A Magyar Referencia Beszédadatbázis és alkalmazása orvosi diktálórendszerek kifejlesztéséhez

Vicsi Klára[1], Kocsor András[2],Tóth László[2], Velkei Szabolcs[1],
Szaszák György[1], Teleki Csaba[1], Bánhalmi András[2] és Paczolay Dénes[2]

[1] BME Távközlési és Médiainformatikai Tanszék,
Beszédakusztikai Kutatólaboratórium, 1111 Budapest, Sztoczek u. 2.
{vicsi, velkei, szaszak, teleki}@tmit.bme.hu

[2] MTA-SZTE Mesterséges Intelligencia Tanszéki Kutatócsoport,
6720 Szeged, Aradi vértanúk tere 1.
{kocsor, tothl, banhalmi, pdenes}@inf.u-szeged.hu

Kivonat: Poszterünk bemutatja a Magyar Referencia Beszédadatbázist, továbbá az erre épülve párhuzamosan fejlesztett két orvosi diktálórendszer jelenlegi szerkezetét és képességeit.

1 A Magyar Referencia Beszédadatbázis

A *Magyar Referencia Beszédadatbázis*t *(MRBA)* a BME TMIT Beszédakusztikai Laboratóriuma és a szegedi SZTE Informatikai Tanszékcsoport együttműködésben hozta létre [1]. A cél egy olyan irodai, otthoni környezetben olvasott folyamatos szöveget tartalmazó beszédadatbázis megalkotása és akusztikai, nyelvi feldolgozása volt, amely alkalmas PC-s beszédfelismerők betanítására, tesztelésére.

Az adatbázis szöveganyagát úgy terveztük meg, hogy lehetőséget adjon különböző típusú beszédfelismerők betanítására és kiértékelésére. Ezek közül a legnagyobb kihívást a folyamatos beszédet felismerő diktáló rendszerek jelentik, amelyeknél a felismerés szónál kisebb felismerési egységek (beszédhangok, difón, trifón egységek) modellezésén alapul. Ezek betanításához olyan folyamatos szöveg összeállítására van szükség, amelyben ezek az elemek elegendően sokszor fordulnak elő, mindamellett a szöveganyag lehetőleg minél rövidebb. Az MRBA szöveganyagának összeállításához újságcikkek szövegét használtuk fel, az adatbázisba bekerülő mondatokat úgy válogatva össze, hogy a leggyakoribb di- és trifónok megfelelő mennyiségben álljanak rendelkezésre. A mondatok mellett fonetikailag gazdag szavakat is kiválasztottunk, az esetlegesen hiányzó vagy nem kellő számban előforduló beszédhangok példányszámának növelése érdekében. Így egy adatközlő 12 különböző mondatot és 12 különböző, a mondatoktól független szót olvas fel, összességében pedig 332 adatközlő hanganyaga került az adatbázisba.

A beszédadatbázis felvételeit különböző helyszíneken: zajos, kevésbé zajos irodai helyiségekben, laborokban, otthonokban rögzítettük. A felvételeknél szinkronban két

különböző rendszerrel dolgoztunk. Az egyik az ún. *referenciarendszer*, amelyben mindig ugyanazt a jó minőségű mikrofont, hangkártyát és laptopot használtunk. A másik rendszer, az ún. *variált rendszer* esetében különböző, jobb, kevésbé jó mikrofonokat, hangkártyákat, PC-ket használtunk, a lehető legnagyobb variáltsággal. A régiók, dialektusok és generációk lefedése céljából a felvételeket Magyarország 4 különböző tájegységében lévő városban rögzítettük: Budapesten, Szegeden, Győrben és Miskolcon, lehetőség szerint különböző életkorú és nemű beszélőket választva.

A felvételek mindegyikét annotáltuk, ami azt jelenti, hogy minden hangfájl mellé egy címkefájlt készítettünk, amely különféle információkat tartalmaz a hangfájl paramétereivel és tartalmával kapcsolatban: az elhangzott szöveg ortografikus lejegyzését, hibás kiejtést, nem érthető szavakat, szótöredékeket, a beszélő nem beszédből származó hangjait, környezeti zajokat, stb. Az adatbázis közel egyharmadán, azaz 100 beszélő anyagán manuálisan fonetikai szintű szegmentálást és címkézést is végeztünk, a fonetikai szegmentumok címkézéséhez a SAMPA nemzetközi kódtáblát használva.

