

A „Beszédmester” csomag bemutatása: informatikai és nyelvi aspektusok

PACZOLAY DÉNES*- KOCSOR ANDRÁS*- SEJTES GYÖRGY^X- HÉGELY GÁBOR*

*Szegedi Tudományegyetem Informatikai Tanszék Mesterséges Intelligencia Kutatócsoport

E-mail: pdenes@inf.u-szeged.hu

E-mail: kocsor@inf.u-szeged.hu

^XSzegedi Tudományegyetem Juhász Gyula Tanárképző Főiskolai Kar

Gyakorló Általános és Alapfokú Művészeti Iskola

E-mail: sejtes@hung.u-szeged.hu

Abstract

The “SpeechMaster” package seeks to apply speech recognition technology to speech therapy and the teaching of reading. The role of speech recognition is to provide visual phonetic feedback. In the first case it is intended to replace the missing auditive feedback of the hearing impaired, while in the latter case it is to reinforce the ‘correct’ association between the phoneme-grapheme pairs. With the aid of a computer, children can practice without the need for the continuous presence of the teacher. Both applications – teaching reading and speech impediment therapy – require a real-time response from the system in the form of an easily comprehensible visual feedback. With the simplest display setting visual feedback is provided by means of flickering letters, their identity and brightness being adjusted to the speech recognizer's output.

The project consists of two main parts: a research part and an application part. The research part is devoted to investigating modern machine learning techniques which can be used in the development of numerous info-communication systems. The machine learning algorithms developed have been made available as open-source software and we hope that they will stimulate the development of systems that requires these kinds of modules. One such example is speech recognition, which is the heart of the “SpeechMaster” program, developed in the second part of our project. Informaticians, surdopedagogues, developer pedagogues, linguists and a graphic designer work in a team to replace the inconvenient “mirror method” and to help children to associate phonemes with graphemes. The software's database with its sound files, recording data, etc. can be used in further informatician and linguistic research. It is important for us that we will make the program freely available on the Internet. The methodology based on our daily work is also available in an appendix. We hope that the software product will help many school-children with learning to read and that teachers will find it useful in giving different tasks to different children with a tool that allows the conscious planning of individual work.

1. Bevezetés

Az MTA szegedi Mesterséges Intelligencia Kutatócsoportja 2002 februárjában egy kétéves projektet kezdett az Oktatási Minisztérium támogatásával. A munka két fő részre bontható. Az első egy alapkutatási feladat, amely olyan modern gépi tanulási módszerek vizsgálatát tűzte ki célul, amelyek alapköveit jelenthetik korszerű info-kommunikációs rendszerek kifejlesztésének. A kitűzött cél második része egy olyan közhasznú szoftver elkészítése, amely az alapkutatás során kifejlesztett algoritmusokra épül, és azok segítségével támogatja a hallás- vagy

részképességükben sérült gyermekek és felnőttek beszédjavítás-terápiáját, illetve a kisiskoláskorúak olvasástanítását. Az automatikus beszédfelismerés összekapcsolását a beszédjavítás-terápiával és az olvasástanítással az indokolja, hogy a cél mindkét esetben valamiféle fonológiai tudatosság kialakítása vagy hiányának pótlása. A számítógépes valós idejű vizuális beszédkiértékelés a hallássérültek esetén pótolhatja a hiányzó auditív kontrollt, míg az olvasástanításban hatékonyabbá teheti a beszédhangok és a betűk helyes összekapcsolását.

A munka összetett; több tudomány (informatika, pedagógia, nyelvészet) szakemberei dolgoznak együtt azért, hogy a „Beszédmester”-nek (röviden „BeMe”-nek) nevezett szoftver az alkalmazás során jól működjön, hatékonyan segítse a gyerekek olvasástanítását, ill. a hallássérültek a beszédoktatását.

1.1. Informatikai feladatok

A BeMe informatikai háttere az SZTE Informatikai Tanszékcsoportjának alapkutatói és szoftverfejlesztési munkájára épül a beszédfelismerés és a gépi tanulás témakörében. A cél olyan gépi tanuló algoritmus kifejlesztése, amely lehetővé teszi a hatékony klasszifikációt, a beszéd alapegységeinek megkülönböztetését.

1.2. Pedagógiai és nyelvészeti feladatok

A statisztikai alapon működő módszerek alkalmazásához nagy méretű beszédatadbázisokra van szükség. Az SZTE JGYTF-i Kar Gyakorló Általános Iskolája kutató munkatársainak (nyelvészek, pedagógusok) feladata a hanganyag megtervezése, a gyakorisági szótár elkészítése, a hanganyag rögzítése és feldolgozása, ill. az adatbázis specifikációja volt. Hívóképanyagot kellett összeállítani, ki kellett dolgozni a számítógéppel segített olvasástanítás metodikáját.

Az Óvoda Általános Iskola Diákotthon és Gyermekotthon (Siketek Iskolája, Kaposvár) szurdopedagógusai a hallássérültek beszédoktatásának anyagából készítettek gyakorisági szótárat, hallássérült gyermekek hanganyagát rögzítették, didaktikát írtak a végleges verzióhoz.

1.3. Grafikai feladatok

A BeMe grafikai megjelenítésének célja a program kiszolgálása. A grafikus olyan közérthető jelrendszert és színszimbolikát használt, amely az alkalmazás során a gyerekek érdeklődését ébren tartja, az esztétikai élmény mellett motivációként is szolgál.

Jelen cikkben a szoftvert mutatjuk be röviden. Az informatikai és nyelvi aspektusokra fókuszálunk azért, hogy segítsük a hasonló kutatások technikai megvalósítását.



1. ábra Beszédmester nyitó képernyő

2. A „Beszédmester”

A programban szétválasztottuk a beszédjavítás-terápiai és az olvasástanítási, ill. olvasásterápiai részt azok speciális didaktikája miatt.

2.1. Beszédjavítás-terápia

A számítógépes beszédjavítás-terápia esetén az elsődleges cél a sérült auditív kontroll helyettesítése vizuális visszacsatolással. A megvalósítás során a mikrofonba bement beszédhangokat a szoftver feldolgozza és a legnagyobb valószínűséggel felismert fonémához tartozó grafémát megjeleníti. A rendszer nem csupán azt határozza meg, hogy melyik fonéma lehetett a bement, hanem azt is,

hogy a felismerés mennyire volt sikeres, mennyire hasonlít az adott fonémára. A felismerés pontosságát a graféma kirajzolásának színével és fényerejével is jelzi a szoftver.

A tapasztalat alapján a magánhangzók helyes kiejtésének elsajátítása nehezebb, mint a mássalhangzóké. A valós idejű fonéma-felismerésen alapuló beszédjavítás-terápiának az a nagyszerűsége, hogy kitartott hang esetén a nyelvállást változtatva folyamatosan és azonnal figyelemmel követhető, hogy épp mely magánhangzót vagy folyamatos mássalhangzót képezi a felhasználó. Ezt az egyértelmű visszacsatolást más könnyen elérhető eszközzel, eddig nem sikerült megvalósítani.

