Alkalmazott Nyelvészeti Közlemények, Miskolc, IX. évfolyam, 1. szám (2014) pp. 144-153.

SZÁMÍTÓGÉPPEL TÁMOGATOTT PROZÓDIAOKTATÓ PROGRAM

COMPUTER-AIDED PROSODY TEACHING PROGRAMME

SZTAHÓ DÁVID[1]–KISS GÁBOR[2]–TULICS MIKLÓS GÁBRIEL[3]–
CZAP LÁSZLÓ[4]–VICSI KLÁRA[5]

Az elmúlt évtized tapasztalata azt mutatja, hogy a számítógéppel támogatott kiejtésoktató alkalmazások (Computer-Assisted Pronunciation Teaching, röviden: CAPT) hasznos és rugalmas eszközök a helyes kiejtés oktatásában és a beszélő beszédének kiértékelésében. A cikk egy újonnan fejlesztett kiejtésoktató alkalmazást mutat be, melynek fő célja olyan szupraszegmentális paraméterek megfelelő tanítása, mint az intonáció, hangsúly és a ritmus. Két modul került megvalósításra: (1) az intonáció és hangsúlyoktató, valamint (2) a dinamikus idővetemítést alkalmazó ritmusoktató modul. A rendszer automatikus visszajelzésének helyességét gyakorló pedagógusok ítéletével hasonlítottuk össze, mégpedig úgy, hogy szubjektív lehallgatási kísérletben siket és nagyothalló gyerekek beszédmintáit a pedagógusok a kiejtés minősége szerint csoportokba sorolták. Az automatikus kiértékelési rendszer visszajelzése összhangban van a tanárok szubjektív döntéseivel. A cikk továbbá bemutatja a megvalósított dinamikus idővetemítésen alapuló vizuális visszajelző rendszert, amely egyszerű és érthető vizuális információt nyújt a beszélő által bemondott intonációról és ritmusról.

Kulcsszavak: prozódia, intonáció, beszédfelismerés, beszédoktatás

The experience of the last decade shows that Computer-Assisted Pronunication Teaching (CAPT) applications are useful and flexible tools in teaching correct pronunciation and evaluating speech. The paper presents a newly developed pronunciation teaching application targeted at the proper teaching of such suprasegmental parameters as intonation, stress and rhythm. Two modules have been implemented: (1) the intonation and stress teaching module, and (2) the rhythm teaching module applying dynamic time warping. The correctness of the automatic feedback of the system was tested using the judgements of practising teachers in the way that in the subjective hearing experiment, the teachers grouped the speech samples of deaf and hearing impaired children according to their

[1] SZTAHÓ DÁVID
Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék
sztaho@tmit.bme.hu
[2] KISS GÁBOR
Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék
kiss.gabor@tmit.bme.hu
[3] TULICS MIKLÓS GÁBRIEL
Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék
tulics@tmit.bme.hu
[4] CZAP LÁSZLÓ
intézetigazgató, egyetemi docens
Miskolci Egyetem, Villamosmérnöki Intézet
Automatizálási és Infokommunikációs Intézeti Tanszék
3515 Miskolc-Egyetemváros
[5] VICSI KLÁRA
Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék
vicsi@tmit.bme.hu

pronunciation quality. The feedback of the automatic evaluation system was in line with the teachers' subjective judgements. The paper also presents the implemented visual feedback system, based on dynamic time warping, which provides simple and clear visual information about the intonation and rhythm produced by the speaker.

Keywords: prosody, intonation, speech recognition, pronunciation teaching

Bevezetés

Az elmúlt évtized tapasztalata azt mutatja, hogy a számítógéppel támogatott, kiejtésoktató alkalmazások hasznos és rugalmas eszközök a kiejtés oktatásában és a beszéd paramétereinek kiértékelésében. Ez azonban (a kiejtésoktatás) számos problémába ütközik, valamint sok pedagógiai és technológiai kérdés is felmerül. Technológiai szempontból nehéz érthető és helyes visszajelzést nyújtani a kiejtési hibákról, szinte lehetetlen száz százalékos pontosságú és automatikus diagnózist adni.

