt a hang helyes megformálására. Részletes vizsgálatok alapján állítottam be, hogy milyen felbontásban milyen színekkel ábrázoljuk a paramétereket, pl. a hangosságot, a spektrális eloszlást, az alaphangváltozást ahhoz, hogy a vizuális megjelenítés minél informatívabb legyen. Ilyen kérdésekre kellett választ keresni: hogy lehet a páciens figyelmét a spektrum maximális energiájú pontjaira irányítani? Hogyan tudja a páciens felismerni, ha rossz ritmusban ejti ki a hangokat?

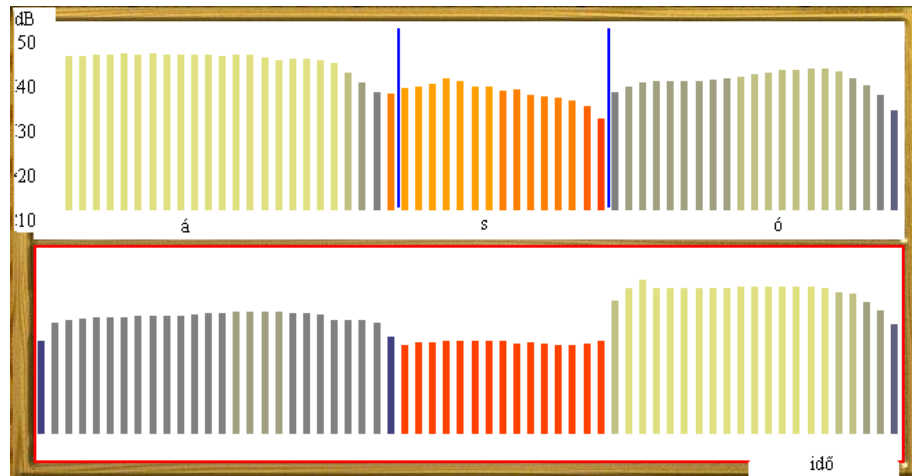
Néhány gyakorlatban a referenciamintával való összehasonlításon alapul a tanulás. Ilyenkor az ablak felső részében a referenciamintát mutatjuk, alsó részben jelenik meg a páciens kiejtéséből származtatott kép. Így a páciens könnyen össze tudja hasonlítani kiejtését a referenciamintával.

A megjelenített hangképek a következők: hangintenzitás szint időbeli változása, alaphang, hanglejtés, hallási spektrum, spektrogram, spektrális különbségek, ritmikai képletek.

3.2.1. Intenzitásváltozás képe

Az intenzitás szinteket 50 dB-es dinamikatarományban mérjük, hiszen az 50 dB intenzitásszint tartomány tökéletesen elegendő a beszéd dinamikaviszonyainak az ábrázolására (Kassai I., 1981). Az integrációs idő 10 msec. A megjelenített színes oszlopsor a beszédintenzitás 10-es alapú logaritmusának időbeli változását követi. Jól mutatja az egyes beszédhangok helyes intenzitásviszonyait. E megjelenítésnek például a megfelelő szóhangsúlyok kialakításánál van jelentősége. Minden fejlesztett nyelv esetében nagyon fontos a helyes hangsúly kialakítása a szavakon belül, hiszen ez adja meg az adott nyelv sajátos ritmusát. A magyar nyelvre jellemző, hogy a szavakban mindig az első szótagon van a hangsúly. Az ásó magyar szó helyes kiejtésekor a 11. ábrán ez jól látható. Vizuális

megjelenítésnél a színek 6 dB-enként változnak, világos árnyalatok a nagyobb intenzitást jelentik, amint azt a 11. ábra mutatja.



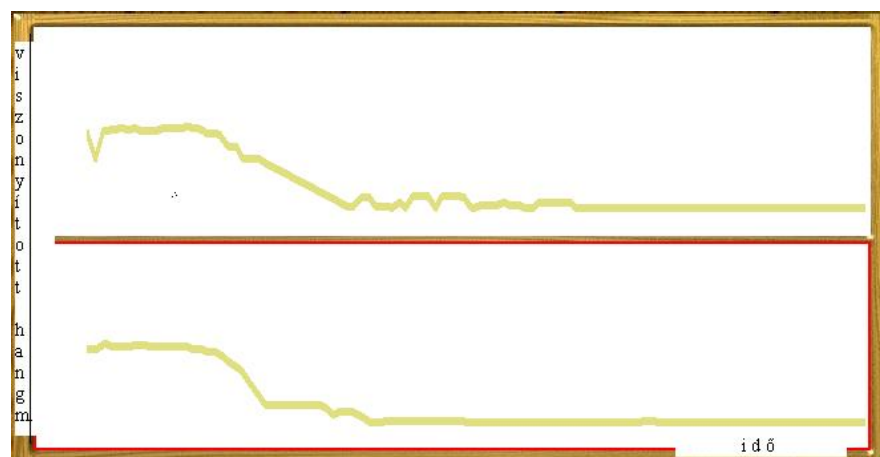
11. ábra

A beszéd dinamikusviszonyaira jellemző beszéd-intenzitásszint változása az „ásó” szó példáján

Fenn: referencia bemondás, lenn: gyermek által bemondott rossz hangsúlyozás. (Az ó hangon van a hangsúly az á helyett)

3.2.2. Hanglejtés képe

Számos normalizációs eljárást kipróbáltam (Gósy, 1980, 1989), (Kassai, 1979), hogy megvizsgáljam, melyik az az eljárás, amellyel a hanglejtés alakulása jól megmutatható. Azt találtam, hogy a legmegfelelőbb, ha az alaphangot az egyenletesen temperált hangmagassági skálán ábrázolom (12. ábra), az alaphangértékeket pedig az első szótag magánhangzójának kvázi stacionárius részén mért alaphanghoz viszonyítom.



12. ábra
A hanglejtés ábrázolása
Fenn. referencia ejtés, lenn: a gyermek által bementett minta.

Vagyis azt mutatom meg, hogy hogyan változik az alaphang a kiejtés során az első szótag alaphangjához viszonyítva. E módszer választásának két oka volt. Először is az intonáció hallási érzékelésénél, agyi feldolgozásánál nem az alaphang konkrét frekvenciaértéke az ami meghatározó, hanem a változás maga, vagyis az, hogy az alapfrekvencia nő, vagy csökken, és azt milyen ütemben teszi. Másodszor vizuális feldolgozás szempontjából is könnyebb a kiértékelés, ha az alaphang görbe mindig ugyanazon a ponton kezdődik. Ugyanakkor előnyös a kép elhelyezése szempontjából is.

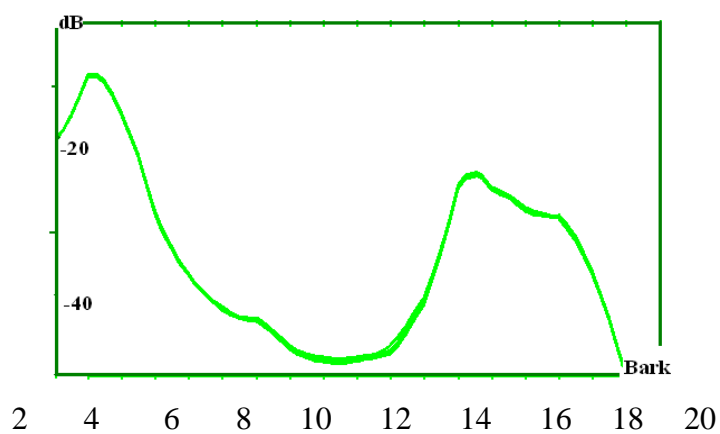
3.2.3. A hallási spektrum képe

A hallási spektrum (színkép) a különböző hallási (Bark) sávokban egy adott időkeretben (10 msec) mért intenzitások. Ábrázolásra legtöbbször az intenzitás logaritmikus léptékben ($10 \log_{10}$) kerül, ezt intenzitás szintnek nevezik.

A hallási spektrum képeken a vízszintes tengelyen a frekvenciát 1-20 Bark sávokban, a függőleges tengelyen az intenzitás szintet mutatom 50 dB-es dinamika tartományban. A szűrőkimeneti szintértékekhez egy simított görbét illeszttek. Minden 10 msec-os időkeretben új spektrum görbe jelenik meg. A megfelelő vizuális ábrázolás érdekében minden spektrum görbéhez a korábbi 9-et is megjelenítem. Példaként a 13.a. ábrán bemutatom egy kitartott é hallási spektrumát. A mérési pontokra illesztett görbét behelyeztem a helyesen képzett é hang megengedett spektrális ingadozási sávjába,* a 13.b. ábrán látható módon.

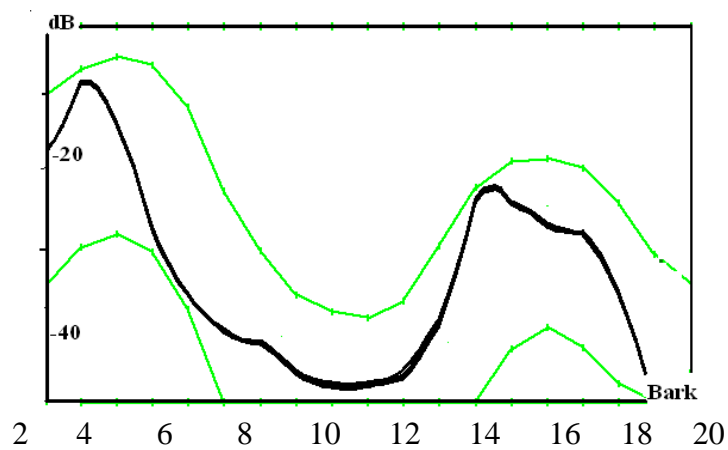
Az 13. a. és 13. b. ábrákon látható mérési görbét a páciens, főleg ha az 5-6 éves kisgyermek, nehezen tudná értelmezni, feldolgozni, de ha egy kedves kis rajzba foglalom (ld. később), akkor a gyermekek számára teljesen érthetővé válik. Az oktatórendszerben használt vizuális ábrázolást a 13. c. ábrán mutatom be. A helyes kiejtéskor a két görbe közötti területen kell a mérési pontoknak lenniük. Az ábrán megjelenített álatoknak is figyelemfelhívó szerepük van, miszerint a mérési pontoknak érinteniük kell ezeket az állatokat. Az állatok a hallási spektrum jellemző pontjait (a jellemző energiamaximum és minimumhelyeket) hangsúlyozzák.

* A hallási spektrum összetevők megengedett ingadozási sávja percepciók kísérletek során került meghatározásra oly módon, hogy vegyes minőségű gyermekbeszéd gyűjteményből a helyes kiejtésű mintákat hallgatók kiválogatták. Az ingadozási sávot a helyesnek ítélt kiejtések alapján határoztuk meg. Részletezése később



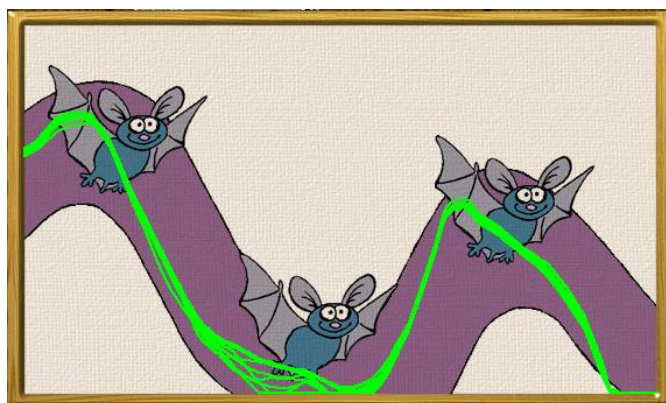
13.a. ábra

Egy gyermek által bemondott kitartott é hang mért hallási spektruma



13.b. ábra

Egy gyermek által bemondott kitartott é hang mért görbéje és a helyesen képzett é hang megengedett hallási spektrumának ingadozási sávja

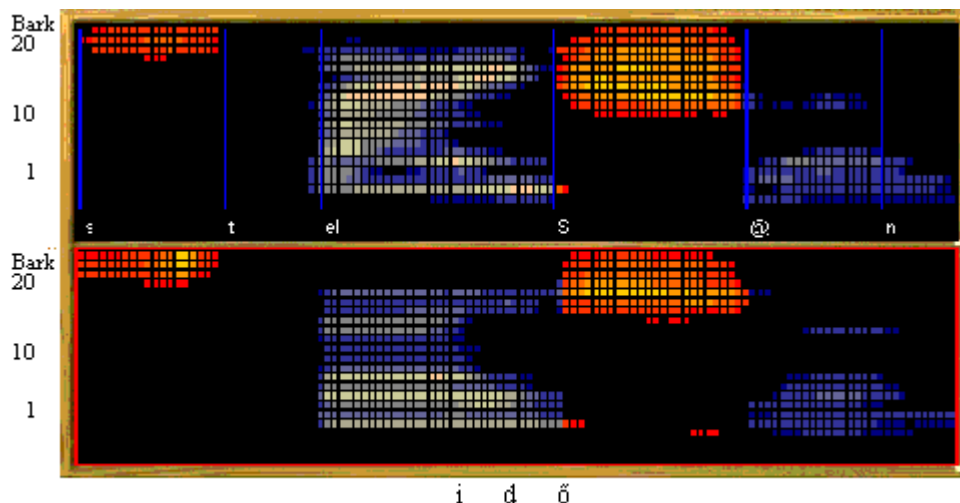


13. c. ábra

Egy gyermek által bemondott kitartott é hang mért görbéi helyes kiejtés esetén, a helyesen képzett é hang megengedett hallási spektrumának ingadozási sávjában, magyarázó háttér képpel.

3.2.4. A hallási spektrogram képe

A hangképnek ebben a megjelenítési formájában a hallási spektrum időbeli változása követhető nyomon. A vízszintes tengely az idő, a függőleges a 20 szűrő (azaz a Bark sávok). Szótagok, szavak és mondatok gyakorlásánál a szűrőkben mért intenzitásszintek időbeli változásait a teljes szótagra, szóra, mondatra ábrázolom úgy, hogy minden 10 msec időkeretben megadom az intenzitásszint értékeket. A színek 6 dB-enként változnak, világos árnyalatok a nagyobb energiát jelentik. A zörej jellegű hangok a piros szín árnyalataival, a rezonáns jellegű hangok a kék különböző árnyalataival jelennek meg.



14. ábra

A station angol szó hallási spektrogramjai

Felső: címkézett, szegmentált (ld. 3.2.7. fejezet) referencia spektrogram, alsó: a gyakorló páciens kiejtése

A képernyő felső felében jelenik meg a referencia hangkép: a szótag, vagy szó helyes ejtésének spektrogramja, amelynél függőleges vonallal bejelölésre kerülnek a beszédhangok határai is, alatta pedig a gyakorló ejtése. A két kép összehasonlításával megállapítható, hogy a tanuló milyen mértékben ejtette helyesen a hangsor hangjait.

A gyakorló a szavakat, mondatokat természetesen más ritmusban ejti, mint a referencia bemondó, tehát a beszédhangok hossza különböző mértékben nyúlik, vagy rövidül. Ez bizonyos határok között természetes. Vizuális összehasonlításnál viszont nagyon zavaró, hogy

nem kerülnek a szavakat felépítő beszédhangok egymás alá a képernyőn. Ennek a problémának a megoldására dinamikus vetemítést alkalmaztam (Sakoe, H. and Chiba, S. 1971), mégpedig a szimmetrikus „dinamikus idővetemítő” (dynamic time warping) eljárást, melyet a beszéd felismerésben először Sakoe és Chiba használt.

Tehát a szótárban lévő referenciaminták egyikét (A) és az aktuálisan bemondott szót (B) kell egymáshoz illeszteni:

$$A = (a_1, a_2, \dots, a_N)$$

$$B = (b_1, b_2, \dots, b_M)$$

ahol a_i és b_j a bemondás i . ill. j . időkeretében mért jellemző vektor (a hallási spektrum intenzitásai a különböző Bark sávokban).

$$a_i, b_j \in R^l; i = 1, 2, \dots, N, j = 1, 2, \dots, M$$

Az, hogy a két mintát különböző ritmusban ejtettük ki, azt jelenti, hogy egyrészt általában $N \neq M$, másrészt általában azt, hogy az A kiejtés i . keretének nem a B kiejtés i . kerete, hanem annak egy másik j . kerete felel meg a kiejtés fonetikai tartalma szempontjából. A két minta közti helyes illesztés leírható (i, j) számpárok sorozataként:

$$F = (i(1), j(1)); (i(2), j(2)); \dots; (i(K), j(K)); \quad (5)$$

ahol az $(i(k), j(k))$ számpár azt fejezi ki, hogy az A kiejtésbeli $i(k)$ kerethez a B kiejtés $j(k)$ kerete tartozik a fonetikai azonosság értelmében. Az (5) számpárok sorozata a kétdimenziós térben egy pontsorozattal szemléletesen ábrázolható. Itt az i -tengely az A referenciamintához tartozó, míg a j -tengely a B-hez, vagyis az aktuális bemondáshoz tartozó időtengely. A k a pont sorszáma.

A kiejtési ritmuskülönbséget leíró (5) pontsorozat pontjainak összekötése egy törtvonallal jellemezhető, ezért az F-et vetemítőgörbének, illetve vetemítőfüggvénynek nevezzük.

A

$$D_F(A, B) = \frac{\sum_{k=1}^K d(a_{i(k)}, b_{j(k)}) w(k)}{\sum_{k=1}^K w(k)} \quad (6)$$

távolságmérték a két kiejtés F vetemítő görbe menti távolságát fejezi ki, ahol $w(k)$ -k alkalmasan megválasztott súlyozó tényezők $k = 1, 2, \dots, K$, és

$$d(a_i, b_j) = \left(\sum_{r=1}^l |a_{i,r} - b_{j,r}|^p \right)^{1/p} \quad (7)$$

ahol $a_{i,r}$ ill. $b_{i,r}$ az A, ill B kiejtés i.ill j. időkeretében mért jellemző vektor r. eleme, míg p szokásos értéke 1 („abszolút értékes távolság”), vagy 2 („euklideszi távolság”).

Feladatunk ezek után „csupán” annyi maradt, hogy meghatározzuk azt az F vetemítőgörbét, amely az A és B kiejtéseik közti ritmuskülönbséget legjobban leírja. Világos, hogy ha egy szó A és B kiejtéseiben fonetikai azonosság szerint az a_i kerethez a b_j tartozik, akkor $d(a_i, b_j)$ kicsi. Éppen ezért érthető, hogy azt az F vetemítőfüggvényt nevezzük optimálisnak, amelyre a (6) szerinti $D_F(A, B)$ minimális, vagyis amelyik a lehető legjobban „egymáshoz vetemíti” a két szómintát (kiejtést). Ebben az értelemben tehát két szóminta távolságának definíciója:

$$D(A, B) = \min_F D_F(A, B) \quad (8)$$

A $D(A, B)$ fenti meghatározása igen számításigényes (kimerítő keresés). Ezért azt közelítő módszerekkel célszerű meghatározni (Sakoe és Chiba, 1978; Hunt és Tomaszewski, 1983; C. Nyers – L.H. Rabiner – A. E. Rosenberg, 1980). Tapasztalataim alapján a Sakoe-Chiba módszerben a „g” részösszegekre alkalmazott „előre lépő részgráf”:

$$g(i, j) = \min \begin{cases} g(i-1, j-2) + 2d(i, j-1) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i-2, j-1) + 2d(i-1, j) + d(i, j) \end{cases} \quad (9)$$

szerinti megválasztása adja a legjobb eredményt.

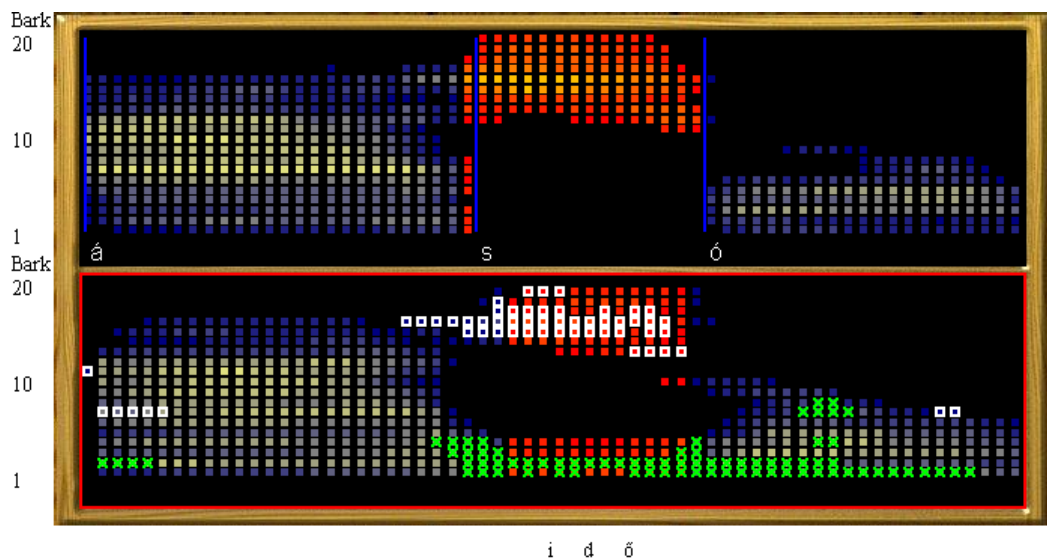
Az optimális vetemítő görbére eső i és j indexek egymás alatti megjelenítésével ábrázolva az alsó spektrogramot, az a felső referencia spektrogrammal már jól összehasonlítható vizuálisan.

A vetemítést csak behatárolt ritmuseltéréseken belül célszerű elvégezni. Az (1,1) és (M,N) pontokat összekötő egyenes körüli „megengedett” vetemítési sávot jól kell beállítani ahhoz, hogy a mintaillesztés ne legyen eltúlzóan hamis. Tehát a normális ingadozáson túli ingadozásokat, például az extrém lassú beszédtempót már nem kell vetemíteni, hadd látszódjon a spektrogramon, hogy milyen nagy az eltérés.

Az összetett hangképből a gyakorlandó hangot, a vizuális értékelés megkönnyítése végett egy háttér kép segítségével mindig kiemeljük. A háttér képet oda tesszük, ahol a szótagban, szóban a gyakorlandó hang van. A sziszegők gyakorlására készült modulban a háttér kép mindig a sziszegő hanggal együtt jelenik meg a képernyőn. A magánhangzók gyakorlására készült modulban a háttér kép mindig a gyakorlandó magánhangzó háttérében van.

