MISKOLCI EGYETEM
GÉPÉSZMÉRNÖKI ÉS INFORMATIKAI KAR

A BESZÉDMINŐSÉG AUTOMATIKUS ÉRTÉKELÉSE

PhD értekezés tézisei

Készítette:
Pintér Judit Mária
okleveles mérnökinformatikus

Hatvany József Informatikai Tudományok Doktori Iskola

Doktori iskola vezetője:
Prof. Dr. Szigeti Jenő
egyetemi tanár

Témavezető:
Dr. Czap László
egyetemi docens

Miskolc, 2015

Köszönetnyilvánítás A kutató munka a Miskolci Egyetem stratégiai kutatási területén működő Mechatronikai és Logisztikai Kiválósági Központ keretében, a TÁMOP-4.2.2. C-11/1/KONV-2012-0002 jelű projekt részeként az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.

A Bíráló Bizottság tagjai

Elnök:

Prof. Dr. Szigeti Jenő	Miskolci Egyetem, egyetemi tanár

Titkár és tag:

Dr. Kovács Szilveszter	Miskolci Egyetem, egyetemi docens

Tagok:

Dr. Kovács László	Miskolci Egyetem, egyetemi docens
Dr. Vicsi Klára, DSc	Budapesti Műszaki és Gazdaságtudományi Egyetem, tudományos tanácsadó,
Dr. Váry Ágnes	Dr. Török Béla Óvoda, Általános Iskola, Speciális Szakiskola, EGYMI és Kollégium, intézményvezető helyettes,

Opponensek:

Dr. Barabás Péter	Miskolci Egyetem, egyetemi adjunktus
Prof. Dr. Takács György	Pázmány Péter Katolikus Egyetem, címzetes egyetemi tanár

Tartalomjegyzék

BEVEZETÉS

Az értekezésben a TÁMOP-4.2.2.C-11/1/KONV-2012-0002 azonosító számú, "Alap- és alkalmazott kutatások hallássérültek Internetes beszédfejlesztésére és az előrehaladás objektív mérésére" c. projekt kapcsán végzett kutatómunkám eredményeit mutatom be. Az értekezés fő fejezetei a beszédkeltés fiziológiáját, a beszédfelismerés alapjait, a siketek és nagyothallók beszédtanulási nehézségeit és lehetőségeit, valamint a kutatási projektben megalkotott beszédminőség automatikus kiértékelésének gyakorlati alkalmazását és megvalósításának kulcslépéseit mutatják be.

A beszéd nem más, mint akusztikus hullámok keltése, azaz beszédhangok, fonémák (hangok olyan elemi, elvont egysége, amely szavakat különböztet meg egymástól, önálló jelentéssel nem rendelkezik) kibocsátása. A beszéd nem csupán fonémák sorozata, hanem fontos a hangsúlyozás, a hanglejtés és számos más szupraszegmentális jellemző is. Ezek alapján egyértelmű, hogy a beszéd az emberek legfőbb kommunikációs eszköze, amiért az akusztikus beszédfelismerést igen sok területen és különböző céloknak megfelelően alkalmazzák [S10]. Ezen célok közül az egyik a hallássérültek oktatása.

Gyógypedagógiai szempontból a hallási fogyatékosság zártabb fogalom, olyan hallási rendellenességet jelent, ahol a sérülés időpontja, mértéke és minősége miatt a beszédbeli kommunikáció spontán kialakulása zavart [10]. *„A hallássérülés gyógypedagógiai fogalma (hallási fogyatékosság) elsősorban a beszédértéshez szükséges hallásterületen közepes vagy annál súlyosabb fokú nagyothallást, siketséggel határos vagy siketségnek diagnosztizált hallásveszteséget jelent. Más megközelítésben a hallássérültek pedagógiája a hallássérült kifejezést olyan hallászcsökkenésre alkalmazza, amelynek következményeként a beszédfejlődés nem indul meg, vagy a beszéd oly mértékben sérült, hogy a beszéd megindításához,*

korrekciójához speciális beszédfejlesztő módszerek alkalmazására van szükség" [13].

A hallássérülés közvetlen következménye a beszéd elsajátításának zavara vagy akár a beszéd teljes hiánya. A beszéd érthetőségét, minőségét elsősorban a szupraszegmentális tulajdonságok befolyásolják. A szupraszegmentális hiányosságok kiterjednek a normálistól eltérő tempóra és ritmusra, a hanglejtés változtatásának problémájára, a hangmagasság korlátozott terjedelmére, és a „siket hangminőségre", ami mind a beszélt nyelv olyan aspektusa, amelyet a hallás, az auditív visszacsatolás és az önirányítás révén lehet megszerezni.

A hallás fontossága azonban nem csak beszédkommunikációs szempontból hangsúlyozandó. A hallássérülés kihatással van a viselkedésre, a kapcsolatok beszűkülését okozhatja, társadalmi szokások, magatartási szabályzók hiányát idézheti elő.

1. A DISSZERTÁCIÓ RÖVID ÁTTEKINTÉSE

1.1 Kutatási projekt hallássérültek internetes beszédfejlesztésére

Az „Alap- és alkalmazott kutatások hallássérültek internetes beszédfejlesztésére és az előrehaladás objektív mérésére” címet viselő projekt a siket és nagyothalló személyek számára – az eddigi eszköztár részbeni megújításával – a sikeres beszédtanulás egyik kulcsát nyújthatja. A projekt gyakorlatban is hasznosítható célja egy komplex rendszer létrehozása, mely a beszédfolyamat audiovizuális megjelenítését szolgáltatja, egyrészről a beszéd hangképeinek másrészről az artikulációnak a vizuális megjelenítésével, egy oktatási keretrendszerbe foglalva. Ezek mellett számos olyan funkciót tartalmaz a rendszer (prozódia megjelenítés. automatikus minősítés, tudásalapú rendszer implementálása), amely a későbbiekben lehetővé teszi az egyéni gyakorlást nem csak számítógépen, hanem mobil eszközön is. A kifejleszteni kívánt technológia audiovizuális transzkódolását végző modulja nyelvfüggetlen, a beszélő fej és az automatikus minősítés újabb neurális hálók betanításával nyelvfüggetlenné tehető.

A beszédasszisztens rendszer egyik szolgáltatása az automatikus minősítés és visszajelzés, hogy a hallássérült diákok önállóan gyakorolhassák a mintaszavak kimondását. A tanulás során a referencia kiejtést a szerver vagy a tanár produkálja. A diák ezt igyekszik utánozni az ő aktuális bemondásával. Ezzel rokon probléma merül fel a beszéd gépi felismerésénél. Előre (modellezés segítségével) eltárolt, valóságos beszédből származó beszédrészletek (hang, hangátmenet, szó, stb.) közül kell a felismerendő beszédrészlethez leghasonlóbbat megtalálni, és ha a hasonlóság elég nagy, akkor a beszédrészlet felismertnek tekinthető. A hallássérültek beszélni tanításánál a hasonlóság automatikus ellenőrzése és a visszajelzés generálása alapkutatás, amely megköveteli egy hasonlósági

mérték kidolgozását. A hasonlósági mértéknek monoton összefüggésben kell lennie a hallássérült és halló bemondók által kiejtett hangok, hangkapcsolatok, szavak szubjektív (épen halló emberek által végzett) tesztek átlagos megítélésével (MOS = Mean Opinion Score). A különböző lényegkiemelési és távolság számítási módok elemzésével kidolgozható a szubjektív értékelésnek megfelelő hasonlósági mérték. Ez az alapja az előrehaladás értékelésének és a visszajelzés generálásának. Az értékelés nyilvánvalóan a korábbi eredményekkel összevetve alakítható ki, hiszen ugyanaz a kiejtés egyik tanulónál siker, a másiknál kudarc lehet. Az automatikus értékelés verifikálása érthetőség vizsgálattal történhet.

