

Gépi tanulás és újszerű alkalmazása a beszédjavítás-terápiában

A new computer aided software package, the „SpeechMaster” was developed for teaching reading and speech impediment therapy. The program is freely available in the Internet at www.inf.u-szeged.hu/beszedmester. Its substance is a machine learning module, which converts the sounds into easily comprehensive graphical elements such as flickering graphemes. This effect efficiently supports the development of the correct articulation of the hearing impaired. The aim of the reading-teaching part of the software is to help children learn, more quickly and easily, the correct phone-grapheme and grapheme-phone associations using the computer as a means of motivation, and by using automatic speech recognition methods. Besides briefly describing the „SpeechMaster” package, in this paper we also summarize our results from the viewpoint of machine learning.

1 Bevezetés

Napjaink felgyorsult, technikai eszközökkel telezsúfolt világában a fogyatékkal élőknek egyre nagyobb gondot okoz lépést tartaniuk az egészségesekkel. A technikai fejlődés legszembetűnőbb tényezője a multimédiás eszközök előretörése: bárhová megyünk, mindenütt képek és hangok bombáznak bennünket. Ezzel párhuzamosan egyre nagyobb szerepet kap a beszédkommunikáció: egyelőre ugyan még mi magunk beszélünk és válaszolunk (vagy hagyunk üzenetet) kommunikációs készülékeinkkel, de a beszédértő és beszéd útján válaszoló intelligens eszközök megjelenése már a küszöbön van. Ezek a technikák kellemesebbé és könnyebbé teszik az átlagember életét, de a beszédértési vagy beszédképzési nehézségekkel küzdőknek újabb és újabb kihívásokat jelentenek. Ésszerűnek tűnik, hogy ugyanazt a technikát hívjuk segítségül, amely a gondokat is kiváltotta. A „Beszédmester” szoftver fejlesztése is így indult: egy olyan program megírását tűztük ki célul, amely segíti a hallássérültek beszédképzésének terápiáját. Ez a munka hagyományosan óriási türelmet és a szurdopedagógus állandó jelenlétét igényli, ugyanis a helyes hangképzés rögzítéséhez rengeteg ismétlés és a pedagógus folyamatos korrekciója szükséges. Ezt az ún. automatizálási folyamatot nagyban segítheti egy megfelelő szoftver, amellyel a gyakorlás részben önállóan is végezhető.

A program tervezése közben jelezték pedagógus kollégáink, hogy ugyanezt a technológiát az olvasás tanításában is fel lehetne használni. Mind két területben a fonológiai tudatosság fejlesztése a cél (Adams 1990). Napjainkban egyre több az olvasási nehézséggel küzdő gyermek, amit sokan annak tulajdonítanak, hogy a televízió és a számítógép háttérbe szorította a könyveket (Péchy 1974). Kézenfekvő lenne, ha a megoldást is a számítógép szolgáltatná, olyan szoftver formájában, amely játékos úton, szinte észrevétlenül gyakoroltatja az olvasást.

A fenti két alkalmazási terület igényeit szem előtt tartva fogtunk hozzá a „Beszédmester” beszédjavítás-terápiái és olvasásfejlesztő rendszer készítéséhez. Abból indultunk tehát ki, hogy a hallás-, vagy részképességükben sérült gyermekek kiejtési problémáit, ha nem is oldaná meg, de lényegesen csökkenthetné egy olyan technika, amellyel a saját hangjukat, kiejtésüket önmaguk ellenőrizhetnék, javíthatnák. A számítógép felruházása automatikus beszédfelismerési képességekkel pótolhatja ezt a hiányzó auditív kontrollt, illetve az ép hallású iskoláskorúak olvasásfejlesztése során hatékonyabbá teheti a beszédhangok és grafémák helyes asszociációját.

A program kulcseleme tehát a beszédtechnológiai modul, amely lehetővé teszi, hogy a rendszer a mikrofonba bemondott beszédhangokat, illetve szavakat valós időben visszajelezz.

Az automatikus beszédfelismerés a gépi tanulás nevű kutatási témakör egyik sajátos részfeladata. Ilyen irányú kutatásaink során olyan modern gépi tanulási módszereket vizsgáltunk meg, amelyek alapköveit jelenthetik más, gépi tanulást igénylő korszerű infokommunikációs rendszerek kifejlesztésének is. Ezekre a technológiákra építve alakítottuk ki „OASIS” névre keresztelt általános célú beszédfelismerő rendszerünket. Az általunk adott eljárás is, mint a beszédfelismerő eljárások általában, a teljes problémát kisebb, a gép által könnyebben megoldható részproblémákra vezeti vissza. Esetünkben ez a beszédhangok felismerését jelenti. A beszédhangokból épített szavak vagy mondatok felismerését egy általunk kifejlesztett, úgynevezett szegmentális alapú dekódolási algoritmusra építettük, amely kellően rugalmas ahhoz, hogy különböző felismerési stratégiákat könnyen ki lehessen próbálni benne. Az eljárás nem a szokásos néhány századmásodperces egységekben gondolkodik, hanem hosszabb időintervallumokban. Ez a kibővített megközelítés eredményezi azt, hogy a módszer magában foglalja a hagyományos technológiákat is, mint speciális esetet.