Neri és munkatársai szerint a CAPT-alkalmazásoknak három elengedhetetlen tényezőnek kell eleget tenniük: (1) elegendő mennyiségű, értelmezhető bemeneti információt szükséges biztosítaniuk, pontos artikulációs útmutatással, (2) legyen cél és motiváció, hogy a tanulók végigkövessék a szabályorientált gyakorlatokat és feladatokat, valamint (3) biztosítsanak azonnali hasznos visszacsatolást, legfőképp azokról a funkciókról, amelyek az érthetőséggel kapcsolatosak (Neri et al. 2002: 441–467). A tanítási folyamatnak tartalmaznia kell olyan szupraszegmentális jellemzőket, mint az intonáció, kiejtési hossz variáció vagy szóhangsúly (Derwing–Rossiter 2003: 1–18).

Bradlow, Pisoni, Akahana-Yamada, valamint Tohkura tanulmányában bemutatta, hogy a percepcióban való fejlődés produkcióban lévő fejlődéshez vezet (Bradlow 1997: 2299–2310), Hirata tanulmánya pedig azt is megmutatta, hogy az ellentéte is igaz, azaz a produkcióban való fejlődés percepcióban való előrehaladáshoz vezet (Hirata 2004: 357–376).

A CAPT-rendszer vizuális visszacsatolása többnyire egy mesterséges arc (Cole et al. 1998; Massaro 1998), valamint a hanghullám, az alaphang és spektrogram megjelenítésével (l. Hiller et al. 1993; Watson et al. 1989; ISLE 1993) [8, 9, 10], továbbá automatikus beszédfelismeréssel (automatic speech recognition, röviden ASR) valósul meg (l. Narusa 1999). Ez segít felismerni a szavakat, továbbá a beszédben rejlő időviszonyok megértéséhez is hozzájárul. Pedagógiai szempontból a spektrogramok és a hullámformák vitatottak. Ezeknek a megjelenítési formáknak a megértése túl sok időbe telik és nem éri meg a befektetett energiát. Ennek ellenére a spektrogram segíthet kiemelni (hangsúlyozni) az energiaváltozást vagy például a különböző szótagok eltérő időbeli megvalósulását. Az alaphang megjelenítése sokkal egyszerűbb és érthetőbb. Nem világos az, hogy az érthetőség javítására mely visszacsatolási formák a leghasznosabbak. Nincs egyetértés abban, hogyan kellene a prozódiát mérni és tanítani.

A kiejtési hibák automatikus diagnózisa a CAPT-rendszerek elvárt funkciója, viszont egy általános, százszázalékos pontosságú rendszer nem reális. Az az elvárás, hogy a rendszer által diagnosztizált hibák korreláljanak az emberi megítéléssel!

Korábban, az európai SPECO projekten belül (szerződésszám: 977126) egy termékorientált, többnyelvű CAPT-rendszert fejlesztettünk ki beszédhibás gyerekek számára. Az alkalmazás a beszéd akusztikai tulajdonságainak vizuális megjelenítését használta fel az akusztikailag szempontból helyes jellemzők alapján, ugyanakkor könnyen érthető és a gyermekek számára érdekes volt (Vicsi et al. 2000; Vicsi 2001; Vicsi et al. 2006). A rendszer elsősorban a hang szegmentális leírását használja.

A jelen cikk egy újonnan fejlesztett CAPT-rendszert mutat be, melynek célja az olyan szupraszegmentális paraméterek helyes oktatása, mint az intonáció, hangsúly és ritmus. Az alkalmazás hallássérült, valamint cochleáris implantátummal rendelkező gyerekek prozódiatanítását célozza meg, de hasznos lesz a számítógéppel támogatott nyelvtanulásban is.

A gyerekek motivációjának fenntartása érdekében a program egy érthető, hasznos és hatékony tanulási módszert kísérel meg megvalósítani, vizuális és automatikus visszajelzés felhasználásával. Két modul került megvalósításra: (1) az intonációt és hangsúlyt oktató, valamint (2) a dinamikus idővetemítést alkalmazó ritmusoktató modul. A rendszer automatikus visszajelzésének helyessége siket és nagyothalló gyerekek beszédmintái segítségével került kiértékelésre, amelynek során a rendszer automatikus válaszát tanárok szubjektív döntéseivel hasonlítottuk össze.

Az alkalmazás egyelőre magyar nyelvű hanganyagot használ a prozódia tanítására, viszont nagyon könnyen átültethető más nyelvekre. A cikk a következőképpen épül fel. Az 1. szakaszban a javasolt CAPT-rendszer kerül bemutatásra, ezt követi a 3. és 4. szakasz, amely részletesen taglalja az értékelési módszereket. Az 5. szakaszban a CAPT-alkalmazás értékelése és a tanárok által formált szubjektív értékelés került összehasonlításra nagyothalló gyerekek beszédmintái alapján.