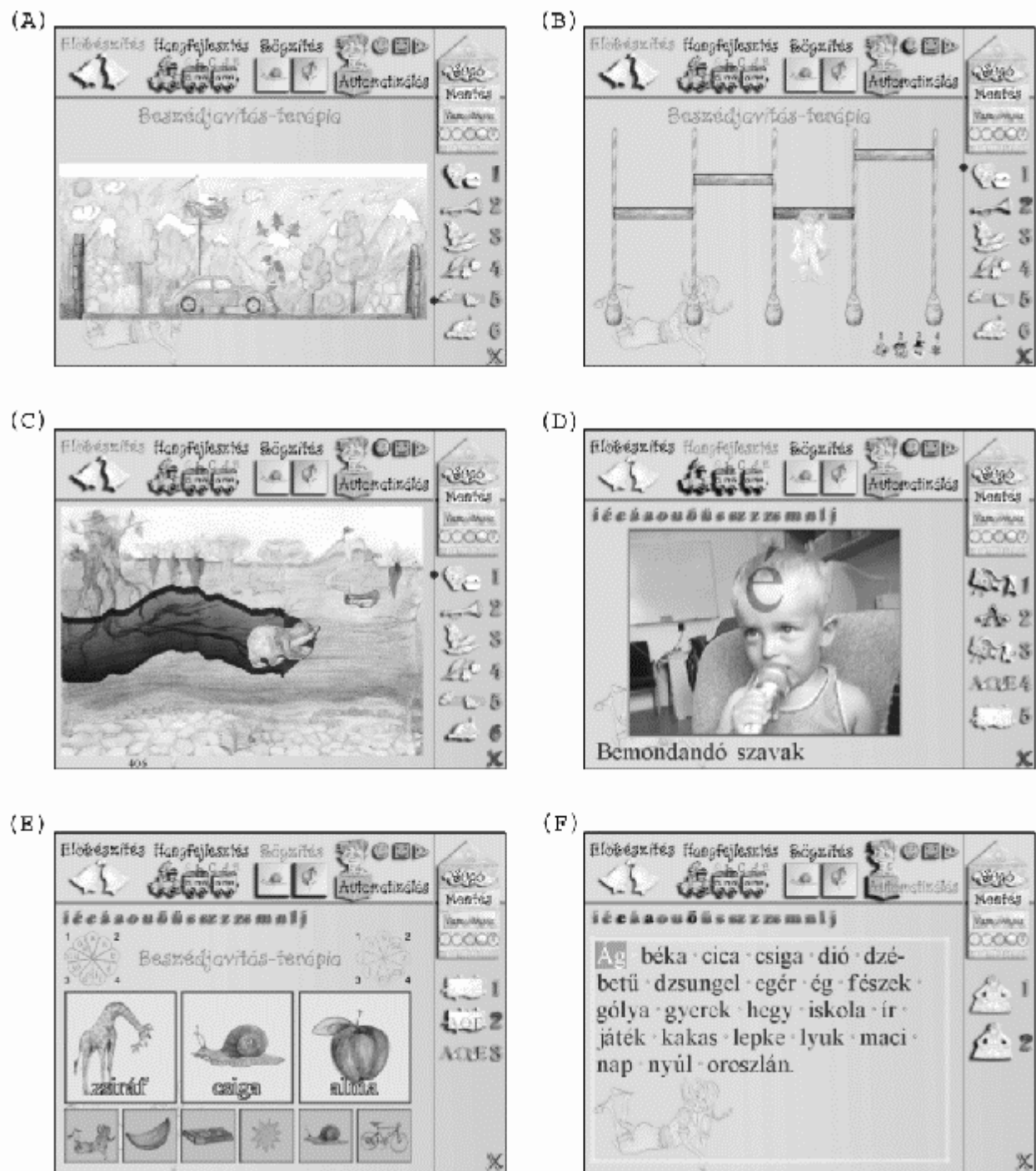
A röviden ejthető mássalhangzók gyakorlásához más módszert kell alkalmazni, hiszen a nagyon rövid vizuális információt – a felvillanó grafémaképet – nem képes az agy feldolgozni. Ezekben az esetekben is segít a „Beszédmester”. A hagyományos siketek oktatásának elengedhetetlen kelléke a fonetikai tükör, amelynek segítségével bemutatható a hang helyes képzésének technikája. A számítógéppel segített terápia során a tükröt webkamerával helyettesítettük. Ezek az olcsó, kisméretű kamerák megfelelő képet biztosítanak, és ügyes megvilágítást alkalmazva kiválóan alkalmazhatók a terápia során. A foglalkozás közben lehetőség van a kameraképben (szájról olvasási kép) áttetszőn megjeleníteni a felismert beszédhangokat, így komplex visszacsatolást biztosítunk a felhasználónak.

A beszédjavítás-terápia általában fiatal korban kezdődik, így szükség van előkészítő játékos feladatokra. A terápia elején ki kell alakítani a tanulóban a tudatos hangerő, ritmus, hangmagasság változtatásának készségét is. Ezek a feladatok a hangfeldolgozásban közismert technikákkal megoldhatók és ötletesen, esztétikusan megjeleníthetők.

2.1.1. A beszédjavítás-terápia részei

A szoftver beszédjavítás-terápia része négy modult tartalmaz. A modulok egymásra épülnek; feltételezzük, hogy a tanulók akkor haladnak előre, ha az egyes részeket jól teljesítik.

A négy fő rész a következő: az „Előkészítés” (a bevezető gyakorlatok), a „Hangfejlesztés” (a hangok helyes artikulációjának kialakítása), a „Rögzítés” (a hangok gyakorlása szavakban) és az „Automatizálás” (a folyamatos beszéd gyakorlása hosszabb mondatokban).



2. ábra. Képek a program beszédjavítás-terápia részéből: (A) zöngeadás gyakorlása (B) a hangerő-változtatás rögzítése (C) játék a hangmagasság finom változtatásával (D) hangfejlesztés valós idejű visszacsatolással (E) hangok rögzítése szavakban (F) hangok rögzítése mondatokban.

Előkészítés

- Zöngeadás: a kezdő képen egy autó áll, ha zöngétlen hangot képzünk; lassan vagy gyorsan halad, ha gyengén vagy erősen zöngés (periodikus) hangot ejtünk. (2.a ábra.)
- A hangerő változtatása: egy bohóc lécek alatt a hangerő függvényében alacsonyabbra vagy magasabbra ugrik. A lécek számát és magasságát (az elvárt hangerőt) a pedagógus előre beállítja. (2.b ábra.)
- A ritmus gyakorlása: a tanító mikrofonba bemozdított ritmikus hangsorát egy ugráló béka nyoma rögzíti a homokos parton. A tanuló feladata a ritmus megismétlése, annak elérése, hogy a béka a nyomokban ugráljon.
- Állathangok utánzása: a pedagógus utasítása és ellenőrzése mellett a kiválasztott állatok hangját utánozni kell a foglalkozáson résztvevő gyermek számára. A gép itt nem értékel, csak képeket kínál fel a gyakorláshoz.
- A hangmagasság változtatása: a feladat első lépésében magas és mély hangok sorozatával két sorban virágokat lehet ültetni. Magas hang esetén a felső sorba, különben az alsó sorba kerül a virág. A gyakorláskor a virágágyást kell meglocsolni úgy, hogy az előző mély és magas hangokból álló sorozatot kell megismételni.
- Hangringatás: a feladat kitűzésekor, egy hosszan kitartott magánhangzó hangmagasságának folyamatos változtatásával egy vakond járatott váj a földben. Magas hang esetén feljebb, mély hang esetén lejjebb halad a földben. Gyakorláskor ebben a járatban kell végigvezetni egy másik kisebb vakondot. A cél, hogy a gyermek megtanuljon gazdálkodni a levegővel. (2.c ábra.)

Hangfejlesztés

Az előkészítést követően kitartott ejtéssel és hangkapcsolatokban kell kialakítani a magánhangzó és a folyamatosan ejthető mássalhangzók helyes artikulációját. A képernyőn megjeleníthető a kamera által közvetített szájról olvasási kép, a kép közepében pedig a gyakorlásra kijelölt és helyesen ejtett hangzók betűképe. A felvillanás fényereje jelzi a megfelelés mértékét. (2.d ábra.) Ez a feladat elvégezhető kamerakép nélkül magas kontrasztos megjelenítéssel, illetve beállítható, hogy hibás hangképzéskor több grafémát is jelenítsen meg a gép (pl. átmeneti hangok esetén kettőt). Lehetőség van arra, hogy a beszédhang-felismerés időbeli változását grafikonok segítségével jelenítsük meg.

Rögzítés

A fejlesztőmunka eredményeként a tanuló elérheti azt a szintet, hogy önállóan gyakorolhat. Ekkor a képernyőn három gyakorolni kívánt szó képe és a fogalom neve jelenik meg. A többi gyakorlóképet kis méretben, betűnként

csoportosítva kínálja fel a gép. Egy-egy szó kimondásakor az aktív képmezőben az előforduló és a kijelölt hangok kiejtésének megfelelő betűképek villannak fel.

Automatizálás

Az automatizálás során a tanulókkal rövidebb, hosszabb szövegek érthető felolvasását gyakoroltatjuk. A gyakorolni kívánt szöveg lehet előre rögzített, de be is gépelhető. A hosszabb mondatok a foglalkozás elején még részekre, blokkokra tagolhatók. A feladatban kiválaszthatók a gyakorolni kívánt beszédhangok és ezek más színnel jelennek meg a szövegben. A gép blokkonként vagy mondatonként kiemeli a szöveget. Ha az adott beszédhangot helyesen képezte a tanuló, a kijelölt grafémák színe megváltozik a felolvasáskor. Ha minden graféma színe eredetire váltott, a következő egységre ugrik a program. Ebben a feladatban nem folyamatos beszédfelismerés történik, mert itt az ejtés tisztaságának objektív értékelése a cél.