A célként létrehozandó rendszerhez hasonló fejlesztés az, amelyet olyan személyeknél alkalmaztak, akiknek gégerák miatt eltávolították a gégéjüket, vagy olyan gyerekeknél, akik ajak- és szájpadhasadékkal születtek. Ezeknél a személyeknél szignifikáns összefüggések érhetők el a szubjektív és az automatikus minősítés között. Az egyes betegségekhez tipikus, jól detektálható beszédhibák társulnak. Ezekhez a beszédhibákhoz a kutatók és fejlesztők rendelkezésére álltak minták, így megalkotható volt az automatikus minősítés. A PEAKS (Program for Evaluation and Analysisof all Kinds of Speech Disorders) egy rögzítő és elemző rendszer hangképzési és beszédzavarok automatikus vagy manuális minősítéséhez. A PEAKS rendszert több kórház is alkalmazza nem csak Németországban. A pedagógusok számára is segítséget nyújt az oktatásban. A félautomatikus módszer a mérsékelttől a jó értékelésig 60%-os korrelációs szintet ér el az összes fonetikai rendellenesség esetén. Beszélők szintjén a percepciós és az automatikus értékelés között 89%-os korrelációt értek el, teljesen automata rendszernél pedig 81%-ot. Ez a korrelációs eredmény az értékelők közötti korrelációs tartományba esik [36], [37]. Az általunk fejlesztett rendszer is hasonló tulajdonságokkal rendelkezik kivéve, hogy a hallássérültek beszédhibái nem definiálhatóak vagy csoportosíthatóak és szűkíthetőek le. Ezért a feladatom egy olyan automatikus minősítés megalkotása volt, ami

független a beszédhiba típusától és az aktuális bemondás mérhető jellemzőin alapszik [3], [7], [28].

1.2 Szóadatbázis az automatikus minősítés megalkotásához

A már említett minősítési skála megalkotásához szükséges adatbázis mintáit eltérő beszédprodukciós fejlettségi fokú gyerekektől gyűjtöttem be. A mintákat laikus hallgatók valamint szurdopedagógusok minősítették. Az artikuláció helyessége a szépen beszélő ép hallóktól az alig érthetően beszélő hallássérültekig terjedt. A minták rögzítését az adott oktatási intézményen belül végeztem egy csendesnek mondható szobában a pedagógusok segítségével. A gyerekek a bemondások előtt egyszer átnézhették a felolvasandó szavakat, hogy a bemondást csak minimális mértékben befolyásolják az olvasási nehézségek.

Az adatbázisban pontosan 2421 szó szerepel (egyes szavak többszörösen is előfordulnak, de a bemondók eltérőek, ezért azok érthetősége is), amit 13 pedagógus és 23 hallgató értékelt. Minden pedagógus csak a másik iskola diákjainak bemondását értékelte, hogy elkerüljük a beszélő felismeréséből eredő előítéleteket. A bemondást többször is meghallgathatták az értékelők és megjegyzéseket is fűzhettek a mintákhoz. Az eredményeket internetes alkalmazáson keresztül rögzítettük. A minősítés alapját a pedagógusok esetén az általuk meghatározott ötfokozatú skála képezte.

A skála értelmezése:

- *Érthetetlen (1):* az artikuláció teljesen torz; felismerhetetlenek a magán- és mássalhangzók; a szótagszám visszaadása sem megfelelő vagy nem kivehető; a levegővétel, a levegővel való gazdálkodás helytelen; rossz a tempó, a ritmus; dallamtalan, dinamikátlan vagy túl feszített a hangadás.
- *Nehezen érthető (2):* súlyos torzítások, hangelhagyások, hangcserék; csak a magánhangzók egy része kivehető; a légzés elégtelensége miatt létrejövő torzítások, pl. túl levegős vagy fojtott; eltérő, zavaró hangszín, ritmus, tempó jellemzi.

- *Közepesen érthető (3):* a magánhangzók ejtése helyes, a szótagszám megfelelő; súlyos beszédhibák előfordulhatnak pl. diszlália (az a beszédzavar, mely szerint egyes hangzók hiányosan képeztetnek, orrhangzósság, fejhangzósság, stb.), prozódiai elégtelenségek.
- *Jól érthető (4):* csekély mértékű beszédhibák; enyhe prozódiai elégtelenségek.
- *Hallók beszédével azonos szinten érthető (5):* legfeljebb 1-2 hanghiba fordulhat elő.

A laikus hallgatóknak a mindennapi nyelvhasználat alapján kellett 1-től 5-ig pontozniuk a bemondásokat. A minősítési skála és az automatikus kiértékelés megvalósításához végzett vizsgálatok során a hallgatók és a pedagógusok általi összevont értékelést tekintettem referenciának és a minősítések átlagát használtam fel a vizsgálatok során.

1.3 Gyenge minőségű beszéd szegmentálása

A beszédtempó beszélőről beszélőre kiejtésről kiejtésre változik. Ezek a nemlineáris megnyúlások és rövidülések nem feltétlenül számítanak hibás ejtésnek. A hallássérültek az átlagos beszédtempónál általában lassabban beszélnek. Az egyes hangok kiejtésének minősítéséhez össze kell párosítani a referencia alakzat és az aktuális kiejtés időszegmenseit. A referencia és az aktuális hullámforma azonos hosszúságúvá tétele lineáris nyújtással, illetve zsugorítással elvégezhető. Ez azonban nem biztosítja az egyes hangok időbeli párhuzamát, mert a kiejtés ritmusa eltérhet a referenciától. Egyes hangokat hosszabban másokat rövidebben ejtve a lineáris vetemítésnél nem azok a hangszegmensek kerülnek fedésbe, amelyekre hasonlítaniuk kell így az összevetés hamis eredményre vezet. Különösen jellemző a hallássérültek beszédére az egyes hangok megszokottól eltérő idejű artikulációja. A referencia és a vizsgált beszéd összehasonlításához tehát dinamikus idővetemítésre van szükség, amire a számítógépes beszédfeldolgozásban kidolgozott eljárások és algoritmusok állnak rendelkezésre. Ezek a

módszerek jó minőségű beszédre, a mindennapi kommunikációban elfogadható kiejtésre megfelelően működnek. A torz hangokra, a rendkívül elnyújtott, akadozó beszédre gyenge eredményt szolgáltatnak. A beszéd minősítésének kulcskérdése a helyes szegmentálás. Egyik kutatási célom a szegmentálásra szolgáló módszerek továbbfejlesztése annak érdekében, hogy a szinte érthetetlen beszédre is használható szegmentálási eredményeket kapjak.

A kutatómunkám során megvizsgált vetemítési eljárásoknál leírt szabályok alapján egy referencia szóhoz az idővetemítés nem bizonyult sikeresnek. A módszereket alkalmazva hallás alapján értékeltem a szegmentálási eredményeket, amik rendkívül torz eredményt mutattak. Felhasználva a bemutatott két szegmentálási eljárás alapszabályait megalkottam egy a gyenge minőségű beszéd szegmentálására alkalmas adaptált dinamikus idővetemítési eljárást (Adapted Dynamic Time Warping – ADTW).

A dinamikus idővetemítésnél a keresett szót egy referencia bemondással vetjük össze és keressük egyes időkeretek megismétlésével, illetve kihagyásával a referencia bemondáshoz leginkább hasonló ütemezést. A hallássérült gyerekek bemondásai közül a nehezen érthetők nem alkalmasak arra, hogy a referencia bemondáshoz valamilyen hasonlósági mérték szerint eléggé hasonlítsanak. Próbálkoztam férfi, női, gyerek bemondáshoz és szintetizált hanghoz is vetemíteni a keresett szavakat, ezek azonban nem voltak sikeresek. A sok bemondóval tanított neurális hálózat statisztikai alapon jobban visszatükrözi az egyes hangokhoz mérhető hasonlóságot.