3.2.5. Spektrális eltérés képe

A referencia ejtés hallási spektrumától való eltérést vizsgáljuk akkor, ha a felső spektrogramot (referenciaminta) az alsóval (a gyakorló által bementett) dinamikus idővetemítés után szemre összehasonlítjuk.



15. ábra

A referencia és a páciens által bementett ejtés közti spektrális eltérés szemléltetése

A referencia és a páciens által bementett ejtés közti eltérést az alsó spektrogramon ábrázolni is tudjuk. Ekkor a két spektrogramot hasonlítjuk össze a dinamikus vetemítés után, pontról pontra, úgy, hogy a szűrők kimeneti értékeit kivonjuk egymásból, a pozitív eltéréseket zöld X-szel, a negatívokat fehér négyzettel jelölve.

Annál nagyobb az eltérés mértéke, minél több X és téglalap jelenik meg. Természetesen egy-két téglalap még nem jelent hibát. Hiba a képzésben akkor van, ha az adott hang hosszában összefüggő zöld vagy fehér terület jelenik meg (15. ábra).

3.2.6. Adatbázis rögzítő egység

Adatbázis rögzítő egység segítségével hoztuk létre a különböző referencia beszéd-mintatárakat: hangokat, szótagokat, szavakat, szópárokat, mondatokat. Minden elemből három minta kerül rögzítésre, amelyből a legjobb a program számára kiválasztható. A gyakorlandó beszédhangokat meghatározott hangkapcsolatokban és a kifejezésen belül különböző helyzetben kellett megtervezni. Minden nyelvre az adott nyelvhez illeszkedő szótárrendszert hoztunk létre.

Az adatbázis rögzítő egység ad lehetőséget a szövegtár és a képtár kialakítására és kezelésére is.

3.2.7. Szegmentáló egység

A beszédhangok helyét, és betűjelét a hallási spektrogramon bejelöltük, amint azt a 14. 15. ábrákon láthatjuk. Ennek a címkézésnek, szegmentálásnak a szerepe kettős: először is vizuális eligazítást ad a gyakorlónak a hallási spektrogramon, másodsor segíti a beszédhangok beazonosítását az automatikus visszajelzésnél.

A szegmentálás fonetikai tudást igénylő, igen időigényes és fárasztó feladat. Mivel a többnyelvű beszédfeldolgozásban igen gyakran szükség van szegmentált és címkézett hanganyagra, ezért megvizsgáltam annak a lehetőségét, hogy vajon több nyelvre is alkalmasan hogy lehet a szegmentálási, címkézési feladatot automatizálni.

3.2.7.1. Szegmentáló egység alapkonceptiója

A beszéd az európai nyelvek többségében ugyanazokból az akusztikai építőelemekből épül fel. Beszédképzésnél ezeknek a különböző akusztikai tulajdonságú hangeseményeknek a sora jön létre, ahol a beszédhangok egy vagy több ilyen akusztikailag kvázi homogén részből állnak. A kvázi homogén akusztikai részek meghatározott 9 akusztó-fonetikus osztályba sorolhatók (Vicsi - Víg 1995, 1998). Amennyiben valamilyen algoritmussal egy ismeretlen hangsor elemeit be tudom sorolni a 9 osztály valamelyikébe, akkor be tudom jelölni az ismeretlen hangsorban a beszédhangok határait (Vicsi, at al., 1990). Amennyiben a bemondott szöveg betűkkel való leírása is ismert, akkor a betűk is automatikusan bejelölhetők a hanganyagban.

A 9 akusztó-fonetikus osztály a következő:

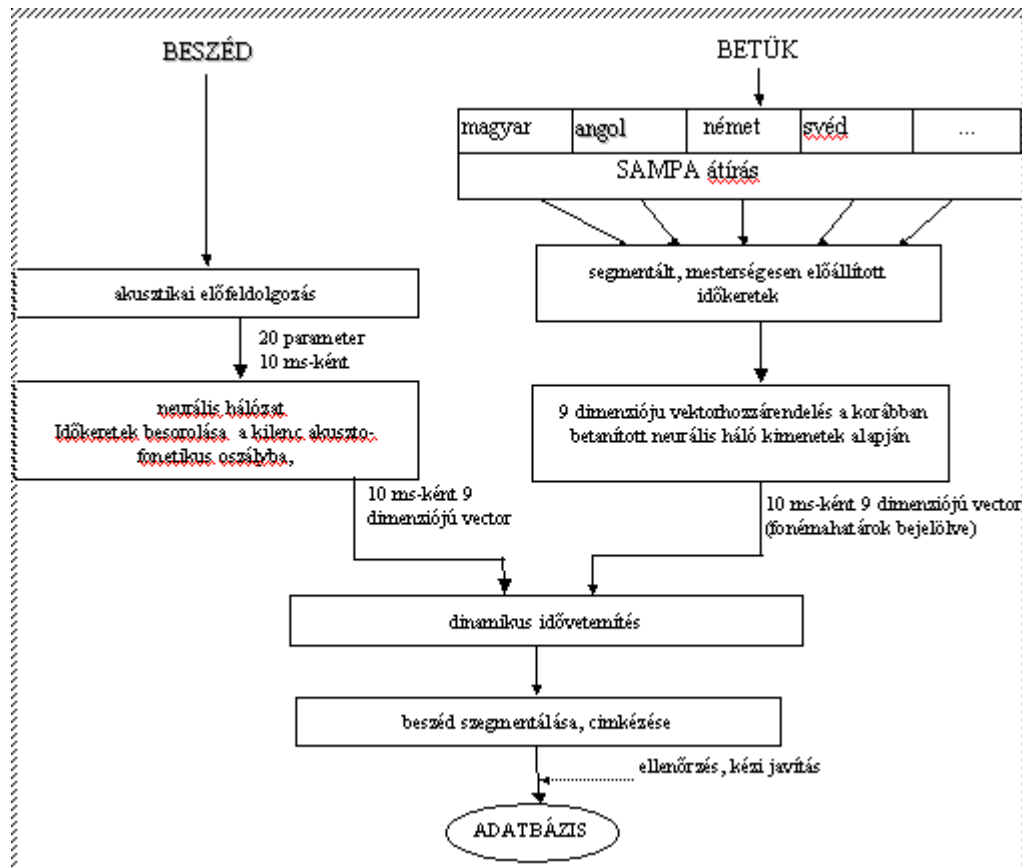
- | | | |
|---------|---|------------------------------|
| 1. mvow | – | mély és közepes magánhangzók |
| 2. hvow | – | magas magánhangzók |

3. +spir	–	zöngés réshangok
4. -spir	–	zöngétlen réshangok
5. res	–	rezonáns mássalhangzók (nazálisok, likvidák)
6. +br	–	zöngés zárfelpattanási zörej
7. -br	–	zöngétlen zárfelpattanási zörej
8. voice	–	zöngés zárhangok
9. sil	–	zöngétlen zárhangok

Az automatikus szegmentálási eljárás blokkvázlata a 16. ábrán látható. A feldolgozás két ágon kezdődik, az akusztikai és a karakteres ágon.

Az akusztikai feldolgozási ágon a bementett mondat időkereteit neurális háló segítségével a 9 akusztikai- fonetikai osztály egyikébe soroltam be, magát az időkeretet egy 9 dimenziójú vektorral jellemezve. A vektor *i*-edik eleme azt mutatja, hogy a neurális háló milyen súlyt ad arra, hogy az éppen vizsgált időkeretben lévő beszédészlet milyen eséllyel esik *i*. osztályba. A háló a visszafelé terjedő algoritmus szerint működött.

A karakterfeldolgozási ágon a bementett mondat betűit automatikusan az adott nyelv fonetikai szabályainak megfelelően beszédhangokká irtam át. Az átíráshoz SAMPA nemzetközi jelölésrendszert használtam (Vicsi, Víg, 1997, 1999). Az átírt anyaghoz modellsorozatot rendeltem úgy, hogy mesterséges időkereteket hoztam létre, (előzetes időstatisztikai vizsgálatok eredményei alapján), amelyeket ezen az ágon is hozzárendeltem a 9 akusztikai- fonetikai osztály valamelyikéhez. Például az **is** szó modelljének első 10 időkerete a magas magánhangzónak megfelelő súllyal, a következő 15 időkeret a zöngétlen réshangnak megfelelő súllyal szerepel. Ez lesz az **is** szó modellje. A két ágon kapott 9 dimenziójú vektorsorozatot dinamikus vetemítéssel egymáshoz illesztettem. Mivel a karakterek mesterséges fonetikai átírásánál pontosan tudom a beszédhangok határait, ezért amikor a hanganyagot illeszttem ehhez a modellhez, akkor a hanganyag hullámformájába is be tudom jelölni a hangok határait.



16. ábra

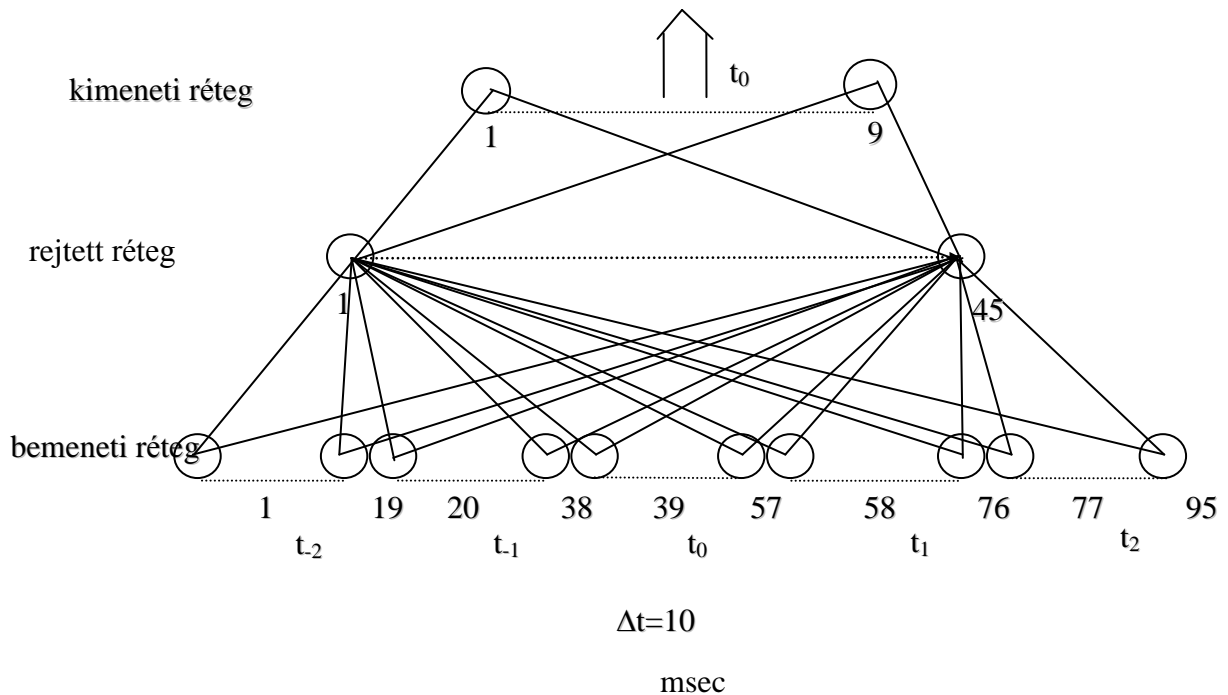
Az automatikus szegmentálási eljárás blokkvázlata
Bal oldal: akusztikai ág, jobb oldal: karakteres ág

Az akusztikai előfeldolgozásnál 10 msec-os időkereteket használunk, így a szegmentálás pontossága sem lehet jobb, mint 10 msec.

Az automatikus szegmentálás nem tökéletes, ezért a kézi javításra mindig szükség van. Azonban a javítás lényegesen kevesebb időbe telik, mint a teljes kézi szegmentálás.

3.2.7.2. Osztályozás neurális hálóval

Az időkeretek besorolását a 9 akusztó-fonetikus osztályba egy rejtett rétegű visszafelé terjedő algoritmussal dolgozó neurális hálóval oldottam meg. A bemeneti csomópontokra időkeretenként a Bark szűrők kimenetein egy időkeretben megjelenő energiák (továbbiakban: „szűrőkimenetek”) kerülnek. Egy t_0 időkeret a bemeneten 5 időkerettel van jellemezve: önmagával, a megelőző kettővel és a követő kettővel, ugyanis az artikuláció egy folyamatos mozgás, ahol az adott állapotot a megelőző és követő állapotok befolyásolnak. Így tehát a háló bemenetére 50 msec-os, keretenként csúsztatott időablakban 5×19 adat kerül.



17. ábra
A szegmentáláshoz használt neurális háló szerkezete

(Itt a szűrősor 19 tagból állt.) A rejtett réteg 45, a kimenet 9 csomópontból állt.

A betanító anyagot méret szerint optimalizáltuk, vagyis meghatároztuk azt a méretet amelyen felül már nem, vagy alig nő a szegmentálás pontossága. Azt találtuk, hogy ez az optimális méret két férfi és két nő bemondása esetén 20 fonetikailag kiegyensúlyozott mondat. A tesztanyag hossza szintén két férfi és két nő bemondásában 20 kiegyensúlyozott mondat volt.

Az osztályok közötti tévesztés a 2. Táblázat tévesztési mátrixából (amely természetesen a főátlóban „találati” és nem „tévesztési” értékeket tartalmaz) jól követhető. Jellemzően a fonetikailag egymáshoz közeli osztályok között fordul elő tévesztés.

A szegmentálás pontosságát úgy ellenőriztük, hogy összehasonlítottuk ugyanazon mondatok kézi szegmentálással kapott anyagát az automatikusan szegmentálttal. A vizsgálati anyag itt is szintén két férfi és két nő bemondásában 20 kiegyensúlyozott mondat volt.

2. Táblázat Tévesztési mátrix a beszéd időkereteinek 9 akusztó-fonetikai osztályba sorolásakor

OUT \ IN	SIL	-SPIR	+SPIR	RES	HVOW	MVOW	-BR	+BR	VOICE
SIL	141	3	2	5	2	1	24	0	1
-SPIR	0	74	0	0	0	0	0	0	0
+SPIR	0	0	23	1	1	0	0	2	0
RES	1	2	10	151	3	7	0	3	10
HVOW	0	1	0	27	156	15	0	0	0
MVOW	0	4	0	12	13	193	0	0	0
-BR	0	0	0	1	1	0	90	1	0
+BR	0	0	0	0	0	0	0	20	0
VOICE	0	0	4	4	2	0	0	5	49

Az automatikus szegmentálás jellemző jegyeit a következők szerint vizsgáltam:

Milyen a szegmentálási pontosság, hogyan függ ez az akusztó-fonetikus osztály típusától? A szegmentálási pontosság függ-e a nyelvtől? A zaj hogyan befolyásolja a szegmentálás pontosságát? A részletes eredmények az ide vonatkozó tanulmányban olvashatók (Vicsi, Vig, 1998). E dolgozatban csak azokat az eredményeket emelem ki, amelyek a Beszédkorrektor szempontjából lényegesek.

Általában nincs beszédhang elhagyás, vagy beszúrás, ez magából az algoritmusból következik. Hiba akkor adódik, amikor a beszélő nem pontosan a kanonikus ejtés szerint beszél, tehát kihagy vagy hozzátesz olyan beszédelemeket, amelyek nincsenek jelölve a SAMPA átírásban. Ilyen esetben a szomszédos beszédhangok elkerülhetetlenül eltolódnak, de a dinamikus idővetemítés segít a dolgon, és egy-két fonémányi idő után korrigálni tud.

A gépi (automatikus) szegmentálással bejelölt beszédhang vagy beszédhang elemek határainak az emberi (kézi) bejelölésű határokhoz viszonyított eltérését vizsgálva a legkisebb eltérés a magánhangzó és csend szegmens, valamint a magánhangzó és zöngétlen spiráns hangok között volt, a legnagyobb eltérés pedig a magánhangzók és a rezonáns mássalhangzók határainál fordult elő. Tendenciájában hasonlóan alakul a szegmentálás bizonytalansága, mint ahogy azt fonetikailag várni lehet. Összefoglalva az átlagos szegmenshatár eltérést a 3 Táblázatban közöljük.

3.Táblázat Az automatikus szegmentálással kapott beszédhanghatár elérések esetei a kézi szegmentáláshoz viszonyítva, angol nyelvre, 4 személy 20 mondatát vizsgálva

eltérések időkeretben mérve	szegmens-határ eltérések eloszlása %-ban
0	35%
±1	70%
±2	83%
±3	88%
±4	92%

Az esetek 35% -ában nem volt a kézi szegmentálás bejelölésétől eltérés , 70%-ban ±1 kereten belül és 92%-ban ±4 kereten belül maradt. Ez annyit jelent, hogy az összes automatikusan bejelölt szegmens 70%-a a kézi bejelöléstől ±1 kereten belülre, 92%-a pedig ±4 kereten belülre esik.

4.Táblázat Tanítóanyag méretének hatása a szegmentálás pontosságára (az adatok a ±2 időkeretet meg nem haladó szegmenshatár eltérésre vonatkoznak)

Nyelv	Angol	Magyar	Angol	Magyar
Tanítóanyag	4 paragrafus (2 nő és 2 férfi, 20 mondat)	4 paragrafus (2 nő és 2 férfi, 20 mondat)	4x4 paragrafus (8 nő és 8 férfi, 80 mondat)	4x4 paragrafus (8 nő és 8 férfi, 80 mondat)
Felismerési anyag	4 paragrafus (2 nő és 2 férfi, 20 mondat)	4 paragrafus (2 nő és 2 férfi, 20 mondat)	4 paragrafus (2 nő és 2 férfi, 20 mondat)	4 paragrafus (2 nő és 2 férfi, 20 mondat)
Rezonáns mássalhangzó szegmenshatár felismerése	83%	76%	85%	77%
Részhang szegmenshatár felismerése	95%	88%	94%	91%
Összes fonéma szegmenshatár felismerése	83%	85%	83%	86%

3.2.7.3. A nyelv befolyása a szegmentálás pontosságára

A célom az automatikus szegmentáló rendszer létrehozásával az volt, hogy minél több európai nyelven használható legyen, megfelelő segítséget adva a tiszta beszédatbázisok szegmentálására. A BABEL és az EUROM 1 adatbázisok különböző nyelveire megvizsgáltam az automata szegmentáló működését és a betanításra egy optimális megoldást alakítottam ki munkatársaimmal. A vizsgálat kezdetén betanításra 4 paragrafust (2 női és 2 férfi bemondóval, ami összesen 20 mondat volt) és tesztelésre másik 4 paragrafust használtam (2 női és 2 férfi bemondóval, ami összesen 20 mondat volt), minden vizsgált nyelv (magyar, angol, német, bolgár, slován, svéd) esetén. Változtattam a betanító anyagot és vizsgáltam az automatikus szegmentálás eredményeit. A neurális háló mindig 4 paragrafussal volt betanítva minden nyelv esetében és vizsgáltam a szegmentálás pontosságát. Például a hálót betanítottuk 4 bolgár, német és magyar paragrafussal külön-külön és ezután megnéztük, milyen a magyar anyag esetében a szegmentálás pontossága. Az eredményeket az 5. táblázatban közlöm. Annak ellenére, hogy a betanító anyag német, vagy magyar volt, az esetek 79-80%-ban a beszédhang határ ± 2 időkereten belülre esett. Ha a tanításnál növeltük a nyelvek számát és ezzel együtt a paragrafusokat (ehhez nyelvenként 4 paragrafust használtunk), a szegmentálás pontossága növekedett egy határig. Ezt mutatom be az 6. és 7. táblázatban. A vizsgált nyelvek közül, ha a neurális hálót 3 nyelvvel, magyarral, angollal és bolgárral tanítottuk be (3 x 4 paragrafus), ez ugyanazt a szegmentálási pontosságot adta, mint amikor a hálót csak 4 paragrafussal tanítottuk be, de a vizsgált nyelven, mint ahogy az az 6. táblázatban látható. Ez az eredmény hasonló volt minden vizsgált nyelv esetében. Ha a betanításhoz még a német nyelvet is hozzávettük, még a németre sem kaptunk jobb szegmentálási eredményt mint a magyar-angol-bolgár anyag együttes betanításával. Ez azért van így, mert a vizsgált nyelveknél az akusztó-fonetikai osztályok jellemző paramétereinek variáltságát e három nyelv paragrafusaival már lefedtük. Az észrt nyelv vizsgálatokor, - ahol sok a hosszú mássalhangzó, és azok hossza jóval nagyobb mint az angol, vagy a bolgár, vagy a magyar nyelv esetében – a szegmentálás pontossága erősen lecsökkent.

Esetünkben, a már korábban angol-magyar-bolgár anyag együttesével betanított háló a svéd, szlovén nyelvre is a többi nyelvhez hasonlóan jó szegmentálási eredményt adott (7. ábra), így a Beszédkorrektor fejlesztéséhez szükséges beszédatbázisok szegmentálására jól használható volt.