2 Orvosi diktálórendszerekről általában

Az automatikus beszédfelismerési technológia jelenleg még nem képes az általános célú folyamatos diktálás tökéletes megoldására, viszont elfogadható pontosságot tud nyújtani olyan feladatok esetében, ahol a szókincs és a nyelvtani felépítés korlátozott. Így lehetővé teheti az ún. beszédalapú dokumentálást olyan szakmák esetében, amelyek szakszöveg-jellegű dokumentációt igényelnek. Kitűnő példa erre az orvosi vizsgálati eredmények rögzítése, amely folyamat felgyorsítása különösen nagy jelentőséggel bír. Ilyen diktálórendszerek a világnyelvekre már léteznek, viszont kisebb és speciális nyelvi tulajdonságokkal rendelkező nyelvekre egyelőre nagyon kevés orvosi diktálószoftver látott ezidáig napvilágot, amely többek között a nyelvi sajátosságokon túl a magas fejlesztési költségeknek tudható be. Az MRBA adatbázisra alapozva mind a BME TMIT Beszédakusztikai Laboratóriuma, az MTA-SZTE Mesterséges Intelligencia Kutatócsoportja belefogott egy orvosi diktálórendszer kifejlesztésébe. A két csoport részben eltérő részfeladatokat tűzött ki maga elé (endoszkópos leletek diktálása illetve pajzsmirigy-scintigráfiás leletek diktálása) és részben eltérő technológiákat alkalmaznak, de természetesen eredményeiket folyamatosan egyeztetik, ami lehetővé teszi a tapasztalatok kicserélését és a technológiák összehasonlítását.

3 Endoszkópos leletek diktálása

Az endoszkópiai leletek gépi beszédfelismerésére és karakteres lejegyzésére képes rendszert a BME TMIT Beszédakusztikai Laboratóriuma készítette el. A laboratóriumban kifejlesztésre került egy Windows XP alatt működő beszédfelismerő fejlesztői környezet, amely alkalmas különböző középszótáras 1000-10000 szavas szövegek betanítására és felismerésére. A felismerő a statisztikai alapon működő HMM akusztikai fonémamodellekkel [2], valamint a statisztikai alapú bi-gram nyelvi modellel működik, nemlineáris simítást használva [3]. Az akusztikai modelleket az MRBA beszédadatbázissal tanítottuk. A nyelvi betanításhoz a budapesti SOTE II. sz. Belgyógyászati Klinikájától (2700 lelet) és a szegedi Orvostudományi Egyetemről (6365

lelet) gyűjtött korábbi leletanyag korpuszt használtuk. Ezen szövegkorpusz alapján elkészítettük el a teljes szóalakszótárat, amely 14331 szót tartalmaz, a kiejtési szótárat és ezek téma szerint osztott kisebb szótárait, valamint a korpusz alapján morfémaszótárat is készítettünk, amelynek nagysága 6824 morfémaelem.
A felismerő optimális működését az akusztikai [4] és nyelvi modellek változtatásával állítottuk be. Lényegében a nyelvi modellhez n-gram modelleket használtunk, de az egyik megoldásban a hagyományos szóalakok az alkotó elemek, a másik megoldásban viszont a morfémák.
Külön súlyt fektettünk a valós idejű felismerés elérésére: a dinamikus címzésen és az akusztikai modellek indirekt megközelítésén túl memóriaelérési optimalizáció, valamint nyalábolt keresésnél (Beam Search) változó terű nyaláb alkalmazásával.

4 Pajzsmirigy-scintigráfiás leletek diktálása

Szegeden kifejlesztettünk egy magyar nyelv automatikus felismerésére alkalmas magmodult, amelyre különböző, speciális feladatokhoz igazított diktálórendszerek építhetőek. A magmodul tartalmazza az un. akusztikai modellt, amely alkalmas a magyar nyelv beszédhangkészletének felismerésére és reprezentatív módon történő modellezésére. A modell felépítésére két egymástól relevánsan eltérő megközelítést alkalmaztunk. Az egyik a beszédfelismerésben közismert és gyakran alkalmazott Rejtett Markov Modell, a másik pedig a Szegeden kifejlesztett újszerű sztochasztikus szegmentális megközelítés. Mindkét modell betanításához és teszteléséhez az MRBA adatbázist használtuk fel.