1. A számítógéppel támogatott kiejtésoktató alkalmazás általános funkcionalitása

A javasolt, számítógéppel támogatott kiejtésoktató alkalmazásnak két modulja van. Az egyik az intonáció automatikus visszajelzésére, azaz az alaphang időbeli változására vonatkozik. A magyar nyelvben a mondatok modalitásai leginkább tipikus intonációs görbék által jelennek meg (alapfrekvencia-görbék). Azoknak a gyerekeknek, akiknek problémáik vannak a helyes intonációjú beszéddel, meg kell tanulniuk a modalitások helyes intonációját, különben kevésbé lesz érthető a beszédük. Az alkalmazásban is használt, a modalitásokhoz tartozó hanglejtéstípusokat az 1. táblázat foglalja össze.

Az alkalmazás másik modulja a ritmussal foglalkozik. Segítségével a felhasználók, például a nagyothalló gyerekek, fejleszthetik a kiejtésüket azáltal, hogy időzítési hibáikról automatikus kiértékelést kapnak. A ritmus kiértékelése két paraméteren alapszik: (1) a magánhangzók időtartamán, valamint (2) a szomszédos magánhangzók időbeli távolságán. A kiejtett fonémák pontos időbeli információinak meghatározására automatikus szegmentálási eljárást használunk kényszerített illesztéssel (Kiss et al. 2003).

1. táblázat

Mondatok modalitásaihoz tartozó tipikus intonációs görbék táblázata

Intonáció típusa	**Mondat típusa**
Ereszkedő	kijelentő mondat
Eső	kiegészítendő kérdés
Emelkedő-eső	eldöntendő kérdés
Eső-ereszkedő	felszólító és felkiáltó mondatok
Lebegő	mellékmondat
Szökő	egyszavas kérdés

2. Az intonáció automatikus értékelése

Annak érdekében, hogy hatékony legyen az intonáció tanulása, helyes alaphangszámító módszerre, az automatikus kiértékelő és vizuális visszajelzés implementációjára van szükség. A CAPT-alkalmazásban javasolt módszereket az alábbiakban mutatjuk be.

Az alaphang kiszámítása autokorrelációval történt 10 ms-os lépésközzel és 100 ms-os nagyságú számítási ablakkal. Az így kapott görbe oktávugrásokat kiküszöbölő szűrőn ment keresztül, majd átlag-szűrővel simítottuk.

Az intonáció helyessége a beszéddallam változásának a függvénye. Ebből következően az automatikus kiértékelőnek csak az intonációs görbe deriváltjának információját kell figyelembe vennie, és nem annak abszolút értékét. A javasolt automatikus intonáció-kiértékelő algoritmus összehasonlítja a bemondó beszédmintáit egy referencia bemondó intonációs görbéjével.

Az intonáció változását a kiejtéstől függetlenül akarjuk megjeleníteni, mivel a bemondások hossza az intonáció helyességét nem befolyásolja. Ezért a mondatokat lineárisan vetemítettük. Az algoritmus részekre bontja a referencia intonációs görbét olyan időpontok szerint, ahol az alaphang függvényben hirtelen változások vannak. A hirtelen változás időpontját (t_{change}) az alaphang függvény második deriváltjának maximumaként definiáltuk.

Az algoritmus összehasonlítja a két beszédminta (beszélő által bemondott és referencia) változásának irányát a referencia t_{change} időpontjai által alkotott szakaszokban. Az i-edik rész változásának iránya a következőképpen van megadva:

$$dir_i = sign \frac{\sum_{t=t_i}^{t_{i+1}} \frac{\partial f_0(t)}{\partial t}}{t_{i+1} - t_i + 1}, \qquad (1)$$

ahol t_i az alaphang görbe hirtelen változásának i-edik időpontja. A beszélő által bemondott intonációs görbe valamely része akkor minősül helyesnek, ha a két beszédminta (beszélő által bemondott és referencia) dir_i értéke egyenlő, ellenkező esetben helytelennek lesz ítélve. Az intonáció helyességének végső pontszáma a helyes és helytelennek ítélt részek aránya.