2.2. Olvasásterápia

A biztos és sikeres olvasáselsajátítás egyik fontos feltétele a kommunikációs készség fejlettsége, ezen belül kiemelt fontosságúak a beszédészlelési, beszédértési folyamatok és a beszédprodukció. Ehhez kapcsolódnak még olyan részterületek, mint a vizuális és verbális memória, ritmuskészség, vizuális és verbális szerialitás, vizuális és hallási differenciáló képesség, tájékozódási képesség, irányfelismerés, formafelismerés, formaállandóság, alak-háttér megkülönböztetése. A fejlett kommunikációs készségű gyermekek könnyebben és sikeresebben tudják az írott nyelvet elsajátítani, mint azok a társaik, akiknek a kommunikációja, vagy a fent említett részképességek közül valamelyik sérült (pl. megkésett beszédfejlődés, egyes típusú diszlexiák, beszédhibák stb.).

A kommunikációs készség fejlesztése, az olvasástanítás megalapozása a fonológiai tudatosság megalapozásával kezdődik. Már az óvodáskort megelőzően a gyermekek elkezdenek játszani, manipulálni a nyelvvel: új értelmes és értelmetlen szavakat hoznak létre, rímelő szavakat keresnek, a szavakból hangokat vesznek el vagy tesznek hozzá, ritmikus mondókákat, versikéket mondogatnak. Ez nem más, mint a fonológiai tudatosság felé vezető út első lépése (Bácsi, Kerekes, 2003). Adams (1990) a fonológiai tudat öt szintjét különbözteti meg:

1. szint: A fonológiai egység nagyságának érzékelése: könnyebb a mondatokat szavakra bontani és a szavakat szótagokra, mint a szótagokat fonémákra.
2. szint: A fonémák számának meghatározása egy szóban: könnyebb felbontani egy fonémák szempontjából rövidebb szót, mint egy hosszú szót.

3. szint: A fonéma helyzetének felismerése egy szóban: a középhelyzetben levő fonémák azonosítása a legnehezebb, a szó utolsó hangjának azonosítása könnyebb, legkönnyebbek a kezdő pozícióban álló hangok.

4. szint: A szóban szereplő hangok fonológiai sajátosságainak felismerése: a folyamatosan hangzó, hangoztatható fonémákat könnyebb azonosítani, mint a csak nagyon röviden ejthető hangokat.

5.szint: A fonológiai tudatosság különleges eseteinek azonosítása: a rímelő és kezdő fonémák azonosítása könnyebb, mint az összeolvasás vagy a hangokra bontás.

Ezek a szintek, vagy legalábbis többségük 6 éves korra rutinszerűen működnek a legtöbb gyermeknél. A szintek gyakorlására a Beszédmesterben különféle feladatok állnak rendelkezésre. A betűtanítás sorrendje sajátos, metodikailag megalapozott.

2.2.1. Az olvasásterápia részei

a) Betűtanítás (magánhangzók és mássalhangzók tanítása)

b) „Fonémapárok” (betű-hangpárok/-hármak) tanítása, gyakorlása

Mindkét rész az olvasástanítás segítségét szolgálja. Hat különböző feladat szolgálja a betűtanítást. A betűnként azonos típusú feladatok biztosítják a gyermek számára az önálló tanulás, gyakorlás lehetőségét a szoftver segítségével. Ez már önmagában sikerélmény mind az egészséges, mind pedig a részképességében sérült gyermek számára.

A betűtanítás feladatai:

Hang- és betűtanítás

Az új betű hívóképe látható a megjelenítőn. A számítógép a magánhangzók esetén a graféma megjelenítésével (szín és élesség) visszajelzi az ejtés/képzés pontosságát, helyességét (3.a ábra). A mássalhangzók esetén a teljes szót jeleníti meg (3.b ábra). A feladat alkalmas a szókincs fejlesztésére, az ejtéshibák megfigyelésére, korrigálására, a gyermekek motiválására.

Új hang keresése tárgyképek nevében

A tárgyképek alatt a szóban található hangok számának megfelelő karika található. A tanult hang helyét kell megtalálni a szóban, és a helyének megfelelő körre kell kattintani az egérrel. A helyes megoldást a gép kipipálja (3.c ábra). A képek összeállításánál az Adams által felállított szinteket vettük figyelembe, így a gyermekek a feladatmegoldás során a könnyebben azonosítható pozícióban álló fonéma azonosítása felől haladnak a nehezebb felé.

Betűk kiválasztása

Az első részben az egér rákattintásával kell kiválasztani a megfelelő betűt az elforgatottak közül. A feladat azért fontos, mert a segítségével korán felismerhető, ha egy gyermek vizuális rotációs problémával küzd. A feladat fejleszti a forma és alakfelismerést, az irányok azonosítását, az iránytartást.

A második részben az újonnan tanult betűt kell kiválasztani a tanultak közül. A feladat fejleszti az alakfelismerést, a gyermekeket a helyes olvasási irány (balról jobbra) megtartására szoktatja.

A helyes kiválasztásoknál a betűk elszíneződnek, s a helytelen betűk vagy a hibás alakok eltűnnek. A képernyőn csak a helyes betűalakok maradnak meg, ezzel is elősegítve azok alakjának rögzülését.

Betű-hang megfeleltetése 1.

Az újonnan tanult betűhöz, hívóképhez kell a betűsorból a megfelelőket kiválasztani és ráhelyezni a hívóképre. A helyes megoldást a gép kipipálja. A gyakorlat rögzíti a hangleválasztást, illetve a betű-hang megfeleltetését.

Négy, már tanult hívóképhez kell a megfelelő kezdőhangokat, illetve az ennek megfelelő betűket kiválasztani és felhelyezni. A gyakorlat elősegíti a hangleválasztás gyakorlását, a betű-hang megfeleltetését.

Betű-hang megfeleltetése 2.

Az előző feladat nehezítését jelenti. Tárgyképek nevében hang keresése, s a megfelelő betűk képre helyezése. A helyes megoldást a gép kipipálja. A gyakorlat fejleszti a hallási diszkriminációs készséget, elősegíti a hangokra bontás gyakorlását, illetve segíti a betű-hang megfeleltetését.

Olvasógyakorlat

Az újonnan tanult betű olvasása 12 elemű betűsorban mikrofon használatával. Magánhangzó esetén csak magánhangzókat kell olvasni, mássalhangzó esetén a hangfelismerő rendszer biztos működése miatt kétbetűs szótagokat olvastatunk. A szótagok hibátlan olvasása után következik a szó, majd a mondatolvasás szakasza. Itt nem használjuk a mikrofont, az ellenőrzést a tanító végzi. (3.d ábra) A feladatsor sikeres elvégzése után jutalomjátékként memóriajátékot játszhatnak a gyermekek. A memóriajáték a program kezdetekor az egérhasználat és a rajzos jelölések tanulására is szolgál.

„Fonémapárok” tanítása, gyakorlása

A betű-hangpárok/-hármaskok tanítása beépül a betűtanítás sorrendjébe. A „Fonémapárok” között azokat a betűcsoportokat találjuk meg, amelyek a képzés módjában, és/vagy helyében, és/vagy vizuális alapon, és/vagy a beszédhanghallási diszkriminációban problémát jelenthetnek, könnyen téveszthetők. A forgalomban levő olvasókönyvekben található betűpárokhoz képest a szoftvernek bővebb a kínálata, ami természetesen nem jelenti azt, hogy mindegyiket tanítani kell, viszont választási lehetőséget biztosít a felhasználónak.

A „Fonémapárok” tanításának feladatai:

A tanult betűk, hangok ismétlése, újratanulása.

A tanult betűk hívóképei láthatók a megjelenítőn. A számítógép a magánhangzók esetén a betű megjelenítésével (szín és élesség) visszajelzi az ejtés, képzés pontosságát, helyességét. A mássalhangzók esetén a teljes szót jeleníti meg (3.e ábra). A feladat alkalmas az ejtészibák megfigyelésére, korrigálására.