A beszédhangok felismerése nem közvetlenül a beszédjelből történik, hanem valamiféle tömör reprezentációt kell előbb előállítani, amely a felismerés szempontjából fontos komponenseket megőrzi, a zavaró tényezőket viszont lehetőség szerint eltávolítja. Az előfeldolgozás után előálló reprezentáció már alkalmas arra, hogy a gépi tanuló algoritmusok hatékonyan működjenek rajtuk. Az eredmények azonban tovább javíthatóak ún. transzformációs algoritmusokkal, amelyek kiválasztják a fontosabb jellemzőkomponenseket, elnyomják a tanuló algoritmusok hatékonyságát zavaró egyéb hatásokat, és olyan térbe viszik át az adatokat, ahol a szeparálhatóság könnyebb.

Amint az a rövid áttekintésünkben látható volt, a felismerési probléma lényegében négy lépésre bontható: a jellemzőkinyerés, a jellemzőtér-transzformáció, a beszédhang-osztályozás és a szószintű felismerés feladataira. Mind a négy lépésben kipróbáltunk több, az irodalomból is ismert eljárást, de új ötletekkel is előálltunk.

A cikk felépítése a következő: a 2. fejezetben röviden ismertetjük a „Beszédmester” rendszert, majd a 3.-ban összefoglaló képet adunk a beszédfelismerésben alkalmazott matematikai módszerekről. A publikációt egy rövid összegzés és irodalomjegyzék zárja.

2 A „Beszédmester”

Ebben a fejezetben röviden áttekintjük a „Beszédmester” szoftver felhasználásai lehetőségeit és funkcióit mind a beszédjavítás-terápia mind az olvasásfejlesztés szempontjából. A részletesebb és több képpel illusztrált leírását egy korábbi cikkben találhatunk programról. (Paczolay 2004)



1.ábra: A „Beszédmester” nyitóképe

2.1 A „Beszédmester” a beszédjavítás-terápiában

A hallás-, vagy részképességükben sérült gyermekek kiejtési problémáinak logopédia kezelése négy egymást követő terápiai lépésben történik: előkészítés, hangfejlesztés, rögzítés, és végül az automatizálás. A program által felkínált, terápiát segítő lehetőségek ugyancsak ezekhez a lépésekhez kapcsolódnak, így a beszédjavítás-terápiai rész tagolódása megegyezik a felsorolt terápiai lépésekkel.

2.1.1 Előkészítés

Ebben a szakaszban, amely a korai életkorban a beszédnevelés kezdetét jelenti, játékos formában kell a kisgyermek figyelmét saját hangadásukra irányítani. Rá kell szoktatni a gyermekeket, hogy figyeljék a képernyőt, igyekezzenek utánozni a bemutatott hangadási mintákat, pl. hangerősség, hangmagasság változtatása, különböző ritmusformák utánzása, zöngés zöngétlen hangok gyakorlása, stb. Az ebben a részben kialakított játékos elemek igen alkalmasak arra, hogy a gyermekek hozzászokjanak és elsajátítsák a program kezelését, megtanuljanak kommunikálni a programmal.

A megvalósított funkciókat a következő felsorolás tartalmazza:

- Zöngeadás: a képzett hangunk zöngésségével irányíthatunk egy autót – zöngétlennél megáll, zöngésnél pedig felgyorsul.
- A hangerő változtatása: egy bohóc a hangerő függvényében alacsonyabbra vagy magasabbra ugrik. Cél a bohóc (a hangerő) lécek által kijelölt szinten tartása meghatározott ideig. A lécek számát és magasságát (az elvárt hangerőt) a pedagógus állítja be előre.
- A ritmus gyakorlása: a tanító mikrofonba bemozdott ritmikus hangsorát egy ugráló béka nyoma rögzíti a homokos parton. A tanuló feladata a ritmus megismétlése, annak elérése, hogy a béka újra a nyomokban ugráljon végig.

- Állathangok utánzása: a pedagógus utasítása és ellenőrzése mellett a kiválasztott állatok hangját utánozni kell a foglalkozáson résztvevő gyermeknek. A gép itt nem értékel, csak képeket kínál fel a gyakorláshoz.
- A hangmagasság változtatása: a feladat első körében virágokat lehet elültetni az alsó, illetve a felső sorba a hang magasságának megfelelő megválasztásával. A második körben a gyakorló személynek a virágágyást kell megloccsolnia úgy, hogy az előző mély és magas hangokból álló sorozatot megpróbálja megismételni.
- Hangringatás: először egy kitarva ejtett hang magasságának folyamatos változtatásával egy vakond járatot váj a földben. Magas hang esetén feljebb, mély hang esetén lejjebb halad az alagút ásásával. Gyakorláskor ebben a járatban kell végigvezetni egy kisebb vakondot.