A fentiek alapján belátható a 7. tézis azon állítása, hogy a 9 akusztó-fonetikus osztályba sorolás alapján **kidolgozott beszédhang szintű automatikus szegmentálási és címkézési eljárás közel nyelvfüggetlen, vagyis a rendszer több európai nyelven a gyakorlat számára elfogadható pontossággal (81-90%) működik, az adott nyelvre történő betanítás nélkül is.**

5. Táblázat Magyar szöveganyag szegmentálásának pontossága a betanító anyag típusa szerint

Nyelv	Magyar (H)					
Betanító anyag típusa	Magyar (H)	Bolgár (B)	Német (G)	Vegyes H_E	Vegyes H_E_B	Vegyes H_E_B_G
Rezonáns mássalhangzó szegmenshatár felismerés	76	72	68	74	78	76
Réshang szegmenshatár felismerés	88	93	88	86	87	90
Összes fonéma szegmenshatár felismerés	85	79	80	83	86	85

6. Táblázat A betanító anyag befolyása a szegmentálási pontosságra különböző vizsgált nyelvek esetében, ha a betanítás saját nyelven, vagy 3 nyelv együttesével történt

Nyelv	Magyar (H)		Német (G)		Angol (E)		Bolgár (B)	
Betanító anyag típusa	Magyar	Vegyes H_E_B	Német	Vegyes H_E_B	Angol	Vegyes H_E_B	Bolgár	Vegyes H_E_B
Rezonáns mássalhangzó szegmenshatár felismerés	76	78	75	69	83	83	86	85
Réshang szegmenshatár felismerés	88	87	96	92	95	95	94	96
Összes fonéma szegmenshatár felismerés	85	86	82	78	83	83	89	89

7. Táblázat Szegmentálási pontosság összehasonlítása különböző nyelveken, vegyes 4 nyelvű betanítás esetén

Nyelv	Magyar (H)	Német (G)	Angol (E)	Bolgár (B)	Szlovén (Sl)	Svéd(Sv)
Betanító anyag típusa	Vegyes H_E_B_G	Vegyes H_E_B_G	Vegyes H_E_B_G	Vegyes H_E_B_G	Vegyes H_E_B_G	Vegyes H_E_B_G
Rezonáns magánhangzó szegmenshatár felismerés	76	73	84	84	68	74
Réshang szegmenshatár felismerés	90	97	98	96	96	97
Összes fonéma szegmenshatár felismerés	85	81	84	90	85	85

3.3. Hangvisszajátszás

A referencia adatbázis minden eleme egyenként behívható és hangszórón keresztül lejátszható. A hallási, látási érzékeltetést mindig együtt alkalmaztam, vagyis a hangképek megjelenésével szinkronban mindig elhangzik az adott kifejezés. A gyermekekkel történt tesztelések alapján világossá vált, hogy igen nagy szükség van arra is, hogy a páciens a saját bemondott hangját is vissza tudja hallani, ezért ezt a lehetőséget is felvettük a rendszerbe.

3.4. Kvalitatív kiértékelés

A terapeuták véleményét, amint ezt már ismertettem, rendszeresen kikértük. A multiszenzoros beszédterápiás környezet kialakítása végeztével felkértük svéd partnerünket, Anne-Marie Östert, aki mint terápiás szakember vett részt a nemzetközi munkacsoportban, hogy rendszerünk átfogó kvalitatív tesztelését végezzék el (Vicsi, 2001.). Szervezésében 8 svéd terapeuta gyermekekkel használta a rendszert 3 hónapig, majd válaszoltak az alábbi teszt kérdéseinkre. A terápiában résztvevő gyermekek 4-12 évesek, 3 különböző hallássérült csoportot alkottak: iskola előkészítő osztályos hallássérült gyermekek, súlyosan hallássérült iskolás korú valamint éphallású organikus (idegrendszeri eredetű) beszédzavarral rendelkező gyermekek. A 8 terapeutának a multiszenzoros környezetre vonatkozó kérdésekre összesített véleményét táblázatos formában az alábbiakban közlöm:

1. Mit gondol, kell valamilyen előzetes számítógépes ismeret a rendszer kezeléséhez?

sok	néhány	semmi
0	6	2

2. Milyen a rendszer együttműködése a gyermekkel?

nagyon jó	jó	nem olyan jó	rossz
5	3	0	0

Megjegyzés

- Az együttműködési aktivitás gyermekfüggő volt, és gyakorlatfüggő is.
- Minden alkalommal a rendszerrel akartak gyakorolni.
- A gyerekek - a hallássérültek is - mind vissza akarták hallani a saját hangjukat a gyakorlásnál

3. Elvárásainak mennyire felelt meg a rendszer?

nagyon jól	jól	nem olyan jól	rosszul
5	3	0	0

Megjegyzés:

- Minnél többet dolgozom az új rendszerrel, annál jobban elfogadom
- Különösen a kötött szótárakat tartom nagyon hasznosnak.

4. Szükség van e arra, hogy a terapeuta rendelkezzen előzetes artikulációra vonatkozó vagy akusztikai - fonetikai ismeretekkel?

sok	néhány	semmi
6	2	0

5. Könnyű volt a rendszerrel dolgozni?

nagyon könnyű	könnyű	nem olyan könnyű	nehéz
3	3	2	0

Megjegyzések:

- Idő kell a megismeréshez.

6. Milyen volt a rendszer pedagógiai szempontból?

nagyon jó	jó	nem olyan jó	rossz
5	3	0	0

Megjegyzések:

- A rendszer nagyon jól átgondolt, és pontosan megtervezett.
- A gyakorlandó beszédhangok mindegyikéhez tervezett háttér kép nagyon hasznos, jó ötlet.
- A rendszerben különböző nehézségi fokú gyakorlat van. A gyerek szellemi képességétől és készségétől függ, milyen fokig lehet vele a rendszerben elmenni.

7. Mi a véleménye a grafikáról?

nagyon vonzó	vonzó	nem nagyon vonzó	nem vonzó
5	3	0	0

Megjegyzések:

- A színek és a képek nagyon szépek
- Minden illusztráció egységesített és nagyon tiszta, világos
- Különösen a rajzos automata visszacsatolást használták a gyerekek gyakran

8. Megértette a gyermek a vizuális visszacsatolást a hallási spektrogrammból?

nagyon jól	jól	nem nagyon jól	rosszul
5	3	0	0

Megjegyzések:

- Annak ellenére, hogy a gyermekek természetesen nem értették az akusztikai ábrát úgy mint fizikailag mért paramétereket, mégis érthető visszacsatolást kaptak azzal, hogy összehasonlították saját hangképeiket a referencia hangképekkel.
- Néhány gyereknél problémát jelentett a legfontosabb dolgokra koncentrálni.

9. Megértette a gyermek a közölni kívánt vizuális információt a hallási spektrumokból?

nagyon jól	jól	nem nagyon jól	rosszul
8	0	0	0

Megjegyzések:

- A feladat teljesen világos volt a gyermekek előtt.

10. A hanglejtés gyakorlására megfelelőnek tartja a rendszer által felkínált hangképi megoldást?

nagyon jól	jól	nem nagyon jól	rosszul
8	0	0	0

Megjegyzések:

- A referenciabeszélő hangját is mindig kéri a gyerekek a hangképpel együtt.

11. A hangerőváltozásokat érzékelik a gyerekek a hangképekben?

nagyon jól	jól	nem nagyon jól	rosszul
8	0	0	0

Megjegyzések:

- Egy kis gyakorlás után szinte minden esetben, a legkifejezettebben a ritmusgyakorlatoknál.

12. Hallási spektrum és hallási spektrogram típusú gyakorlatoknál, mind ahogy önnek tudomása van róla, a háttér képek ki-be kapcsolhatóak. Hogyan használta ön a hangképeket? A háttér bekapcsolásával, vagy kikapcsolt helyzetben?

bekapcsolt helyzetben	felváltva	kikapcsolt állapotban
8	0	0

Megjegyzések:

- Üres, nincs értelme, ha kihagyom.

A 12 kérdés közül 5 a rendszer használhatóságára vonatkozó általános kérdés, 7 kérdés viszont csak a hangképekre vonatkozik. Az általános kérdésekre adott válaszok a pozitív irányban ugyan, de szórnak, viszont a hangképekre vonatkozó kérdések teljesen egyértelműek mindig a hangképek maximális elfogadását tükröző válaszokat adtak. A háttér képeket, a terapeuták válaszaiból következtetve, hasznosnak tartják. Ez a kvalitatív kiértékelés, valamint a 9. fejezetben részletezett, számszerűsített érthetőségvizsgálat együttesen szolgálnak az 1. és 2. tézis alátámasztására.

4. Adatbázisok létrehozása

A kialakított multiszenzoros beszédterápiás környezet egyben adatbázis szerkesztő is, amely lehetőséget ad a különböző beszédoktatási feladatok (beszédjavítás, nyelvtanulás) ellátásához szükséges különböző típusú szöveg, hang és kép adatbázisok előállítására (különböző nyelvekre). A beszédhibás gyermekek beszédjavítására létrehozott rendszerben a következő adatbázisok szerepelnek.

Szövegadatbázisok: a gyakorlandó hanganyag szövegadatbázisa nemzeti betűkészlettel, ill. SAMPA* szimbólumokkal (Vicsi, Víg, 1997). A teljes szöveg és képadatbázis rendszerterve mind a 4 nyelvre az 1. sz. mellékletben található.

Referencia beszédmintatár: Minden gyakorlandó hang a rendszerben izolált formában, továbbá különböző hangkapcsolatokban, szavakban, mondatokban és minimálpárokban szerepel különböző helyzetben. (minimálpárok például: szár – sár, szár – zár). Ezeket a szótárban lévő helyes kiejtésű beszédmintákat nevezzük referencia beszédmintáknak, a hozzájuk rendelt beszédkészletet pedig referencia beszédmintatárnak. A referencia mintákat egy gondosan kiválasztott személy mondta be, akinek a beszédje szép, érthető. A tanulónak a terápia során ezekhez a referencia beszédmintákhoz hasonló beszédet kell előállítania. Külön adatbázist hoztam létre a magyar Beszédkorrektornál gyermek és felnőtt nő, valamint férfi beszélőkre. Egyéb nyelvekre csak gyermekadatbázis készült.

A referencia beszédminták felvételénél igen körültekintően kellett eljárni, hiszen jórészt nagyothalló gyermekek részére készült a rendszer. A referencia mintatárakat erősen megválogatott kizárólag szépen beszélő fiatal mondta be. A mintákat a gyakorlásnál nagy hangerővel hallgatják, így az eleve zajos, vagy torz felvétel hibái erősen előjönnek. Ugyanakkor legtöbbjük nagyothalló készüléket visel, ami rossz felvételek esetén még további torzítást vihet be a fülbe. Ezért a felvételeket süket, vagy csendes helyiségben, kis zajú hangkártyával vettük fel. A jel – zaj viszony min 40 dB volt minden nyelv esetén. Jó felvételkor nemcsak a hangfelvétel kellett hogy tiszta, kiszajú legyen, hanem a megjelenő hangképeknek is tisztának kellett lenni.

Referencia hangképtár: Minden referencia beszédmintához tartozik képi megjelenítés is: a referencia hallási spektrogram képe, az intenzitás szint alakulása, a hanglejtés. A mért paraméterek ábrázolása mellett minden beszédhangmintához magyarázó háttér kép is tartozik. A hangképtárba helyeztük el a hívóképeket és a beszédhangok szimbólumképeit is.

Gyermekbeszéd adatbázis: Az adott nyelv fonémáit reprezentáló beszédhangok statisztikai vizsgálatához külön, sok bemondóval készített adatbázisra volt szükség. Meg kellett határozni a helyes kiejtésű beszéd szórását, szélsőértékeit. A 4 különböző nyelvű

* SAMPA nemzetközi fonetikai szimbólumrendszer, amely a számítógép billentyűzetéhez igazodik (<http://www.phon.ucl.ac.uk/home/sampa/hungaria.htm>)

adatbázis szöveganyaga tartalmazta a gyakorlásra külön kiemelt beszédhangokat, azaz a réshangokat, affrikátákat és a magánhangzókat izolált ejtésben, hangkapcsolatokban, szavakban és mondatokban, különböző pozícióban. Egy beszélő bemondási ideje 10-15 perc volt. A bemondók életkora 5 és 10 év között változott. A beszélő és a felvételvezető teljes beszélgetését rögzítettük. Ebből válogattuk ki a helyesen kiejtett mintákat. Nyelvenként a beszélők száma különbözött (35 és 80 között) a helyi lehetőségek szerint.. A bemondók száma nyelvenként változott. Az adatbázisok szöveganyagát, nyelvenkénti leírását a 2 sz. melléklet tartalmazza. (Az adatbázisok mérete nem elegendő igazán korrekt statisztikai vizsgálathoz, de sajnos, sem időben sem pénzben nem volt lehetőség nagyobb anyag létrehozására, ezért ezt a hiányosságot el kellett fogadnunk.)

Magyar nyelvre 72 beszélő mondott be összesen 80 szót, és 29 mondatot. Az 5-6 éves gyerekeket óvodából, a 7-10 éveseket egy általános iskola alsó tagozatából gyűjtöttük össze (10 ötéves, 10 hatéves, 18 hét, 16 nyolc, 6 kilenc és 12 tízéves). Azoknál a gyerekeknél, akik nem tudtak még olvasni, a felvételnkészítő előmondta a felveendő anyagot.

5. Az adott nyelv fonémáit reprezentáló beszédhangok spektrális modelljeinek létrehozása

A jelen fejezetben és később is fontos szerepe lesz a fonémák statisztikus szinképi modelljének. Egy fonéma statisztikus szinképi modelljét az alábbi lépésekben építjük fel, ill. az alábbi paraméterek alkotják.

Legyen egy fonéma k . helyes kiejtésében (l. 3.2.4. fejezet) lévő i . időkerethez (pl. 10 msec) tartozó hallási spektrum vektor a_{ki} . Ennek f . eleme az f . Bark sávban a k . kiejtés i . időkeretében mért energia (intenzitás): a_{fki} legyen $1 \leq k \leq K$, $1 \leq i \leq I$, $1 \leq f \leq F$. (Megjegyezzük, hogy F értéke tipikusan 19).

A statisztikus spektrális modellben

- a fonéma átlagos hallási spektrum vektora (röviden átlag vektora)

$$a = \frac{\sum_k \sum_i a_{ki}}{K + I} \quad (10a)$$

melynek elemei ($a_1, a_2, \dots, a_f, \dots, a_F$),

- a fonéma hallási spektruma kiterjesztésének maximum vektora (röviden: max)

$$a_{\max} = \max_{k,i} a_{ki} \quad (10b)$$

- a fonéma hallási spektruma kiterjesztésének minimum vektora (röviden: min)

$$a_{\min} = \min_{k,i} a_{ki} \quad (10c)$$

- a fonéma hallási spektrum kiterjedésének átlagos pozitív eltérés vektora (röviden: poz)

$$a_{\text{poz}} = \frac{\sum_{k,i} (a_{ki} - a)}{K_1 + I_1} + a \quad (10d)$$

azon k és i indexekre, amelyekre $a_{ki} > a$ és amelyek számossága K_1 , ill. I_1

- a fonéma hallási spektrum átlagos negatív vektora (röviden: neg)

$$a_{\text{neg}} = \frac{\sum_{k,i} (a - a_{ki})}{K_2 + I_2} + a \quad (10e)$$

azon k és i indexekre, amelyekre $a > a_{ki}$ és amelyek számossága K_2 . ill. I_2 .

(Megjegyzés: az a_{ki} mennyiségek hisztogramja az a - (átlag) ra nem szimmetrikus, ezért célszerű a pos és neg használata, pl. a szórás helyett).

A táblázatokban legtöbbször a mennyiségek „szint” változatban, dB-ben kifejezve jelennek meg (pl. $a^{\text{dB}} = 10 \log_{10} a$).

A helyes kiejtésű fonémák statisztikus spektrális modelljeinek létrehozását több cél vezérelte.

- Először is tudományos cél. A magyar beszédhangok hallási szűrősávokban mért átfogó spektrális elemzése még nem történt meg. Különösen nem gyermekek, nők és férfiak összehasonlításában. Ezért mindenképpen célszerű volt ezt a vizsgálatot elvégezni.

- Másodszer a multiszenzoros oktató, fejlesztő rendszeremben mind a hallási spektrum, mind pedig a hallási spektrogram megjelenítés esetében értelmezést segítő háttér képeket alkalmazok. Ezeket a háttér képeket a beszédhangok spektrális átlagértéke és átlagos eltérése szerint szerkesztettem meg.

- Harmadszor az automatikus kiejtési jóság megítéléséhez szintén szükség van e modellekre, mivel az értékelésnél az aktuális bemondás szinképét a spektrális modell adataival vetem össze. A spektrális modellek kiszámításához a gyermekbeszéd adatbázist használtam fel.

Az adatbázisban, mivel válogatás nélkül gyűjtöttük a felvételeket, voltak szép, kevésbé szép és hibás bemondások. A helyes kiejtésű mintákat ki kellett az adatbázisból választani. Ezt a válogatást szubjektív lehallgatási kísérletekkel oldottuk meg.

5.1. Szubjektív lehallgatási kísérletek

Felkértünk 20 beszédterápiával foglalkozó szakembert, és 13 egyetemistát, akik nem ezen a területen tanultak, hogy hallgassák meg az adatbázis elemeit és sorolják be a hallott beszédmintákat 3 csoportba: jó, elfogadható és rossz kiejtés. Koncentráljanak a megadott beszédhangra, és állapítsák meg, a három csoport közül melyikbe tartozik. A kitöltendő űrlapot a 8. Táblázatban mutatjuk meg, ahol a szótárkészlet elemeit tüntettük fel a besorolási oszlopokkal. A vastagon kiemelt beszédhang az amire a hallgatóknak figyelni kell.

8. Táblázat. A szavak egy példalistája a szubjektív lehallgatási kísérletekben

	kiejtés	rossz	elfogadható	jó
1	cs íp	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	rú zs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	fő z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	tes z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	tí z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	tü z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	kés z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	há z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	fő z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	ő sz	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	z seb	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	bo cs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	z sír	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

14	z ár	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	rú zs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16	fő z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17	cs íp	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18	szí v	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19	z öld	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20	z öld	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21	nyol c	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22	tü z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23	fő z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24	fő z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25	mé z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26	mé z	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27	rú zs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

28	Fö <u>z</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29	<u>z</u> öld	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30	K <u>é</u> s	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31	<u>s</u> zúr	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32	Ko <u>s</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

33	rú <u>z</u> s	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34	szí <u>v</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35	<u>z</u> sír	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

A lehallgatók az egy adott beszélő egy adott fonémájára vonatkozó egyedi döntéseiből átlagos döntés kiszámolásához jelölje:

N_u a „jó (helyes)” egyedi döntések számát,

N_a az „elfogadható” egyedi döntések számát,

N_g a „nem elfogadható” egyedi döntések számát.

Egy adott beszélő egy adott fonémára vonatkozó átlagos döntés:

$$\text{Átlagos Döntés} = \frac{3 * N_u + 2 * N_a + 1 * N_g}{N_u + N_a + N_g}$$

(Az összefüggésben szereplő súlyozó tényezők jól szolgálják a skála széthúzását).

Az átlagos döntések alapján három csoportba soroltam a mintákat. Jó csoportba került az a minta, ahol az átlagos döntés érték 2.8-3 közötti, elfogadható csoportba, ahol az érték 1.8-2.2 közötti, és nem elfogadható csoportba kerül az a minta, ahol 1-1,5 közötti az átlagos döntésérték.

A jó kiejtésű csoport mintáiból számítottuk ki az egyes beszédhangok statisztikusmodelljének jellemzőit. Ehhez természetesen az egyes beszédhang minták határait a 3. fejezetben részletezett automatikus szegmentálás segítségével ki kellett jelölni, majd a hallási spektrumot meg kellett határozni.

5.2. A helyes kiejtésű spektrális modellek kialakítása

Folyamatos beszédben a beszédhangok kvázi stacioner állapotai közötti átmenetek folyamatosak, hiszen a hangképző szervek mozgása is folytonos, amíg az egyik hang képzésére jellemző állapotból egy másikba eljutnak. Az átmeneti szakasz időtartama 20-30 msec, a beszéd tempótól, hangsúlyozástól lényegében függetlenül, míg a kvázi stacioner rész időtartama a beszédtempó, a hangsúlyozás függvénye. (Ezzel kapcsolatban megemlítendő, hogy az akusztikailag többkomponensű zárhangok és zárreshangok esetében a zárszakasz hossza változik a beszédtempóval, viszont kötött a zár-felpattanási időtartam).

A helyes kiejtésű magyar beszédhangok spektrális modelljeit a beszédhangok akusztikailag kvázi stacioner részeiből (amennyiben van ilyen) hoztam létre, kiindulva abból, hogy a hangképzésnél ezen állandósult állapot elérése a cél a tranziens képzésű

beszédhangokat kivéve. Mindig az adatbázis jó kiejtésűnek ítélt mintáival számoltam. Az általános beszédjellemzők mellett, részletes terápiára két fonetikai csoportot jelöltem ki: a magánhangzókat és a sziszegő hangokat, amelyek akusztikailag egykomponensű hangok és amelyekre helyes kiejtésben és normál beszédritmusban jellemző, hogy a hang elején és végén lévő átmeneti részek között jelentős kvázi stacioner részt ejtünk.

A beszédhangok átmeneti tranziens részeit itt most nem kívántam figyelembe venni. A tranziens részek színeképek időbeli alakulása a környező – elsősorban szomszédos - hangok képzési helyének és módjának a függvénye.* Ezért a feldolgozás során minden hang elejéről és végéről 2 keretet (20 msec) elhagytam és csak az így maradó tiszta fázist vettem figyelembe. Mindezekért egy fonéma átlagos hallási spektrum vektorának kiszámítása az alábbiak szerint módosul:

$$a = \frac{1}{K} \sum_{i=1}^K \left(\frac{1}{I-4} \sum_{i=3}^{I-2} a_{ki} \right) \quad (11)$$

Amennyiben a beszédhang vagy beszédhangrész hossza kisebb, mint 50 msec, akkor értelemszerűen a rendelkezésre álló összes időkeretet figyelembe vettem (azaz (10a) szerint számoltam). Ez az eset állt elő a zárhangoknál a zárfelpattanási zörej esetében. Ezeknél a hangoknál a spektrális modell pontatlan, de a 10 msec-os időkeret sűrűség, már eleve nagyon eltorzítja a zárfelpattanási zörejeket a jelben. (A zárhangok pontosabb modellezéséhez, akár Rejtett Markov Modellről, akár spektrális modellről van szó, finomabb időfelbontásra volna szükség a fejlesztési folyamatban.) Jelenleg viszont a terápiára kijelölt magánhangzók és sziszegők esetében a kvázi stacioner rész stabil, a spektrális modell jól meghatározott.

Meghatároztam az összes magyar fonéma statisztikus spektrális modelljét gyermekekre vonatkozóan a gyermekbeszéd adatbázis alapján (72 beszélő mondott be összesen 80 szót és 29 mondatot), míg felnőtt férfiakra és nőkre vonatkozóan a BABEL felnőtt beszédatadabázis segítségével (30 férfi, 30 nő, 30 perces szöveganyag). A kapott átlag vektor értékeket a 3. sz. mellékletben grafikusán közöljük. Az összehasonlíthatóság érdekében fonémaként egy ábrára vettük fel a női, a férfi, és a gyermek beszéd spektrális átlagértékeit. Tudomásom szerint magyar nyelv fonémáira ilyen részletes statisztikai vizsgálat ez ideig nem történt.