Betűk kiválasztása

A feladat elsődleges célja, hogy a tanult betűk közül a gyermek kiválassza a sor elején kijelölt betűt. A feladat nehézsége, hogy a betűsorban megtalálható a kritikus betűcsoport többi eleme is. Egy szósorból kell kiválasztani azokat a szavakat, amelyben a sor elején kiemelt betű megtalálható. Cél a betűk felismerése, megkülönböztetése különböző pozícióban (szó elején, közepén, végén). Fontos, hogy a gyermek a grafémákat ne csak különálló helyzetben, betűszinten tudja megkülönböztetni egymástól, hanem szószinten is.

Hang azonosítása tárgyképek nevében

A képernyőn mindig két kép jelenik meg (összesen 12) és a betűpár vagy-hármask. A megfelelő betűt kell a megfelelő képre helyezni, aszerint, hogy melyik kép nevében, melyik hangot halljuk. A gyakorlat a hallási diszkriminációs készséget, a hangokra bontást fejleszti.

Betűfeltöltés

A képernyő közepén nagy kitöltetlen (üres) betűket találunk, a képernyő alján pedig ezeknek megfelelő kis, színes betűket. Az egér mozgatásával kell a kis, színes betűket ráhelyezni a nagy kitöltetlenekre. Minden jó megoldással egy-egy réteggel töltődik a nagy betű (3.f ábra). A betűk kiválasztása, felhelyezése tetszőleges sorrendben történik. Cél, az összes betű elhelyezése, feltöltése. A gyakorlat alkalmas a vizuális differenciáló képesség fejlesztésére,

gyakoroltatására, a különbségek megfigyeltetésére. (Ékezetes betűk esetén azonos időben töltődik az ékezet és a betűforma, kétjegyűek esetében pedig mindkét betűjegy, így kiválóan alkalmas az optikai különbségek megfigyeltetésére.)

Olvasógyakorlat

Betűsorokat (magánhangzók esetén), illetve szótagsorokat kell felolvasni. Az olvasógyakorlat felépítése a következő:

Kezdő pozícióban a kritikus betűcsoport egyik tagja szerepel. Az első sorban kis, a második sorban nagy nyomtatott betűkkel jelenítettük meg.

Végpozícióban szerepel a kritikus betűcsoport egyik tagja. Az első sorban kis, a második sorban nagy nyomtatott betűkkel jelenítettük meg.

Mindkét csoporthoz (pozíció szerint) tartozik egy olyan sor, melyben a kritikus betűcsoport mindegyike megtalálható.

A gyakorlat célja a graféma-fonéma megfeleltetés gyakorlása, rögzítése, a tiszta beszédhangejtés megszilárdítása.

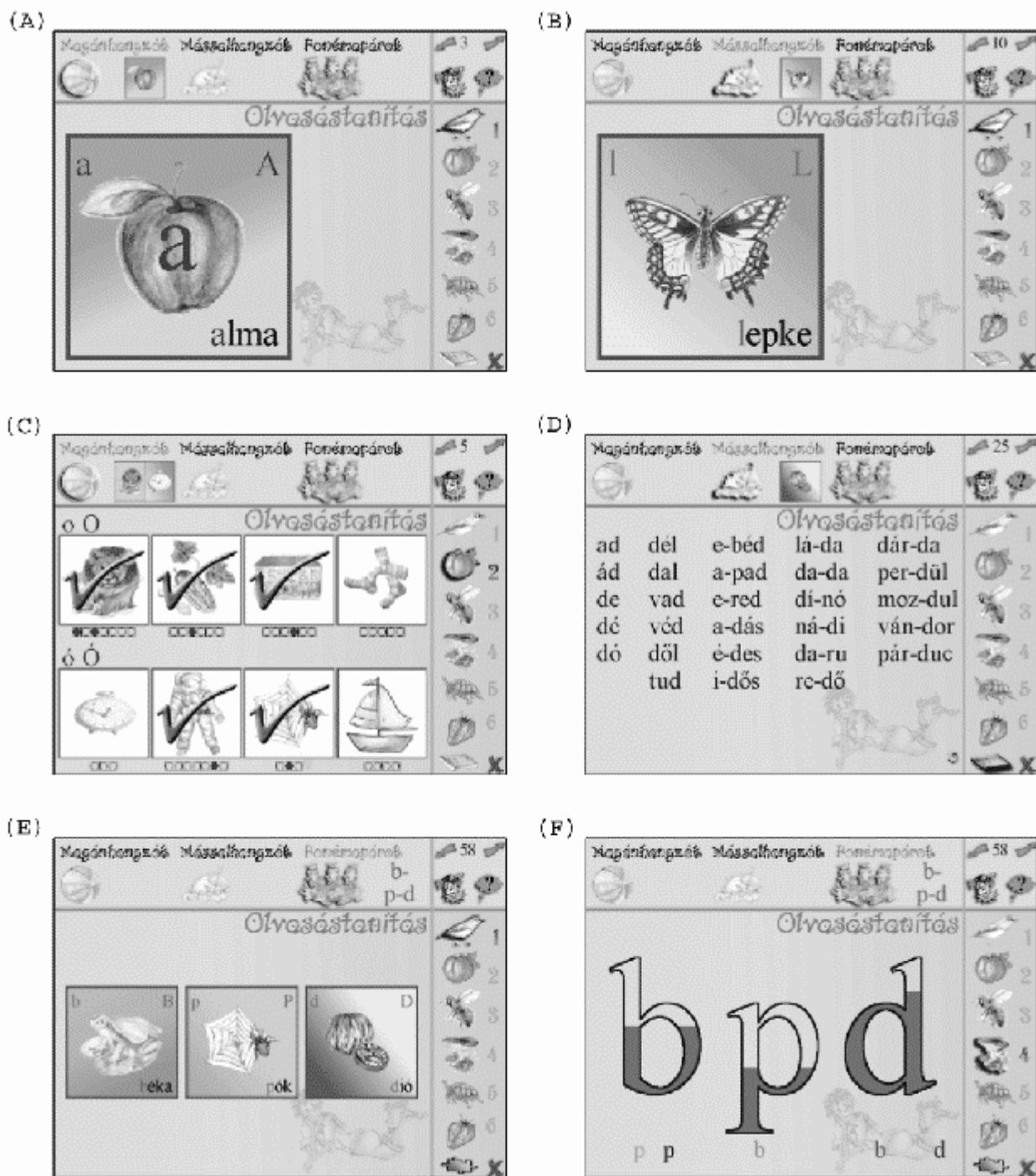
Betűpótlás

A képernyőn hiányos két-, illetve hárombetűs szótagok, szavak találhatók. A képernyő felső részén a szótagokból, szavakból hiányzó betűket találjuk meg. Az egér mozgatásával tetszőleges sorrendben választhatjuk ki a szavakat. A kiválasztott szóhoz érve elhangzik a teljes szó, ennek alapján kell kiválasztani a megfelelő betűt és pótolni a szóban. Helytelen kiválasztás esetén a betű visszaugrik, helyes megoldás esetén a betűk a szóban maradnak. A gyakorlat a hang-betű megfeleltetését gyakoroltatja, fejleszti a tiszta fonémahallást, a hallási diszkriminációs képességet és a hallási figyelmet.

A feladatsorok sikeres elvégzése után a számítógép játékkal jutalmazza a gyermekeket. A játék nem öncélú: fejleszti a fonémahallást, illetve gyakoroltatja a vizuális differenciálást.

Dominó: a dominó egyik oldalán kép van, a másik oldalán betű. A betűhöz olyan képet kell választani, azaz alá/fölé olyan dominót helyezni, amelyen olyan kép található, melynek nevében (tetszőleges pozícióban) benne van a betűnek megfelelő beszédhang.

Színező: egy egyszerű képet puzzleszerűen részekre bontottunk. A különböző részekben, darabkákon különböző, de vizuális alapon téveszthető betűk vannak (pl. a-e). Egy-egy betűhöz egy-egy szín tartozik, s ennek alapján kell az egér rákattintásával a képet kiszínezni.



3. ábra. Képek az olvasásterápia részéből: (A) magánhangzó-tanítás valós idejű fonéma-felismeréssel (B) mássalhangzó-tanítás gépi szófelismeréssel (C) fonéma helyének keresése (D) olvasási gyakorlat (E) kritikus párok újratanítása gépi szófelismeréssel (F) betűfeltöltés a kritikus párok írott formájának a rögzítéséhez.

