

NÉGYESI IMRE¹**A mesterséges intelligencia és a hadsereg III.
(Beszédfelismerő szoftverek II.)****The artificial intelligence and the army III.
(Speech recognition software II.)****Absztrakt**

A mesterséges intelligenciára számos terület vizsgálatával foglalkozik, mint a szakértői rendszerek, látás és képfeldolgozás, természetes nyelvek feldolgozása, beszédfelismerés stb. Ennek a cikknek a témája a beszédfelismerés, amelynek területe nem maradhat ki a hadseregekben sem a kutatási irányokból. Az első részben² megtörtént az elméleti alapok tisztázása, ebben a második részben a beszédfelismerő rendszerek működését és a gyakorlati megvalósítás kérdéseit vizsgáltam.

Kulcsszavak: informatika, informatikai rendszer, mesterséges intelligencia, beszédfelismerés

Abstract

Artificial intelligence deals with a wide range of areas such as expert systems, visual and image processing, natural language processing, speech recognition, and so on. The subject of this article is speech recognition, the field of which can not be left out of the military either from the research directions. In the first part the clarification of the theoretical bases was made, in this second part I examined the functioning of the speech recognition systems and the practical implementation issues.

Key words: Informatics, It System, Artificial Intelligence, Voice Recognition

¹ Nemzeti Közszolgálati Egyetem, Hadtudományi és Honvédtisztképző Kar, egyetemi docens – National University of Public Service, Faculty of Military Science and Officer Training, associate professor, E-mail: negyesi.imre@uni-nke.hu, ORCID: 0000-0003-1144-1912

² Négyesi Imre: A mesterséges intelligencia és a hadsereg II. Hadtudományi Szemle, 2017. X. évf. 2. szám, ISSN: 2060-0437

HADTUDOMÁNYI SZEMLE

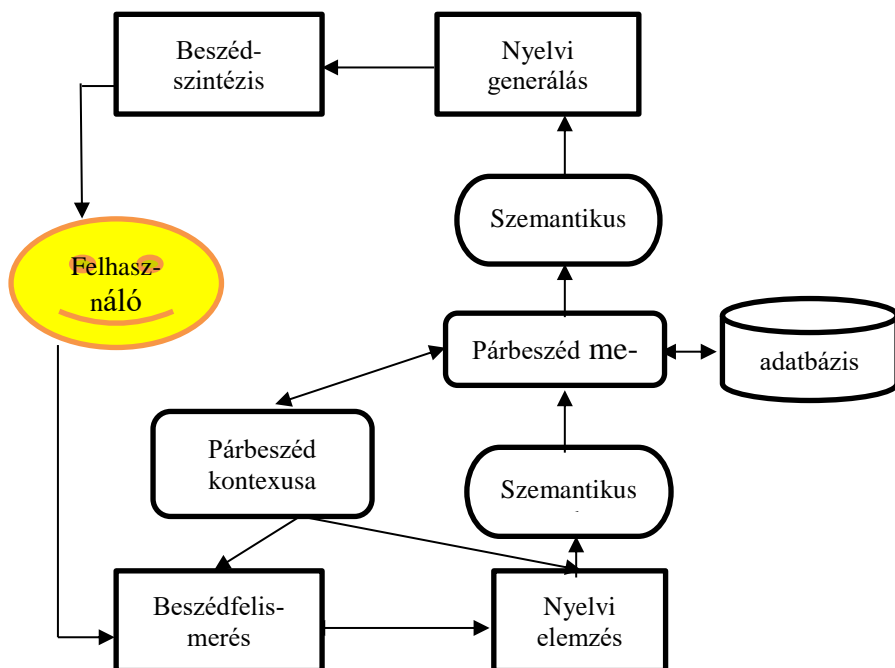
2017. X. évfolyam 4. szám

BEVEZETÉS

A mesterséges intelligenciák alkalmazási terület szerinti osztályai között egy viszonylag új, a különböző típusú munkák hatékonyságát növelő lehetőség lehet a beszédfelismerő rendszerek alkalmazása a beérkező információk kezeléséhez. A beszédfelismerő rendszer alkalmazásának alapelveit az előző cikkemben már tisztáztuk, így most következhetnek a gyakorlati alkalmazás alapjai és a megvalósítás története, valamint kitekintés a jelenlegi helyzetre. A mesterséges intelligencia alkalmazása a beszédfelismerésben a katonai feladatok végrehajtásánál is jelentős távlatokat jelenthet, ezért a cikk kapcsolódik a folyamatban levő katonai kutatásokhoz is.