3. Ritmus

A helyes ritmus a folyékony beszéd alapkövetelménye. A túl hosszú vagy túl rövid fonéma időtartamok csökkenthetik a beszéd érthetőségét, vagy teljesen érthetetlenné tehetik azt. Emiatt a fonémák helyes időzítése nagyon fontos. A javasolt CAPT-rendszerben a ritmus kiértékelése két időinformáción alapszik: (1) a magánhangzók időtartamán, valamint (2) a szomszédos magánhangzók időbeli távolságán. A kiejtett fonémák pontos időbeli információinak meghatározása kényszerített illesztéssel kiegészített szegmentálási eljárással történik.

3.1. Kényszerített illesztéssel kiegészített szegmentálási eljárás

A beszédfelismerő rendszerek általános rendeltetése a folyamatos beszéd felismerése. Tehát nem arra tervezték őket, hogy a fonémák pontos időbeli pozícióját megállapítsák, ami bármely beszédtanító alkalmazás alapkövetelménye. Olyan felismerési eljárást mutatunk be, amely fonémák felismerése helyett a fonémák különböző akusztikai jellemzői szerint fonémacsoportokat képes elkülöníteni egymástól. Ezek a következők: mély (mély és középállású) magánhangzók, magas magánhangzók, zöngés és zöngétlen mássalhangzók, spiránsok és a csend. Ezek a kategóriák némi egyszerűsítéssel elegendők ahhoz, hogy

minden európai nyelvet lefedjünk, ezáltal ez a módszer szinte nyelvfüggetlen megoldásnak tekinthető. A bemondott fonémák előzetes ismeretével, valamint kényszerített illesztéssel képesek vagyunk a pontos időzítési információkat meghatározni.

A kényszerített illesztés fő célja, hogy a bemondott fonémasorozat ismeretében megtalálja a fonémák pozícióját a beszédjelben. Az ismert fonémasorozat előbb fonémacsoportok szerint kerül átírásra, mivel a felismerő a fonémák különböző akusztikai jellemzők szerinti csoportosításával dolgozik.

Legyen $h_S = h_{s_1}^{j_0,j_1}, h_{s_2}^{j_1,j_2}, \dots, h_{s_n}^{j_{n-1},j_n}$ egy valószínűségsorozat, ahol $h_{s_n}^{t_1,t_2}$ az s_n fonémacsoport előfordulási valószínűsége t_1 időpillanatról t_2-re, valamint n az adott fonémacsoport összes fonémájának a száma. Ha figyelembe vesszük a $j_0, j_1, \dots, j_n \in J$ időindexek összes kombinációs lehetőségét, azzal a megkötéssel, hogy $0 \leq j_0 \leq j_1 \leq \cdots \leq j_n \leq t_{max}$, akkor a cél a legnagyobb valószínűségű S sorozat megtalálása:

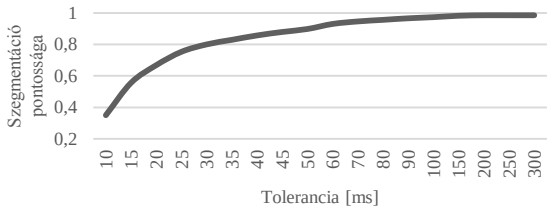
$$S_{best} = \underset{S}{argmax}\, h_S = \underset{S}{argmax}\, h_{s_1}^{j_0,j_1} \cdot h_{s_2}^{j_1,j_2} \cdot \dots \cdot h_{s_n}^{j_{n-1},j_n} . \quad (2)$$

A $h_{s_n}^{t_1,t_2}$ valószínűségeket a következőképpen számoljuk ki:

$$h_{s_n}^{t_1,t_2} = \prod_{t=t_1}^{t_2} a_{s_n,b_t} , \quad (3)$$

ahol b_t a bemutatott eljárás által felismert fonémaosztály t időpontban, a_{s_n,b_t} pedig az előzetes állapotátmeneti valószínűség s_n fonémaosztályból b_t fonémaosztályba.

A kényszerített illesztés után meghatározzuk az eredeti fonémasorozatot a legjobb illesztett fonémaosztály-sorozat alapján. Az 1. ábra mutatja a szegmentáció pontosságát annak függvényében, hogy a kiértékeléskor mekkora időbeli toleranciát engedünk meg.