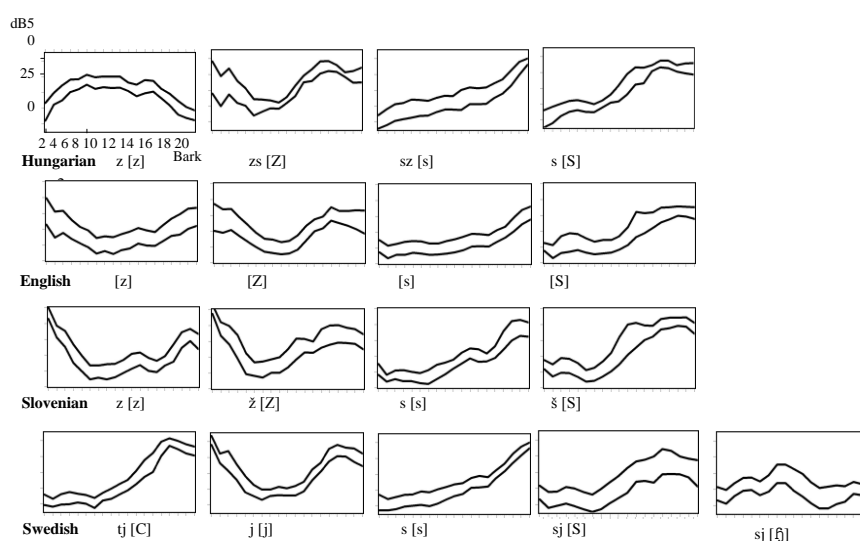
* A környező hangok képzési helye szerinti beszédhang változatok spektrális modelljeit elkészítve pontosabb modelleket lehetett volna létrehozni, de ehhez jóval nagyobb adatbázisra lett volna szükség. A jövőbeli fejlesztés célja mindenképpen a pontosabb modellek létrehozása, vagy az említett módon, vagy eleve nem beszédhang alapon, hanem hangkapcsolat modellekkel kell dolgozni.

9. Táblázat. A helyes kiejtésű magyar **i** beszédhang spektrális modelljének adatai folyamatos szövegben, férfiak, nők és gyermekek esetében

Szűrősorszám	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
										[dB]									
Férfi Min.	13	5.8	20	-6	-9	-11	-12	-13	-12	-12	-12	-8	-8	-11	-12	-11	-12	-13	-13
Férfi Átlag	38	35	28	17	8.3	3.9	1.4	0.6	0	2.2	8.3	15	16	13	12	7.7	0.6	-1	-1
Férfi Max.	49	49	46	40	28	29	26	25	19	28	33	32	34	35	33	29	22	16	17
Férfi Neg.	34	29	22	11	3.1	-2	-4	-4	-5	-3	0.9	8.8	9.4	6.4	6.1	2	-4	-5	-5
Férfi Poz.	42	39	33	23	14	9.6	6.9	5.8	5.3	8.5	16	21	21	18	18	13	5.8	4.4	4.2
Női Min.	19	12	8.8	6.4	-1	-10	-11	-12	-13	-14	-14	-13	-7	-7	-8	-12	-10	-13	-15
Női Átlag	39	36	34	26	15	8.5	4.7	2.8	0	1.1	3.1	7.2	15	16	10	9.6	6.6	-1	-1
Női Max.	50	49	47	41	32	32	25	22	20	20	24	33	34	36	32	31	28	19	17
Női Neg.	35	31	29	21	10	2	-1	-2	-5	-4	-2	-0	8.8	9.9	5	3.3	0.3	-5	-6
Női Poz.	43	41	39	31	20	15	10	8	5.3	6.6	9.6	15	21	21	17	16	13	5	3.3
Gyerek Min.	21	15	17	8.8	3.3	-1	-2	-3	-4	-4	-7	-5	-2	6.1	8.3	7.5	2	1.7	-2
Gyerek Átlag	38	31	33	30	21	14	7.2	7.5	6.4	5	3.6	7.2	12	21	23	21	18	10	7.5
Gyerek Max.	48	40	43	43	35	22	19	21	18	17	16	21	29	32	34	32	29	22	17
Gyerek Neg.	32	24	27	24	15	5.8	3.1	2.8	1.4	1.1	-1	2.8	7.7	16	18	14	12	6.1	1.4
Gyerek Poz.	43	35	39	35	25	18	12	12	9.4	8.5	8.8	13	21	26	28	25	23	15	12

Angol, svéd és szlovén nyelvre a partnerkutatók ugyanezen eljárás első fázisait - a szubjektív akusztikai lehallgatásokat, a beszédhangok szegmentálását - hajtották végre, a spektrális statisztikai modellek meghatározása viszont a Beszédakusztikai laboratórium munkája. Példaként a sziszegő hangokra számolt átlagos eltérés értékeket a négy különböző nyelv esetében a 18. ábrán mutatom meg.

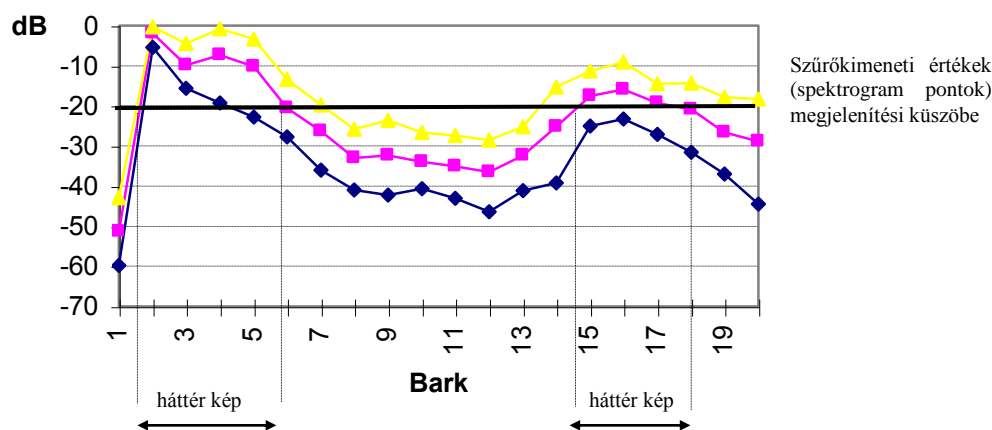
A rendszerben a **kitartott hangok** gyakorlására a modell átlagos spektrális eltérés adatait, mint referencia háttér képeket használjuk, a 13. b. és a 13. c. ábrákon közölt módon és a 3.2.3. Hallási spektrum c. fejezetben leírtak szerint.



18. ábra

A modellek átlagos spektrális eltérései spektrum típusú megjelenítésben, 4 nyelvre sziszegő beszédhangok esetében

Szavakban, mondatokban történő **folyamatos beszéd** gyakorláskor a háttér képeket a spektrális átlagból származtattam. Egy keret példaképpen színekpi jellemzőit a 19. ábra mutatja. Azokra a spektrális területekre (Bark sávokra) hívom fel rajzosan a figyelmet, ahol az adott jó kiejtésű fonéma modell spektrális átlaga az alkalmasan választott megjelenítési küszöbnél (a 19. ábrán -20 dB) nagyobb intenzitásszint értékeket vesz fel.



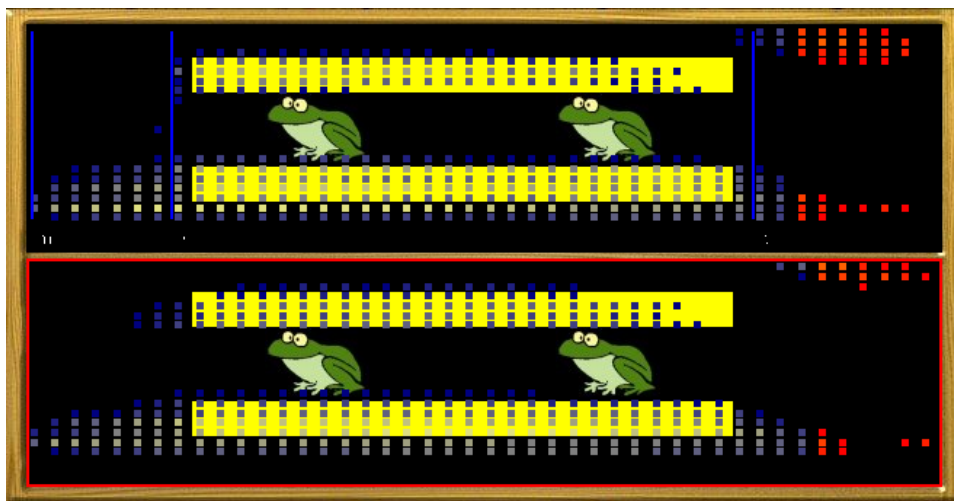
19. ábra Háttér képek berajzolási helyei a hallási spektrogram típusú megjelenítési formában.

Az é beszédhang spektrális átlaga (középső görbe), és átlagos eltérése (szélső görbék).

A 20. ábrán mutatom be egyrészt a beszéd környezetben felhangzó, hosszan tartott é hang hallási spektrogramjából a megjelenítési küszöb fölé eső részeket. Másrészt, világos színnel feltüntettem a megjelenítési küszöbnél nagyobb színekpi területeket, amelyek együttesen világos téglalappá állnak össze, így képezvén a háttér képeket. A háttér kép a magánhangzóknál a megvalósított programban színes téglalap. A spektrális súlypontok helyét a hallási spektrogram típusú hangképben magánhangzóknál ezekkel a téglalapokkal emelem ki. Harmadrészt, állatfigurákat helyezek el a spektrogramban ott, ahol jellemzően a „szűrőkimenetek” intenzitásszint értékei a megjelenítési küszöb alá esnek. A paciensnek el kell kerülni azt, hogy az állatfigurákra az itt ábrázolt hallási spektrogram pontjai ráessenek.

A háttér képek alkalmazása segíti a hangképek helyes értelmezését az oktatás és gyakorlás során. Gyermek számára érdekessé teszi a mért beszédparaméterek ábrázolását. Segíti az önálló gyakorlást, mivel útmutatást ad a gyakorló gyermek számára.

A háttér képpel és állatfigurákkal kiegészített hallási spektrogram típusú ábrázolás használhatósága döntő mértékben a megjelenítési küszöb helyes megválasztásán múlik. Ez a megválasztás sok kísérleti munkával volt lehetséges. Megjegyzem, hogy a megjelenítés a megjelenített beszéd részlet hallási spektrogramjának intenzitásszint maximumát értelmezi 0 dB-nek (ezzel normálást érünk el).



20. ábra

Az é hang hallási spektrogram típusú megjelenítése a háttér képpel és állatfigurákkal. Magánhangzóknál a háttér képek (téglalapok) területén kell a legtöbb spektrumpontnak megjelennie.

A háttér képek lényegkiemelő szerepének helyességét a 3.4. fejezet kvalitatív vizsgálatával, valamint a 9 fejezetben leírt érthetőség vizsgálattal bizonyítom.

5.3. Az életkor szerepe a fonémák kiejtésének fejlődésében

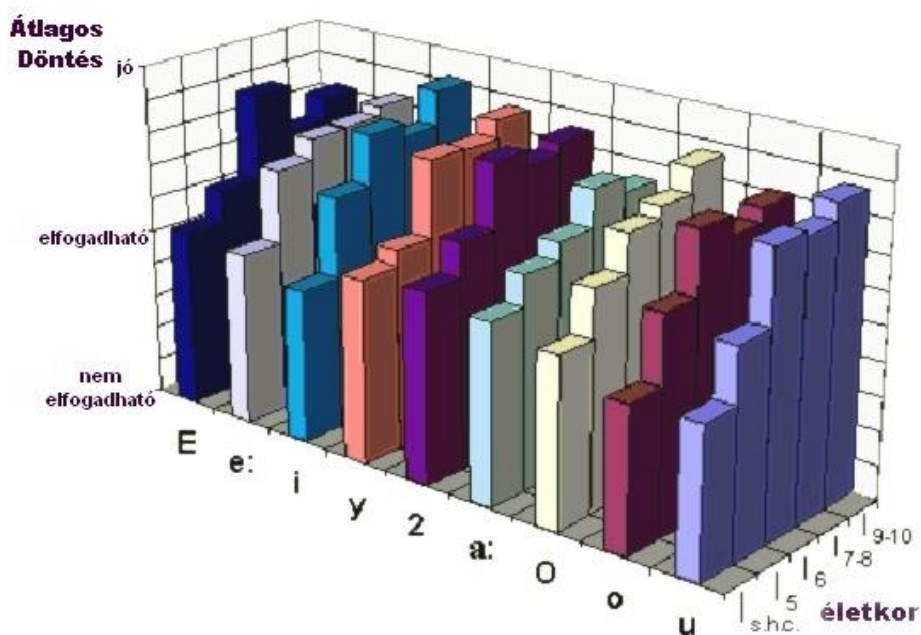
E kutatási szakasz célja az volt, hogy meghatározzam a fonémák kialakulását 5-10 éves korban. Az 5-10 éves életkor ugyanis az, amelyre az oktatási rendszeremet optimalizáltam, tehát tudnunk kellett azt, hogy mit várhatunk a gyermekektől a terápia során.

A vizsgálathoz új kutatási módszert dolgoztam ki. Az új módszert alkalmazva megállapítottam, hogy a gyermeki beszédfejlődés megfelelően kialakított adatbázisok segítségével jól meghatározható, az adatbázis elemeinek szubjektív megítélésével.

A gyermekbeszédre akusztikailag jellemző, hogy a kiejtéskor képzett paraméterek még nem stabilak, változnak, formálódnak (Lee, et al., 1999). Szórásértékük is nagyobb, mint a felnőtteké, akiknek már rögzült artikulációs mozgása van. Gyermekbeszéd kialakulására magyar nyelven korábban Gósy végzett kutatásokat, de ő a korai 2 év alatti fejlődési szakaszt vizsgálta, egyéni megfigyelésekkel (Gósy, 1981), valamint Kassai a magyar gyermeknyelvi dallamot (Kassai, 1979) és a hangsúly kialakulását vizsgálta (Kassai, 1981 a).

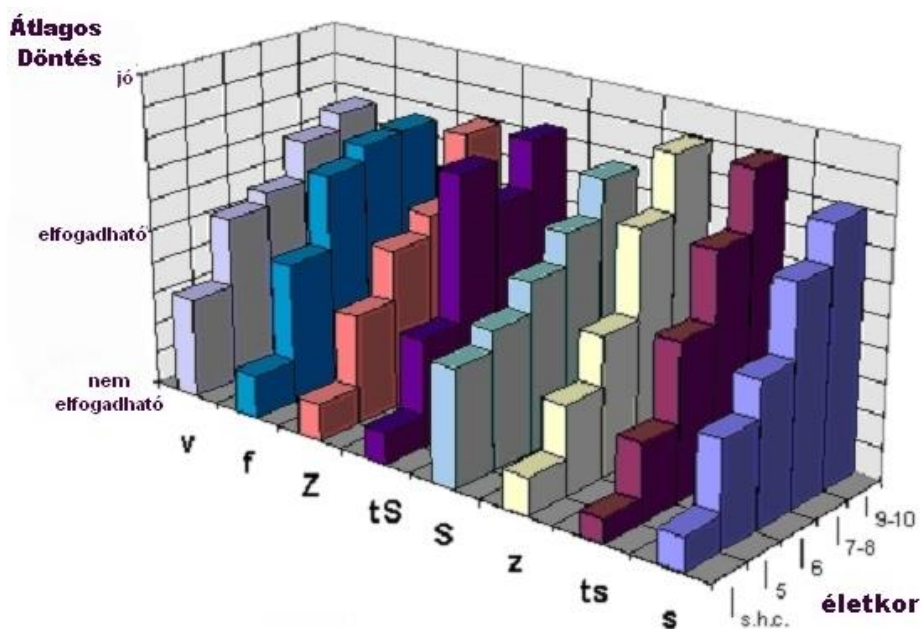
A gyermekbeszéd adatbázis beszéd elemeinek szubjektív megítélése után életkor függvényében vizsgálható a gyerekek által kialakított hangok minősége. Az 5.1. fejezetben leírt módon kiértékeljük a különböző fonémák kiejtésének minőségét a gyermekek életkora

szerint. Az eredményt az egyes magánhangzókra a 21. ábra, az egyes mássalhangzókra a 22. ábra, a mássalhangzók átlagára a 23. ábra, míg a magánhangzók átlagára a 24. ábra mutatja.



21. ábra.

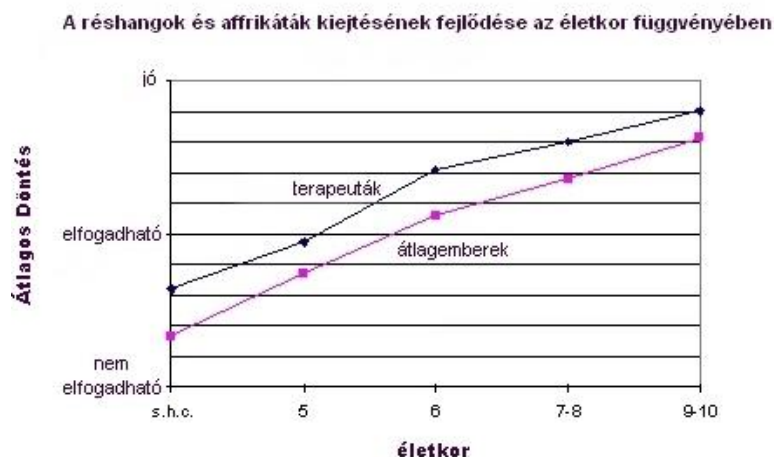
A magánhangzók kiejtésének fejlődése az életkor függvényében. (Az s.h.c. az adatbázisban szereplő beszédhibás gyermekek mintáit jelenti.) A beszédhangok jelölésére SAMPA szimbólumokat használunk az ábrán



22. ábra

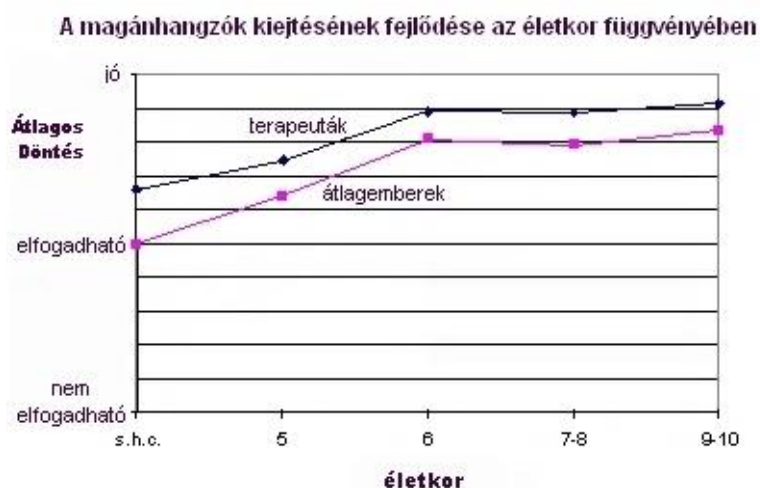
A réshangok és affrikáták kiejtésének fejlődése az életkor függvényében. (Az s.h.c. az adatbázisban szereplő beszédhibás gyermekek mintáit jelenti.) A beszédhangok jelölésére SAMPA szimbólumokat használunk az ábrán

A magánhangzók határozott kiejtése korábban alakul ki, mint a mássalhangzóké. A 5.1. fejezetben leírt hallgatói teszteredmények azt mutatták, hogy a magánhangzók minősége 5 éves korban nem találtatott elég tökéletesnek, ellenben a 6 éveseknél a magánhangzókat határozottan jónak ítélték. Ezt mutatja a 21. és 24. ábra.



23. ábra.

A réshangok és affrikáták kiejtésének fejlődése az életkor függvényében. A két görbe közül a felső a terapeuták átlagos véleményét mutatja, az alsó pedig az átlagemberekét. (Az s.h.c. az adatbázisban szereplő beszédhibás gyermekek mintáiról alkotott vélemény.)



24. ábra

A magánhangzók kiejtésének fejlődése az életkor függvényében. A két görbe közül a felső a terapeuták átlagos véleményét mutatja, az alsó pedig az átlagemberekét. (Az s.h.c. az adatbázisban szereplő beszédhibás gyermekek mintáiról alkotott vélemény)

Érdeemes megemlíteni, hogy a beszédterapeuták sokkal engedékenyebbek voltak az érthetőség megítélésénél, mint a nem szakértők, amint azt a 23. és 24. ábrák mutatják.

Azonban nekem a nem szakértők döntését kellett figyelembe vennem, hiszen a gyermekeknek a mindennapi életben is általában beszédterápiához nem értő emberekkel kell magukat megértetniük.

A gyermeki beszédfejlődés vizsgálata során kapott összefüggések új tudományos eredmények.

6. A kiejtés jóságának automatikus megítélése

Gyakorlaskor a páciens először a pedagógus segítségével, majd önállóan dönti el, hogy a gyakorlandó szótagot, szót helyesen ejtette-e ki, vagy nem. A kidolgozott rendszerben lehetőség van automatikus értékelésre, amelyet automatikus visszajelzésnek nevezünk. Ez az automatikus visszajelzés segíti a gyakorlót kimondása helyességének megítélésében. Önálló gyakorlásnál ez a funkció sokat segít. Hallássérült gyermekeknél lélektani jelentősége van az automatikus visszajelzésnek, hiszen nincs rászorulva a gyermek egy külső személy segítségére ahhoz, hogy megfelelő visszajelzést kapjon.

A rendszertani leírásban, tanfolyamokon azonban erősen hangsúlyozom a beszédterapeuták részére, hogy nagyon vigyázzanak arra, hogy a gyakorló a hangképet figyelje, az az elsődleges visszacsatolás, ne pedig az automata visszajelzést. A visszajelzés ugyanis csak annyit mond, hogy jó, vagy nem jó a kiejtés.