3. Informatikai aspektusok

3.1. A szoftver tervezése

A program tervezésekor figyelembe vettük, hogy a felhasználók többsége kisgyerek, ezért esztétikus, motiváló és könnyen kezelhető grafikus felületet készítettünk.

A másik tervezési szempont a felhasználás helye volt. A programot főként általános iskolákban és otthonokban fogják használni. A gépigényt - amennyire lehetett - igyekeztünk csökkenteni, hiszen míg a fejlesztés modern számítógépeken történhetett, addig a felhasználók többsége néhány éve vásárolt számítógépen fogja futtatni a programot.

A harmadik szempont az operációs rendszer megválasztása volt. A Microsoft Windows mellett döntöttünk, mert ez a legelterjedtebb az előző pontban említett felhasználási területeken, továbbá jelenleg tanulók számára a tanulás ideje alatt ingyenesen használható. A választást az is indokolta, hogy ezen az operációs rendszeren bőséges a szoftvertámogatás, illetve a programozási segítség a webkamerák többségéhez.

3.2. A szoftver kifejlesztése

A Beszédmester fejlesztéséhez a Microsoft Visual C++ fejlesztőkörnyezetet választottuk. A fejlesztés során felhasználtuk néhány korábbi saját segédprogramunkat, illetve több segédprogramot írtunk a projekt folyamán a részfeladatok elkészítéséhez, és az adatbázis teszteléshez is. A klasszifikációs algoritmusokat külön fejlesztettük, ezeket teszteltük, és forrás szinten nyilvánossá tettük. A klasszifikációs eljárások azonos felépítésű adatbázist várnak, így könnyen tesztelhetők, versenyeztethetők és kombinálhatók egymással.

3.2.1. A fonémaklasszifikációs probléma

A felismerés adott osztályba való besorolást, azaz klasszifikációt jelent. Fonémafelismerés esetén a feladat az, hogy az egyes beszédhangokról döntsük el, hogy azok milyen fonémának (beszédegységnek) felelnek meg.

A munka nagyméretű beszédatadbázis készítésével kezdődött: beszélők szavakat mondtak mikrofonba és ezeket számítógéppel rögzítettük. A bemondandó szavakat úgy állítottuk össze, hogy bennük minden fonéma kellő számban és hasonló gyakoriságban forduljon elő, hasonlóan a beszélt nyelvhez. Ezt a Beszédmester céljaihoz igazodó gyakorisági szótár készítésével biztosítottuk. A szavakat nem fonémákra bontottuk, hanem kvázifonetikus egységekre. Az általunk használt

kvázifonetikus egységek a legtöbb esetben megfelelnek egy-egy fonémának, de előfordul, az is hogy annál kisebb egységet képeznek. Az elkészült adatbázisban elegendő számú minta halmozódott fel az egyes kvázifonetikus egységekre. A feladat ezután az, hogy olyan alkalmas reprezentációt találjunk a mintáknak, ahol az azonos kvázifonetikus egységnek megfelelő példányok egymáshoz közel esnek, a különbözők pedig távol. A nyelvészetben a különböző bemozdók magánhangzóit az első és a második formásuk helyével szokás jellemezni. Az összes fonéma sikeres elkülönítése céljából több és más jellegű jellemzőket is érdemes felhasználni. A jellemzők összességét jellemzőtérnek hívjuk, ahol a minták a tér egy-egy pontjának felelnek meg. A jellemzőtér minőségét az karakterizálja, hogy az azonos osztályba tartozó elemek szórása mennyire kicsi, illetve az osztályok súlypontjainak a szórása mennyire nagy.

Az azonos kvázifonetikus egységekhez tartozó minták egy-egy ponthalmazt alkotnak. A gépi tanuló algoritmusok feladata, hogy valamilyen módon elhatárolják ezeket a ponthalmazokat egymástól, azaz felületekkel válasszák el a tér azonos kvázifonetikus egységeket tartalmazó részeit. A különböző algoritmusok elválasztó felületeit leíró függvények megfelelően rugalmasak ugyan, de általában nem tudnak tetszőleges formát leírni. Az algoritmusok szeparációs képességén – a megfelelő jellemzőtér-választás mellett – nagymértékben segíthet a különféle jellemzőtér-transzformációk alkalmazása is. A jellemzőtér-transzformációk a teret úgy igyekeznek deformálni, hogy az elválasztó felületek simábbak lehessenek. A program fejlesztése közben több jellemzőteret, gépi tanulóalgoritmust és jellemzőtér-transzformációt alkalmaztunk:

Klasszifikációra szolgáló gépi tanulási módszerek:

Neuronháló (Artificial Neural Nets), Szupport vektor gép (Support Vector Machine), Gaussi keverékmodell (Gaussian Mixture Modelling), Vetületkereső tanuló (Projection Pursuit Learner)

Lineáris jellemzőtranszformációs eljárások:

Főkomponens Analízis (Principal Component Analysis), Független komponens analízis (Independent Component Analysis), Lineáris diszkrimináns analízis (Linear Discriminant Analysis), Rugó alapú diszkrimináns analízis (Springy Discriminant Analysis)

Nemlineáris jellemzőtranszformációs eljárások:

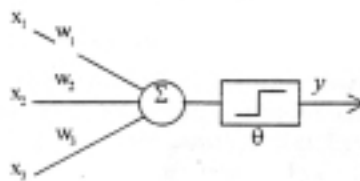
Kernel főkomponens analízis (Kernel Principal Component Analysis), Kernel független komponens analízis (Kernel Independent Component Analysis), Kernel lineáris diszkrimináns analízis (Kernel Linear Discriminant Analysis), Kernel rugó alapú diszkrimináns analízis (Kernel Springy Discriminant Analysis)

Ebben a cikkben a számunkra leginkább megfelelő kombinációt, a Neuronhálót és a Lineáris diszkrimináns analízist fogjuk bemutatni részletesebben.

3.3. Neuronháló (Artificial Neural Network - ANN)

A mesterséges neurális hálózatok, mint gépi tanuló algoritmusok használata rendkívül elterjedt. Az egyik leggyakrabban használt ANN modell az úgynevezett többrétegű előrecsatolt hálózat (Multilayer Feed-forward Neural Network) [0], amellyel a továbbiakban foglalkozunk.

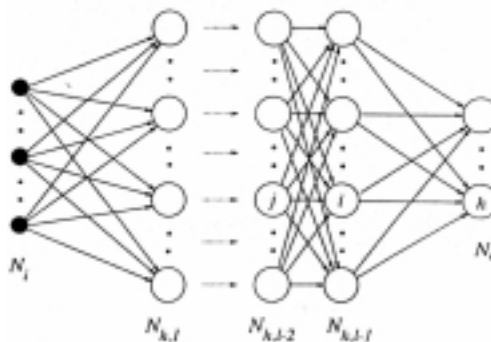
A többrétegű előrecsatolt neuronhálózatokban használt neuronmodell a McCulloch-Pitts-féle neuronmodell [0] egy módosított változata:



Egy neuron bemeneti értéke az összes, inputjához kapcsolódó neuron aktivációjának (kimenetének) $w_1, \dots, w_m$ -mel súlyozott összege, míg a neuron saját aktivációja az így számolt bemenet valamilyen rögzített függvénye. Ezt a függvényt aktivációs függvénynek nevezzük, ami a valós számok halmazát a $[0,1]$ intervallumba képezi:

$$y = g\left(\sum_{j=1}^n w_j x_j - \Theta\right) \quad g: \mathbb{R} \rightarrow [0,1]$$

A többrétegű előrecsatolt neuronhálózatokban a neuronok rétegekbe szervezettek:



A hálózat több réteggel rendelkezik: egy input, néhány közbenső és egy output réteggel. Az input réteg bemenő értékeit kívülről kapja, az outputja pedig a közbenső réteg inputjaihoz kapcsolódik. Az előrecsatolt jelző azt jelenti, hogy minden neuron kimenete kizárólag a következő rétegbeli neuronok bemenetéhez

kapcsolódhat. Az utolsó réteg, az output réteg kimenete maga a neuronháló kimenete. Adott input esetén a hálózat kiértékelését úgy végezzük, hogy rétegenként sorban haladva (az input réteg utáni rétegtől kezdve), kiszámoljuk a neuronok aktivációját az előző réteg neuronjainak aktivációjából, amíg el nem érjük a kimeneti réteget.