1. ábra. Szegmentáció pontossága a tolerancia függvényében

3.2. A ritmus automatikus kiértékelése

A program a szegmentációs eljárás által generált fonémák időbeli pozíciója alapján kétfajta automatikus pontszámot számol ki, amely a bemondó ritmusának helyességére vonatkozik. Mindkét pontozási eljárás a Magyar Referencia Beszédadatbázisból kinyert referencia fonéma-hosszúságokat használja.

Az első mérési pontozás a magánhangzó-hosszúságon alapszik. A bemondó beszédmintájának minden v_m magánhangzójára a következő sc_{v_m} értéket számítjuk ki:

$$sc_{v_m} = \begin{cases} 1 & \text{if } d_{min} \le d_{v_m} \le d_{max} \\ 0 & \text{if } d_{v_m} < w \cdot d_{min} \text{ or } d_{v_m} > \dfrac{d_{max}}{w} \\ 1 - \dfrac{d_{max} - d_{v_m}}{\frac{d_{max}}{w}} & \text{if } d_{max} < d_{v_m} < \dfrac{d_{max}}{w} \\ 1 - \dfrac{d_{v_m} - d_{min}}{d_{min} \cdot w} & \text{if } d_{min} \cdot w < d_{v_m} < d_{min} \end{cases}, \quad (4)$$

ahol d_{v_m} a v_m magánhangzó időtartama; d_{min} és d_{max} rendre a referencia adatbázisból mért minimális és maximális időtartam v_m magánhangzóra. w 0 és 1 közötti konstans, amellyel a kiértékelés szigorúságát lehet beállítani (jelenlegi értéke: 0.8). A hangminta magánhangzó- időtartamra vonatkozó végleges pontértékét a következő összefüggés adja meg:

$$sc_v = \frac{\sum_{v_m} sc_{v_m}}{n_v}, \quad (5)$$

ahol n_v az összes magánhangzó száma.

A második pontozási eljárás a szomszédos magánhangzók időbeli távolságán alapszik. A (4) képlethez hasonló módon számolható a pontérték, az egyes intervallumok viszont itt az egyes magánhangzók között elhelyezkedő mássalhangzók hosszúságának felelnek meg. Az intervallumokra a következő módon határozható meg a pontozás:

$$sc_{c_m} = \begin{cases} 1 & \text{if } d_{min} \le d_{c_m} \le d_{max} \\ 0 & \text{if } d_{c_m} < w \cdot d_{min} \text{ or } d_{c_m} > \dfrac{d_{max}}{w} \\ 1 - \dfrac{d_{max} - d_{c_m}}{\frac{d_{max}}{w}} & \text{if } d_{max} < d_{c_m} < \dfrac{d_{max}}{w} \\ 1 - \dfrac{d_{c_m} - d_{min}}{d_{min} \cdot w} & \text{if } d_{min} \cdot w < d_{c_m} < d_{min} \end{cases} \quad (6)$$

ahol d_{c_m} a v_m és v_{m+1} magánhangzó közötti időtartam; d_{min} és d_{max} a referencia adatbázisból számított referencia minimum és maximum időtartam. w pedig ugyanaz a konstans, mint a (4) összefüggésben. A magánhangzók közötti intervallumokra vonatkozó végleges pontértéket a következő képlet adja meg:

$$sc_c = \frac{\sum_{c_m} sc_{c_m}}{n_c}, \quad (7)$$

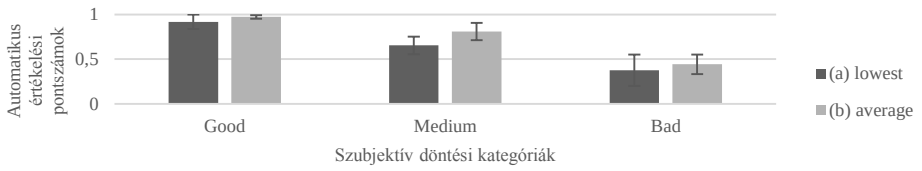
ahol n_c a magánhangzók közötti intervallumokat jelenti.