A hangképre figyelve viszont azonnal látja mi a rossz, például ha túl hangosan beszél, csökkenteni kell a hangerőt, vagy a gyakorlás során megtanulja látni, hogy nem képez zöngét a zöngés réshangoknál, vagy nem áramoltat elég levegőt, vagy pl. magánhangzóknál nem eléggé nyitott az állkapcsa az é hangnál, vagy nem eléggé kerekít az u hangnál stb. Gyakorlás során ugyanis a páciens meg kell hogy tanulja a hangképek értelmezését. Közös gyakorlásoknál akár ki is kapcsolhatjuk az automata visszajelzést, hogy ne vonja el a tanuló figyelmét a hangképről.

6.1. Spektrális távolságszámító eljárás kiválasztása

A beszédfejlesztés során használunk olyan gyakorlatokat, amelyekben a gyakorlandó mintát a tárolt statisztikai modellel, vagy a referencia mintával kell összehasonlítani. A páciens ezt az összehasonlítást hallás, látás és tapintás (a beszédképző szervek, például a nyelv helyzetének belső érzékelése) együttesével teszi meg a kialakított rendszerben. Különösen önálló gyakorlaskor segítheti őt az értékelésben, ha egy jól működő automatikus visszajelzést is alkalmazunk. A tanuló kiejtését ekkor egy ötfokozatú skáláján értékelem, azaz megadom, hogy a kiejtése milyen messze van a szép kiejtésből származtatott modelltől.

A vizsgálat célja egy olyan kiejtés helyességét értékelő automatikus módszer kidolgozása, amely segítséget nyújt az érthető szépbeszéd kialakulásához. Módszerűl spektrális távolságszámító eljárásokat hasonlítottam össze, ahol a viszonyítási alapok a fonémák helyes kiejtésű spektrális modelljei voltak. Vizsgáltam, hogy az aktuális kiejtés beszédhangjainak paraméterei milyen távolságra vannak a megfelelő fonéma statisztikus színeképi modelljétől.

A vizuális megjelenítésnél a referencia bemondáshoz dinamikus idővetemítéssel illeszttem az aktuális bemondást, és mivel a referencia bemondásban a beszédhangokat és azok határait már korábban bejelöltem, az illesztésnél az aktuális bemondásba is bekerülnek a beszédhanghatárok. Távolságértékeket célszerű beszédhangonként megadni, ha az egyes beszédhangok kiejtésének a jóságát akarjuk automatikusan megítélni. A távolság számításnál a bemondásokból is elhagytam a beszédhangkezdő és záró 2-2 időkeretet.

Több távolságszámítási eljárást próbáltam (Vicsi, 1999), amelyek közül a legjobbnak a módosított négyzetes (euklidészi) spektrális eltérés, a második legjobbnak a módosított spektrális eltérést találtam.

6.1.1. Módosított spektrális távolság

Mivel a modellt helyes ejtésű mintákból építettük fel, kézenfekvő volt a minták szélsőértékét, mint határt felvennünk, melyen belül eső távolságérték 0, így jutunk az egylépcsős módosított spektrális távolsághoz. Mindezt figyelembe véve az m jelű fonéma statisztikai modellje és ugyanezen fonéma egy aktuális kiejtése közötti távolság kiszámításához értelmezzük az aktuális kiejtés i -ik időkerete és a modell közötti távolságot:

$$d_{if}(m) = \begin{cases} (b_{if} - a_{\max f}) & \text{ha } b_{if} > a_{\max f} \\ (a_{\min f} - b_{if}) & \text{ha } b_{if} < a_{\min f} \\ 0 & \text{ha } d_{\min} \leq b_{if} \leq a_{\max f} \end{cases} \quad (12)$$

ahol, $a_{\max f}$ az a_{\max} vektor f -edik eleme,

b_{if} az aktuális kiejtés i -edik időkerethez tartozó hallási spektrum f -edik eleme.

Ezzel az m jelű fonémamodell és az aktuális bemondás távolságok (elég hosszú aktuális bemondás esetén)

$$D(m) = \frac{1}{I-4} \sum_{f=1}^F \sum_{i=3}^{I-2} d_{if}(m) \quad (13)$$

ahol I az aktuális kiejtés időkereteinek száma.

A számolt eredmények alapján ez a távolságszámítás nem volt elég érzékeny, amint az adatbázis szavaiban, mondataiban lévő magánhangzókra és sziszegőkre kapott 4. sz. melléklet 1. Táblázatában látható. Célszerűnek látszott egy töblépcsős távolságszámítási eljárás kialakítása.

6.1.2. Módosított spektrális négyzetes távolság

Olyan lépcsőhatárokat, elemi távolság értelmezés és lépcső-súlyozást kerestünk, amely a „módosított szinképi távolság” hiányosságait jelentősen enyhíti. Így jutunk a többlépcsős „módosított szinképi négyzetes távolság”-hoz. Az átlagos spektrális eltérésnél és annál kisebb eltérésekre a távolságérték 0. az átlagos spektrális eltérés és a szélsőértékek között az átlagtól való eltérés négyzetével számolunk, ennél nagyobb értékeknél súlyozott négyzetes eltérést számolunk a (14) képletnek megfelelően.

$$d_{if}(m) = \begin{cases} w \cdot (b_{if} - a_f)^2 & \text{ha } b_{if} > a_{\max f} \\ w \cdot (a_f - b_{if})^2 & \text{ha } b_{if} < a_{\min f} \\ (b_{if} - a_f)^2 & \text{ha } a_{\text{poz}} \leq b_{if} \leq a_{\max f} \\ (a_f - b_{if})^2 & \text{ha } a_{\min} \leq b_{if} \leq a_{\text{negf}} \\ 0 & \text{ha } a_{\text{negf}} < b_{if} < a_{\max f} \end{cases} \quad (14)$$

ahol, $a_{\text{neg } f}$ az a_{neg} , míg $a_{\text{poz } f}$ az a_{poz} vektor f -edik eleme,

b_{if} az aktuális kiejtés i -edik időkerethez tartozó hallási spektrum f -edik eleme.

w = súlyozó tényező, számításaimban $w = 100$.

Ezzel a módosított spektrális négyzetes távolság (15)

$$D^{(2)}(m) = \frac{1}{I-4} \sum_{f=1}^F \sum_{i=3}^{I-2} d^{(2)}_{if}(m)$$

Fentieket szó és mondat szintre általánosítva célszerű még a szómodell, illetve a mondatmodell és az aktuális szó- vagy mondat közötti távolság értelmezése:

$$D^{(2)}(\text{szó}) = \frac{1}{M} \sum_{m=1}^M D^{(2)}(m) \quad (16)$$

$$D^{(2)}(\text{mondat}) = \frac{1}{L} \sum_{l=1}^L D^{(2)}(\text{szó})$$

ahol M a fonémák száma a szóban, L a szavak száma a mondatban. A kapott távolságértékek a 4. sz. melléklet 2. Táblázatban találhatóak.

A távolságszámító eljárások összehasonlítására vizsgáltuk a fonémák közötti és a fonémákon belüli távolságarányt (R). Ezeket egy referencia fonémahalmaz és egy teszt fonémahalmaz között értelmezzük

$$R(m, n) = \frac{\sum_{f=1}^F \left[\left(a_{\text{poz}f}^{(m)} - b_{\text{poz}f}^{(n)} \right)^2 + \left(a_{\text{neg}f}^{(m)} - b_{\text{neg}f}^{(n)} \right)^2 \right]}{\sum_{f=1}^F \left[\left(a_{\text{poz}f}^{(m)} - b_{\text{poz}f}^{(m)} \right) + \left(a_{\text{neg}f}^{(m)} - b_{\text{neg}f}^{(m)} \right) \right]^2} \quad (17)$$

ahol az a referenciamodell, b a tesztmodell megfelelő jellemzője
 m és n a fonémák jelei.

Olyan módszert kerestünk, ahol a fonémák közötti távolság minél nagyobb, a fonémákon belüli minél kisebb, tehát az R minél nagyobb.

Összehasonlításként elemezzük a 4. sz. melléklet 3., 4. táblázatokat, ahol magánhangzókra és sziszegő hangokra vizsgáltuk a távolságarányokat. A fonémák sokkal jobban szétválnak a módosított négyzetes távolságszámítás esetében, mint az egyszerűbb spektrális távolságnál. Persze itt is vannak részleges átfedések, de ez természetes is, a közeli színeképű beszédhangoknál.

6.1.3. A kész rendszer, a Beszédkorrektor tesztelése

A kész rendszerbe a módosított négyzetes távolságszámítást alkalmaztam, ahol végeredményben a páciens kiejtésének helyességét 5 fokozatban ítélt meg (választhatóan osztályzattal, vagy egy kacska mozgásával, vagy virág kinyílásával). Ennek az automatikus visszajelzésnek a magyar nyelvű vizsgálatához 20 gyakorlót kértünk fel, mondjon be szavakat a Beszédkorrektor szótárából, gyakorlatonként egyet-egyet helyesen, majd úgy hogy a vizsgált beszédhangot kicseréli először egy fonetikailag távolálló, majd egy közelálló hangra és lejegyeztük a visszajelzéseket.

Például az é magánhangzó esetében közeli beszédhang az í hang, távoli az á (szél, szil, szál). Sziszegő hangoknál sz hang esetén közeli az sz hang zöngés párja a z, távoli az s (szár, sár, zár).

Minden beszédhangot 3x12 szóban vizsgáltunk. A magánhangzókra és a sziszegőkre a rendszer által adott osztályzatokat a 10., 11. Táblázat mutatja.

10. Táblázat. A kész rendszer automatikus visszajelzésének tesztelése magyar magánhangzókra

Magánhangzók:					
azonos		közeli		távoli	
osztályzat	%	osztályzat	%	osztályzat	%
5	79	5	4	5	0
4	12	4	8	4	0
3	8	3	12	3	0
2	1	2	32	2	1
1	0	1	44	1	99

Magánhangzók esetén 4-es vagy 5-ös osztályzatot kapott minden saját osztályába eső beszédhang az esetek 90%-ában. Egy távoli magánhangzót 100%-ban utasít vissza 1-es, 2-es osztályzat együttesével. A közeli magánhangzókat 76%-ban ítéli 1-es, 2-esnek. A közepes ítéletet is beszámítva 88%-ban ad visszautasító ítéletet.

11. Táblázat. A kész rendszer automatikus visszajelzésének tesztelése magyar sziszegőkre

Sziszegők:					
azonos		közeli		távoli	
osztályzat	%	osztályzat	%	osztályzat	%
5	93	5	5	5	0
4	4	4	7	4	0
3	3	3	7	3	1
2	0	2	24	2	3
1	0	1	57	1	96

Sziszegőknél még jobb eredményt is kaptunk, mint a magánhangzóknál.

Összegezve megállapítható, hogy a módosított négyzetes távolságszámításon alapuló eljárás jó az automatikus kiejtés értékelés számára. Az értékelés a gyakorlatban jól alkalmazható.

7. A kezelt beszédhangok akusztikai tulajdonságai és képi megjelenítésük

A terápia módszertanát a magánhangzókra és a sziszegő hangokra dolgoztam ki teljes részletességgel, persze a szavak, szópárok, mondatok gyakorlásánál minden beszédhang előfordul. Az angol, svéd, szlovén rendszer fejlesztésekor, a külföldi kollégákkal együtt, terapeuták bevonásával állítottuk össze a módszertant az adott nyelv és a nemzeti oktatási sajátosságok figyelembe vételével. Rövid, a feladathoz illeszkedő akusztikai leírása ennek a két fonetikai csoportnak minden nyelvre elkészült. A dolgozatban a magyar nyelvű összefoglalást közlöm, mivel minden partner ezt az összefoglalót az anyanyelvén készítette el.

Az alaphang, (a zöngé) férfiak, nők és gyermekek esetében erősen különbözik, de jellemzően a hallásmező alsó részébe 80 és 500 Hz között (az első 4 hallási sávban) esik. A magánhangzók, a rezonáns jellegű mássalhangzók jellemző spektrális összetevői megközelítőleg 250 Hz és 3400 Hz között (a 2.-tól a 15.-ig hallási sávban), a sziszegőké jellemzően 3000 és 8000 Hz között (a 15.-től a 20.-ig hallási sávokban) vannak.

8.1. A magyar sziszegő hangok akusztikai leírása

A szigmatizmus, amely az sz, z, s, zs spiránsok és a c, cs, dz, dzs affrikáták, összefoglaló néven: a *sziszegők* helytelen képzését jelöli, egyik leggyakoribb beszédhiba a halló emberek között is, hallássérült gyermekeknél pedig szinte minden esetben előfordul.

A sziszegőknél a hallássérült gyermekek értékelési és kiejtési nehézségeit éppen az indokolja, hogy a hallásveszteség leggyakrabban a hallásmező felső, éppen a sziszegőkre jellemző frekvenciatartományban a leggyakoribb.

Korábban a réshangokat, affrikátákat Szende, Olasz, Kassai vizsgálták (Szende, 1976), (Olasz, 1989), (Kassai, 1998). A kezelt sziszegők akusztikai jellemzőinek Bark-sávokban általam történt vizsgálatának rövid összefoglalása a következő:

Az sz hang jellemző frekvenciasávja 6000-8000 Hz között van. Helyes ejtésékor a hangot hallási frekvenciasávban elemezve a 18., 19., 20. szűrőbe esik az energia nagy része. Mivel a hallássérültek többségénél a magas frekvenciák területén jellemző a halláshiány, így a magas frekvenciákon hallható hangokat gyakran még erősítővel sem képesek differenciálni. Hangossága kisebb, mint a környező magánhangzóké.

A z hang az sz zöngés párja, amely viszont zöngéssége következtében jobban hallható, jól erősíthető. A zöngé jelenléte miatt jellemzően az 1., 2., illetve 3. hallási sávba (Bark szűrősávba) esik energia, valamint a magas összetevőjű spiráns zöreje miatt, a 19. és/vagy a 20. sávba. Hangossága nagyobb, mint az sz hangé, de a magánhangzók hangosságát nem éri el.

Az **sz** hang jellemző frekvenciasávja 3000-6000 Hz között van. Helyes ejtésekor hallási frekvenciasávokban elemezve a 14.-19. sávokba esik az energia nagy része. Helyes ejtésnél hangossága általában kisebb, mint a környező magánhangzóké, de több energiát hordoz mint az **sz** hang.

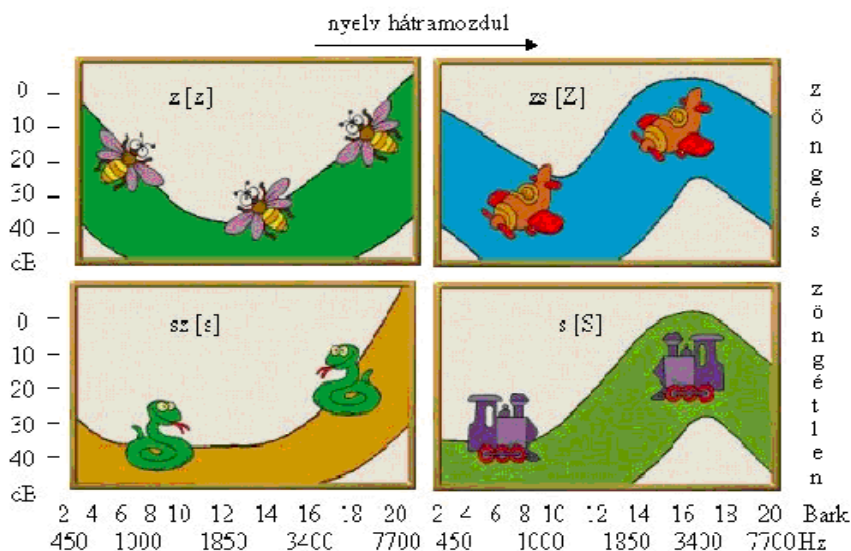
A **zs** hang az **sz** hang zöngés párja. Zöngéssége miatt a 2., illetve 3. hallási sávba esik energia, a képzési helyére jellemző frekvenciasávja pedig szintén 3000 és 6000 Hz között helyezkedik el.

A **cs** hang összetett hang. A hang első felét, a zárszakaszt, energiamentes rész jellemzi. Az ezt követő rész, a zár felpattanása, és a kiáramló levegő által okozott zörej, amelynek frekvenciasávja szintén magasan, 6000-8000 Hz között helyezkedik el. Hallássérülteknél ez a hang alig erősíthető, nehezen hallható.

A **cs** hang összetett hang. A hang első felét, a zárszakaszt, szintén energiamentes rész jellemzi, mint a **cs** hangnál. Az ezt követő részre a zár felpattanása, és a kiáramló levegő által okozott zörej jellemző. A zörej spektrális összetevői alacsonyabb frekvenciasávban helyezkednek el, mint a **cs**-nél, 3000-6000 Hz között.

A hallási spektrum megjelenítése

A vizsgált sziszegő hangok spektrális összetevőjének alakulása egy adott, közel állandósult állapotú szakaszban a 25. ábrán látható. Az ábrákon a spektrális összetevők intenzitásának a hallgatók által elfogadott ingadozási sávja van feltüntetve. Helyes ejtés esetén a szűrő kimeneti energia értékeinek ezen a sávon belül kell maradni.



25. ábra

Spektrális modellek max és min eltérései (felső görbe: max, alsó görbe: min), mint háttér képek, hallási spektrum típusú megjelenítésben, magyar **sz**, **z**, **s**, **zs** hangokra

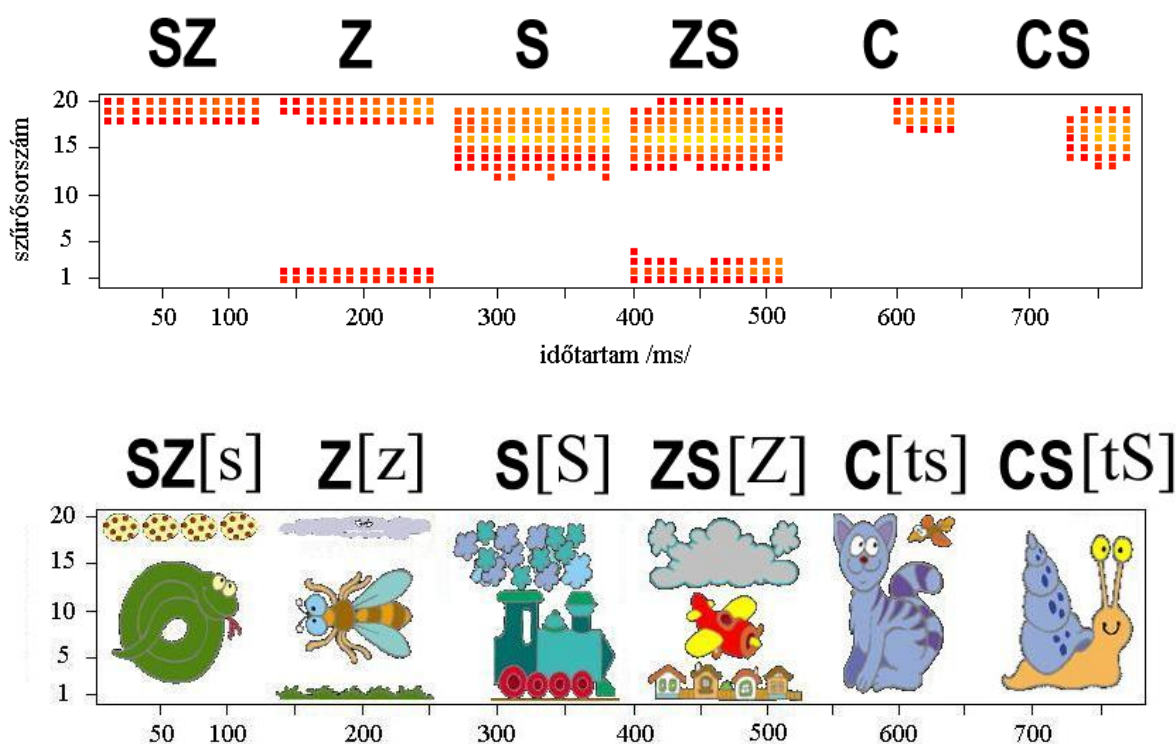
A hallási spektrogram megjelenítése

A vizsgált hangok esetében az egyes szűrősávokban, a kvázi stacioner részben mért energiaértékeket a 26. ábrán bemutatott példaképpeni hallási spektrogram típusú ábrázolásban mutatjuk be. Az ábrán a minta spektrogram és azok értelmezését segítő háttérképek láthatók.

A sziszegő hangok közül a réshangok állandósult állapotú szakasza folyamatos, szépen artikulált, átlagos tempójú beszédben 100-300 msec közötti (Vicsi, 1981). Egy példa az **sz** hang rossz, laterális ejtésének a megjelenítése a 26.b. ábrán látható.

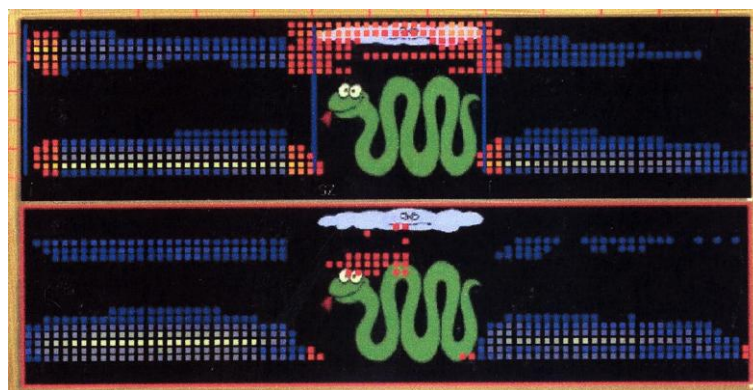
Az affrikáták zárfelpattanást követő szakaszának színe megegyezik az azonos képzési helyen keltett sziszegő hangok színeivel, csak időtartamuk kisebb, 50-100 msec.

A hangkörnyezet jelentősen befolyásolja a színeképet. Az **u** hang például a sziszegők jellegzetes energiamaximumait a mélyebb frekvenciák felé mozdítja, az **i** hang viszont felfelé, a magasabb frekvenciák irányába húzza el az energiamaximumot.