2.1.2 Hangfejlesztés



2. ábra: A „Beszédmester” beszédjavítás-terápiai részének egyik feladata

A sikeres előkészítési szakaszt követően kitarított ejtéssel és hangkapcsolatokban kell kialakítani a magánhangzók és a folyamatosan ejthető mássalhangzók helyes artikulációját. A kiejtés és az olvasás tanítása egyszerre történik: a kiválasztott hangzók gyakorlása során a helyesen ejtett hangzó képe az ejtés ideje alatt megjelenik a képernyőn úgy, hogy a kijelzés fényereje a megfelelés mértékével arányos (azonos idejű vizuális visszacsatolás). Ha rendelkezünk számítógéphez csatlakoztatott kamerával, akkor a kamera által közvetített kép is megjelenik a kijelzés háttérében, hasonlóan – megfelelő beállításokkal – a tükör módszer szájáról olvasási képéhez (2. ábra). A gyakorolt hangok, vagy hangkapcsolatok írott képe a képernyő alsó sávjába gépelhető, illetve a későbbi felhasználás érdekében el is tárolható.

A gyakorlásra kijelölt hangzók kiejtésében előforduló hibák elemzéséhez beállítható, hogy hibás hangképzéskor több grafémát is jelenítsen meg a gép (pl. átmeneti hangok esetén kettőt), illetve lehetőség kínálkozik arra, hogy a beszédhang-felismerés időbeli változását grafikonok segítségével ábrázoljuk.

2.1.3 Rögzítés

A fejlesztőmunka eredményeként a tanuló elérheti azt a szintet, hogy önállóan gyakorolhat. Ekkor a képernyőn három kiválasztott, gyakorolni kívánt szó képe és a fogalom neve jelenik meg. A többi gyakorlóképet kis méretben, betűnként csoportosítva kínálja fel a gép. A gyakorlásra kiválasztott hangzók egy-egy szó kimondásakor az aktív képmezőben a kiejtésnek megfelelően betűképek formájában, a megfelelés mértékével arányos módon villannak fel.

2.1.4 Automatizálás

Az automatizálási lépés célja rövidebb, hosszabb szövegek érthető felolvasása, illetve ennek gyakoroltatása a tanulókkal. Maga a szöveg lehet előre rögzített, de lehetőség van tetszőleges szöveg begépelésére és eltárolására is. A gyakorlás kezdeti fázisában a tanuló számára hosszú mondatok további, általuk is feldolgozható blokkokra tagolhatók. A gyakorlásra kiválasztott beszédhangok a szövegben egyedi színnel jelennek meg, csakúgy mint az aktuálisan soron következő blokk vagy mondat. A feladatban nem folyamatos beszédfelismerés történik, hanem az ejtés tisztaságának objektív értékelése a cél. Ha a kijelölt beszédhangok közül a következőt helyesen képzi a tanuló, akkor a graféma kijelölése eltűnik, és a következő kijelölt hangzóra fókuszál a beszédfelismerő.

2.2 A „Beszédmester” az olvasásfejlesztésben

A „Beszédmester” szoftver az olvasástanítás segítségét, az olvasás terápiáját, fejlesztését is célul tűzte ki (Bácsi 2003). A szoftver segítségével játékos úton, szinte észrevétlenül lehet gyakoroltatni az olvasást (3. ábra). Használható az iskolai olvasástanítás során és egyéni gyakorlásra. A részképességükben sérült gyermekek fejlesztő terápiájában komplex készségfejlesztést biztosít: memóriafejlesztés, figyelemfejlesztés, irányfelismerés, iránytartás kialakítása, finommotorika fejlesztése, hallási diszkriminációs készség, hallási figyelem, vizuális differenciáló képesség fejlesztése. Segítheti a diszlexiaterápiát, hiszen a betűk újratanításának feltételei biztosítottak, a fonémapárok feladatai pedig célzottan a kritikus párok gyakorlására, differenciálására készültek (4. ábra). A tesztelési eredmények azt mutatják, hogy nemcsak az első osztályosok fejlesztésére, tanítására alkalmas a szoftver, hanem a 8-10 évesek terápiájában is kiválóan használható.

A „Beszédmester” szoftver olvasástanítást segítő funkciói betűtanítási (magán-, és mássalhangzók) és fonémapár tanítási, illetve gyakoroltatási részre tagolódnak. Az olvasástanítás folyamatához hat különböző feladat kapcsolódik, amely a betűk tanításán a saját, metodikailag megalapozott sorrendjében halad végig. Az önálló tanulást és gyakorlást a betűnként azonos típusú feladatok biztosítják a gyermekek számára, amely önálló munka már önmagában is sikerélményt nyújt mind az egészséges, mind a részképességükben sérült gyermekek számára. A következőkben végigtekintjük a szoftver betűtanítási feladattípusait.

2.2.1 Betűtanítás

Hang- és betűtanítás

A képernyőn a tanult új betű hívóképe jelenik meg, és a szoftver a bemondás alapján vagy a graféma képét rajzolja ki a megfelelés mértékével arányos módon magánhangzók esetében, vagy mássalhangzók esetében a helyesen bemondott teljes szót írja ki. A feladat így alkalmas a szókincs fejlesztésére, az ejtéshibák megfigyelésére, illetve korrigálására.