Egy 300 bemondótól származó 4 és fél órás hangadatbázis alapján PLP lényegkiemelést alkalmazva meghatároztam az egyes hangok stacionárius szakaszaihoz tartozó együtthatók átlagát. Majd Euklideszi távolságot képeztem a magyar beszédhangok átlagai között [15]. A szünet és a négy osztály valamint a 32 hanghoz tartozó outputot használtam a távolság meghatározására. A normalizált távolság megfordításával hasonlóságmértéket képeztem az egyes hangok között. Az egyes hangok

referencia időtartamát az adott beszédhang átlagos hosszára állítottam be [42]. A neurális hálózatok által képezett osztályok:

- szünet;
- magánhangzó (*a, á, e, é, i, o, u, ü*);
- fél magánhangzó (*m, n, ny, r, l, j*);
- réshang (*f, sz, s, h, v, z, zs*);
- zárhang. (*p, t, ty, k, b, d, gy, g*).

A hallássérült gyerekek bemondásában gyakran találkoztam hangok között több tized másodperces szünetekkel és hosszan ejtett hangokkal. Ezért minden hang után beiktattam a referencia előállítása során egy szünetet, a szünet pedig akárhányszor ismétlődhet. Értekezésemben az ismert eljárásoknál ismertetett szabályok szerint egy időintervallumot maximum kétszeresére lehet nyújtani. A hallássérült gyerekek bemondásában azonban ennél hosszabban ejtett hangokkal is gyakran találkoztam. Ezért egy időkeret kétszeri ismétlése is megengedhető, ezzel egy időintervallum háromszorosára nyújtható.

Összevetettem az általam fejleszt ADTW eljárást az értekezésben részletesen bemutatott és a gyakorlatban is alkalmazott módszerekkel. Az eredmények alapján az eljárásom teljesített a legjobban.

I. Tézis

Megvizsgáltam a nemlineáris idővetemítés különböző módjait, és a gyenge minőségű beszédre módosítottam a dinamikus idővetemítés kapcsolódási szabályait, ezzel az általam vizsgált eljárásoknál lényegesen több határérték esett az egyes hangintervallumok belsejébe.

Tézishez tartozó publikációim: [S1], [S2], [S10], [S13]

Újdonság

Jó minőségű beszédre kidolgozott eljárások nem megfelelően működnek nagyon torz és akadozó beszédre. Az eljárás újdonsága a referencia generálás és az alkalmazott kapcsolódási szabályok.

Mérések

A szegmentálás pontosságát különböző mutatók alapján vizsgáltam, és a gyenge minőségű beszédhez nem adaptált rendszereknél pontosabb eredményeket kaptam.

Érvényességi korlátok

Az alkalmazás korlátai: Az alkalmazott neurális hálózat nyelvfüggő, a tanításra használt beszédadatbázis magyar nyelvű. Más nyelvekre a neurális hálózatot be kell tanítani.

Konklúzió

A gyenge minőségű beszédre kidolgozott szegmentálási eljárás alapját képezi az automatikus minősítésnek, hiszen kulcsszerepe van a referencia hang és az elemzés alatt álló hang megfeleltetésében.

1.4 Hangsúlydetektálás relatív intenzitás alapján

Mivel a siketek nem hallják a saját hangjukat, külön nehézséget okoz a szupraszegmentális jellemzők megtanulása. A beszédminősítés egyik fontos eleme a hangsúly detektálása. A hangsúlyos szótagnál a hangmagasság és a hangerő emelkedik. A szótagon belül általában a magánhangzó a legnagyobb energiájú, ezért az intenzitás mérésekor célszerűnek tűnik a magánhangzó energiájának vizsgálata. Ha megmérjük az egyes magánhangzók átlagos energiáit, egy sok beszélős hosszú hangmintákat tartalmazó adatbázison, kiderül, hogy a magánhangzók igen eltérő átlagos energiával rendelkeznek. A hangsúlydetektálásnál az energia az egyik vizsgált jellemző, de egy hangsúlytalan nagy átlagenergiájú hang (pl. *a, e*) energiája jelentősen meghaladhatja a hangsúlyos gyengébb hangok (pl. *i, u*) pillanatnyi energiáját.

A hangsúly detektálását általában az energia és az alapfrekvencia alapján végzik. Megvizsgálták a hossz, az amplitúdó és a spektrális változások különböző módokon normalizált értékeit [51]. Több esetben mély neurális hálók betanításával és alkalmazásával valósították meg az automatikus hangsúlydetektálást angol nyelvre [41], [51], [52].

A hangsúlydetektálást célzó vizsgálataim során ezért a magánhangzó intenzitását nem egyszerűen az energiával azonosítom, hanem relatív intenzitás értéket határozok meg a magánhangzó átlagenergiájához viszonyítva a pillanatnyi energiát. A módszer eredményességét a hangmagasságot is figyelembe vevő hangsúlydetektálásra betanított neurális hálózattal verifikáltam. A vizsgálatokhoz egy speciálisan hangsúlyvizsgálatokra létrehozott hangsúlyadatbázist használtam fel [1], [44]. Kutatásomhoz 10 személytől egyenként 50 bemondott mondatot használtam fel. A hangsúlyos és hangsúlytalan bejegyzés nem az aktuális bemondásra vonatkozik, hanem a mondat értelmezése alapján hangsúlyosnak és hangsúlytalannak ítélt szavak kapták a H illetve – jelzést. A bemondók nem mindig az elvárt hangsúlyozási minta alapján olvasták fel a mondatokat, így felkértem egy nyelvészeti szakértőt, hogy elemezze a mondatokat egyenként és alkossa meg a mondatok hangsúlyképletét. Az elemzés során a szakértő külön jelölte a hangsúlyos, a félhangsúlyos és hangsúlytalan szavakat. A tanító és tesztelő mondatok kiválasztása véletlenszerű volt. Tesztelő mintáknak a rendelkezésre álló mondatok 25%-t választottam ki véletlenszerűen.

Az általam magyar nyelvre létrehozott hangsúlydetektálási módszer több komponensből tevődik össze. A detektálás egyik fő összetevője a relatív intenzitás. A módszer a magánhangzók átlagenergiájához viszonyítja a pillanatnyi energiát. A mondatok elejétől vége felé haladva csökkenő relatív intenzitást tapasztalható. A csökkenő tendencia korrigálására kiegyenlítem a mondatok átlagos amplitúdóját. Az értékeket a magánhangzó közepének 50 milisecundumos (800 minta) környezetére számoltam.

Az energiát az alábbi képlet alapján határozom meg:

$$E_n = log \sqrt{\sum_{i=n-N+1}^{n} x_i^2} \qquad (10)$$

A pillanatnyi és az átlag energiát is a (10) képlet alapján számolom, a relatív intenzitást pedig a két logaritmus különbségeként kapom meg.

A másik hangsúlyt befolyásoló jellemző a beszéd alapfrekvenciája. A hangsúlyos szótagoknál az alapfrekvencia megemelkedik. A relatív intenzitás alkalmazásának a pillanatnyi energiájával szemben a verifikálása érdekében neurális hálózatot tanítottam be az alapfrekvencia (F0) és a relatív intenzitás felhasználásával. Az alapfrekvencia meghatározásához kipróbáltam az ismert beszédelemző alkalmazásokat (Praat, Wavesurfer, Opensmile). Egyes magánhangzókat – különösen a mondatvégi mélyebb és halkabb hangokat – ezek a rendszerek gyakran zöngétlennek mutatták. Az F0 alapján a hangsúlyos szótagok felismerése nem volt elég sikeres. Ezért saját alapfrekvencia meghatározó algoritmust fejlesztettem ki.