A BESZÉDFELISMERŐ RENDSZER ESZKÖZEINEK MŰKÖDÉSI ELVE

A beszédfelismerő rendszerek eszközei viszonylag egyszerűek, de működési elvük meglehetősen összetettek. A beszédfelismerő rendszert használók számára feltétlen szükséges ismereteket tekintjük át, amely beszédfelismerő rendszerek technikai eszközökből és megfelelő szoftverből (esetleg több szoftverből) állnak. Az egyes párbeszéd informatikai rendszerek architektúrája igen különböző lehet.



1. sz. ábra: A párbeszédés rendszerek általános architektúrája
(Forrás az irodalomjegyzékben.)

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 4. szám

Az ábrán azok a modulok láthatók, melyeknek mindegyik párbeszédés rendszerben elő kell fordulniuk. Mint látható, a felhasználó a beszédfelismerést végző modulon keresztül közül információt a rendszerrel, és a beszédszintézist végző modul szolgáltat számára adatokat. Az egyes modulokról, komponensekről nézzünk egy részletesebb leírást, kiemelve az esetleges működési különbségeket a katonai felhasználás esetében.

A „Beszédfelismerés” modul a felhasználó által kimondott hanginputot a beszédfelismerő dolgozza fel. A célja a szöveg, azaz szavak, mondatok kinyerése az inputból. Mindez egy igen bonyolult elemzési folyamat eredményeként áll elő. A „Nyelvi elemzés” modul a beszédfelismerő által kinyert szöveget sokszor nyelvtani elemzésnek vetik alá. A nyelvi elemző által szolgáltatott információk sokszor segítenek a rendszernek tisztázni a felhasználó által mondott szöveget, hiszen az zajjal terhelt és rosszul artikulált lehet, illetve kétértelműségeket hordozhat magában. Éppen ezért a nyelvi elemző visszahathat a beszédfelismerő működésére, mint azt az ábrán a köztük húzott oda-vissza nyíl is illusztrálja. A nyelvi elemző a szöveget szintaktikusan elemzi, azaz pl. megállapítja, hogy mi volt az állítmány a mondatban vagy mi volt a tárgy. Egy beszélgetés során azonban az emberek a beszéd szemantikáját, azaz a jelentését próbálják kinyerni. Hogy egy informatikai rendszerben mit tekintünk egy szöveg jelentésének, az nagyban függ az egyéni értelmezésektől, a rendszer robusztusságától, az alkalmazási területtől stb. Katonai jellegű szövegek esetében ezeknek a moduloknak a működése megegyezik az általános szövegekkel. A „Szemantikai tartalom” modul szerepe azonban gyakorlatilag minimális lesz, hiszen egy katonai szöveg (parancs, utasítás, nyugtázás stb.) esetében nem függhet a szövegek jelentése egyéni értelmezéstől és egyéb más tényezőktől sem.

A „Párbeszéd menedzser” modul a párbeszédés rendszer motorja. A szövegből kinyert szemantikus tartalom az inputja, ez alapján kell döntéseket hoznia olyan kérdésekben, mint hogy a tudásbázisból (adatbázisból) milyen adatokat kérjen le, milyen adatok írjon oda-vissza, illetve a felhasználó felé szolgáltatandó outputnak mi legyen a jelentése (szemantikus tartalma). A párbeszéd menedzser karbantartja az adatoknak egy kollekcióját, melyet a párbeszéd kontextusának nevezünk. Ebben olyan átmeneti információk tárolódnak, mint hogy például kiről/miről szól aktuálisan a beszélgetés, vagy hogy milyen adatok pontosítására kell a felhasználót a közeljövőben felszólítanunk. Nyilvánvalóan ez a kontextus visszahathat mind a beszédfelismerő, mind a nyelvi elemző működésére.

A „Nyelvi generálás” modul a párbeszéd menedzser által szolgáltatott szemantikus tartalom alapján a nyelvi generálnak kell helyes és értelmes mondatokat, szöveget alkotnia. A „Beszédszintézis” modul a nyelvi generáló által szolgáltatott szöveget a beszédszintetizátor szóllaltatja meg, amit a felhasználó meghallgathat.

A beszédfelismerő rendszer alkalmazásához elengedhetetlen, hogy annak legfontosabb alapjaival a felhasználók megismerkedjenek. Ezek közül az első, hogy miként kezelik ezek a rendszerek a különböző karaktereket (hangokat, betűket). A beszédfelismerés során a technikai eszközök érzékelik a beszéd hangjait és felfogják a beszélő által írott különböző írásjelekkel és ezek feldolgozásával alakítják ki az írott szöveget. A beszédfelismerő rendszer gördülékeny alkalmazásához megfelelő gyakorlatra van szükség, mely az alkalmazás

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 4. szám

során folyamatosan bővül. A beszédfelismerő rendszert használók kezdetben nehezen tudják rugalmasan összeilleszteni a gépelendő beszédszöveget és a szerkesztésre vonatkozó szavakat, ezért célszerű a szerkesztésre vonatkozó szavak gyűjteményét jól láthatóan elhelyezni, addig amíg a szavak felismerése automatikussá válik. Ez ezért is szükséges, mert a beszédfelismerő rendszer által írt szöveg a legjobb felismerési ráta mellett is tartalmazhat hibákat, tehát azt rendszerint javítani kell, bár a hibák a szöveg értelmezését általában nem nehezítik meg.