4. Az automatikus értékelés minősítése

A korábbi automatikus értékelési pontszámokat 19 hallássérült gyermek beszédmintája segítségével értékeltük ki. Valamennyi kategóriából minden gyerek három mondatot olvasott fel. A felvételek a Dr. Török Béla Általános Iskolában készültek, Budapesten. 48 véletlenszerűen kiválasztott mondatot (az 1. táblázat kategóriái szerint egyenletes eloszlással) három tanár értékelt szubjektíven a kiejtés helyessége szerint, majd három osztályba sorolták őket: jó, közepes és rossz (a minták száma az egyes osztályokban: a

felvételek 10, 40 és 50%-a). A pedagógusok a beszédkárosult gyerekek tanításának szakértői. Ugyanezek a mondatok egészséges gyerekekkel is rögzítésre kerültek, és a „jó" értékelést kapták. A szubjektív értékeléseket összehasonlítottuk az előző fejezetekben bemutatott kiértékelésekkel (intonáció, egészséges hangmintát használva referenciaként; magánhangzók időtartama; a szomszédos magánhangzók időbeli távolsága) két módon: (a) a háromfajta kiértékelési módszerrel kapott pontszámok közül a legalacsonyabbat és (b) azok átlagát használtuk az összehasonlítás során. A szubjektív és az automatikus értékelés viszonyát a 2. ábra mutatja.

A számítógéppel támogatott kiejtésoktató alkalmazás által adott pontszám megfelel a három szubjektív kategóriának. Az automatikus pontozási értékek (az automatikus visszajelzés legalacsonyabb és átlagos pontozása) az egyes szubjektív kategóriák esetében eltérnek, az automatikus visszajelzés legalacsonyabb pontszáma szigorúbb ítéletet jelent, amely jobban megfelel a tanárok által megadott szubjektív értékelésnek. Ez maga után vonja azt a következtetést, hogy a teljes kiejtés észlelésének helyességét akár a kiejtés egyetlen helytelen paramétere is meghatározhatja.



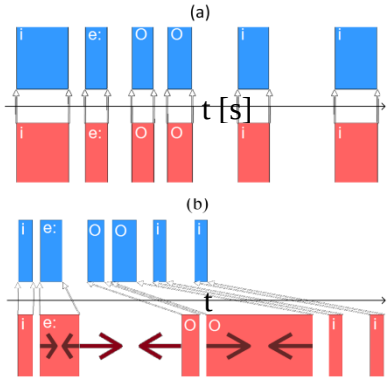
2. ábra. Automatikus és szubjektív értékelés korrelációja nagyothalló gyerekek hangmintáira; Az automatikus kiértékelés (a) legalacsonyabb (lowest) és (b) átlagos (average) pontszáma.

5. Vizuális visszacsatolás

A megfelelő vizuális visszacsatolás minden CAPT-rendszerben alapkövetelmény. A vizuális visszacsatolásnak fenn kell tartania a felhasználó figyelmét és motivációját, ugyanakkor elég hatásosnak kell lennie ahhoz, hogy tanítson.

Javaslatunk kétfajta vizuális visszacsatolást foglal magában: intonációs görbék összehasonlítását, valamint a referenciahang és a bemondó ritmusának összehasonlítását. A két hangminta hossza különböző, ezért dinamikus idővetemítést használunk. A ritmus pontszámait a (4) és (6) összefüggésekkel számoltuk valamennyi magánhangzóra, valamint magánhangzók közötti intervallumra. Minden olyan magánhangzó hosszát, amire a (4) összefüggés 1.0 értéket ad, a referenciában szereplő hozzátartozó fonémának a hosszértékére állítjuk. Hasonlóan, ha egy intervallumra a (6) összefüggés 1.0 értéket ad, a magánhangzók közötti intervallum hossza a referenciában szereplő hozzátartozó intervallum hosszértékére van állítva. Ha (4) és (6) értéke 1.0-től különböző, akkor a magánhangzóhosszak és a köztes intervallumok időértékeit változatlanul hagyjuk.

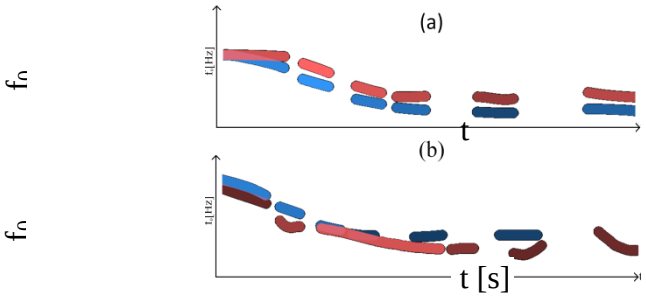
Ennek az lesz az eredménye, hogy ha a bemondó beszédmintáját a program jónak ítéli, akkor az időbeli paraméterei meg fognak egyezni a referencia időbeli paramétereivel. Ellenkező esetben, ha helytelen ritmust észlelünk, akkor a hiba megfelelő pozíciója a bemondó számára nyilvánvaló lesz. A 3. ábrán két, eltérőnek ítélt bemondást tüntettünk fel.