Ismert tény, hogy a lineáris predikció hibája a zöngés hangoknál a periódus elején kiugróan magas. A periódus kezdetén a predikciós hibát erős aluláteresztős szűréssel csökkentettem.

A hang rekedtessé válása vagy az autokorrelációs függvényben egy felharmonikusnál jelentkező maximum oktávugrást okozhat az alapfrekvenciában. Ezt a hibát a mondat magánhangzóinak alapfrekvenciáit elemezve oktávszűréssel korrigálom. Az alapfrekvencia becslésére három érték átlagát használom fel:

1. az előző magánhangzó alapfrekvenciáját;
2. a mondat F0 menetének regressziós egyenesét;
3. az F0 mediánszűrését.

Amennyiben ehhez a becsült értékhez közelebb áll az adott magánhangzóra kapott alapfrekvencia fele vagy kétszerese, az F0 értékét az adott magánhangzóra a közelebb álló értékre változtatom.

A relatív energia és az alapfrekvencia felhasználásával neurális hálózatot tanítottam be, a hangadatbázis mondatainak megosztásával a tanítás a szakértői hangsúlyképlet alapján történt. Egy szótagot:

- relatív energiájával;
- alapfrekvenciájával;
- az aktuális szótag és a következő szótag aktuális energiájának különbségével (az utolsó szótagnál 0);
- az aktuális szótag és az előző szótag aktuális energiájának különbségével (az első szótagnál 0);
- az aktuális szótag és a következő szótag aktuális alapfrekvenciájának különbségével (az utolsó szótagnál 0);
- az aktuális szótag és az előző szótag aktuális alapfrekvenciájának különbségével (az első szótagnál 0) jellemeztem.

A tanításhoz és a teszteléshez az itt felsorolt 6 jellemző aktuális, az előző és következő szótagjának konkatenált 18 jellemzőjét használtam. Az első szótagnál a megelőző szótag jellemzőit nullával helyettesítem, az utolsó szótagnál a következő szótag jellemzőit helyettesítem nullával, hogy minden esetben 18 jellemzőt kapjak.

Az egyes mondatokhoz tartozó hangsúlyképletet, a tanító minták esetén úgy alkalmaztam, hogy ha a képlet alapján egy szó hangsúlyos, akkor az adott szó első szótagjának súlyozása 1 a többi szótagjának és a hangsúlytalan szavak összes szótagjának a súlyozása 0. A mondatokhoz szegmentált anyag is tartozik, így adott volt az egyes hangok időzítése, ami alapján számolhatóak voltak az alapfrekvencia és relatív intenzitás értékek.

A tesztelés során a hangsúlymintázatok referencia szerepet töltöttek be. A tanítási folyamat után a tesztelésre kiválasztott mondatok eredményeit a hangsúlymintázataikkal vetettem össze, amikre a Pearson-korrelációs

eredmény 60,1%-ra adódott. Amennyiben a relatív intenzitást a pillanatnyi energiával helyettesítettem, a korreláció 54,6%-ra csökkent.

II. Tézis

Megvizsgáltam a hangsúlydetektáláshoz használt jellemzők hatékonyságát. Megállapítottam, hogy ha a magánhangzó pillanatnyi energiája helyett a relatív intenzitását használtam, mintegy 10%-kal nagyobb korrelációt értem el a vizsgált adatbázison a mondatok hangsúlyképletéhez viszonyítva.

Tézishez tartozó publikációim: [S1], [S10], [S11], [S12]

Újdonság

A hangsúlyadatbázisban az adatbázis megalkotói a mondat értelme alapján jelölték ki a hangsúlyos szavakat. Ezek általában nem esnek egybe a felolvasó hangsúlyozásával. Ugyannak a mondatnak a több bemondó által felolvasott változataihoz egyénileg határoztam meg a hangsúlyokat.

Mérések

A hangsúlydetektálás hatékonyságát a hangsúlyképlettel vetettem össze és Pearson-korrelációt számoltam. A hangsúlydetektálásánál általában használt pillanatnyi energiával szemben a magánhangzók relatív energiáját használva érdemi hatékonyság javulást értem el.

Érvényességi korlátok

A betanítás mindössze 10 beszélővel és beszélőnként 50 mondattal történt, bővebb adatbázis használatával a beszélő függetlenség és a hatékonyság javítható.

Következtetések

A szótag (magánhangzó) intenzitásának vizsgálatánál érdemes figyelembe venni az adott magánhangzó átlagos energiáját és ehhez

viszonyítani az aktuális energiát, ezzel elérhető, hogy a kis átlagenergiájú hangsúlyos hangok is megfelelő nyomatékot kapjanak.

1.5 A minősítési skála megalkotása

Az automatikus minősítés megvalósításának következő lépése a minősítési skála definiálása. Ehhez elengedhetetlen az 3.3 fejezetben ismertetett hangadatbázis, aminek mintái a beszédprodukció különböző fokán álló hallássérült diákoktól kerültek rögzítésre és az érthetetlentől az észrevehetetlen kiejtési hibáig átfogják a beszédminőség különböző szintjeit. A szurdopedagógusok szakmai szempontok szerint értékelték a bemondásokat. Nem szakértő (naiv) egyetemi hallgatók a hétköznapi nyelvhasználat szempontjából pontozták ugyanezeket a bemondásokat. Ezek a pontszámok képezik az egyes bemondások szubjektív minősítését. Az egyes szavak automatikus minősítésének célja, hogy a szubjektív értékelés eredményét minél jobban megközelítse.

Elvégeztem néhány elemzést és kiértékelést a teljes adatbázison, kiszámoltam a szórásokat és a minősítések átlagát. Összességében megállapítható volt, hogy az értékelések alapján a beszéd minőségének megítélésére nem fogalmazható meg olyan kritérium, amely alapján a minősítés egyértelmű lenne, ezért felkértem egy szakértőt, akinek a szakterülete a beszédfeldolgozás, hogy a minősített szavak egy szűkített csoportját elemezze szakmai szemszögből. Az kiértékelést végző szakembert megkértük, hogy elemezzen egy 300 szóból álló mintahalmazt. Tesztelései során a hanghiba és a ritmushiba súlyát próbálta meghatározni.

A minősítési skála meghatározásánál a legkisebb négyzetes hibát a hanghiba és a ritmushiba figyelembevételével a hanghiba súlyozó együtthatójára -2,78 és ritmushibáéra -0,51 adódott. Lineáris illesztést végeztem a szubjektív eredmények átlagára és a hanghiba és ritmushiba együttes alkalmazására, így az optimális együtthatók:

$$y=ax+bz+c, \text{ ahol } a=-2{,}78 \;\; b=-0{,}51 \;\; c=4{,}25 \qquad (11)$$

A negatív szorzók azt fejezik ki, hogy minél nagyobbak a hibák, annál gyengébb a minőség. A tapasztalatok és az eredmények azt mutatják, hogy hanghibára jóval érzékenyebb a szubjektív értékelő, mint a ritmushibára.

III. Tézis

Megvizsgáltam a hanghiba és ritmushiba hatását a szubjektív értékelés eredményeire. Meghatároztam a hanghiba és ritmushiba súlyozó együtthatóit az optimális lineáris illesztéshez.