A beszédfelismerő rendszer azon az elven alapul, hogy a beszéd hangját, megfelelő mikrofonnal elektronikus jelekké alakítjuk, majd ezeket az elektronikus jeleket szavanként különböző írásjelekkel leírható karakter-sorokká alakítjuk át. A beszéd elektronikus jelekké történő alakítása a távközlésben már több, mint egy évszázad óta ismert megoldás, ami azóta is folyamatosan fejlődött. A beszéd elektronikus jelekké történő átalakításának eszköze a mikrofon, ami a hang mechanikus hullámait képes elektronikus hullámokká alakítani. Az elektronikus hanghullámok megfelelő hangszórók alkalmazásával vissza állíthatók mechanikus hanghullámokká és ezáltal ismét hallhatóak lesznek. A mai korszerű technikai eszközök igen nagy érzékenységgel képesek a mechanikus hanghatásokat felfogni, azokat elektronikus jelekké alakítani, az elektronikus hangjeleket a legkülönbözőbb célokból formálni és alakítani. Ezek a műveletek teszik lehetővé, hogy távközlésben és a szórakoztató elektronikában a nagyon sok igényt kielégítő hanghatásokat létrehozzuk.

Az elektronikus hangjelek feldolgozása ma már legtöbb esetben digitalizált formában történik, ami nagyon sok előnnyel jár. A digitalizált elektronikus jelek torzítástól szinte mentesen tetszés szerinti alkalommal lejátszhatók, erősíthetők, szétoszthatók és hihetetlen távolságokra továbbíthatók. A digitalizált elektronikus hangjeleket, leegyszerűsítetten úgy tekinthetjük, mintha az egyes hanghatásokat megfelelő számokkal fejeznék ki.

A beszédfelismerő rendszerek legmeghatározóbb folyamata az elektronikus hangjeleknek karakterekké történő átalakítása. A folyamatnak sokféle megoldása ismert, de a legáltalánosabb használt módszerek nem az egyes hangelemeket, hanem az egyes szavakat alakítják át megfelelő karakter-sorokká. A leírt szövegek lényegében a beszélt szavak hangösszetevőit tartalmazó karakter-sorokból, leírt szavakból állnak. Egy-egy szó karakter-sorának kialakítása viszonylag egyszerű elektronikai és számítástechnikai feladat. Az értelmes mondatokat alkotó szavak együttesének karakter-sorokkal történő leírása azonban ennél jóval összetettebb folyamat, melyhez a nyelvi törvényszerűségeket figyelembe vevő szoftvereket kell alkalmazni.

A beszédfelismerő rendszerekkel szembeni igazán nagy kihívás, hogy a folyamatos beszédet legyenek képesek, lehetőleg a beszéd elhangzásával egy időben, folyamatos karakter-sorokká átalakítani, leírni. Szerencsére a beszédfelismerő rendszert felhasználóknak az itt ismertetett, bonyolult folyamatot csak olyan mélységig kell megismerniük, hogy a beszédfelismerő rendszert működtetni tudják és értsék, hogy mit, miért kell csinálni. A beszédfelismerő rendszerek alkalmazási kézikönyvei ezt az egész folyamatot nagyon leegyszerűsítve mutatják be és a felhasználó számára egyszerű lépések megtételében adják meg.

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 4. szám

A SZÓKÉSZLET ÖSSZEÁLLÍTÁSÁNAK FELADATAI

Amint azt már fontossága miatt többször is említettük a beszédfelismerő rendszerek bevezetésének legfontosabb és egyben legmunkaigényesebb része a megfelelő szókészlet összeállítása. A beszédfelismerés pontossága, és ezáltal használhatósága döntően attól függ, hogy az alkalmazott szókészlet milyen mértékben tartalmazza a szakterületben előforduló szavakat. A beszédfelismerő rendszer csak azokat a szavakat képes felismerni és megfelelően alkalmazni, melyek a szakmai szókészletben megtalálhatók.

Általában meglehetősen nehéz feladat meghatározni azokat a részterületeket, melyeknek szókészletét a szakmai szókészletbe be kell venni. Az a szakmai szókészlet melyből hiányoznak kisebb, nagyobb részterületek szókészletei, azáltal lassítja a beszédfelismerés folyamatát, mert gyakran fordulnak elő olyan szavak, melyeket a rendszer nem ismer fel. Ezzel szemben viszont az a szakmai szókészlet, mely sokkal szélesebb részterületek szókészletét is tartalmazza, azért lassítja a beszédfelismerés folyamatát, mert a szükségesnél nagyobb szókészletből kell kiválasztani a megfelelő szót.