A rendszerhez jelenleg egy olyan nyelvi modellt fejlesztünk, amely pajzsmirigy-scintigráfiás leletek diktálását teszi lehetővé. A nyelvi modellt 9231 írott pajzsmirigy lelet és több mint 2500 szóalak alapján építettünk fel. A nyelvi modellezésre többféle technológiát kipróbáltunk. A legegyszerűbb ezek közül az ún. szó N-gram modell. Ez megadja, hogy milyen valószínű egy adott szó az N-1 darab előtte álló szó ismeretében. Az N-gram modell kiszámításakor az ún. előrehozott N-gramm kiértékelési technológiát használjuk, amelynek segítségével a keresés során a hipotézisek száma lecsökkenthető, így a felismerés gyorsabbá tehető.

A magyar nyelv szabad szórendűsége miatt az N-gram technika nem olyan hatásos, mint pl. az angol esetében, ezért a hosszú távú kapcsolatok leírásához más modellekre is szükség van. Egy ilyen lehetőség az MSD-kód (morfoszintaktikai kód) alapú szabályok alkalmazása. Az MSD kódos leírásnál a szavak jelentése eltűnik, csak a szavak mondattani szerepe marad meg, így az MSD-kódon alapuló nyelvtanok segítségével modellezhető a mondatok felépítése.

Mind az MSD-kódokon alapuló nyelvtanok, mind a szó-N-grammok esetén komoly gondot okoz a memóriaigény, illetve az modellek pontos betanítása/kialakítása. Egy lehetséges megoldás az, hogy az osztályokra készítünk egy nagyobb (4-, vagy 5-gramm) és szavakra egy kisebb (2-, vagy 3-gramm) szótár alapú nyelvtant. A nyelvtannak az osztály-N-gramm része szintaktikai szabályokat, míg a szó-N-gramm része inkább szemantikai szabályokat szolgáltat. Jelenleg ez a kombinált megoldás tűnik a legígéretesebbnek nyelvi modelljeink közül.

Folyamatos beszéd felismerésekor további problémát jelent a hasonulás, ami akusztikailag megváltoztathatja a szavak végét vagy elejét. A hasonulás kezelését bonyolítja az is, hogy nem tudjuk, hogy a beszélő tart-e szünetet a szavak között vagy

sem, ezért ezeknek a hasonulást leíró szabályoknak, mint alternatíváknak kell megjelenniük a nyelvtanban. A hasonulást leíró szabályok összetettsége és nagy száma miatt a felismerés során a hasonulás megfelelő sebességgel való kezelése speciális problémaként jelentkezik.

Jelenleg a rendszerünk 95% körüli szó-szintű találati pontosság elérésére képes (a konkrét értékek természetesen függenek a teszt-adatbázistól és a használt nyelvi modelltől).

Bibliográfia

1. Vicsi Klára, Kocsor András, Teleki Csaba, Tóth László: Beszédadatbázis irodai számítógép-felhasználói környezetben, II. Magyar Számítógépes Nyelvészet Konferencia, 2004.
2. Claudio Becchetti, Lucio Prina Ricotti. Speech Recogntion,Theory and C++ implementation. Fondazione Ugo Bordoni, Rome, 1999. ISBN 0-471-97730-6
3. Ney, H., Essen, U., Kneser, R. On Structuring Probabilistic Dependencies in Stochastic Language Modeling. Computer Speech and Language, 8:1-38. oldal
4. Velkei Szabolcs, Vicsi Klára: Beszédfelismerő modellépítési kísérletek akusztikai, fonetikai szinten, kórházi leletező beszédfelismerő kifejlesztése céljából, MSZNY 2004.
5. Kocsor, A., Bánhalmi, A., Paczolay, D., Csirik, J., Pávics, L.: The OASIS Speech Recognition System for Dictating Medical Reports, Annual Congress of the European Association of Nuclear Medicine, EANM'05, 15-19 October, Istanbul, 2005.
6. Kocsor, A., Bánhalmi, A., Paczolay, D.: Informatikai és matematikai módszerek egy pajzsmirigy scintigráfiás leletek diktálására alkalmas rendszerben, IV. Alkalmazott Informatika Konferencia, AIK 2005, Május 27, Kaposvár, 2005.