Tézishez tartozó publikációim: [S1], [S2], [S3]

Újdonság

Hallássérült gyerekek beszédének elemzéséhez ilyen méretű adatbázist és ilyen részletes elemzést nem ismerek. A pedagógusok által meghatározott szakmai szempontok és a mérhető jellemzők összekapcsolására nem találtam irodalmi utalást.

Mérések

Gradiens módszerrel megállapítottam az optimális együtthatókat.

Érvényességi korlátok

A hangadatbázisban a nagyon rossz minőségű beszédre kevesebb példa található, mint a jobb minőségűekre. Bővítve a kevésbé érthető szavak listáját a megállapítás pontosítható.

Következtetések

Az automatikus minősítés kiindulópontja lehet a hanghibára és a ritmushibára vonatkozó súlyozó együttható.

1.6 Automatikus értékelés megalkotása

A bemutatott beszédasszisztens rendszer automatikus minősítési szolgáltatásának célja, hogy a hallássérült diákok önállóan gyakorolhassák a

mintaszavak kimondását. A tanulás során a referencia kiejtést a szerver vagy a tanár produkálja, a diák pedig ezt igyekszik utánozni az ő aktuális bemondásával. Az értékelés nyilvánvalóan a korábbi eredményekkel összevetve alakítható ki, hiszen ugyanaz a kiejtés egyik tanulónál siker, a másiknál kudarc lehet. Az automatikus értékelés verifikálása érthetőség vizsgálattal történhet.

Megvizsgáltam szokványos lényegkiemelési eljárásokat (MFCC, PLP, BARK) a hallássérült gyerekek bemondásainak elemzésére. A szegmentálási adatok alapján kijelöltem a hangok stacionárius szakaszát (amennyiben értelmezhető) és elvégeztem a lényegkiemelést. A stacionárius szakaszt a hanghoz tartozó időintervallum közepére helyeztem. Az aktuális jellemző vektorokat az értekezésben kifejtett teljes adatbázisra kiszámított jellemző vektorok átlagával vetettem össze. Így minden hangra kaptam egy távolság értéket. A szó jellemzésére a hangokra kapott távolság értékek átlagát vettem. A szavakra kapott átlagokat a szubjektív értékelés alapján kapott a csoportba tartozó szavakra átlagoltam.

A kapott távolságok nem követik következetesen az osztályok monotonitását. Nincsenek monoton összefüggésben az osztályok minősítésével, nem következetesek. MFCC ezredekben tér el és nem is teljesül a monotonitás.

A több mint 300 beszélős 4,5 órás hangadatbázissal betanított HMM modell HTK implementációjából kiolvashatók, hogy egy adott hangot milyen valószínűséggel generál a hozzátartozó HMM modell. A hallássérült gyerekek bemondásainak felismerési eredményeiből kiolvasott valószínűségek és a szó szubjektív értékelése között nem fedeztem fel korrelációt. Ennek oka lehet, hogy a felismerő szegmentálása nem volt megfelelő.

A jellemzők átlagához viszonyított távolság elemzés sikertelensége után ismét a már betanított neurális hálózat kimeneti aktivitásait kezdtem el vizsgálni. A neurális hálózat ideális esetben az adott hangra egységnyi, merőben eltérő hangra 0 kimeneti aktivitással válaszol. A helyesen artikulált

hangra nagy, a hibásan artikulált hangra kis kimeneti aktivitást produkál. A hasonlósági mértéket az adott hanghoz tartozó outputtal azonosítom. Megvizsgáltam a tanulmányozott szavak egyes hangjaihoz tartozó kimenetek átlagát. A szubjektív tesztek minősítésével vizsgált Pearson - korreláció 55,3% lett az átlagra. A szakértői hanghiba és a szubjektív pontok korrelációja -0,7515. A szubjektív pontok és a ritmus hiba korrelációja -0,2648.

A szubjektív tesztekkel az összehasonlítást részben korrelációs részben a számított pontszámok különbsége szerint vizsgáltam. Az összehasonlításhoz lineáris illesztést végeztem a szubjektív tesztek pontjai és a hasonlóságmérték között. A hasonlóság 0 és 1 közötti kimeneteket produkál, a szubjektív tesztek pontjai 1 és 5 közé esnek. Gradiens módszerrel megkeresem azt a szorzót és eltolást, amellyel a hasonlóság értéket korrigálva a szubjektív pontszámokkal a legkisebb négyzetes hibát adja. Az automatikus minősítés jóságát a szakértői értékeléssel vetem össze. A szakértői értékelés a hanghibát és a ritmushibát jelölte meg a minősítés alapjaként. A hanghibára és a ritmushibára megállapítom az optimális együtthatókat az $y = ax + bz + c$ kifejezésben, ahol x a hanghiba, z a ritmushiba 0 és 1 között értelmezett az értéke. Az előző 9. fejezetből adódóan a=-2,78 b=-0,51. Mivel ezek a hibák annál nagyobbak minél gyengébb minőségű az artikuláció az a és b szorzók negatívra adódnak. A legkisebb négyzetes hibát eredményező együtthatók meghatározása után megvizsgáltam, hogy a szakértői minősítés korrigált pontszámai mennyiben térnek el a szubjektív minősítés eredményétől. A részletes elemzés alá vetett 294 szó közül megvizsgáltam, hogy hány szó pontszám különbsége hány szónál kisebb a 11. táblázat oszlopaiban szereplő értékeknél. Az eredmények átlagosan 88,5 százalékos egyezőséget mutatnak.
Mivel a szubjektív pontszámok elég nagy szórást mutatnak, azt is megvizsgáltam, hogy a szakértői és az automatikus minősítés pontszáma hány szónál esik a hallgatói és a pedagógusi pontszám átlagok közé.

1. táblázat *A szakértői és az automatikus pontszámok referenciához mért különbsége intervallumokra bontva*

	<=0,1	<=0,2	<=0,5	<=1	<=1,5	<=2
Szakértői	*21*	*44*	*131*	*253*	*285*	*291*
Automatikus	*21*	*35*	*101*	*207*	*268*	*287*
(Automatikus/Szakértői)	*100%*	*80%*	*77%*	*82%*	*94%*	*99%*

A szakértői hanghiba és a ritmushiba optimális illesztésekor a 294 szóból a hallgatói és a pedagógusi átlagok közé 54 szó esik. Ugyanez a neurális hálózatok kimeneteinek szavankénti átlagnál és a ritmushibánál, vagyis az automatikus minősítésnél 44 szó.

Az automatikus minősítéshez használt átlagos neurális hálózat kimeneti aktivitás és a ritmushiba együtthatói a legkisebb négyzetes hiba esetén: 2,92 és -0,76. Az együtthatókból megállapítható, hogy a hanghibánál kevésbé megbízható neurális hálózat outputtal párosítva a ritmushiba nagyobb súlyozó együtthatót kap.

IV. Tézis

Több módszert megvizsgáltam a hallássérült gyerekek hangfelvételeinek elemzésére. Csak a hangfelismerésre betanított neurális hálózatok kimeneti aktivitására találtam differenciált és monoton eredményeket a különböző minőségi osztályokra. Módszeremmel a szubjektív értékelést a tolerancia tartományokban a szakértői becsléshez képest átlagosan 88,5 százalékos pontossággal közelítettem meg.

Tézishez tartozó publikációim: [S1], [S2], [S3]

Újdonság

Nem találtam irodalmi utalást arra, hogy a beszédfelismerésre betanított neurális hálózatok kimeneti aktivitását beszédminőség becslésre használták volna.

Mérések

Az automatikus értékelés eredményét a szakértői minősítéssel vetettem össze. Referenciaként a szubjektív értékelés pontszámait használtam.

Érvényességi korlátok

A betanított neurális hálózat nyelvfüggő a szegmentálás kritikus eleme az eljárásnak.