Új fonéma keresése tárgyképek nevében

A megjelenített tárgyképek alatt a szóban található hangok számának megfelelő „karika” található. A tanult hang előfordulásait kell megkeresni a tárgy nevében, és a vélt előfordulások helyén lévő karikákat ki kell színezni egérkattintással. A helyes megoldásokat a gép kipipálja. A képek összeállításánál az Adams által felállított szintek lettek figyelembe véve, így a feladatmegoldás során a könnyebben azonosítható pozíciótól haladunk a nehezebb felé.

Graféma kiválasztása

A feladat első részében a tanult betű és annak különböző elforgatott változatai jelennek meg, és ebből a sorozatból kell a tanulónak kiválasztani a tanult betű eredeti alakját. A feladat segítségével már korán felismerhető, ha egy gyermek vizuális rotálási problémákkal küzd, ezen kívül fejleszti a forma és alakfelismerést, az irányok azonosítását és az iránytartást.

A második részben az eddig tanult grafémák egy véletlen sorozatából kell kiválasztani a tanult betűt. Egyrészt a feladat segíti a betű alakjának további rögzülését, de főként a gyermeket a helyes olvasási irány megtartására szoktatja.

Graféma-fonéma megfeleltetés 1.

Az első körben a képernyőn megjelenik a tanult graféma hívóképe az eddig tanult betűk egy sorozatával együtt. A betűsorból az összes, a tanult betűnek megfelelő grafémát rá kell húzni a hívóképre ahhoz, hogy a feladat befejeződjön. Második körben már négy, eddig tanult hívókép kerül megjelenítésre a hozzájuk tartozó, betűkkel együtt. Minden hívóképre fel kell helyezni a neki megfelelő kezdőhangokat. A gyakorlatok rögzítik a hangleválasztást, illetve segítik a fonéma-graféma megfeleltetést.

Graféma-fonéma megfeleltetés 2.

Az előző feladat nehezítéseként hívóképek helyett tárgyképek jelennek meg, így a hangokat a tárgyak nevében kell megkeresni, majd a hangokat a megfelelő helyre húzni. A feladatok menete megegyezik az előzőével. A gyakorlat a hallási diszkrimináció készséget fejleszti, elősegíti a hangokra való bontást, illetve a graféma-fonéma megfeleltetést.

Olvasógyakorlat

Az újonnan tanult graféma olvasása 12 elemű betűsorban mikrofon használatával. Magánhangzó esetén csak a magánhangzókat kell olvasni, míg mássalhangzók esetén kétbetűs szótagokat olvastatunk. A szótagok hibátlan olvasása után következik a szó, majd a mondatolvasás szakasza, ahol már az ellenőrzést a tanító végzi.

A feladatsor sikeres teljesítése után a gyermekek jutalomjátékként memóriajátékot játszhatnak, amely a program kezdetekor az egérhasználat és a rajzos jelölések tanulására is szolgál.

2.2.2 Fonémapárok tanítása

A tanult betűk, hangok ismételése, újratanulása.

A képernyőn a fonémapárokhoz tartozó betűk hívóképei jelennek meg, és a szoftver a bemondás alapján vagy a graféma képét rajzolja ki a megfelelés mértékével arányos módon magánhangzók esetében, vagy mássalhangzók esetében a helyesen bemondott teljes szót írja ki. A feladat így alkalmas az ejtészabályok megfigyelésére, illetve korrigálására.

Betűk kiválasztása

A feladat célja, hogy a sor elején megjelölt betűt a gyermek kiválassza a többi, már tanult betű közül. A feladat nehézségét jelenti, hogy a graféma mindig a kritikus betűcsoport betűi között található meg. Második részben a kiemelt grafémát már nem betűk közül kell kiválasztani, hanem azokat a szavakat kell megjelölni, amelyben megtalálható. A feladat során a különböző pozícióban (szó elején, közepén, végén) lévő grafémákat felismerését lehet gyakorolni, így a betűket nem csak különálló helyzetben, betűszinten tudja majd a gyermek megkülönböztetni, hanem szószinten is.

Hang azonosítása tárgyképek nevében

A képernyőn egyszerre mindig két kép jelenik meg, illetve a kiválasztott fonémapár betűi. A betűket a megfelelő képekre kell ráhelyezni aszerint, hogy melyik képen lévő tárgy nevében melyik hangot halljuk. A gyakorlat a hallási diszkriminációs képességet, a hangokra bontást fejleszti.

Betűfeltöltés

A fonémapár betűi a képernyő közepén nagyban, csak körvonalukkal jelennek meg, a képernyő alján az ezeknek megfelelő betűk egy sorozata. Az eger segítségével a kis, színes betűket rá kell húzni a neki megfelelő, középen található kitöltetlenre: minden jó választás esetén egy-egy újabb réteggel töltődik fel a körvonalas betű. A gyakorlat alkalmas a vizuális differenciálós képesség fejlesztésére, gyakoroltatására, illetve az optikai különbségek megfigyeltetésére.