Ha a lehetséges inputok egy véges halmazához ismerjük a megfelelő függvényértékeket:

$$\underline{x}_i \rightarrow y_i \quad i:1,\dots,n,$$

akkor egy rögzített struktúrájú neuronhálózat súlyait beállíthatjuk úgy, hogy a neuronhálózat outputjainak az eltérése a tényleges függvényértékektől:

$$\varepsilon = \sum_p \varepsilon_p = \sum_p (y_p - out_p)^2$$

a lehető legkisebb legyen.

Többváltozós függvények minimalizálása többféleképpen történhet. A részletektől eltekintve megemlítjük, hogy ehhez többek között rendelkezésre állnak különféle gradiens alapú Newton illetve kvázi-Newton eljárások. Mi a tanítás során a back-propagation algoritmust használtuk. A back-propagation algoritmus [0] a gradiens módszer egy olyan változata, amely minden lépésben a minta pontjainak csak egy halmazát használja fel a súlyok megváltoztatására. A neurális hálózat erőssége, hogy nemlineáris input-output megvalósítására képes.

3.4. LDA (Linear Discriminant Analysis)

A tanuló algoritmusok abban az esetben működnek leghatékonyabban, ha az egyes osztályokba tartozó pontok egymáshoz közel, míg a különböző osztályok egymástól távol elkülönülve helyezkednek el. Matematikailag az, hogy a pontok egy gócba sűrűsödnek vagy elkülönülve távol állnak egymástól azt jelenti, hogy a pontok variációjára, azaz a súlyközéptől vett eltérése kicsi, illetve nagy. Mérje a σ_1 mérték azt, hogy az azonos osztályba tartozó pontok mennyire sűrűsödnek egy csomóba. Hasonlóan σ_2 legyen az a mérték, amely megadja azt, hogy az osztályok súlypontjai mennyire állnak közel egymáshoz. Tehát matematikailag a feladat abban a jellemzőtérben tanulható jól, ahol kis σ_1 mellett a σ_2 értéke nagy. A σ_1 és σ_2 hányadosából előállító $\sigma = \sigma_2 / \sigma_1$ mérték pedig a tér „jóóságát” karakterizálja, hiszen egy tört akkor nagy, ha a nevezője kicsi ugyanakkor a számlálója nagy, ami a kívánalmainkkal teljesen összhangban áll. A Lineáris diszkrimináns analízis (LDA) [0] célja az, hogy olyan lineáris transzformációt keressen, ahol a σ mérték maximális. A keresett transzformáció és így az új jellemzőtér optimalizáció segítségével kapható meg.

4. Nyelvi aspektusok

4.1. A beszédadatbázis

A beszédtechnológiai alkalmazások fejlesztéséhez megfelelően rögzített és feldolgozott beszédadatbázisokra van szükség. A beszédfelismerésnél a statisztikai alapú technikák különösen nagyméretű adatbázisokat igényelnek. A speciális szempontok alapján kialakított, áttekinthetően megszerkesztett és fonetikailag feldolgozott szóadatbázis ad alapot a program háttérében működő valós idejű, gépi fonémafelismerés kialakítására.

A beszédjavítás-terápiához összesen 200 ép és sérült hallású ember hangját rögzítettük. Minden adatközlő 204 rögzített szót mondott a számítógép mikrofonjába. Az olvasástanításhoz 250 fiú és 250 lány bemondását rögzítettünk, akik egyenként 100 szót mondtak be. A bemondók első 40 szava rögzített, a betűtanításnál használt hívóképek nevei voltak. A maradék 60 szó bemondónként változott, a gyakorisági szótár első 2000 szava közül. Végül rögzítettük 100 felnőtt beszélő kitarított magánhangzóját és folyamatos mássalhangzóját, hogy a program a tanárookra is megfelelően működjön. A beszélőfüggetlenséget még további algoritmusok felhasználásával is szeretnénk javítani.

Az így elkészült adatbázisok (50000 és 40800 szavas) egyedülállók az országban.

4.2. Az adatbázis szóanyaga

A BeMe olyan beszélőfüggetlen szoftver, melynek célja, hogy a beszédképzési, ill. olvasási nehézségekkel küszködő gyerekek megfelelő fejlesztésben részesüljenek. Az olvasásterápia rész szóanyagát egy olyan gyakorisági szótár közel 2000 leggyakrabban használt tartalmaz (É. Kiss, *et al.*, 1998: 20) szava adja, amelyet 14 első osztályos olvasókönyv szóanyagából állítottunk össze (Bácsi, Kerekes, 2002).

Az 500 szóbemondó lap felépítése a következő:

- 14 kitarított magyar magánhangzó
- 40 (leggyakoribb) tartalmaz szó betűrendben
- 60 szó (előfordulási gyakoriság alapján).

A magánhangzók és a leggyakoribb 40 szó minden adatközlőnél azonosak, míg a 60 szó adatközlőnként változó.

A beszédjavítás-terápia szóanyagát a gyakorisági „szó-tár” leggyakoribb 2000 szava alkotja, amelyek tartalmazzak:

- magánhangzókat szókezdő pozícióban
- folyamatosan képzett mássalhangzókat szókezdő pozícióban

- magánhangzó-magánhangzó kapcsolatot az artikulációs sorban
- palatális-veláris oppozíciókat
- magánhangzókat szomszédos artikulációs pozícióban
- magánhangzókat spektrumuk hasonlósága szerint
- labiális-illabiális oppozíciókat
- azonos magánhangzókat rövid-hosszú oppozícióban
- 30 leggyakoribb difónt szókezdő pozícióban
- 30 leggyakoribb difónt szózáró pozícióban.

A beszédjavítás-terápia rész a hallássérültek beszédnevelésében előforduló 204 leggyakoribb szóból és 28 szintagmából áll, amely egy szűkített gyakorisági szótár szóanyaga; tartalmaz állathangokat, egy beszédérthetőségi vizsgálat szóanyagát, a szurdologopédiai automatizálás bázisszavait, magánhangzókat mássalhangzóval kitartott ejtésben és magánhangzósorokat folyamatos ejtésben a képzési sor szerint.

A szóanyagot véletlenszerű keveréssel 40 sorba osztottuk. A sorok 4-5 szót, szókapcsolatot tartalmaznak.

Az adatközlők kiválasztásáról, a hangfelvételek helyszínének megtervezéséről, az anyaggyűjtés módjáról a „*Beszédmester*” szoftver adatbázis-specifikációja című kapcsolódó cikkünkben részletesen írunk.

4.3. Az adatbázis hanganyaga

A felvett hanganyag kisebb része tartalmaz köznyelvi fonémarendszertől való eltéréseket. Ezek elsősorban a betanítási szakaszban jelentkeztek, mivel olyan szavakban fordultak elő, amelyek kevésbé gyakoriak, s emiatt nem kerültek a hívóképek közé. E tanulmánynak nem célja a nyelvjárási jelenségek kimerítő vizsgálata, de néhány érdekességre felhívjuk a figyelmet. A magánhangzós jelenségek közül többször előfordult az *á* előtti, ill. utáni *o*-zás: *modár*, *holász*, *váloszol*. Több helyütt talákoztunk a felső nyelvéllésű hosszú magánhangzók rövidülésével: *viz*, *tüz*, *hid*. Megfigyelhető az anyagban az intervokális mássalhangzónyúlás a dunántúli felvételeknél: *tanítottó*. Hiátusos adatok a hívóképanyagban is szerepelnek. A gép a hiátusos bemondást és a *j*-vel való kitöltést is elfogadja: *di-ó* ~ *dijó*, *fi-ú* ~ *fijú*... (Kiss, 2001).

A szupraszegmentális elemekről (tempó, ritmus, emfatikus nyújtás, dallam, szünet, hangszínezet, hangsúly, hangerő) csak annyit jegyzünk meg, hogy természetesen a gyerekek olvasástudása befolyásolta a hangerőt, a tempót, a szünetet. A szósorok felolvasásakor a sor utolsó szava általában halk, az utolsó szótag gyakran elnyelődik. Ha egymást követő két szó határán azonos mássalhangzó (játék, kakas) állt, akkor az egymás utáni két rövid, külön-külön felpattanó *k* helyett időnként egy hosszú zárt egy felpattanás követett (*játé[kk]akas*

helyett *játé[k:]akas*). Az első szótagi, kötött hangsúly használata jellemző az anyag egészére.