26.a. ábra

Spektrális modellek spektrogram típusú megjelenítése (fent) a megfelelő háttér képekkel (lent), magyar sziszegők esetén, folyamatos szöveg gyakorlásához



26.b. ábra

Az iszi hangkapcsolat

Fenn: referencia kiejtés, lenn sz hang hibás ejtésben, laterális képzés

7.2. A magyar magánhangzók akusztikai leírása

A magyar nyelvben 14 magánhangzó van. Amennyiben a hosszú és rövid párokat egy csoportba soroljuk, akkor kilenc csoportot különböztetünk meg. Ezek a következők (l. a 27. ábrát is):

i-í, ü-ű, u-ú,

é, ö-ő, o-ó,

e, á, a

Rendszerünkben a hosszú magánhangzókat együtt kezeljük a rövid párjukkal, mivel ebben az elemzési felbontásban spektrális eltérés nincs köztük, csak időtartamban

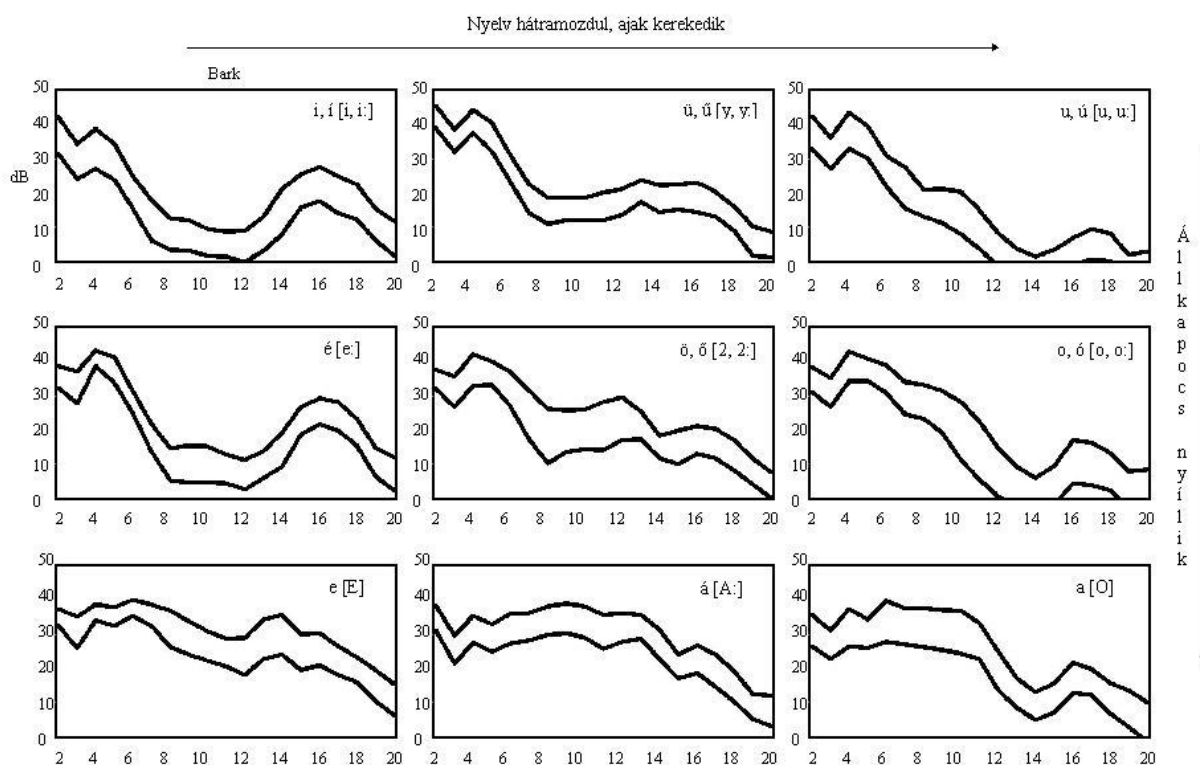
A magánhangzók hangképző szerveinek helyzete szerint több csoportot különböztetünk meg:

- a) Ajakműködés szerint
ajakréssel - ajakkerekítéssel képzett magánhangzók.
- b) A nyelv vízszintes mozgása szerint
elölképzett (palatális) és hátulképzett (veláris) magánhangzók
- c) A nyelv függőleges mozgása, az álkapocs nyitása szerint
alsó, középső és felső nyelvállású magánhangzók

Mind a három hangképző szerv a szájüreg alakját és ezen keresztül a kialakítandó hang színeképét befolyásolja (Tarnóczy, 1974).

Spektrális megjelenítés

A 27. ábrán a magánhangzók egy adott időpillanatban mért színképe látható, folyamatos beszédben a magánhangzó közel állandósult szakaszában.



27. ábra.

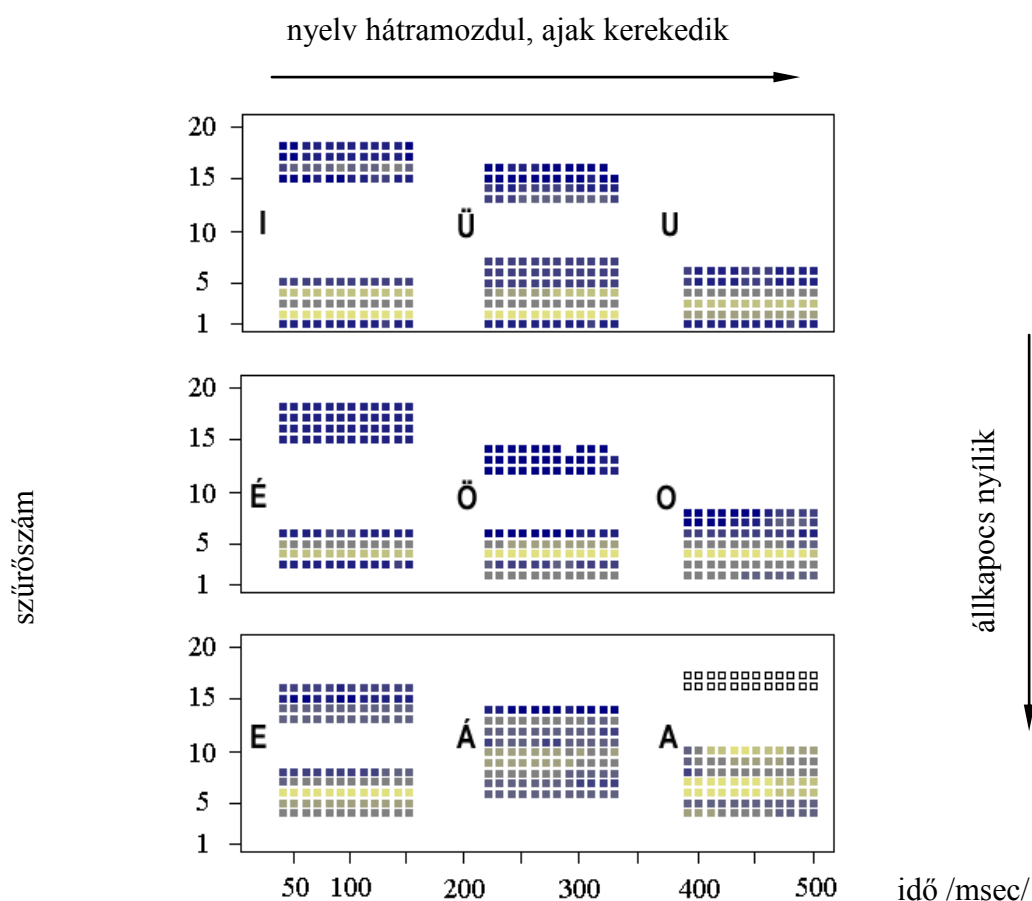
A magyar magánhangzók spektrális modelljei, az átlagos spektrális eltérés feltüntetésével (a_{poz} felső görbe; a_{neg} alsó görbe) (folyamatos szöveg kvázi stacionárius részeinél mért adatok átlagolása alapja).

A szórási sáv vonulata igen jellemző az adott magánhangzóra. A hang képzésekor ezen a tartományon belül kell maradni ahhoz, hogy a magánhangzót egy átlagos hallgató a megadott magánhangzónak hallja.

Az artikulációs szervek helyzetének változtatásával a kapott görbe ezekbe a sávokba jól illeszthető. Például a nyelv hátramozdításával (a nyelvhat velum felé való domborításával), ill. az ajkak kerekítésével a második energiamaximum eltolódik a kis frekvenciák irányába, végül egybeolvad az első energiamaximummal. Ez történik egy adott állkapocs nyílásnál pl. az **i-u** vagy az **é-o** átmenetben. Az állkapocs nyitása, a nyelv lefelé mozgatása, jellemzően az első energiamaximum helyét mozdtítja el a nagyobb frekvenciák felé, pl. **i-e** átmenetben.

Nemcsak az energiamaximumok jellemzik az egyes magánhangzókat, hanem az energiamaximumok és minimumok egymáshoz viszonyított nagysága, helyzete is. Az e és az é hang energiamaximum helyében például kicsi az eltérés, viszont a két maximum közötti minimumhely az i hangnál lényegesen mélyebben van (az energia itt jóval kisebb), mint az é hang esetében. Nem a konkrét energiamaximum értékek azok, amelyek kizárólag fontosak. Ezek egy bizonyos értékhatáron belül változnak a személyek, az alaphang, stb. függvényében. Lényeg a teljes összkép. (pl. a maximumok, minimumok egymáshoz viszonyított aránya.)

A 28. ábrán a magánhangzók példaképpen hallási spektrogram típusú megjelenítései láthatók. Abban az esetben, ha a nyelv hátramozdul, és/vagy az ajak kerekedik, a második energia maximum lefele tolódik. Az u, o és á hangoknál az elsővel egybe is olvad. Az állkapocs nyitásával az első energia maximum emelkedik.

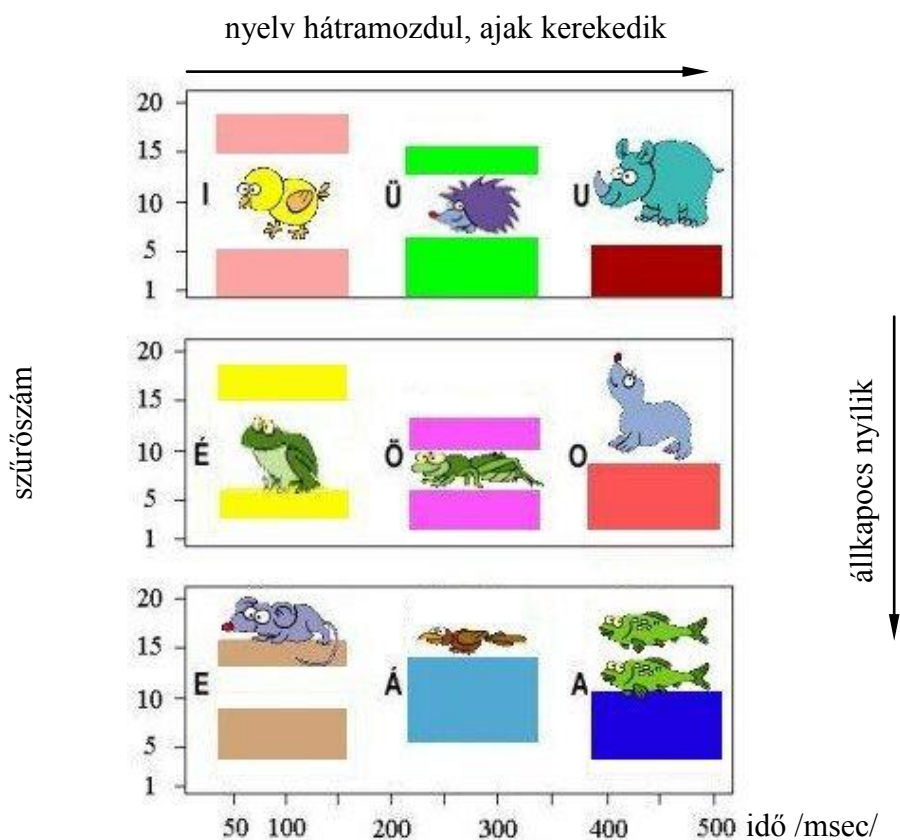


28.a. ábra

Példaképpen magyar magánhangzók hallási spektrogram típusú ábrázolásai

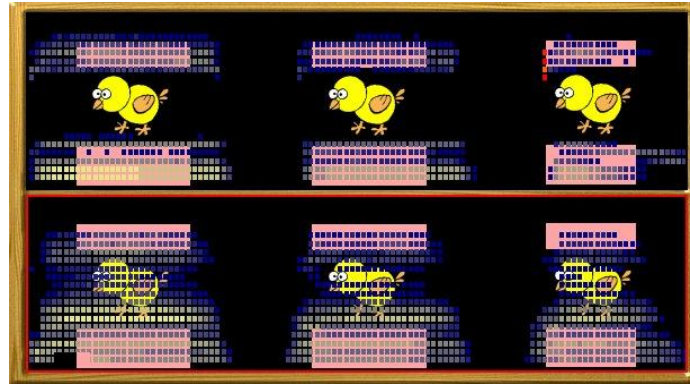
A hallási spektrogram típusú megjelenítés egyik fontos vonása, hogy egy alkalmasan választott megjelenítési küszöb alatti részletek elhagyásával lehetőséget ad a lényegre való figyelésre, másrészt a szokásos frekvencia-tengely Bark-sávokban skálázott tengelyre való kicserélésével szubjektíven jobban értékelhető megjelenítést ad, mint a lineáris vagy logaritmikus frekvenciaskála.

A többnyelvű fejlesztéskor általános módszerként vezettem be, hogy a kísérleteket mindig először magyar nyelvre végeztük el, és a már kidolgozott módszert adaptáltuk a partnerek nyelvére. Végül is 2001 szeptemberére elkészült a magyar, angol, svéd, és szlovén változat. A



28.b. ábra

A magyar magánhangzók jellegzetes háttérképei hallási spektrogram típusú megjelenítésben, folyamatos beszédnél



28.c. ábra

A pipipi hangkapcsolat hallási spektrogram típusú megjelenítése és háttérképei
fenn: referencia kiejtés,
lenn: az i hang képzésekor hibás ejtés, túlságosan nyitott álkapocs

kész programok demo változatát mind a 4 nyelvre CD-n a dolgozathoz mellékeljük. A partner országokban a terapeuták örömmel fogadták a rendszert. Hatékonyság vizsgálatot minden nyelvre végeztünk, de a magyar nyelv esetében a vizsgálat a Beszédkorrektorral végzett oktatásra fél évig folyt, míg a partnereknél csak három hónapig.

8. Beszédoktatási módszertan

A beszédterápia általános módszertana gyermekek és felnőttek kezelésére is alkalmas. A játékos háttér képek kikapcsolásával a rendszer felnőttek számára is jól használható, mivel azonban a helyes beszéd kialakítását minél előbb el kell kezdeni, jellemzően még beiskolázás előtt, ezért a megjelenítésnél, a gyakorlatok kialakításánál, a szótárkészlet összeállításánál elsősorban az 5-7 éves, szellemileg egészséges gyermekek átlagos teljesítményszintjéhez alkalmazkodtam. Ebben a korban még ki lehet alakítani stabil fiziológiás szenzomotoros csatolásokat, így a megtanult, helyes hangadás, artikuláció, hanglejtés stabilan megmaradhat.

Részletes oktatási módszert dolgoztam ki vezető logopédus, szurdopedagógus szakemberekkel a hangkiadás fejlesztésétől a folyamatos beszéd kialakításáig. A mérési, számolási algoritmusok a legújabb technológiára épülnek, de az oktatási módszer kidolgozásánál figyelembe vettük a hagyományos beszédterápia már bevált módszereit (Vassné Kovács E. 1974, 1983.), (Csányi, 1990), (Kassai I., 1998), (Farkas, 1996). A többnyelvű rendszereknél pedig a nemzeti sajátosságokat. A tanulás során állandó gyakorlásra, ismétlésre van szükség. Minden gyakorlandó hang izolált formában, szótagokban, szavakban, mondatokban és minimál párokban szerepel különböző helyzetben és különböző hangkapcsolatokban. Az elegendő és változatos gyakorlás érdekében nagy szótárkészletet használtam (ld. 12. Táblázat). Minden gyakorló lépésnél, minden pozícióban legalább 10 elem (mondat, szó, szótag) szerepel. A gyakorlásnál először megjelenik a szótag vagy szó, mondat betűsora, elhangzik a bemozdandó elem, megjelenik szinkronban a hangkép. Tehát a betűk, és a hallott, látott információ együttesen jelenik meg.

A gyakorlatok felépítésénél mindig a legkönnyebb kiejtéstől haladtunk az egyre nehezebb felé. A gyakorlatok sorrendje is egymásra épülve az egyre bonyolultabb szöveg kiejtése felé halad. A végcél a szép folyamatos beszéd automatikus kiejtésének az elsajátítása. Külön gyakorlatsorozatot terveztem egy-egy szupraszegmentális beszédjellemező (a hangerő, a ritmus, és a hanglejtés) helyes kialakítására is.

Végig minden nyelven nagy figyelmet fordítottam a betűk, hívóképek (beszédterapeuták által használt, az egyes fonémáknak megfelelő rajzos szimbólumok), háttérképek összhangjának megteremtésére. A fejlesztés előtt nyelvenként a fonetikus szakembereknek meg kellett tervezni, hogy milyen lépésekben milyen szótárkészlettel alakítják ki a programot, vagyis milyen lesz a program felépítése, a referencia adatbázisba milyen felvételek kerülnek. Meg kellett tervezni a hívóképeket, a háttérképeket. A teljes szöveg és képadatbázis rendszerterve mind a 4 nyelvre az 1. sz. mellékletben található: a. magyar, b. angol, c. svéd, és

d. szlovén nyelvre. A 4 nyelvre elkészített oktató programok demo változatát a dolgozathoz mellékelem.

8.1. A gyakorlatok sorrendjének jelentősége

A különböző beszédjellemzők gyakorlására változatos lehetőségeket dolgoztam ki, de itt is, mint a hagyományos terápia során, nagyon fontos a lépésről lépésre haladás. A könnyebb gyakorlatok felől kell haladni a nehezebbek felé. Erre fel kell hívni a terapeuta és a szülő figyelmét. A használt hangképekkel egymás után fokozatosan kell megismertetni a gyermeket, hiszen meg kell tanulnia a látottakat értelmezni, és miután ez az értelmezés automatikussá válik, akkor lehet már a hangképzésre összpontosítani. Persze ez a folyamat, ahogy haladunk a gyakorlatokkal automatikusan végbemegy, hiszen az egyszerű vizuális megjelenítéstől haladunk a bonyolultabb felé. A legösszetettebb képi megjelenítés a szavak, mondatok gyakorlásánál van, de mire ide érünk a terápia során, addigra semmi problémát már nem jelent a látott hangkép értelmezése. Például, amikor egy új hang kialakítása a feladat, célszerű a kitartott hangokkal indulni, majd a hangkapcsolatokkal folytatni. A szavak gyakorlása akkor célszerű, amikor a hangkapcsolatokban (szótagokban) már jól be lettek gyakorolva az adott szó hangjai, és a gyermekek jól megismerték a hang és a hanghoz tartozó hangkörnyezet hangképét, a háttérrel együtt.

8.2. Az előkészítés

Igen nagy szükség van az úgynevezett „előkészítő” gyakorlatokra egy terápiásorozat elkezdése előtt, de minden foglalkozás előtt is. Ezek az úgynevezett bemelegítő gyakorlatok.

Egyszerű, olyan általános beszédjellemzőket (szupraszegmentumokat) kell beállítani a páciensnél, amelyek az egyes beszédhangoktól függetlenek, viszont a természetes beszéd kialakításának alapparaméterei. Ilyenek a hangerő, a ritmus, a hangszín, az alaphang. Az előkészítés gyakorlataival szoktatjuk rá a páciens, hogy a képernyőre figyeljen. Fel kell hívni a figyelmüket arra, hogy amint játszanak a hangjukkal, pl. hosszabb vagy rövidebb ideig, vagy nagyobb, vagy kisebb energiával képeznek egy hangot, stb. azonnal változik a görbék alakja a képernyőn. Itt kell megszokniuk azt is, hogy az artikulációs szervek különböző beállítása különböző hangképet eredményez a képernyőn. A hangerő, a ritmus, a hangszín helyes beállítására meghatározott gyakorlatsorokat alakítottunk ki minden nyelvre (ld. 1. sz. melléklet). Persze a szabad gyakorlásnak is helye van.

Ritmusgyakorlatok

A természetes folyamatos beszédet figyelve azonnal feltűnik annak ritmikus tagoltsága. A ritmikus tagolódás legkisebb egysége a szótag. (A természetes szótagolás élettani eredetű jelenség, a beszédszervek működése során megnyilvánuló ritmus, amely önálló tartalommal nem rendelkezik, a gondolatok közlésében nem vesz részt, mégis minden nyelvben fellelhető). Alapesetben a hosszú és rövid szótagok szabályszerű változása adja a ritmust, de a ritmusképletet befolyásolja a hangsúly. Az a szótag mondható hangsúlyosnak, amelynek a magánhangzóját nagyobb nyomatékkal (rendszerint nagyobb hangerővel) ejtjük ki a környező szótagok magánhangzóihoz képest. A magyar nyelvre az a szabály érvényes, hogy a szavak első szótagja kap hangsúlyt. Angol nyelvben a hangsúly helye a szavakban a szóra jellemző. Hosszabb szavaknál különböző mellékhangsúlyok is jellemzőek.

Ritmikai gyakorlatokra jól használható a beszédintenzitás időbeli változásának kirajzolása. Ebben az ábrázolási módban például jól látható, hogy az egyik hang a másikhoz képest túlságosan megrövidül vagy meghosszabbodik.

A zöngégyakorlása

A helyes hangfekvés beállítására több szempontból is szükség van. Először is a természetes hangzás elérése miatt fontos, másodsorban a természetellenesen magas hangfekvés gyors hangkifáradást, majd esetleg biológiai elváltozást eredményezhet.

A hangfekvés beállítása köhécseles, ásítási hang és más spontán hangadási megnyilvánulások frekvenciájának megfigyelése és értékelése alapján történhet.

8.3. A beszédhang kialakítása

Beszédhang kialakításakor egy hang megformálását a helyes artikuláció beállításával célszerű kezdeni. A helyes artikuláció képét a programban bármikor be lehet hívni.

Artikulációs ábrák

Ebben a gyakorlatban megmutathatjuk a gyermeknek a kialakítandó hang esetében a hangképző szervek: a száj, a nyelv, a fogak stb. helyes elhelyezkedését.