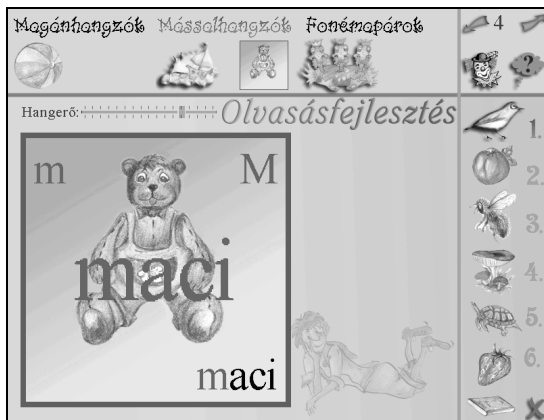
Olvasógyakorlat

A képernyőn megjelenő szótagsorokat, illetve magánhangzók esetén betűsorokat kell felolvasni. A megjelenő „szavak” kezdő-, vég-, illetve mindkét pozíciójában a kritikus betűcsoport egy-egy tagja található. A gyakorlat célja a graféma-fonéma megfeleltetés gyakorlása, rögzítése, a tiszta beszédhangjejtés megszilárdítása.

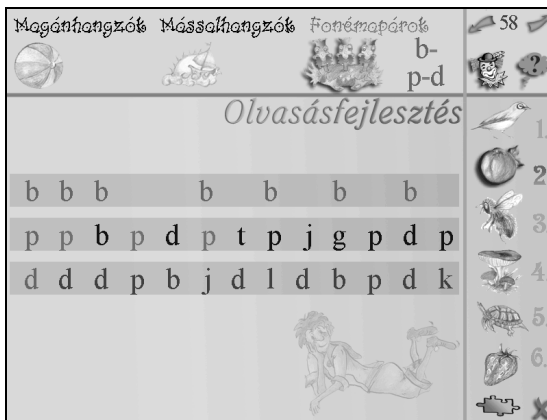
Betűpótlás

A képernyő felső részén a fonémapár betűi, alatta két-, illetve hárombetűs hiányos szavak és szótagok jelennek meg. Az egeret a hiányos szavak fölé mozdítva elhangzik a teljes szó, ez alapján kell a megfelelő grafémával pótolni a hiányzó részt. A gyakorlat a fonéma-graféma megfeleltetést gyakoroltatja, fejleszti a tiszta fonémahallást, a hallási diszkriminációs képességet és a hallási figyelmet.

A feladatsorok sikeres elvégzése után a számítógép játékkal jutalmazza a gyermeket. A játék nem öncélú: fejleszti a fonémahallást, illetve gyakoroltatja a vizuális differenciálást. A dominó játéokban az egyes dominók egyik oldalán kép van, a másik oldalán graféma. A graféma mellé olyan dominót kell helyezni, amelynek a képén található tárgy nevében megtalálható a grafémának megfelelő beszédhang tetszőleges pozícióban. A színező játékban egy képet kirakószerűen részekre bontottunk. Az egyes darabkákon különböző, de vizuális alapon téveszthető grafémák vannak. Az egyes grafémákhoz a palettán színek vannak rendelve, ezekkel kell a megfelelő darabkákat kiszínezni az egér használatával.



3. ábra: A „Beszédmester” olvasástanítási része.



4. ábra: Kritikus fonémahármas gyakoroltatása

3 Matematikai és számítástudományi alapok

A gépi beszédfelismerés célja egy tetszőleges beszédjelhez a hozzá legjobban illeszkedő szimbólumsorozat megtalálása, ahol az egyes szimbólumok beszédhangokat, vagy valamely más, alkalmasan megválasztott kódolási egységeket jelölnek. Ez az általános megfogalmazás algoritmikailag egy kiterjedt keresési feladatot jelent, amely feltétel nélküli végrehajtása még a mai modern számítástechnikai eszközök segítségével is rendkívül idő-, és műveletigényes. Emiatt a manapság hatékonyan működő automatikus beszédfelismerő rendszerek a feladatot valamilyen szempontból egyszerűsítik, ami állhat a szimbólumok halmazának szűkítéséből, illetve a keresési tér redukciójából (Kocsor 2004).

Jelenleg a beszédfelismerésben szinte kizárólag az ún. rejtett Markov-modellezést használják. Az utóbbi évtizedben történt rendkívüli elterjedését a modell viszonylagos egyszerűségének, matematikai megalapozottságának és főként hatékonyságának köszönheti, ami számos feltevés, köztük irreális feltételezések következménye. Ezeknek a megszorításoknak a célja, hogy a modellben szereplő szabad paraméterek számát és szerepét olyan szintre szorítsák, hogy már statisztikai alapú matematikai becslésekkel számíthatóak legyenek az értékeik. Ezek a feltételezések irreális voltak ellenére robusztus matematikai modell használatát teszik lehetővé, ami végső soron hatékony beszédfelismeréshez vezet. Azonban ezek a kikötések jelentik egyben a modell korlátait is: a modellel szembeni főbb kifogások a beszéd időbeliségének gyenge reprezentálása, és a mindössze elsőrendű kapcsolatok feltételezése.