Következtetések

Az automatikus minősítés alkalmasnak látszik arra, hogy a beszédasszisztens rendszerben a gyakorló minták sorrendjének és nehézségi fokának meghatározásához inputként szolgáljon. Ugyancsak szándékomban áll a minősítés eredménye alapján audiovizuális visszajelzést generálni dicsérő és további gyakorlásra ösztönző üzenetek kiválasztására.

2. ÖSSZEFOGLALÁS ÉS TERVEZETT KUTATÁSI IRÁNYOK

Feladataim elsősorban a beszédprodukció minőségének automatikus értékelése köré csoportosultak. A beszédhangok helyes hangzásának értékeléséhez nélkülözhetetlen a hangzó beszéd szegmentálása beszédhangokra. Ez a feltétele annak, hogy a beszédnek azt a szegmensét hasonlítsuk össze egy referenciával, amely az illető hanghoz tartozik. A rejtett Markov–modell természeténél fogva kezelni tudja állapotainak többszöri ismétlődését, így egyes beszédintervallumok nyújtására és zsugorítására kiválóan alkalmas. A torz és gyakran érthetetlen beszéd hangjai azonban olyan távol esnek a Markov-modell állapotaitól, hogy a tiszta beszéddel tanított HMM felismerő nem volt képes a megfelelő pontosságú szegmentálásra. Amikor a HMM felismerőt neurális hálózat outputokkal tanítottuk, valószínűleg a sok ellentmondó adat miatt nem tudtuk betanítani a felismerőt. A beszédhangok fonetikai osztályaira és az osztályon belüli hangok megkülönböztetésére betanított neurális hálózatok kimeneti aktivitása lehetővé tette a torz beszédnél is a szegmentáláshoz elegendő hasonlósági értékek generálását. Az akadozó és ritmushibás beszéd szegmentáláshoz a dinamikus vetemítés algoritmusát célszerűen módosítottam. Ezzel a gyenge minőségű beszédre is jó szegmentálási eredményeket értem el.

Az automatikus minősítés referencia adatainak felvételéhez hangfelvételt készítettünk hallássérült gyerekek bemondásaival. A hangadatbázis minőségét a szurdopedagógusok által megadott szakmai szempontok alapján pedagógus és naiv értékelőkkel pontoztattam. Referenciának a szubjektív pontszámok átlagát tekintettük. A szubjektív minősítés nehézségét jelzi, hogy mind a szakértői mind a hallgatói pontszámok nagy szórást mutatnak. A pedagógus és a hallgatói pontszámok közötti különbség is gyakran jelentős. Szűkített szókészletre részletes elemzést adott egy erre felkért szakértő. Ezzel a több hetes munkával készült hang- és ritmushibát

számszerűsítő értékeléssel vetettem össze az automatikus minősítés eredményét. Az automatikus minősítéshez megkíséreltem szokványos távolság mértékeket generálni, ezek azonban nem mutattak egyértelmű kapcsolatot a különböző minőségi kategóriákkal. A szegmentálásnál alkalmazott neurális hálózatok kimeneti aktivitása differenciált és monoton összefüggést mutatott a minőségi osztályokkal. A hiba négyzetes középértékét minimalizálva illesztettem az automatikus minősítés pontszámait a szubjektív minősítés pontjaihoz. Eredményeimet szakértői értékeléssel hasonlítottam össze.

A szegmentálási jellemzőkön túl a szupraszegmentális tényezők is befolyásolják a beszéd érthetőségét. A prozódia értékelésének egyik szempontja a hangsúly a megfelelő szótagra helyezése. A hangsúly automatikus detektálására az alapfrekvencia mellett a magánhangzók relatív intenzitását használtam fel, amit az adott magánhangzó átlagenergiájához viszonyított pillanatnyi energiával értelmezek.

Munkám során, mivel a nagy adatbázissal tanított neurális hálózatok betanítása több órát vesz igénybe, így a neurális hálózat optimalizálását nem végeztem el. A fejlesztés során a hallássérült gyerekek bemondásairól felvételeket készítünk. Ezeknek a felvételeknek az elemzése a pedagógusi értékeléssel minősítve további tanító mintaként szolgálhat. Az automatikus minősítéshez a formáns struktúra elemzése hasznos kiegészítő lehet, ha a formánsok kinyerésére megbízható módszert találunk.

3. SUMMARY

Summary and future research directions

My main task has been to find a way for the program to be able to automatically evaluate the speech-reproduction of these children. For the proper evaluation the speech needs to be segmented into phonemes. This needs to be done in order to be able to compare the corresponding segment of the speech of the children to a reference speech. The hidden Markov-model can handle the continuous repetition of its states, which means that it is perfect for the elongation and shrinkage of the speech intervals. As the states of the Markov-model are very different from the malformed and – in most of the cases – incomprehensible speech, the HMM speech recognizer is not able to segment the speech to an adequate accuracy. When training the HMM speech recognizer with neural network outputs the contradictory data caused the training to fail. The output activity of the phoneme have been recognized by neural networks, which have been trained to be able to distinguish between the different phonetic classes of speech and the different phonemes in a given class, that has made it possible to generate similarity values for malformed speech patterns as well. The algorithm of the dynamic time warping has been modified in order to be able to segment the erratic and rhythm defective speech patterns as well. This has provided good segmenting results even in case of poor quality speech.

We have made voice records from speech of hearing impaired children to define the reference data for the automatic evaluation. The quality of the voice database has been classified with experts and non-experts considering technical aspects given from teachers. We considered the average of the subjective scores as a reference. Both the expert's and non-expert's scores show a large deviation which indicates the difficulty of subjective rating. The difference between the scores of the experts and non-experts are also often significant. I have attempted to generate standard distance scales for

the automatic evaluation but these did not show unequivocal relationship with different categories. The output activity of the neural networks applied in the segmentation phase has shown differentiated and monotone correlation with the evaluation classes. I have framed the automatic evaluation scores into the subjective evaluation scores by using minimum mean square error estimator. I have compared the results with that of experts' reviews.

The legibility of speech is influenced not only by the segmentation properties, but by the supra-segmentation factors as well. One of the evaluation criteria of the prosody is to place the emphasis on the proper syllable. To be able to automatically detect the emphasis, beside the fundamental frequency, I have used the relative intensity of vowels, which has been interpreted by comparing the average energy of a given vowel to its momentary energy.

As the training of neural networks with big databases takes several hours, the optimization of the neural network has not been performed. During the development phase voice recordings of hearing impaired children has been made. The analysis of these recordings, evaluated by pedagogues can be used as further training examples. The analysis of formant structure can be a useful additional tool for the automatic evaluation, if we find a reliable method for the extraction of formants.

I. Thesis

I have examined different methods of non-linear time warping and modified the connectivity rules of dynamic time warping for low quality speech by which significantly more limit values have fallen into each voice interval than in case of the methods I have examined.

II. Thesis

I have examined the efficiency of stress detection features. I have concluded that when using relative intensity of vowel instead of instantaneous energy,

then regarding the test database I have achieved almost 10% correlation in relation to the emphasis formula of the sentences.

III. Thesis

I have analyzed the effect of word error and tone error on results of subjective evaluation. I have determined the weight factor of word error and tone error to the optimal linear fitting.

IV. Thesis

I have tested several methods for analysing the voice recordings of hearing impaired children. Only the output activity of the neural networks trained for phoneme recognizing has given differentiated and monotone results for the various quality classes. As regards subjective rating in tolerant ranges, my method has reached 88.5 % accuracy in relation to expert rating.