Kitartott hang gyakorlása

A kitartott hang gyakorlásakor a hangot vagy önmagában ejtve lehet gyakoroltatni, vagy lassan ejtett hangkapcsolatokban.

Három különböző nehézségi fokú gyakorlási lehetőséget alakítottam ki. A leggyengébb fokozatot főleg az erősen hallássérült gyermekek kezdő gyakorlása számára készült. A harmadik fokozat az adott nyelvre jellemző szép kiejtést mutatja.

A kitartott hang gyakorlásánál is amilyen hamar csak lehet, össze kell kötni a gyakorlandó hangot más hangokkal és együtt kell hangoztatni.

Gyakorlás hangkapcsolatokban, szavakban, mondatokban

A gyakorlat kiválasztásakor a hangkapcsolat hallási spektrogramja megjelenik a képernyő felső részén. A kimondott szótag képe a képernyő alsó részén jelenik meg.

A gyakorlandó beszédhangot a hangkapcsolat, vagy szótag elején, belsejében, végén minden magánhangzóval összekötve gyakoroljuk az adott hangot, majd ismételve, a hívóképeken megadott szövegvariációkban. A hívóképek sorrendben először a könnyebb, majd az egyre nehezebb hangkapcsolatokat mutatják, amit célszerű követni, de a sorrend szabadon választható.



29. ábra

Magyar **zs** hang gyakorlása a **Józsi** szóban.

Fenn a referencia bemondás hallási spektrogramja, lenn a páciens általi bemondás.

A **zs** hang hallási spektrogramja mögötti háttér kép repülővel, felhővel, házsorral a spektrogram értelmezését segíti a gyermeknél. Helyes ejtés esetében a spektrum pontoknak a felhőt el kell fedniük, viszont a repülőt nem szabad elfedni. A házsor a zöngé spektrális helyét mutatja, amelynek nagy részét szintén el kell fedni.

Az **intenzitás** időbeli változása hangképének használatát szótagok, szavak, mondatok gyakorlásánál minden olyan esetben javasoljuk, ha a hangsúlyozás megvalósításában valamilyen problémát fedezünk fel (l. 11. ábra).

A **hanglejtés** tetszés szerint bármikor használható, szótagok, szavak, mondatok gyakorlásakor, de igazán jelentősége a mondatok gyakorlásában van (l. 12. ábra).

Gyakorlás szópárokban

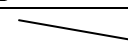

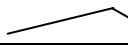
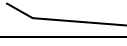
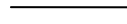

A szópárokban történő gyakorlás alkalmas arra, hogy felhívjuk a figyelmet a hasonló képzésű hangokban a meglévő különbségekre és egyezésekre, a hangképekben lévő különbségekre, majd ezután kezdjük el a gyakorlást.

Demonstrálni lehet a tanuló számára, hogy a beszédprodukción apró változásai jelentésváltozással járnak. Valódi értelmes szavakat állítottunk párba, hangsúlyozva a fonémák megkülönböztetésének beszélt nyelvi jelentőségét. Például azoknál, akik hajlamosak például **s** és **sz** hangokat felcserélni, olyan szópárokat választhatunk, ahol ezek a hangok vannak párba választva.

8.4 Hanglejtés

A hanglejtés a beszéd alaphangjának relatív változása. A Beszédkorrektorban viszonyítási alap az első szótag magánhangzójának hangmagassága. Az alaphang mérésről, hanglejtés megjelenítéséről már a 3. fejezetben beszéltem.

Táblázat 8: Leggyakoribb hanglejtéstípusok

Hanglejtés típusa	példák
ereszkedő 	A fán mókus volt.
gyors eső 	Melyik?
emelkedő – eső 	Balázs hol van?
gyors eső – ereszkedő 	Kati zenél?
lebegő 	Nem, már hazament.
szökő 	Áll?

Tehát, ha az első szótag magánhangzójának hangmagasságához képest a hangmagasság emelkedik, akkor emelkedő hanglejtésről beszélünk, amikor az első szótag magánhangzójának hangmagasságához képest a hangmagasság csökken, ereszkedő

hanglejtésről van szó. Tipikus, leggyakoribb 5 magyar hanglejtésforma számára bemutatunk referencia-kiejtések, és azok rajzolatát a 8. táblázatban:

Mivel a hanglejtés helyes megformálása fontos a természetes beszéd kialakításában, ezért szavak és mondatok gyakorlásánál is megjeleníthetjük a hanglejtést, a megfelelő funkciógomb használatával.

8.3. Szabad gyakorlás

A terápia során mindig felmerülhet olyan gyakorlat, ami előre nincs a rendszerbe beépítve, de a felhasználó oktató kívánatosnak tartaná a használatát. Ilyen esetben javasoljuk, hogy válassza a szabad gyakorlást, ahol maga dönthet, hogy mit fognak gyakorolni.

Itt lehet új hangcsoportokat, például a nazális hangokat gyakorolni, vagy olyan új szavakat, mondatokat, amelyek a program szótárában nincsenek. A terapeuta bemond egy szót, vagy kifejezést, amelynek spektrogramja a képernyő felső részén jelenik meg, majd a páciens mondja be ugyanazt a kifejezést, amelynek spektrogramja a képernyő alsó részében látható. A két kép összehasonlításával eldönthető, hogy jó volt-e a kiejtés, vagy nem.

8.4. Hallásfejlesztés

A képernyőn megjelenő referenciaminták, vagy a gyermek által kimondott minták tetszés szerint többször is meghallgathatóak. A hang és a kép együttes használata alapeleme a kidolgozott módszereknek, és együttes használatuk a leghatékonyabb.

Hallássérült, vagy implantátumot viselő gyermekeknél ez az együttes használat is fejleszti a hallást. Ezeknél a gyermekeknél viszont külön hallás tréningre van szükség. A program lehetőséget ad hallástréningre a teljes szótárkészletével oly módon, hogy a képernyőtől elfordulva, vagy a hangképet lefedve, csak a hallásra koncentrálnak, és visszamondatjuk a gyermekkel a hallott kifejezést. A gyermek próbálkozásai ilyenkor is lementhetőek a tárból és később visszahívhatóak.

A közeli beszédhangok hallás alapján történő megkülönböztetésének fejlesztésére a szópárokban végzett gyakorlás kiválóan alkalmas.

A kialakított teljes menürendszer minden nyelvre közelíti a 12. Táblázatban mutatott menürendszert.

12. Táblázat. A Beszédkorrektor kialakított teljes menürendszer

+ Előkészítés	+ Sziszegők	+ Magánhangzók	Hanglejtés
• Hangerősség	• Beszédhang kialakítás	• Beszédhang kialakítás	Ereszkedő 5 minta
• Ritmus	* Artikuláció	* Artikuláció	Eső 5 minta
• Színkép	* Kitartott hang	* Kitartott hang	Emelkedő- eső 5 minta
• Zöngé	* Hangkapcsolat	* Hangkapcsolat 10 minta/fonéma	Eső- ereszkedő 5 minta
* Alaphang	- Elején 10 minta/fonéma	• Szavak	Lebegő 5 minta
* Zöngés-zöngétlen	- Belsejében 10 minta/fonéma	* Egyszótagú 10 minta/fonéma	Szökő 5 minta
	- Végén 10 minta/fonéma	* Többszótagú 10 minta/fonéma	
	• Szavak	• Mondatok 10 minta/fonéma	
	* Elején 10 minta/fonéma	• Szópárok 10 minta/fonéma	
	* Belsejében 10 minta/fonéma	+ Hanglejtés ~ 30 minta	
	* Végén 10 minta/fonéma		
	• Mondatok 10 minta/fonéma		
	• Szópárok 10 minta/fonéma		

8.5. Kapcsolat a betűkkel

A módszert hangsúlyozottan közvetlenül iskolakezdés előtt álló gyermekek számára optimalizáltuk. Azokra, akik még nem tudnak olvasni, de szellemileg már érettek a különböző szimbólumok megkülönböztetésére és memorizálására. Minden gyakorlandó beszédhangnak a programban állandó szimbólumképe van. Az **sz** hang szimbólumképe pl. a kígyó, az **i** hangé a kicsibe stb. Mindig ezek a szimbólumképek jelennek meg az adott hang hívóképeként, vagy a háttér-képecskéken. Így a gyermek megtanulhatja csak a hangképeket látva, hogy melyik beszédhangról van szó. Mivel kimondott hang, szótag, szó, kifejezés betűje, betűsora is mindig megjelenik a képernyőn, a terápia végére a gyerekek megismerkednek a magyar

betűkészlet nagy részével. Az iskolai oktatás során ezt a kialakult új ismeretet tudatosan megerősíthetjük.

A szavak hívóképeire a szavak betűsorát, valamint a szavak értelméhez kötődő, azt kifejező rajzokat tüntettük fel. A szavak választékát adó kártyák kiváló alkalmat adnak arra, hogy a kártyán lévő rajzokról beszélgetni lehessen.

9. Hatékonyságvizsgálat

A magyar rendszer részletes hatékonyság-vizsgálatát a budapesti Török Béla Általános Iskola és Diákotthonban végeztük el közösen Váry Ágnes vezető szurdopedagógus együttműködésével. A kiváló beszédterapeuta vállalta, hogy a SPECO program keretében gyakorlatilag egy évig tartó hatásvizsgálatot végez a Beszédkorrektossal, velünk együttműködve.*

A vizsgálati minta a Dr. Török Béla Óvoda, Általános Iskola, Speciális Szakiskola, Diákotthon, Gyermekotthon és Pedagógiai Szakszolgálat által és az egyéb keretek között logopédiai ellátásban részesülő gyermekekből állt össze.

A minta kiválasztásánál két fő szempontot vettünk elsődlegesen figyelembe: a hallássérülés mértéke szerinti csoportba sorolást és az életkort.

A vizsgálatban résztvevő gyermekek az 5. és 8. életév közötti populációt prezentálják. Azért ezt a korosztályt választottuk, mert hallássérültek esetében a beszédfejlődés szempontjából ez az utolsó szenzitív időszak.

A 40 fős mintából ezután 5, egyenként 8 főből álló csoportot alkottunk. A beszédhibás és a három, különböző súlyosságú hallássérült csoportból két-két gyermeket a kontrollcsoportba soroltunk. A nyolcfős kontrollcsoport speciális ellátását ugyanúgy egyéni fejlesztési diagnosztika alapján, logopédiai, illetve szurdologopédiai foglalkozások keretében, ugyanolyan időbeosztással terveztük megvalósítani a Beszédkorrektor programrendszer bevonása nélkül, hagyományos, de természetesen egyénre szabott módszerekkel. A másik négy csoportban a fejlesztési tervet a Beszédkorrektorra alapozva készítettük el. A Beszédkorrektort megvalósító szoftver rendszer („program”) részét képező ún. „számítógépes napló”-ban megjelöltük előre a feladatcsoportokat a későbbi módosítás lehetőségét fenntartva.

A kutatás fő része a program hatékonyságvizsgálata volt. Szubjektív lehallgatással történt annak a hanganyagnak a percepció vizsgálat, amelyet három időpontban hangfelvétellel készítettünk a negyven tanulókból álló mintán. A több mint egyéves kutatási időszak 1999 októberétől 2000 decemberéig tartott, az eredmények statisztikai feldolgozása, a kiértékelés, az összefüggések feltárása 2001 őszén fejeződött be.

A hatékonyságvizsgálat különböző, egymásra épülő fázisokban történt. A fejlesztés fázisát megelőzte és követte egy-egy teljesítménymérés.

* Váry Ágnes PhD dolgozatát 2003. 06. 05-én adta le “Multimédia alapú beszédfejlesztő és gyakorló rendszer differenciált hatékonyságvizsgálata – terápiás ajánlások” címmel.

Az első hangfelvételeket a gyermekekkel '99 októberében készítettük. Ezután következett egy intenzív fejlesztési fázis. Majd a második hangfelvétel-sorozat következett 2000 májusában. A nyári szünet után következett a harmadik hangfelvétel-sorozat 2000 novemberében. A felvételi mintában szereplő gyermekek nyelvi állapotát különböző, az előzőekben tárgyalt szempontok szerint felmérve olyan szóanyagot állítottunk össze a vizsgálathoz, amely minden gyermek számára már ismert szavakat tartalmazott. A szavak sorrendje a hangfelvételek alkalmával mindig ugyanaz maradt. A szavak összeválogatásánál elsődleges szempont az **s**, **zs**, **cs**, **sz**, **z**, **c** hangok elhelyezkedése volt. Így a mindössze 18 szóban a hat „sziszegő” a szó elején, belsejében, vagy a végén jelenet meg. Ez a válogatás a logopédiai vizsgálatok során évtizedek óta bevált gyakorlat szerint történt (Kovács, 1974), (Vinczéné, 1978). A felvételt szétvágtuk úgy, hogy minden gyermek produkcióját külön, monogrammal megjelölt fájlba mentettük el. Az egyes gyermekek produkcióját is feldaraboltuk úgy, hogy minden kimondott szó külön fájlba került mentésre.

Az első hangfelvétel után a kiválasztott alcsoportokban intenzív beszédfejlesztés történt a Beszédkorrektor program bevonásával. Kivételt képezett a kontrollcsoport, ahol a terápiás munka ugyanolyan intenzitással, de a hagyományos eszközökkel történt. A gyermekek hetente két-három alkalommal 25 percen keresztül, egyénileg vettek részt a fejlesztő foglalkozásokon. Váry Ágnes minden gyermekkel differenciáltan, a számára legszükségesebb gyakorlatsorokat megjelölve foglalkozott.

A fejlesztés során már a korai szakaszban kiderült, hogy a tervezett terápiás lépéseket rugalmasan kell kezelni, és módosítani kell aszerint, hogy az adott gyermek esetében az optimális fejleszthetőség mely irányba mutat. A program rugalmassága erre gazdag lehetőségeket kínált.

Fő feladataink a következők kialakítása:

- helyes beszédlelés
- természetes alapzöngé képzése
- helyes beszédritmus
- megfelelő dinamika
- megfelelő hangsúly, hangleadás
- a hangok helyes ejtése.

A fejlesztési időszakban a Beszédkorrektor programot használók a szótár és mondatkészlet bőséges anyagának segítségével gyakoroltak folyamatosan, míg a kontrollcsoportnál a már korábban is meglévő memóriakártyákat, képeskönyveket, gyakorlóanyagokat hasznosították.

A hangfelvételeket nem szakemberekből álló csoport bevonásával kívántuk minősíteni. Ehhez a három különböző időpontban felvett hanganyag a BME TTT Beszédakusztikai Kutatólaboratóriumában került preparálásra. Minden gyermek első, második és harmadik hangfelvételének megfelelő szavaiból szópárokat alkottunk. Így alakultak ki az 1-2, 2-3 és 1-3 párok. Ezt követte egy véletlen keverés, amely utána az egyes párok sorrendje és a párokon belüli szórend is átrendeződött. A negyven gyermek három, egyenként 18 szót tartalmazó hangfelvételtől összesen 18x3x40, azaz 2160 szópár készült. Az összekevert hangfájlokat egy nagy fájlban egyesítettük, úgy, hogy az egyes szópárokat minden esetben 3 másodperc szünet válassza el egymástól. A teljes hanganyag meghallgatásával bizonyosodtunk meg arról, hogy a keverés eredményeként a hallgatónak semmilyen támpontja nem maradt azt illetően, hogy melyik gyermek mely időpontból származó hangfelvételének szavait hallja éppen.

A szubjektív lehallgatásra harminc személyt (20-28 életév közötti férfit, nőt vegyesen, akik nem ezen a szakterületen dolgoznak) kértünk fel. Feladatuk az volt, hogy a hallott szópárokról eldöntsék, hogy az elsőnek, vagy másodiknak hallott szót tartják érthetőbbnek, majd az előre elkészített kiértékelő lapokon rögzítsék azt. Harmadik lehetőségként minden esetben bejelölhették azt is, ha nem tudtak dönteni (bizonytalan). A 2160 szópárból álló hangfelvétel lehallgatása és értékelése több órát vett igénybe, ezért a figyelem ébrentartása és a korrekt ítéletek érdekében három szünetet kellett beiktatunk.

A 13. Táblázatban a szubjektív lehallgatás eredményeinek összesítése látható a vizsgálat egyes csoportjai szerint összesítve. A szemléletesség kedvéért az eredményeket a 30. és 31. ábrákon lévő diagrammokon is közöljük. A 30. ábra az 1999. október és 2000. május közötti fejlesztési szakasz első és második hangfelvétele szubjektív lehallgatási eredményeiken összefoglaló értékelését mutatja. A négy Beszédkorrektor programmal fejlesztett csoport (SPECO csoport) eredményeit oly módon ábrázoltuk, hogy a második, illetve az első szót érthetőbbnek jelölő döntések arányát jelenítettük meg. Ezek az arányok valószínűsíthetővé tették a fejlesztési szakasz alatt elért javulást. A legnagyobb fejlődést az ép hallású beszédhibás csoportnál tapasztaltuk: megközelítően ötször annyian döntöttek úgy, hogy jobb lett a szó érthetősége. Eszerint ezeknél a gyermekeknél gyorsabb, kifejezettebb fejlődés volt tapasztalható, mint pl. a súlyos hallássérült vizsgálati csoportnál.

A további terápiák alatt pedig a rögzítés, a szavakban, mondatokban való automatizálás történt. A gyermekek a bemondott szavakban már helyes képzésű, a normál ejtés szórásának megfelelő sziszegőket hangoztattak, és ezzel az első bemondáskor még hibásnak vélt szó jól érthetővé, javult kiejtésű hangsorrá vált.

A súlyos fokban hallássérült gyermekeknél lehet, hogy még erőteljesebb a pozitív irányú elmozdulás, hiszen náluk nem csak a hang kiejtési hibájával, hanem a hangszínnel, hangerővel, ritmussal, stb., tehát a szupraszegmentális elemekkel is problémák voltak.

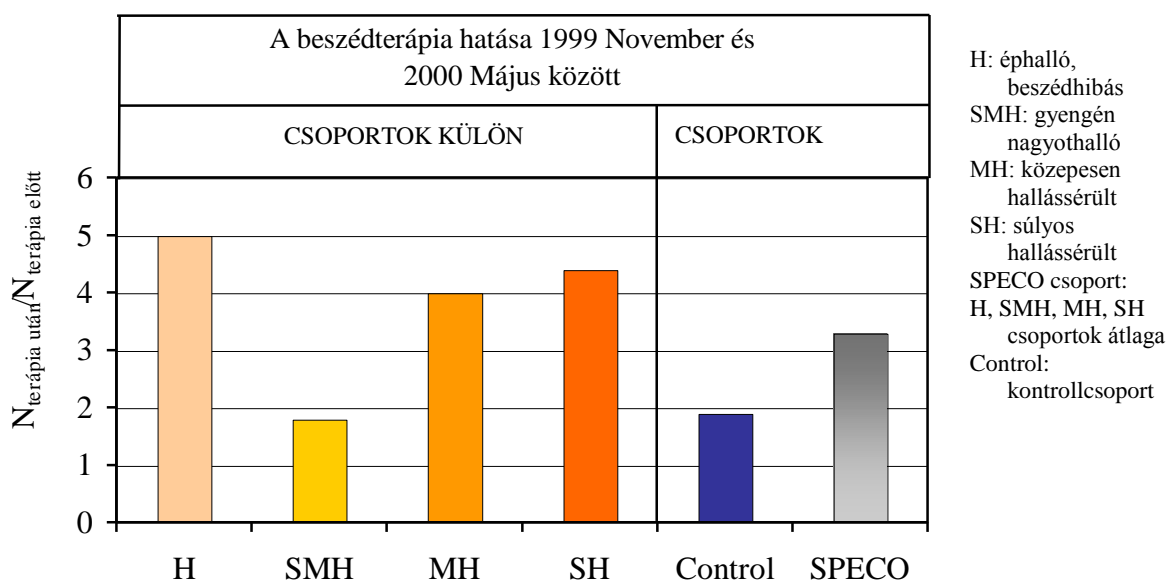
13. Táblázat. A szubjektív lehallgatás eredményeinek összesítése

Gyerek-csoportok	Lehallgatott párok *	A szubjektív lehallgatás döntései		
		A párban az ELSŐ szó a jobb [%]	A párban a MÁSODIK szó a jobb [%]	Bizonytalan [%]
az összes csoport átlaga	1-2	18	55	26
	1-3	14	66	20
	2-3	22	40	39
Kontroll csoport	1-2	24	45	31
	1-3	14	58	28
	2-3	20	35	44
SPECO csoport	1-2	17	57	26
	1-3	14	67	19
	2-3	22	40	38
Ép-halló, beszédhibás (H)	1-2	15	74	11
	1-3	16	73	12
	2-3	24	22	54
Gyengén nagyothalló (SMH)	1-2	25	46	29
	1-3	21	64	14
	2-3	24	48	28
Közepesen hallássérült (MH)	1-2	16	62	22
	1-3	9	78	13
	2-3	20	52	27
Súlyos hallássérült(SH)	1-2	12	53	35
	1-3	12	57	31
	2-3	22	35	44

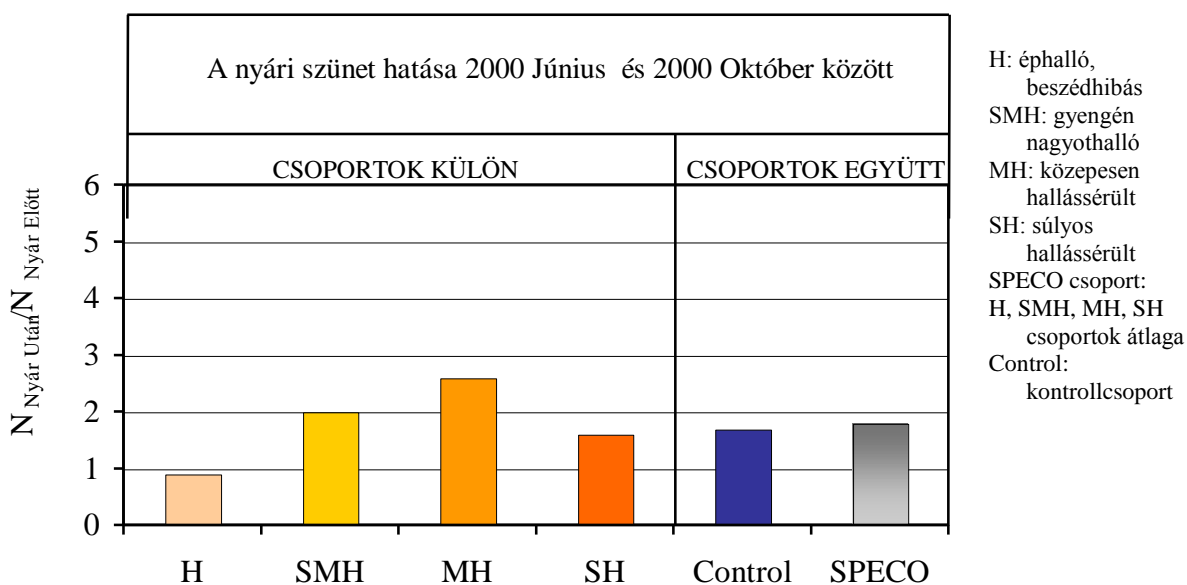
- 1.hangfelvétel: 1999. október
- 2. hangfelvétel: 2000. május, terápia után
- 3.hangfelvétel: 2000. október, nyár után

Az intenzív fejlődés ellenére is visszamaradt még annyi javításra szoruló szegmens, amelyek következtében a hallgatóság számára a beszéd érethetősége nem javult olyan érzékelhetően. A szakember számára ennek igazolásául szolgál a bizonytalan értékelések aránya is. A 13. Táblázatból kiolvasható: míg az ép hallású beszédhibás csoportnál csak a

hallgatók 11%-a nem tudta eldönteni, hogy melyik szó érthetőbb, addig a súlyos fokban hallássérültek csoportjánál 35%-ban tudott dönteni!