Ezért a „Beszédmester” szoftver automatikus beszédfelismerő rendszerét egy általunk fejlesztett, úgynevezett szegmentális típusú dekódolási algoritmusra építettük, amely kellően

rugalmas ahhoz, hogy könnyen ki lehessen próbálni benne különböző stratégiákat. Az eljárás már nem keretekben dolgozza fel a beszédjelet, hanem nagyobb egységekben, időintervallumokban gondolkodik, és ez a kibővített megközelítés eredményezi azt, hogy a módszer magában foglalja a rejtett Markov-modellezést is, mint speciális esetet.

Természetesen az új eljárás, mint a beszédfelismerő eljárások általában, a teljes problémát kisebb, gép által könnyebben megoldható részproblémákra vezeti vissza, mint a betűk, vagy fonémák egyedi felismerésére. A fonéma-felismerés hatékony gépi megvalósítása paraméteres matematikai modellek felírását, illetve a paraméterek statisztikai alapú optimalizálását követeli meg, amit a szakirodalom gépi tanuló algoritmusként ismer. Az ilyen típusú statisztikai módszerek használata szükségessé teszi gyakorisági szótárak és nagyméretű annotált és fonetikailag szegmentált beszédadatbázisok létrehozását is, amelyek a későbbiekben további kutatásokat és fejlesztéseket implikálhatnak – természetesen nem csak a fonémák vizsgálatával kapcsolatban (Sejtes 2004).

A gépi tanuló algoritmusok fonéma felismerésben történő alkalmazásához szükséges a rendkívül nagy mennyiségű információt szolgáltató beszédjel tömör reprezentálása úgy, hogy csak a fonémák felismeréséhez nélkülözhetetlen információkat emeljük ki jellemzőkinyerési technikák segítségével. A beszédadatbázisok alapján készített, a fonémákat leíró jellemző tér már a gépi tanuló algoritmusok számára feldolgozható, így lehetőség adódik arra, hogy a kinyert adatokon alkalmazzuk azt a rengeteg fajta modellezési technikát, amelyet valaha felvetettek rögzített dimenziószámú terekben való osztályozásra, illetve valószínűségi regresszióra. A beszédfelismerés alapját képező, tömören reprezentált fonémák felismerése még tovább finomítható a gépi tanulás általános transzformációs eljárásaival, amelyek működésük során kiválaszthatják a fontosabb térkomponenseket, elnyomhatják a tanuló algoritmusok hatékonyságát zavaró egyéb hatásokat, és olyan térbe vihetik át az adatokat, ahol a szeparálhatóság könnyebb (Kocsor 2004).

3.1 Jellemzőkinyerés

A beszédfelismerést megelőző lépés minden esetben a beszédjel digitális feldolgozása, az ún. jellemzőkinyerés. Ennek célja egy olyan tömör reprezentáció előállítása, amely a felismerés szempontjából fontos komponenseket megőrzi, a zavaróakat viszont lehetőség szerint eltávolítja. A szakirodalom rengeteg előfeldolgozási módszert ismer, azonban olyan kijelentést nem lehet tenni, hogy az egyik módszer egyértelműen jobb lenne a másikinál. Az optimalitás ugyanis függ a rákövetkező felismerési módszertől, sőt, maguktól az adatoktól (tkp. a felismerési feladattól) is. Ezért egy adott felismerési feladat esetében mindenképpen érdemes többféle módszert is kipróbálni. Az elmúlt években számos előfeldolgozási módszert alkalmaztunk, kombinálva a később ismertető diszkriminatív szegmens-alapú felismerési sémával (Suykens 2000). A vizsgált felismerési feladatok szinte minden esetben magyar nyelvű felismerési problémák voltak, úgymint kapcsolt számok felismerése, telefonos általános szöveg felismerése illetve gyermekhangok felismerése. Egy esetben angol nyelvű felismeréssel is próbálkoztunk (Kocsor 2004), és eredményeink összemérhetőnek bizonyultak a nemzetközi szakirodalomban közöltekkel.

A jellemzőkinyerés hagyományosan a beszédjel 10-30 ezredmásodperces darabkáiból történik. Ebben a lépésben főleg a szakirodalomból már ismert módszereket alkalmaztuk. A jól bevált megoldások mellett egy pályázat keretében implementáltunk egy ún. hallásmodellt is, amely egyelőre még nem számít sztenderdnek a beszédfelismerésben. Ezen túl egy saját

algoritmust is kifejlesztettünk jelek szinuszos alapú dekompozíciójára, amely jóval pontosabb reprezentációt tesz lehetővé, mint a Fourier technikák (Kocsor 1999), (Kovács 2002).