AZ ÉRTEKEZÉS TÉMAKÖRÉBEN KÉSZÍTETT SAJÁT PUBLIKÁCIÓK

Folyóiratcikkek

[S1] Dr. Czap László, Pintér Judit Mária: A beszédasszisztens koncepció, Multidiszciplináris tudományok- A Miskolci Egyetem közleménye, (2013) 3. kötet. 1 sz. pp. 241–250. HU ISSN 2062-9737

[S2] Bodnár Ildikó, Czap László, Pintér Judit: Kutatási projekt a hallássérültek internetes beszédfejlesztésére, Alkalmazott Nyelvészeti Közlemények 2014. VIII. évfolyam 2. szám, pp. 19–32. ISSN 1788-9979

[S3] Csetneki Sándorné Dr. Bodnár Ildikó, Czap László, Pintér Judit: Számítógéppel segített beszédfejlesztés, Modern Nyelvoktatás (2014) XX. évfolyam, 4.szám: pp. 75–86. ISSN 1219-628X

[S4] Dr. Czap László, Pintér Judit Mária: Az akusztikus és vizuális jel aszinkronitása a beszédben, Multidiszciplináris Tudományok- A Miskolci Egyetem közleménye (2014), 4. kötet 1.szám, pp. 67–76., HU ISSN 2062-9737

[S5] Czap László, Pintér Judit: A hangsúly egyik jellemző modalitásának vizsgálata, Alkalmazott Nyelvészeti Közlemények 2014. IX. évfolyam 1. szám, pp. 114–121., ISSN 1788-9979

Konferenciaközlemények

[S6] Dr Czap László, Pintér Judit Mária: Beszédfelismerés hatékonyságának vizsgálata különböző nyelvtanokkal, XVII. Fiatal Műszakiak Tudományos Ülésszaka, Műszaki Tudományos füzetek, 2012, pp. 71–74, ISSN 2067-6 808

[S7] Dr Czap László, Pintér Judit Mária: A szavakon túli kommunikáció az audiovizuális beszédszintézisben cikk, XVII. Fiatal Műszakiak Tudományos Ülésszaka, Műszaki Tudományos füzetek, 2012, pp. 67–70, ISSN 2067-6 808

[S8] Czap, L.; Pinter, J. M.: Improving Performance of Talking Heads by Expressing Emotions, Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference, 2012, pp. 523–526, E-ISBN : 978-1-4673-5188-1; Print ISBN: 978-1-4673-5187-4

[S9] Laszlo Czap, Judit Maria Pinter: Multimodality in a Speech Aid System - International Conference on Human Machine Interaction (ICHMI 2013) VOLUME 01 pp.6–11, ISBN-13: 978-81-925233-1-6, ISBN-10: 81-925233-1-4

[S10] Dr. Czap László, Pintér Judit Mária: A beszédprodukció automatikus minősítése hallássérültek beszélni tanításához, XVIII. Fiatal Műszakiak Tudományos Ülésszaka, Műszaki Tudományos füzetek, 2013, pp. 99–102, ISSN 2067-6 808

[S11] László Czap, Judit Mária Pintér: Relative Intensity for Stress Detection; 8. International Scientific Conference on Mechanical Engineering COMEC 2014, Cuba, 5 p., Paper ISBN: 978-959-250-997-9

[S12] Dr Czap László, Pintér Judit Mária: Beszédfelismerés hatékonyságának vizsgálata különböző nyelvtanokkal, XVII. Fiatal Műszakiak Tudományos Ülésszaka, Műszaki Tudományos füzetek, 2012, pp. 71–74, ISSN 2067-6 808

[S13] Dr. Czap László, Pintér Judit Mária: Gyenge minőségű beszéd szegmentálása, XX. Fiatal Műszakiak Tudományos Ülésszaka, Műszaki Tudományos füzetek, 2015, pp. 119–122, ISSN 2067-6 808

Független hivatkozások

[H1] Illésné Kovács Mária: Pozitív és negatív visszajelzések hallássérültek internetes beszédfejlesztésében, Alkalmazott Nyelvészeti Közlemények IX. évfolyam 1. szám, ISSN 1788-9979, 2014. pp. 135–143.

[S1] Dr. Czap László, Pintér Judit Mária: A beszédasszisztens koncepció, Multidiszciplináris tudományok- A Miskolci Egyetem közleménye, 3. kötet. 1 sz. HU ISSN 2062-9737, 2013. pp. 241–250.

[S2] Bodnár Ildikó, Czap László, Pintér Judit: Kutatási projekt a hallássérültek internetes beszédfejlesztésére, Alkalmazott Nyelvészeti Közlemények VIII. évfolyam 2. szám, ISSN 1788–9979, 2014. pp. 19–32.

[S3] Csetneki Sándorné Dr. Bodnár Ildikó, Czap László, Pintér Judit: Számítógéppel segített beszédfejlesztés, Modern Nyelvoktatás XX. évfolyam, ISSN 1219-628X, 4.szám. 2014.pp. 75–86.

IRODALOMJEGYZÉK

[1] Abari K., Olaszy G.: *Magyar hangsúlyadatbázis az interneten kutatáshoz és oktatáshoz*, MSZNY 2014. pp.347–356.

[2] Ali Zilouchian: *Fundamentals of Neural Networks*, CRC Press LLC, 2001.

[3] Anne-Marie Öster: *Computer-Based Speech Therapy Using Visual Feedback with Focus on Children with Profound Hearing Impairments*, [Doctoral Thesis], Stockholm Sweden: KTH Computer Science and Communication, 2006.

[4] Baker, J., Deng, L., Glass, J., Khudanpur, S., Lee, C. H., Morgan, N.: *Updated MINDS Report on Speech Recognition and Understanding*, Part I. IEEE Signal Processing Magazine 26/3. 75–80. Part II. IEEE Signal Processing Magazine 26/4. 2009. pp.76–85.

[5] Ben. K., Patrick, S.: *An Introduction to Neural Networks*, University of Amsterdam, 1996.

[6] Bolla Kálmán: *A Phonetic Conspectus of Hungarian*, Tankönyvkiadó, Budapest. 1995.

[7] C.Jeyalakshmi, Dr.V.Krishnamurthi , Dr.A.Revathy: *Deaf Speech Assessment Using Digital Processing Techniques*, Signal & Image Processing : An International Journal(SIPIJ) Vol.1, No.1, 2010. pp.14–25.

[8] Crystal David: *A nyelv enciklopédiája,* Budapest: Osiris. 1998.

[9] Czap L.: *Audio-Visual Speech Recognition And Synthesis*, Phd Thesis, Budapest University Of Technology And Economics, 2004.

[10] Czap, L., Kovács, Zs., Tóth, Á., Váry, Á.: *A beszédasszisztens használata hallássérültek egyéni beszédfejlesztésében* (Közlés alatt)

[11] Csányi Yvonne: *Bevezetés a hallássérültek pedagógiájába*, B. G. Gyógypedagógiai Tanárképző Főiskola Budapest, 1998.

[12] Csányi Yvonne: *Tanulmányok a hallássérültek beszéd-érthetőségének fejlesztéséről*, Bárczi Gusztáv Gyógypedagógiai Tanárképző Főiskola Budapest, 1995.

[13] Csányi, Y., Zsoldos, M., Perlusz, A.: *Hallássérült (hallásfogyatékos) gyermekek, tanulók komplex vizsgálatának diagnosztikus protokollja*, Budapest: Educatio Társadalmi Szolgáltató Nonprofit Kft., 2012.

[14] Denkinger Géza: *Valószínűségszámítás*, Nemzeti Tankönyvkiadó, 2001.

[15] Deza, E, Deza, M.: *Dictionary of Distances*, Elsevier, ISBN 0444520872, 2006.

[16] Faragó, A., Fülöp, T., Gordos, G., Magyar, G., Osváth, L, Takács, Gy.: *Egyszerű izolált szavas beszédfelismerő,* Kutatási anyag, 1985.