4.4. Az adatbázis felépítése

Az adatbázisban megtalálható a digitálisan tárolt hanganyag. A hullámforma mellett a hozzájuk tartozó ortografikus karaktereket is rögzítettük. A teljes hanganyagot annotáltuk és szegmentáltuk.

Az általánosan elfogadott szempontokon túl specifikus mutatókat is felhasználtunk. A hanganyag annotálásának eredményeként elkészült 700 gyerek adatrögzítő lapja, amely a hangfájl paramétereivel és tartalmával kapcsolatos változókat tartalmazza.

Az adatlap szerkesztésénél ügyeltünk arra, hogy a gyerekek személyiségi jogai ne sérüljenek, ugyanakkor az adatlapok felhasználhatóak legyenek szociolingvisztikai kutatásokra.

A BeMe informatikai, nyelvi és társadalmi változóit a SpeechDat (Pollak, *et al.*, 2000) és a The Budapest Sociolinguistic Interview (Kontra, Váradi, 1997) felhasználásával készítettük el.

Példa egy adatlapra

SES:	s0001	* (SD) Session number - a felvétel sorszáma
ENV:	Iskola	(SD)Environment- környezet
SEX:	Lány	(SD és BUSZI) Speaker sex - az adatközlő neme
AGE:	7 éves	(SD és BUSZI) Speaker age - az adatközlő kora
ACC:	Dél-alföldi r. Szeged környéki cs.	(SD) Speaker accent - az adatközlő kiejtése
REP:	Szeged	(SD és BUSZI) Recording place - a felvétel helye
RED:	b01-b06: 2002. 06. 12. b07-b012: 2002. 09. 27.	(SD és BUSZI) Recording date a felvétel dátuma
RET:	b01-b06: 12:06-12:10 b07-b12: 12:03-12:04	(SD és BUSZI) Recording time - a felvétel időpontja
LBR:	b02: alma, ág, béka, cica, csiga, dió, dzé betű, dzsungel b03: egér, ég, fészek, gólya, gyerek, hegy, iskola, ír b04: játék, kakas, lepke, lyuk, maci, nap, nyúl, oroszlán b05: óra, ördög, őz, pók, róka, süni, szőlő, tulipán b06: tyúk, uborka, út, üveg, úrhajós, virág, zokni, zsiráf	(SD és BUSZI) Orthographic text prompt- az adatközlőtől várt szóanyag ortografikus formában

LBO:		(SD) Orthographic labelling - az adatközlő által bemondott szóanyag ortografikus formában
CMT:		(SD) Comment row - megjegyzés: Sampa-átirat
SNL:	Alacsony	(SD) Subjective noise level - szubjektív zajszint
EDU:	Óvoda	(BUSZI) Education - végzettség
BRP:	Szeged	(BUSZI) Birthplace - születési hely
MOT:	Magyar	(BUSZI) Mother tongue - anyanyelv
TIA:	-	(BUSZI) Time abroad - külföldön töltött idő
KFL:	-	(BUSZI) Knowledge of foreign language - idegennyelv-tudás
MIC:	Hama Dynamic microphone IMP. 600	(BeMe) Microphone type - a mikrofon típusa
SPD:	-	(BeMe) Speech defect - beszédhiba
POC:	apja: gépészmérnök anyja: fogorvos	(BeMe) Parents' occupation – a szülők foglalkozása
OCC:	-	(BUSZI) Occupation - foglalkozás
HLT:	éphallású	(SD) Healthy - a beszélő egészségi állapota

* Az adatlap ezt az oszlopot nem tartalmazza, csak magyarázatként szolgál.

4.5. A szegmentálás

Az adatbázis tartalmazza a korlátozottan elérhető szegmentált hanganyagot is. A szegmentálás a beszéd folyamat lineáris tagolása, vagyis a hangtest hangegységekre bontása. A hangtestben előforduló diszkrét metszetek kijelölésével jutunk el a beszédhanghoz, a hangtest legkisebb szegmentális szerkezeti egységéhez (Kiss, 2001). A beszéd időfüggvényében bejelöljük a fizikailag megfigyelhető beszédhangokat és azok határait. A szegmentálás célja, hogy az adatbázis gépi feldolgozásához megadjuk a beszédjel és a fonetikai átírat közti időbeli kapcsolatot, azt, hogy melyik szimbólum a beszédjel mely időintervallumának felel meg. A szegmentálás egységei a beszédhangok, ezekből absztraháljuk a fonémákat (Vicsi, *et al.*, 2002).

4.5.1. A szegmentálás menete

A munkát a rögzített hanganyag tisztításával kezdtük. Az első meghallgatás során a Cool Edit program segítségével töröltük a nagyobb szüneteket, eltávolítottuk az ismétlődő részeket, ill. a szó bemozdó lapon nem szereplő szavakat a wav fájlokból. Ezután kezdtük magát a szegmentálási műveletet. A munkához az MTA-SZTE Mesterséges Intelligencia Kutatócsoportja által kifejlesztett HotSA programot használtuk. Az ismételt meghallgatás célja már az volt, hogy kiszűrjük a

tisztítás után maradt esetleges zajokat, szótévesztéseket, ill. a fonetikus átírat hibáit. Ezután a spektrogram erősebb nagyítása mellett jelöltük a beszédhanghatárokat. A következő meghallgatás során pontosítottuk, véglegesítettük a határokat, javítottuk az esetleges ejtési eltéréseket. Így létrehoztuk a huo és a wav fájl mellé a dat fájlokat is. Jegyeztük az adatközlő által elkövetett hibákat és a felvétel során hallható, nem környezeti zajokat. Az adatbázis tartalmazza a beszédhangok fonetikai átírását is, ez a beszédrészlet hangjait jelölő speciális fonetikai kódok sorozata. Az elhangzott szövegek lejegyzésére a nyelvészetben általában az IPA szimbólumrendszert használják. Ez a szimbólumrendszer azonban nem illeszkedik a számítógép billentyűzetéhez, ezért a nemzetközi SAMPA (Wels, 1992) kódolással dolgoztunk. A SAMPA alapja az IPA jelölésrendszer, de az előbbi jobban alkalmazkodik a számítógéppel kezelhető karakterkészlethez; a segítségével könnyebb megoldani a gépelést és a továbbítást.

4.5.2. A szegmentálás szempontjai

A beszédhangok a lineáris hangszerkezetben hangkapcsolatokat képeznek. A hangok gyakran nem különíthetők el élesen egymástól, rövidebbek-hosszabbak köztük az átmenetek, ezért a munka során azt tapasztaltuk, hogy a beszédhanghatárt nem lehet mindig egyértelműen bejelölni. Gyakran a spektrogram és a hullámforma összehasonlítása, de még a meghallgatás sem adott objektív fogódzót a határ pontos bejelölésére.

Ezekben az esetekben az alábbi elvekhez (Vicsi, *et al.*, 2002) tartottuk magunkat.

- A szegmentumhatárokat igyekeztünk a null átmenetekhez igazítani.
- Zöngés hangok esetében ez a pozitív null átmenetet jelenti. Igyekeztünk nagyon precízen jelölni a határt. Zöngétlen hangoknál 1 ms pontossággal jelöltük a hang kezdetét.
- A magánhangzó kezdetét a zöngé indulásánál jelöltük (zöngétlen hang után).
- A zárhangok, affrikáták kezdetét a megelőző hang utolsó periódusa előtt jelöltük.
- A magánhangzó-magánhangzó vagy magánhangzó-rezonáns mássalhangzó kapcsolatokban a határt az átmeneti rész 50%-ánál jelöltük be. Ilyenkor némileg pontatlan a bejelölés, mert a hangok kettéválasztása bizonytalan (Vicsi, Víg, 1998).

Többször előfordult, hogy egy-egy hangot többszöri visszahallgatás után sem lehetett azonosítani. Ennek jelzésére a [cut] kódot használtuk.

4.5.3. Az automata szegmentálás

A beszédhang szintű szegmentálás nagy figyelmet igénylő, igen fáradságos, monoton munka. Egy adatközlő anyagát a gyakorlott szakember is min. egy óra alatt dolgozta fel.