3.2 Jellemzőtér-transzformációk

Egy klasszifikációs vagy gépi tanulási módszer célja, hogy előre ismert vagy kiszámítható jellemzők alapján eldöntse, hogy az adott jellemzőkkel leírt egyed milyen valószínűséggel tartozik a lehetséges besorolásokhoz. A tanuló algoritmusok alkalmazása előtt lehetőségünk van arra, hogy a reprezentációra használt téren bizonyos transzformációkat végezzünk. Az alkalmazott transzformáció kiválaszthatja a fontosabb térkomponenseket, elnyomhatja a tanuló algoritmus hatékonyságát zavaró egyéb hatásokat, és olyan térbe viheti át az adatokat, ahol a szeparálhatóság könnyebb.

3.3 Gépi tanulási módszerek

A gépi tanulás leginkább tanulmányozott területe az úgynevezett függvénytanulási probléma. Ennek lényege, hogy az első, tanulási fázisban az algoritmus $(\underline{x}, f(\underline{x}))$ alakú mintákat kap valamely $f(\underline{x})$ függvényből, amelyek alapján valamilyen hipotézist állít fel a függvény alakjára vonatkozólag. A felhasználási fázis során pedig $f(\underline{y})=?$ típusú kérdéseket tesznek fel a rendszernek, $\underline{y}$ értékek valamely sorozatára nézve. A tanulás pontossága $\underline{y}$ teszt-pontok egy olyan halmazán mérhető, amely esetében ismerjük a valódi $f(\underline{y})$ függvényértékeket – így azokat össze tudjuk vetni a gép által becsültekkel.

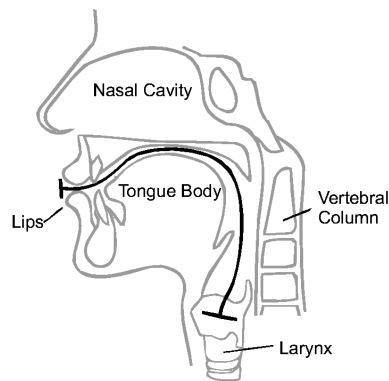
A függvénytanulási probléma a gyakorlatban legtöbbször osztályozási feladat formájában vetődik fel. Ez speciális esete a függvénytanulásnak olyan értelemben, hogy az $f(\underline{x})$ függvényértékek egy véges készletből, az osztálycímkék készletéből kell, hogy kikerüljenek. Az osztályozási feladat további szűkítése az ún. fogalom-tanulási feladat, amikor is a függvényérték bináris, vagyis csak két lehetséges osztály van. Bizonyos algoritmusok csak a kétosztályos feladattal képesek megbirkózni, de az n -osztályos tanulás ilyenkor is megoldható kétosztályos feladatok egy sorozatára való visszavezetéssel.

Az osztályozási feladat megközelítésére alapvetően kétfajta szemléletmód, illetve irányvonal jellemző. Az egyik a formális, logikai szemlélet (Kerns 1994), amely olyan elméletekhez vezetett, mint pl. a PAC-tanulás illetve az Occam-tanulás. A másik megközelítés a statisztikai alakfelismerés (Duda 2001), amelynek alapköve a Bayes-döntélmélet.

3.4 Beszélőnormalizálás

A beszédfelismerési alkalmazások széleskörű elterjedésének előfeltétele a rendszerek beszélőfüggetlensége, azaz különböző felhasználók esetén is megbízhatóan és pontosan működő felismerési mechanizmus. A rendszerek általánosítási képessége természetes módon fokozható a tanítási folyamat során felhasznált beszédatadabázisok nagy méretével és sokszínűségével, de ez hosszadalmas és költséges munkát követel meg a módszer korlátozott hatékonysága mellett. Egy másik megközelítés ezért az egyes beszélők hangját változtatja meg úgy, hogy az a lehető legjobban hasonlítson az alkalmazott beszédfelismerő rendszer „ideálisan” felismerhető hangjához.

A beszélők hangjai közti eltéréseket a garat-, száj- és orrüreg alkotta artikulációs csatorna egyedekre jellemző különbségei okozzák. Ez a csatorna a gégeből kibocsátott hanghullámot, mint áramló levegőt közvetíti a külvilág felé. Működése során a kiinduló hang bizonyos frekvenciáit felerősíti, másokat elnyom, de a jelenlévő frekvenciák szerkezetét nem módosítja. Emiatt a csatorna jelfeldolgozási szempontból szűrőként viselkedik, azaz matematikailag egy átviteli függvénnyel jellemezhető (megadja az egyes frekvenciák erősítésének mértékét). Az átviteli függvény maximumhelyeit formánsoknak nevezzük. Fonetikai kísérletek alapján kijelenthető, hogy a hang első három formánsa a fonetikai minőséget, míg a többi a beszélő egyedi jellemzőit (pl. hangszín) határozza meg. Beszédfelismerés szempontjából az itt jelentkező különbségeket célszerű kompenzálni, létrehozva így egy „idealizált” beszédhangot.



7. ábra Az emberi artikulációs csatorna.