[17] Farkas Miklós: *A hallássérültek kiejtés- és beszédfejlesztésének elmélete és gyakorlata*, B. G. Gyógypedagógiai Tanárképző Főiskola Budapest, 1996.

[18] Farkas, M., Perlusz, A.: *A hallássérült gyermekek óvodai és iskolai nevelése és oktatása*, In: Illyés Sándor (szerk.) Gyógypedagógiai alapismeretek. Budapest: ELTE Bárczi Gusztáv Gyógypedagógiai Főiskola, 2000.

[19] Fujisaki, H., Ohno, S.: *The Use of a Generative Model of F0 Contours for Multilingual Speech Synthesis*, Fourth International Conference on Signal Processing, Vol. 1, 1998. pp. 714–717.

[20] Futó Iván: *Mesterséges intelligencia*, Aula Kiadó, 1999.

[21] Gordos, G., Takács, Gy.: *Digitális beszédfeldolgozás*, Műszaki Könyvkiadó, Budapest, 1983.

[22] Gósy Mária: *Fonetika, a beszéd tudománya,* Osiris, Budapest, 2004. pp.182–243.

[23] Győrffy, P., Bődör, J.: *1978. Gyógypedagógiai ismeretek a hallási fogyatékosság köréből*, Budapest: Tankönyvkiadó, 1978

[24] Hermansky Hynek: *Perceptual linear predictive (PLP) analysis for speech*, Journal of the Acoustical Society of America, 87(4), 1990. pp.1738–1752.

[25] Huang, X., Deng, L.: *Overview of modern speech recognition*. In Indurkhia, Nitin – Damerau. Fred (eds.): Handbook of natural language processing. CRC Press Boca Raton, London–New York. 2010.

[26] Illyés Sándor: *Gyógypedagógiai alapismeretek*, ELTE Bárczi Gusztáv Gyógypedagógiai Főiskolai Kar Budapest, 2000.

[27] Imai, S.: *Cepstral analysis synthesis on the mel frequency scale*, IEEE International Conference on ICASSP '83. (Volume:8), Yokohama, Japan , 1983. pp. 93–96.

[28] J.D. Subtelny: *Speech assessment of the deaf adult*, J. Acad. Rehab. Audiol., 8 (1&2), 1975. pp. 110–116.

[29] Kálmán, Zs., Könczei Gy.: *A Taigetosztól az esélyegyenlőségig*. Budapest: Osiris Kiadó. 2002.

[30] Karl Pearson: *Notes on regression and inheritance in the case of two parents*, Proceedings of the Royal Society of London (58), 1895. pp. 240–242.

[31] Kárpáti Árpád Zoltán *Gondolat – ébresztő! Egy apa feljegyzései a siket gyermekek oktatásáról,* Budapest: Underground Kiadó. 2011.

[32] Kassai Ilona: *Fonetika*, Nemzeti Tankönyvkiadó, Budapest, 1998.

[33] Kiss, G., Vicsi, K.: *Akusztikai hangosztályok felismerésén alapuló, nemlineáris idővetemítés megvalósítása a mondathanglejtés és a*

szóhangsúlyozás oktatásához; Beszédkutatás 21, 2013. pp.247–261.

[34] Kun L., Xiaojun Q., Shiyin K., Helen M.: *Lexical Stress Detection for L2 English Speech Using Deep Belief Networks*, INTERSPEECH, 2013. pp. 1811–1815.

[35] Mády Katalin: *Beszédpercepció és pszicholingvisztika*, Pszicholingvisztikai kézikönyv, 2008.

[36] Maier A., Haderlein T., Eysholdt U., Rosanowski F., Batliner A., Schuster M., Nöth E.: *Peaks – A System for the automatic evaluation of voice and speech disorders*, Speech Communication, 2009.

[37] Maier A., Hönig F., Hacker C., Schuster M., Nöth E.,: *Automatic evaluation of characteristic speech disorders in children with cleft lip and palate*, Proc. of 11th Int. Conf. on Spoken Language Processing, Brisbane, Australia, pp. 1757–1760.

[38] Markó Alexandra: *A spontán beszéd néhány szupraszegmentális jellegzetessége*, PhD értekezés, Eötvös Lóránd Tudományegyetem, Budapest, 2005.

[39] Martin Kroul: *Automatic Detection of Emphasized Words for Performance Enhancement of a Czech ASR System*, SPECOM'2009, St. Petersburg, 21-25 June 2009. pp.470–473.

[40] Molnár József: *The Map of Hungarian Sounds*, Tankönyvkiadó, Budapest, 1986.

[41] Neha P. Dhole, Dr. Ajay A.Gurjar: *Detection of Speech under Stress*, A Review , International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 10, 2013. pp.36–38.

[42] Németh, G., Olaszy, G.: *A Magyar Beszéd*, Akadémiai Kiadó, Budapest, 2010.

[43] Olaszy Gábor: *Az alapfrekvencia és a hangsúlyozás kapcsolata a magyarban*, In: Kísérleti fonetika - Laboratóriumi fonológia 2002. (szerk.: Hunyadi László) Kossuth Egyetemi Kiadó, Debrecen, 2002.

[44] Olaszy, G., Abari, K., Bartalis, M.: *Magyar hangsúlyjelölési szöveges adatbázis fejlesztése és referenciavizsgálata*, Beszédkutatás 2014. pp. 205–219.

[45] Pytel József: *Audiológia*, Budapest: Victoria Kft., 1996.

[46] Rabiner, L. R.: *A tutorial on hidden Markov models and selected applications in speech recognition*, Proceedings of the IEEE, 1989. 77(2):257–286.

[47] Rabiner, L., R.: *On the use of autocorrelation analysis for pitch detection*. IEEE Transactions on Acoustics, Speech, and Signal Processing, Volume 25, No. 1. 1977. pp. 24–33.

[48] Ross, M. et al.: *Average magnitude difference function pitch extractor*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Volume 22, No. 5. 1974. pp. 353–62.

[49] Szaszák György: *A szupraszegmentális jellemzők szerepe és felhasználása a beszédfelismerésben,* PhD értekezés, Budapesti Műszaki és Gazdaságtudományi Egyetem, Budapest, 2008.

[50] Tarnóczy Tamás: *Zenei akusztika*, Zeneműkiadó, Budapest, 1982. pp. 151–82.

[51] Van Kuijk, D., Boves, L.: *Acoustic characteristics of lexical stress in continuous telephone speech*, Speech Communication, 27(2), 1999. pp. 95–112.

[52] Ververidis, C. Kotropoulos: *Emotional speech recognition: resources, features and methods*, Speech Commun., 48 (9), 2006. pp. 1162–1181.

[53] Vicsi, K., Szaszák, Gy.: *Automatic Segmentation of Continuous Speech on Word Level Based on Supra-segmental Features*,

International Journal of Speech Technology, Vol. 8, Num. 4, 2005. pp. 363–70.

[54] Waibel Alex: *Prosody and Speech Recognition*, Pitman, London, UK. 1988.

[55] William M. Hartmann: *Signals, Sound, and Sensation*, American Institute Of Physics, 2004. ISBN 1-56396-283-7.

[56] Xie, H., Andreae, P., Zhang, M., Warren, P.: *Detecting stress in spoken English using decision trees and support vector machines*, In: Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation, Australian Computer Society, Inc. 2004. 145–150.

[57] Young, S. et al.: *The HTK Book (For Version 3.3)*, Cambridge University, 2005.

[58] Zwicker, E.: *Subdivision of the audible frequency range into critical bands*, The Journal of the Acoustical Society of America, 33, Feb., 1961.