108 kisgyerek szegmentált anyagának segítségével készült el a kézi szegmentálást könnyítő automata szegmentáló program. Ez a HotSa részeként működő olyan speciális algoritmus, amely automatikusan jelöli a hanghatárokat, ezzel könnyíti és gyorsítja a kézi szegmentáló munkáját. Tapasztalataink szerint nem működik elég pontosan, így az algoritmust még fejlesztenünk kell.

4.5.4. Szegmentálási tapasztalataink

- A koartikuláció jelensége miatti átmeneti részek, vagy az egyik vagy mindkét hang észlelésének tisztaságát rontották. A továbbiakban ezeket az átmeneteket vizsgáljuk majd a pontosabb jelölés érdekében.
- Az *l* és a *j* nehezen ismerhető fel. Sem a hullámformán, sem a spektrumképen nem különíthető el a hanghatár, pedig az auditív ellenőrzéskor jól hallható az *l* és a *j* hang is. Ennek valószínűleg az az oka, hogy a mássalhangzók azonosításához eddig használt szempontjaink (felpattanás, frikatív zörej, orrhangúság) egyike sem alkalmas a *j* és az *l* megragadására. Azért nem, mert a *j* és az *l* a magánhangzókhoz hasonlóan zengőhang, ráadásul artikulációjukat rendkívül erősen befolyásolja a mellettük lévő magánhangzó (pl. ajakkerekítéses magánhangzó mellett a *j* és az *l* is ajakkerekítéses. Mindezek miatt nincs éles határ a *j* és az *l* valamint a mellettük lévő magánhangzó hangszínképében. Ezeket a hangokat tehát nem a mássalhangzóra alkalmazott szempontok segítségével lehetséges megtalálni, hanem a magánhangzószéleket kell megfigyelni.
- A spiránsok közül az *sz-z-s-zs* jól felismerhető a spektogramon, mivel nagy zörejrel járnak. A magánhangzók közötti *v-t* nehéz megtalálni, mert kevés benne a zörej. Az *m-n* folyamatos zárhangok általában egyértelmű spektrumképet mutattak, de az *m-b* kapcsolatban (gomba, bim-bam) gyakran nem tudtuk a felpattanó zárhang zöngés zárát külön szegmentálni, hanem csak a felpattanást.
- A zöngés és zöngétlen *h* megkülönböztetése nem volt egyértelmű, bár a fonológusok által számontartott allofónokat igyekeztünk figyelembe venni.
- Szó végén, szótag végén gyakran nem pattannak fel a zárhangok (*k, g, t*), de mivel mi a szegmentáláskor a felpattanásnál lokalizáljuk a zárhangokat, ezzel a módszerrel nem vagyunk képesek az ilyen fel nem pattanó zárhangokat behatárolni. Füllel ezeket érzékelhetjük, mert az ember a záralkotási szakaszból is azonosítani tudja a zárhangokat.

- A *g* hang esetében többször találkoztunk azzal, hogy a zöngés zárt nem követte felpattanás.
- A hosszú mássalhangzó két hosszú magánhangzó között megrövidül.
- Mindezen problémák ellenére az adatbázis alkalmas arra, hogy fonológiai kutatások anyagául szolgáljon.

5. Összegzés

A projekt eredményeként az alapkutatás során kidolgozott, számos gépi tanulási algoritmus vált elérhetővé nyílt forráskódú szoftver formájában, ami reményeink szerint alapot biztosíthat az ilyen jellegű fejlesztések széles skálájának. A vizsgált új módszerek hatékonyságát a beszédfelismerés területén alkalmaztuk. Olyan közhasznú szoftver első prototípusát készítettük el, amely az alapkutatás során kifejlesztett algoritmusokra épül, és azok segítségével támogatja a hallás- vagy részképességükben sérült gyermekek beszédjavítás-terápiáját, illetve az iskoláskorúak olvasásterápiáját.

A BeMe hanganyaga korlátozottan elérhető, az adatbázis struktúrájának specifikációja nyilvános. A program fejlesztése során létrejöttek gyakorisági szótárak, nagyméretű annotált és szegmentált adatbázisok. Ezek célja a közvetlen felhasználás mellett, hogy az adatbázis alkalmazható legyen más vizsgálatokban pl.: a gépi beszédfelismerés más alkalmazásaiban, és nyelvészeti (fonetikai, fonológiai, szociolingvisztikai...) kutatásokban.

A Beszédmester szoftver 2004 februárjától ingyenesen hozzáférhető az interneten. A program mellékleteként automatikusan használható az olvasástanítás, ill. a beszédjavítás-terápia metodikája. Az olvasástanítás tematikáját tapasztalt pedagógusok állították össze, a beszédjavítás képköreit gyakorlott szurdopedagógusok válogatták, illetve a program képeit elismert grafikus készítette, de az első prototípussal nem zártuk ki a változtatás lehetőségét. A beszédjavítás-terápia és az olvasástanítás feladatait szöveges (xml) fájlban tároljuk, így könnyen módosítható a tesztelések után a tematika is. A programban használt képeket BMP formátumban tároljuk, így egyszerűen újra rajzolhatók. A képekhez tartozó szavakat is szöveges fájlban tároljuk, így ez is bővíthető.

Az általunk kifejlesztett Beszédmester szoftver azzal a céllal készült, hogy segítse az iskolások betűtanulását, illetve az otthoni gyakorlást, a beszédjavítás-terápiát. Előnye, hogy játékosan, színes képekkel, a számítógép motivációs erejét felhasználva próbálja meg a kisiskolásokat az olvasás rejtelmeire megtanítani, és a hallássérültek helyes beszédartikulációját - kiemelten a magánhangzók kiejtését -

gyakoroltatni. A szoftver célirányos, tudatos, oktató-fejlesztő program, mely figyelembe veszi az életkori sajátosságokat, komplex készségfejlesztést, nagyfokú önállóságot biztosít.

Irodalom:

- Bácsi, J., Kerekes J.** (2003) Az első osztályos olvasókönyvek szóanyagából készült Gyakorisági szótár „Van szó”. *Módszertani Közlemények* (Szeged) 2003/2.sz. 52-57.
- Bryson, A. E. Ho, Y-C.** (1975): *Applied Optimal Control: Optimization, estimation, and control*. Washington, Hemisphere-Wiley Pub.
- É. Kiss, K., Kiefer, F., Siptár, P.**(1998) *Új magyar nyelvtan*. Budapest: Osiris Kiadó.
- Fukunaga, K.** (1990) *Introduction to Statistical Pattern Recognition*. Academic Press.
- Hornik, K., Stinchcombe, M. and White, H.** (1989) *Multilayer feedforward networks are universal approximators*. *Neural Networks*, vol. 2(5). 359-366.
- Kiss, J.** (2001): *Magyar dialektológia*. Budapest: Osiris Kiadó.
- Kontra, M., Váradi, T.** (1997) *The Budapest Sociolinguistic Interview: Version 3. Linguistics Institute*. Budapest: Hungarian Academy of Science.
- McCulloch W. S., Pitts, W.**(1943) *A logical calculus of the ideas immanent in nervous activity*. *Bulletin of Mathematical Biophysics*, vol. 5. 115-133.
- Pollak, P., Cernocky, J., Boudy, J., Choukri, K., Heuvel, H., Vicsi, K., Virag, A., Siemund, R., Majewski, W., Sadowski, J., Staroniewicz, P., Trops, H., Kochanina, J., Ostrouchov, A., Rusko, M., Trnka, M.** (29 May 2000): SpeechDat(E) –Eastern European Telephone Speech Databases. Proceeding LREC’ 2000 Satellite workshop XLDB – Very large Telephone Speech Databases , Athens.
- Vicsi K., Víg A.** (1998) Az első magyar nyelvű beszédatbázis. Budapest: Beszédkutatás ’98, MTA Nyelvtudományi Intézete.
- Vicsi, K., Tóth, L., Kocsor, A., Gordos, G., Csirik, J.** MTBA-Magyar nyelvű telefonbeszédatbázis, *Híradástechnika*, LVII. 2002/8, Budapest, 35-43.
- Wels, J. et al.** (1992) Standard Computer-Compatible Transcription. Esprit Project 2589 (SAM), Doc.no. SAM-UCL-037. London: Phonetics and Linguistics Dep., UCL.