A különböző beszélők által ejtett, de azonos fonetikai csoportba tartozó hangok formánsszerkezete hasonló. Így minden fonetikai csoport jellemezhető egy átlagos formánsszerkezettel. A legelterjedtebb megközelítés szerint az egyes beszélők formánsainak átlagostól vett eltérése korrelál az artikulációs csatorna hosszával (Vocal Tract Length – VTL), ezért lehetőség nyílik a hangok normalizálására a beszélő VTL hossza szerint (VTLN). A VTLN során egy transzformációs függvény segítségével próbáljuk a beszélők formánsait az átlagos helyre leképezni (Edie 1997), (Uebel 1999). A beszélőnormalizálás feladata egy megfelelő típusú transzformációs függvény kiválasztása, és a beszélőre jellemző optimális paraméterek beállítása.

A VTLN során a transzformációs paraméterek beállítása történhet néhány másodperces bemondás alapján, vagy a teljes hanganyag segítségével is. Mindkét esetben számottevően csökkenthető a fonémafelismerés hibája. A „Beszédmester” működése során azonban valós idejű fonémafelismerésre van szükség, így nincs lehetőség az előző technikák alkalmazására az eredeti formájukban. A probléma megoldására a szerzők korábban megadtak egy valósidejű beszédnormalizációt lehetővé tévő VTLN eljárást, amely a bemondás pillanatában hatékonyan megbecsüli az alkalmazott transzformációs függvény paramétereit (Paczolay 2003).

4 Összefoglalás

Ebben a cikkben összefoglaltuk a „Beszédmester” beszédjavítás-terápia és olvasásfejlesztő szoftver kifejlesztésének eredményeit mind alkalmazás, mind az elméleti

eredmények szempontjából. A gyakorlati alkalmazást tekintve elmondhatjuk, hogy a siketek és nagyothallók beszédterápiája, bizonyos logopédiai problémák kezelése, olvasástanítás, illetve részképességükben sérült gyerekek olvasásterápiája szempontjából egy újszerű eszközt sikerült kifejleszteni. Az elméleti eredmények a mesterséges intelligenciai kutatások olyan modellépítési technológiáit jelentik, amelyek a „Beszédmesteren” túl más hasonló számítógép-ember interakcióra épülő tanító / terápiai csomag kifejlesztését tehetik lehetővé.

5 Irodalomjegyzék

- M. I. Adams, Beginning to read: Thinking and learning about print. MIT Press, Cambridge, 1990
- J. Bácsi, J. Kerekes: Az első osztályos olvasókönyvek szöanyagából készült Gyakorisági szótár „Van szó”. Módszertani Közlemények (Szeged) 2003/2. sz. 52-57., 2003
- R. O. Duda, P. E. Hart, D. G. Stork: Pattern Classification, Wiley, 2001.
- E. Eide, H. Gish: A Parametric Approach to Vocal Tract Length Normalization, Proc. ICASSP '97, pp. 1039-1042, Munich, Germany, 1997.
- G. Héegly Látható beszéd – kiejtéstanítás Kommunikáció – nehezített társadalmi beilleszkedés
- M. J. Kerns: An introduction to computational learning theory, 1994.
- A. Kocsor, L. Tóth: Kernel-Based Feature Extraction with a Speech Technology Application, IEEE Transaction on Signal Processing, Vol. 52, No. 8, pp. 2250-2263, 2004.
- A. Kocsor, L. Tóth: Application of Kernel-Based Feature Space Transformations and Learning Methods to Phoneme Classification, Applied Intelligence, Vol. 21, No. 2, pp. 129-142, 2004.
- A. Kocsor, L. Tóth, I. Bálint: On the Optimal Parameters of a Sinusoidal Representation of Signals, Acta Cybernetica, Vol. 14, pp. 315-330, 1999.
- K. Kovács, A. Kocsor, L. Toth: Hungarian Speech Synthesis Using a Phase Exact HNM Approach, in: Grosky, W. I., Plasil, F. (eds.): Proc. of SOFSEM 2002: Theory and Practice of Informatics, LNCS 2540, pp. 181-185, Springer Verlag, 2002.
- D. Paczolay, A. Kocsor, L. Tóth: Real-Time Vocal Tract Length Normalization in a Phonological Awareness Teaching System, Matousek, V., Mautner, P. (Eds.): Proc. of TSD'2003, pp. 309-314, 2003.
- D. Paczolay, A. Kocsor, Gy. Sejtes, G. Héegly: A „Beszédmester” csomag bemutatása, informatikai és nyelvi aspektusok, Alkalmazott Nyelvtudomány, IV. évf. , 1. sz., 57-79, 2004.
- B. Péchy: Beszélni nehéz! Budapest, 1974
- Gy. Sejtes, A. Kocsor: A „Beszédmester” adatbázis-specifikációja, Alkalmazott Nyelvtudomány, IV. évf. , 1. sz., 81-89, 2004.
- J. A. K. Suykens, J. Vandewalle: Recurrent least squares support vector machine, IEEE Transactions on Circuits and Systems, vol. 47, No. 7, pp. 1109-1114, Jul. 2000.
- L. F. Uebel, P. C. Woodland: An Investigation into Vocal Tract Length Normalisation, Proc. EUROSPEECH-99, Vol. 6, pp 2527 - 2530, Hungary, 1999.