3. ábra. Ritmus vizuális visszajelzése. (a) helyes ritmus; (b) példa helytelen magánhangzóhosszra és magánhangzók közötti intervallumhosszra. A magánhangzók időbeli pozícióját téglalapokkal ábrázoltuk (a fonémák Sampa-karakterekkel vannak jelölve). Az ábrák mindegyikén a referenciahang magánhangzóit a felső sor, a bemondó beszédmintáit pedig az alsó sor tartalmazza. A bemondó ritmusában a hibák helye és iránya nyilakkal van jelölve.

Az intonáció vizuális visszajelzése az intonációs görbék időbeli változásán alapszik. Az intonációs görbéket az alaphang első három értéke átlagának kivonásával normalizáltuk (minden görbénél).

A referenciát és a beszélő beszédmintájának normalizált alaphangját ugyanazzal az idővetemítési eljárással jelenítettük meg, mint a ritmus esetében. A vizuális visszacsatolás a felhasználó bemondásának helyes és helytelen időzítési paramétereit (időbeli pozíció, időtartam), valamint intonációját is képes megjeleníteni. A 4. ábra helyes és helytelen intonációkat mutat.



4. ábra. Intonáció vizuális visszajelzése (a) helyes intonáció; (b) példa helytelen intonációra. Színjelölések: kék – referencia, vörös – a beszélő beszédmintája. A vonalak a magánhangzók időbeli elhelyezkedését jelölik.

Az általunk használt vizuális visszacsatolás, amely egyszerre jeleníti meg az intonációt és a ritmust (valamint ezek helyességét), újszerű és nagyon kifejező. Nagyothalló gyerekekkel szerzett első tapasztalataink alapján a rendszer a prozódia tanulására egyszerű módszert biztosít.

6. Tárgyalás

Az 5. szakasz értelmében az intonáció- és ritmuskiértékelési módszerek összhangban vannak a tanárok szubjektív döntéseivel. Számos nyitott kérdés van azonban a számítógéppel támogatott kiejtésoktató alkalmazás használatát illetően, főképp, ha gyerekek a bemondók. Ilyen például a bemondó motivációjának fenntartása, valamint az elvégzendő feladatok kérdése.

A 3. és 4. ábra olyan egyszerű, de könnyen érthető elemeket tartalmaz, amelyek ahhoz szükségesek, hogy a gyerekek ne veszítsék el az érdeklődésüket az oktatási rendszer használata során. Amellett, hogy megfelelő vizuális és automatikus visszajelzést használunk, az alkalmazásnak olyan célzott, jól felépített feladatokat kell tartalmaznia, amelyek a gyermekek figyelmét fenntartják, ugyanakkor elősegítik kiejtésük fejlődését. Ezeket a feladatokat a tanárokkal együttműködve terveztük meg. Számos gyakorlat a gyerekek intonáció- és ritmustanulását különböző mondattípusok beépítésével segíti. A program kötött és szabad párbeszédet is tartalmaz.

A nagyothallók iskolájában, a Dr. Török Béla Általános Iskolában a siket és nagyothalló gyerekek sikeresen használják az alkalmazást, és az eddigi visszajelzések szerint szeretik azt. A tanulórendszer, valamint a vizuális visszajelzés hatékonyságának hivatalos hatásvizsgálata még zajlik, eredmény néhány hónap múlva várható.