30. ábra
A féléves intenzív oktatás hatása



31. ábra
A nyári szünet beszédterápiás hatása

Azon hallgatók száma, akik határozottan nem érzékeltek javulást, szinte azonos a két csoportnál, 15, illetve 12%.

A következő grafikus ábrázolás a 2-3 bemondás szópárjai szubjektív lehallgatási eredményeinek összehasonlítása. A nyár folyamán gyakorlatilag nem történt visszaesés, a hallgatóság szinte egyenletesen további pozitív elmozdulást érzékelt minden vizsgálati csoportban. Ez azt mutatja, hogy egyrészt az élettanilag determinált spontán nyelvi fejlődés hatott mind a Beszédkorrektor, mind a kontrollcsoport esetében, másrészt, hogy a kiejtésben, a beszéd kiejtési minőségében (lényegében az érthetőségében) végbement korrekciós folyamat nem csak eredményes, hanem maradandó is volt.

A kapott eredmény szerint a Beszédkorrektor hatékonysága függ a sérülés típusától, de minden csoportnál jobb eredményt adott, mint a kontrollcsoportban, ahol hagyományos beszédterápia folyt.

A hatásvizsgálat eredményeiből levont következtetések

A számszerű eredményeket és a kutatás során felgyűlt benyomásokat, tapasztalatokat összesítve az általam létrehozott új oktató és rehabilitációs rendszerről beláthatók a következők:

- Az új típusú multiszenzoros beszéd-oktatás, -fejlesztés bevonása a beszéd érthetőségnek fejlesztésébe, javításába az általánosnál gyorsabb, látványosabb eredményekhez vezet.

Az igazán gyors, látványos eredmény a hatékonyságvizsgálatban a halló pösze, illetve megkésett beszédfejlődésű gyermekeknél volt tapasztalható. Mindez nem mond ellent annak a feltevésnek, hogy a vizuális megerősítés hallássérültek esetében kompenzál a legnyomatékosabban, ugyanakkor bizonyosságot nyújt arra nézve, hogy a vizualitás akkor is jelentős teljesítménymotivációs tényező, ha az auditív csatornák épek.

- Általánosságban az egyes hangokat hamarabb formálták meg a gyermekek a rendszert használva, mint a kontroll csoportban (Vicsi, Hacki 1996). Azokban az esetekben, amikor egy-egy hang kiejtésénél a hagyományos terápiával nem tudtak segíteni, a beszédkorrektor rendszer használata eredményre vezetett.

- A Beszédkorrektor erőteljes motiváló hatásának nagy szerepe van a figyelem felkeltésében, a figyelem tartósságának fokozásában.

- A Beszédkorrektor hasznos oktatási segédeszköz a terapeuta kezében, valamint alkalmas az oktatás mellett az önálló otthoni gyakorlásra is, mivel a hangképek a háttér képekkel együtt alkalmazva, érthető visszacsatolást nyújtanak, továbbá az automatikus visszajelzés segíti a hangképek értelmezését.

- **A tanuló önálló, tudatos gyakorlásra való felkészítése, önmaga értékelésének megalapozása elengedhetetlen feltétel a további egyéni tanulásnak.**

A felsoroltak az 1. 2. és 3. tézis alátámasztásának összefoglalása.

Természetesen az új módszer nem helyettesíti a beszédterápiás szakemberek munkáját, de hatékonyan elősegíti az oktatást és könnyebbé teszi a beszédoktató szakemberek munkáját.

Kitekintés

E dolgozatban leírt kutatási eredmények hozzájárulnak a multiszenzoros beszédtanító eljárások egyre jobb megvalósításához. A nyelvtanulásban, a kiejtésoktatásban érdemes lesz az itt kialakított módszereket alkalmazni, a rendszert továbbfejleszteni.

A nyelvfüggetlen szegmentálási módszer alapelvét, és aktuális megvalósítását a beszédadatbázisok feldolgozásánál lehet hasznosítani.

A fonémák kiejtésének fejlődéskutatása gyermekeknél olyan eredményeket szolgáltatott amelyet először is az oktatásmatika kidolgozásánál, másodsor egyéb beszédfeldolgozási eljárásoknál lehet hasznosítani, mint például a gyermeki beszéd gépi felismerése.

A beszédhangok statisztikus spektrális modelljei általános beszéd kutatásokhoz adhatnak további alapot.

Köszönetnyilvánítás

Az itt leírt többnyelvű beszédoktatási módszer gyakorlati megvalósítása, a Beszédkorrektor, egy felhasználó orientált programrendszer, igen nagy adatbázissal, amelynek létrehozásában az évek során számos külső és belső munkatárs vett részt: Barczikay Péter és Tantos András – RCS KFT, Csatári Ferenc, Valyon Zoltán – BME TTT, dr. Vassné Kovács Emőke és dr. Keresztessy Éva – ELTE Bárczi Gusztáv Gyógypedagógiai Kar, Váry Ágnes és dr. Sári Jánosné – Török Béla Általános Iskola és Diákotthon, Nemes Zoltán – grafikus. Köszönetemet szeretném kifejezni lelkes és odaadó munkájukért.

Szeretném megköszönni továbbá Professor Gordos Gézának, hogy a BME Távközlési és Telematikai Tanszék vezetőjeként, és dr. Illényi Andrásnak, hogy a Békési György Akusztikai Kutatólaboratórium vezetőjeként hosszú éveken át aktívan támogatták kutatásaimat. A kutatás az Országos Tudományos Kutatási Alap, a BME Távközlési és Telematikai Tanszék valamint az Európai Közösség támogatásával folyt az INCO-Copernicus programja keretében (SPECO Contract no. 977126).

Irodalom

- Campbell, R. et al. (1998).** Hearing by Eye II. *Psychology Press*. p.-102
- Cole, R. (1998) et al:** Intelligent Animated Agents for Interactive Language Training, *ESCA-STILL 98*, Marholmen, Sweden, pp. 163-166.
- Csányi, Y.(1990):** Hallás-beszéd nevelés, *Tankönyvkiadó*, Budapest.
- Esser, G. (1987):** Sprach-Farbbild-Transformation (SFT). In: *Fellbaum K. R. (HRSZ.): Elektronische Kommunikationshilfen*. Weidler Verlag, Berlin.
- Farkas, M. (1996):** A hallássérültek kiejtés- és beszédfejlesztésének elmélete és gyakorlata, *BGGYTF*, Budapest.
- Gáspár, Á. (1971):** Chordoscop, mint hallási fogyatékosok hangbeszédtanításának segédeszköze, *A Gyógypedagógiai Tanárképző Főiskola Évkönyve IV.*, Budapest.
- Gibbon, F. et al.(1998):** Deviant Articulation in a Cleft Plate Child following Late Repair of the Hard Palate: A Description and Remediation Procedure using Electropalatography *Clinical Ling . and Phonetics 3*, pp. 93-110.
- Gordos G. (1983):** Digitális beszédfeldolgozás, *Műszaki Könyvkiadó*, Budapest, 239.
- Gósy, M. (1980):** Az intonáció percepciója összehasonlító vizsgálatban, *Magyar Fonetikai Füzetek*, 5:110-119.
- Gósy, M. (1981):** Beszédhang kialakulása a gyermeknyelvben, *Magyar Fonetikai Füzetek*, 7:67-89.
- Gósy, M. (1989):** Beszédészlelés, *MTA Nyelvtudományi Intézet*, Budapest.
- Hardcastle, W. J., et al (1999):** Visual Display of Tongue-palate Contact: Electropalatography in the Assessment and Remediation of Speech Disorders, *British Journal of Disorders Communication 26*, pp. 41-74.
- Hunt, M. H, Tomaschewski, H. (1983.):** Improvements in isolated word recognition, *IEEE Trans. Acoust. Speech, Signal Process*, Vol. ASSP-31, Feb.
- Illényi, A., Vicsi, K. (1987):** A Device for Measuring Loudness and Complex Acoustical Signals. *Proc. of Inter-Noise'87* Tokio, pp. 1355-1358.
- ISLE** *Interactive spoken Language Education*, ecisle@nats.informatik.uni-hamburg.de
- ISTRA** *Indiana Speech TRaining Aid features*, *Communication, Disorders Technology, Inc.*
Indiana University Research Park, 501 N. Morton, Suite 215, Bloomington, IN 47404
http://www.comdistec.com/istra_faq.shtm
- Jelinek, F.(1996):** Continous Speech Recognition by Statistical Methods. *IEEE*, 1976 64 (4), pp. 532

- Jones, D. (1961):** The Pronunciation of English, *Cambridge University Press*.
- Kassai, I. (1979):** Gyermeknyelvi dallamminták, *Magyar Fonetikai Füzetek*, 4:147-171.
- Kassai, I. (1981):** A hangsúly kialakulása a gyermeknyelvben, *Magyar Fonetikai Füzetek*, 7:156-174.
- Kassai, I. (1998):** Fonetika, *Nemzeti Tankönyvkiadó*, Budapest.
- Kawai, G. et al (1999):** A CALL System using Speech Recognition to Teach the Pronunciation of Japan, *Eurospeech '99* Budapest
- Kay Elemetrics Ltd:** <http://www.kayelemetrics.com/ProductInfo/ProductPages/>
- Kempelen, F. (1989):** Az emberi beszéd mechanizmusa, *Szépirodalmi Kiadó*, Budapest.
- Kent, R.D. and Read, C. (1992):** The Acoustic Analysis of Speech, San Diego, *California, Singular Publishing*, pp. 158-164.
- Keresztessy, É. (1994):** A beszédkorrektor-program használata. *MAGYE XXII. Országos Szakmai Konf.* 206-207. o.
- Kovács-Vass, E.(1974):** Logopédiai jegyzet, *Tankönyvkiadó*, Budapest
- Kovács-Vass, E. (1983):** Experimental Phonetic Research of Oral Stigmatisms. *Dissertation for the degree of Ph.D.* Prague.
- Köster, J-P., Kiefer, S. (1981):** FB 2 – eine Filterbank mit Visualisierungseinheit für Experimentalphonetik und Ausspracheunterricht. *Hamburger Phonetisch Beiträge* 35, pp. 59-62.
- Lee, S. et. al. (1999):** Acoustics of Children Speech: Development Changes of Temporal and Spectral Parameters, *JASA*, Vol. 3. pp. 1455-1468.
- Leysieffer, H. (1986):** Sprachübertragung mit einer vibrotaktilen Sinnesprothese für Gehörlose. *In: Workshop Elektronische Kommunikationshilfen*, BIG-Tech Berlin '86, Weidler, Berlin, pp. 287-298.
- Massaro, D. W.: (1998):** Perceiving Talking Faces: From Speech Perception to a Behavioural Principle, *Cambridge, MA, MIT Press*. (Information about the facial animation software developed at the University of California, Santa Cruz, is available at <http://mambo.ucsc.edu/psl/pslfan.html>)
- Myers, C., Rabiner, L. H., Rosenberg A., E (1980):** Performance tradeoffs in dynamic time warping algorithms for isolated word recognition, *IEEE Trans. Acoust. Speech, Signal Process*, Vol. ASSP-28, Dec. 1980.
- Narusa, J. (1999):** Computer Aided Spoken Language Training with Enhanced Visual and auditory Feedback, *Eurospeech '99* Budapest pp. 183-186

- Neppert, J. M. H. (1989):** Funktionsgeschichtlicher Abriss der Entwicklung von apparativen Sprechunterrichtshilfen vom Neubeginn nach dem Zweiten Weltkrieg bis zu den heutigen Computer-Systemen. *In. Hörgeschädigtenpädagogik* No. 4., pp 208-217.
- Nikléczy, P., Olaszy, G. (2002):** Kempelen Farkas beszélőgépeznek rekonstrukciója, *Beszédkutató 2002*, MTA Nyelvtudományi Intézet, pp. 5-18.
- PCLX** *The NEW electrolaryngograph system for the PC*, Millgrant Wells Limited 7 Stanley Road, Rugby CV21 3UF, England.
- Potter, R. K., Kopp, G. A., Green, H. C. (1947):** VISIBLE SPEECH, *van Nostran CO.*, New York.
- Sakoe, H., Chiba, S. (1978):** Dynamic programming algorithm optimization for spoken word recognition, *IEEE Trans. Acoust. Speech, Signal Process*, Vol. ASSP-26.
- Siemens.** <http://schulen.eduhi.at/spzfreistadt/gehoer.htm>
- Speech Viewer** Psychological Corporation, Harcourt Brace Jovanovich, INC. (IBM Business Partner), 555 Academic court, San Antonio, Texas.
- Strauss, H.-C. (1982):** Technical development and pedagogical conceptualization of the Mini-Fonator System (dt. Übersetzung in: DB 29). *In: Feuchte, Herbert et al. (eds): Proceedings of the International Congress on Education of the Deaf in Hamburg*, 1980. Vol. 3. Heidelberg: Groos– pp. 138-142.
- Szende T. (1974):** A beszéd folyamat alaptényezői, *Akadémiai Kiadó*, Budapest.
- Talk to me**, Tell me more CD, 134, Avenue Joseph Kessel 78960 Voisins le Bretonneux, Auralog, France <http://www.auralog.com/us/schools.html>
- Tarnóczy T. (1974):** A magánhangzók vizsgálatának akusztikai problémái, *Általános Nyelvészeti Tudományok*, 10:181-196.
- Tarnóczy T. (1984):** Kempelen Farkas beszélőgépe és a mai beszédösszeállító berendezések, *Magyar Fonetikai Füzetek*, 13:22-33.
- Váry Á., Erdős, T. (1996):** A beszédkorrektor szerepe a nagyothalló gyermeke egyéni anyanyelvi nevelésében. *Fejlesztő Pedagógia*, 2-3, pp. 41-44.
- Váry, Á. (2000):** Multimédiás beszédfejlesztő program hatékonysága, pedagógiai tapasztalatai. *X. Magyar Alkalmazott Nyelvészeti Kongresszus*, Székesfehérvár, 200. pp. 206.
- Váry, Á. (2001):** Multimédia alkalmazása hallás- és beszéd-sérült gyermekek speciális terápiájában. *Média – Informatika – Kommunikáció Nemzetközi Konferencia*, Veszprém, MTA Veszprémi Akadémiai Bizottságok kiadványa, pp. 182-185.

- Váry, Á. (2001):** Multimédiás beszédfejlesztő program – új módszerek, lehetőségek a speciális pedagógiában. *Új Pedagógiai Közlemények: Doktoranduszok bemutatkozása, ELTE-BTK, Neveléstudományi Intézet, Budapest*, pp. 52-60.
- Váry, Á (2001):** Egy multimédiás többnyelvű beszédfejlesztő rendszer használatának pedagógiai tapasztalatai. *Nyelv*Info*, IX. évf. 2001/1-2. sz. pp. 34-37.
- Váry. Á. (2002):** Varázsdoboz – Audiovizuális beszédoktató és gyakorló rendszer differenciált alkalmazása. *Demoszthenész Hirmondó, Beszédhibások és Segítőik Országos Érdekvédelmi Egyesületének kiadványa*, 2002. április, 13. sz., Budapest, 13-15.
- Vicsi, K. (1981).** Az időtartam szerepe néhány mássalhangzó-típus hallás alapján történő megkülönböztetésében, *Magyar Fonetikai Füzetek*, 7:59-67.
- Vicsi, K. (1981).** The Most Relevant Acoustical Microsegment and Its Duration Necessary for the Recognition of Unvoiced Stops, *ACOUSTICA* Vol. 48, pp. 53-58.
- K. Vicsi and P. Berényi (1989):** Speech Recognition Systems in Acoustical Research Laboratory. *Proc. on the Speech Research '89*, Budapest, Hungary, p.301,
- Vicsi, K. Matilla, M. Berényi, P. (1990):** Continuous Speech Segmentation Using Different Methods, *Acoustica*, Vol. 71, 152-156.
- Video Voice**, Micro Video, 210 Collingwood, Suite 100. P. O. Box 7357 Ann Arbor, MI 48107.
- Vicsi, K. – Farkas, A. – Vassné, K. E. – Keresztessy, É. (1993):** Beszédkorrektor. *Beszédgyógyítás*, 93/2-3. 13-21.
- Vicsi, K.- Víg, A. (1995):** Neural network/rule based hybrid, text independent continuous speech recognition. In: *Eurospeech '95*, Madrid, 1995. pp. 2201-2204.
- Vicsi, K. (1995):** Beszédkorrektor: Audio-vizuális számítógépes beszédoktató rendszer. *Bárcki Gusztáv Gyógypedagógiai Főiskola*, jegyzettrészlet, Budapest, p. 265-288,
- K. Vicsi (1995):** A product Oriented Teaching and Training System for Speech Handicapped Children. *Journal of Microcomputer Applications*, 18, Academic Press Limited, p. 287-297.
- K. Vicsi and T. Hacki (1996):** “CoKo” - Computergestützter Sprechkorrektor mit audiovisueller Selbstkontrolle für artikulationsgestörte und hörbehinderte Kinder. *Sprache-Stimme - Gehör*, Georg Thieme Verlag, Stuttgart. Germany, p.141-149, September 1996.
- K. Vicsi and A. Víg (1998):** LIAS: Language Independent Automatic Segmantation Technique Using SAMPA Labelling of Phonemes, *First International Conference on*

Language Resources and Evaluation, Granada, Spain, 28-30 May, 1998, Vol II. pp. 1317-1323.

Vicsi, K. – Víg, A.(1998): Az első magyar nyelvű beszédatbázis. *In: Beszédkutatás '98.* Szerk.: Gósy Mária. MTA Nyelvtudományi Intézet. Budapest, pp. 163-177.

Vicsi, K. (1999): Speechdat – Hungarian speech database for creation of voice driven teleservices. *Technical report.* <http://luna.ttt.bme.hu/speech/speechdt.htm>. „György Békésy” *Acoustics Research Laboratory of the Budapest University of Technology and Economics*, Budapest.

Vicsi, K., Hacki T.: (1996) „CoKo” - Computergestützter Sprechkorrektor mit audiovisueller Selbstkontrolle für artikulationsgestörte und hörbehinderte Kinder. *Sprache-Stimme - Gehör, Georg Thieme Verlag, Stuttgart.* September. 141-149.

F. Csatári, Zs. Bakcsi, K. Vicsi (1999): A Hungarian Child Database for Speech Processing Applications, *Proc. on the Eurospeech '99*, Budapest, Hungary, p. 2231

Vicsi, K. - Csatári F. - Bakcsi Zs. - Tantos A. (1999): Distance Score Evaluation of the Visualised Speech Spectra at Audio-visual Articulation Training, *ESCA, Eurospeech '99*, Budapest, Hungary, ISSN 1018-4074, 1911-1914.

Vicsi, K., Roach, P., Öster, A., Kačič, Z., Barczikay, P., Sinka, I.(1999): A Multimedia Multilingual Teaching and Training System for Speech handicapped Children, *Proc. on the Eurospeech '99*, Budapest, Hungary, p. 859.

Vicsi, K. et al. (2000): A Multimedia Multilingual Teaching and Training System For Speech Handicapped Children. *Int. Journal of Speech Technology* Vol. 3 pp. 289-300. http://luna.ttt.bme.hu/speech/inter_co.htm

Vicsi, K. (2001): Multimedia Multilingual Teaching and Training System for Speech Handicapped children. *Final SPECO report of the project coordinator*, European Commission, Inco-Copernicus (pn. 977126), 31.08.

Vicsi, K. (2002): Computer Assisted Speech Training Methods based on Auditory and Visual Feedback and their Psycho-Acoustical Evaluation. Invited paper, Proc.on the *NATO Advanced Study Institute (ASI) on the Dynamics of Speech Production and Perception*, Summer University, June 25- July 6, Il Ciocco, Italy, p.2807

Vicsi, K. (2004): An Overview of Speechtraining Methods Based on Multi Modal Feedback. "Dynamics of speech production and perception "editors": *Pierre Divenyi & Georg Meyer. "* IOS Press, Amsterdam, Könyvrészlet elfogadva, elküldve, 2004-ben jelenik meg.

Video Voice, Micro Video, 210 Collingwood, Suite 100. P. O. Box 7357 Ann Arbor, MI 48107.

Vincéné, B. E. (1978): A fonetikai hibák javításának dinamizmusa. *Magyar Fonetikai Füzetek*, 1:107-112.

Vincéné, B. E. (1996): A számítógép logopédiai terápiás alkalmazásáról. *Fejlesztő Pedagógia*, 2-3. 19-21.o.

Wallace, J. et al. (1998): Applications of Speech Recognition in the Primary School Classroom *ESCA-STILL 98*, Marholmen, Sweden, pp. 21-24.

Zwicker, E. and Terhardt, E. (1980): Analytical expressions for band rate and critical bandwidth as a function of frequency, *J. Soc. Am.* Vol. 68, 1523.

Zwicker, E. (1982): Psychoakustik, *Springer Verlag*, Berlin, p.67.