7. Következtetések

Cikkünk egy újonnan fejlesztett kiejtésoktató alkalmazást mutat be, melynek fő célja olyan szupraszegmentális paraméterek megfelelő tanítása, mint az intonáció, a hangsúly és a ritmus. Két modul került megvalósításra: (1) az intonáció- és hangsúlyoktató, valamint (2) a dinamikus idővetemítést alkalmazó ritmusoktató modul. A rendszer automatikus visszajelzését siket és nagyothalló gyerekek beszédmintái segítségével értékeltük ki. A kiértékelő rendszer automatikus visszajelzése és pontozása összhangban van a tanárok szubjektív döntéseivel. Az eredmény azt mutatta, hogy a rendszer szigorú (az automatikus pontszámok legalacsonyabb értéke) minősítése a leginkább megfeleltethető a szubjektív ítéleteknek. Egy dinamikus idővetemitésen alapuló vizuális visszajelzést is bemutattunk, amely a beszélő által bemondott intonációról és ritmusról egyszerű és érthető vizuális információt nyújt. A megtervezett gyakorlatokat tartalmazó alkalmazás a nagyothallók iskolájában, a Dr. Török Béla Általános Iskolában kerül bevezetésre Budapesten, de rendszeres használatát és szakszerű értékelését Egerben és Debrecenben is tervezzük, a gyerekek fejlődését több hónapon keresztül figyelemmel kísérve.

Irodalom

BRADLOW, A.–PISONI, D.–YAMADA, R. A.–TOHKURA, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*. 101 (4): 2299–2310.

COLE, R. et al. 1998. *Intelligent animated agents for interactive language training*. ESCA-STILL 98, Marholmen, Sweden, 1998. 163–166.

DERWING, T.–ROSSITER, M. 2003. The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning*. 13 (1): 1–18.

HILLER, S.–ROONEY, E.–LEFEVRE, J. P.–JACK, M. 1993. SPELL: an automated system for computer-aided pronunciation teaching. *Proceedings of Eurospeech*. 93. Berlin.

HIRATA, Y. 2004. Computer-assisted pronunciation training for native English speakers learning Japanese pitch and duration contrasts. *Computer Assisted Language Learning*. 17(3–4): 357–376.

ISLE. *"Interactive Spoken Language Education" Annual Report* 1999. http://www.ec-isle.org. 1993.

KISS, G.–SZTAHÓ, D.–VICSI, K. 2013. Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features. *4th IEEE International Conference on Cognitive Infocommunications* – CogInfoCom 2013. 579–582.

MASSARO, D. W. 1998. *Perceiving talking faces: from speech perception to a behavioural principle*. Cambridge, MA: MIT Press.

NARUSA, J. 1999. Computer-aided spoken language training with enhanced visual and auditory feedback. *Eurospeech*'99. Budapest, 183–186.

NERI, A.–CUCCHIARINI, C.–STRIK, H.–BOVES, L. 2002. The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*. 15(5): 441–467.

VICSI, K.–ROACH, P.–ÖSTER, A.–KACIC, Z.–BARCZIKAY, P.–TANTOS, A.–CATÁRI, F.–BAKCSI, Zs,–SFAKIANAKI, A. 2000. A multimedia, multilingual teaching and training system for children with speech disorders. *International Journal of Speech Technology*. Vol. 3. Kluwer Academic Publisher, 289–300.

VICSI, K. et al. 2001. A multilingual multimodal Speech Training System SPECO *Eurospeech* 2001. Aalborg, Sweden. 2807–2810.

VICSI, K. 2006. Computer Assisted Pronunciation Teaching and Training Methods Based on the Dynamic Spectro-Temporal Characteristics of Speech. In: Divenyi, P.–Greenberg, S.–Meyer, G. (eds.) *Dynamics of Speech Production and Perception*. Amsterdam: IOS Press, 283–307.

VICSI, K.–KOCSOR, A.–Teleki Cs.–Tóth, L. Hungarian Reference Speech Database (MRBA)
http://alpha.tmit.bme.hu/speech.

WANG, X.–MUNRO, M. 2004. Computer-based training for learning English vowel contrasts. *System*. 32: 539–552.

WATSON, C. S.–REED, D. J.–KEWLEY-PORT, D.–MAKI, D. 1989. The Indiana Speech Training Aid features. *Journal of Speech and Hearing Research*. 07/1989; 32(2): 245–251.

Köszönetnyilvánítás

Ezt a munkát részben az Európai Unió és az Európai Szociális Alap támogatta a FuturICT.hu TAMOP-4.2.2.C-11/1/KONV-2012-0002. és a TAMOP-4.2.2.C-11/1/KONV-2012-0013 számú projekten keresztül, melyet a VIKING Zrt., Balatonfüred szervezett.

Köszönjük Dr. Váry Ágnesnek és a Dr. Török Béla Általános Iskola munkatársainak: Tóth Ágnesnek, Kovács Zsuzsannának, valamint Szőlősiné Sípos Virágnak a program fejlesztésében és kiértékelésében nyújtott segítségüket.