Az első feladat, hogy határozzuk meg azokat a szakterületeket, amelyek részterületként meghatározzák a szókészlet összeállítását. Az alkalmazandó szakkifejezések köre természetesen a tesztelések során és az „éles” feladatok végrehajtása során egyaránt bővíthet. Az ily módon áttekintett szakterületek nagyon széleskörűek, ezért a szakmai szókészlet is meglehetősen nagy lesz, ami lassíthatja a beszédfelismerő rendszer működését, azonban a szűk szókészlet miatt a pontosság romlana. A katonai feladatokhoz kapcsolódó szókészletek összeállítása egy viszonylag egyszerűbb feladat, mert a hadseregekben a szakterületek elhatárolása mindig is jelen volt.

A szaktárgyi szövegből történő szókészlet összeállítás menete az alábbi lehet:³

- ki kell jelölni azt a személyt, vagy személyeket, akik a témakör jó ismerői;
- össze kell gyűjteni a kiválasztott szakértők dokumentáció gyűjteményének elektronikus változatát;
- a dokumentáció gyűjteményből a beszédfelismerő rendszer eszközeivel ki kell választani a szókészletbe bekerülő szavakat;
- rendszeres vizsgálatokat (tesztek) kell folytatni a már rendelkezésre álló szókészlettel;
- ha a tesztek során a felismerési ráta eléri a minimálisan szükséges szintet (kb.: 92-95%), el lehet kezdeni a rendszer gyakorlati alkalmazását és a gyakorlati eredmények alapján kell a szókészletbővítést folytatni;
- a 95% feletti helyes beszédfelismeréshez tartozó szókészletet legalább két egymástól független biztonságos helyen tároljuk;
- a 95% feletti pontosságú szakmai szókészlet összeállítása több hónapot, különleges esetekben akár egy évet is igényelhet.

³ A beszédfelismerő rendszer szakirodalma ezt a folyamatot, „adaptációnak” nevezi, abból kiindulva, hogy folyamat során újabb és újabb szavakat adunk, a már korábban kialakított szókészlethez.

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 4. szám

A szakmai szókészlet összeállításának folyamata viszonylag egyszerű, de meglehetősen hosszadalmas és nagy figyelmet igénylő munka. A szakmai szókészlet összeállítása akkor kezdődhet meg, amikor az alkalmazandó beszédfelismerő rendszert kiválasztottuk és azt a számítógépre telepítettük. A szakmai szókészlet összeállítását tehát meg kell, hogy előzze az alkalmazandó rendszer kiválasztása és telepítése, valamint annak a szállító által történő bemutatása, a munkába bevont szakértők betanítása.

A továbbiakban tekintsük át, hogy melyek lehetnek a megvalósítás ütemezési folyamatainak főbb szakaszai:

- a részterületek kiválasztása, meghatározása;
- az egyes részterületekhez tartozó jelentősebb dokumentumok összegyűjtése;
- a szakterületek dokumentumaiban szereplő szavak adaptálása a szakmai szókészlethez;
- a folyamatosan bővülő szókészlet felismerési próbái (tesztek);
- a tesztek alapján a szakmai szókészlet pontosító korrekcióinak elvégzése;
- a szakmai szókészlet korrekciók utáni zárása, a lezárt szókészlet biztonsági tartalékainak elhelyezése.

Miután az új szavak adaptálása több szakterületet érinthet, ezért több szakértő bevonásával hajtható végre, azonban mindenképpen szükséges egy olyan szakértő kijelölése, aki a munkát megfelelően koordinálja, vezeti. A szakmai szókészlet összeállítása, a ráfordított napi időtől függően akár több hónapig is eltarthat. Az adaptációs folyamat alatt a beszédfelismerő rendszer már használható, azzal a megjegyzéssel, hogy kezdetben igen magas lesz a felismerési pontatlanság és ezért a végleges szöveg előállítás, a javítások miatt több időt vesz igénybe. Az adaptáció előrehaladásával a felismerési pontosság fokozatosan nő és egyre gyorsabban lehet a felhasználásra alkalmas gépelt szöveget előállítani. Az előírt felismerési pontosság (minimum 90%) eléréséig a rendszer „éles” használata nem javasolt. Katonai feladatok esetében a végleges szöveg előállítás előtt a beszédfelismerők „éles” használata nem javasolható. Az előírt felismerési pontosságot a szabályzatokban szükséges megjelölni.

A BESZÉDFELISMERŐ RENDSZER GYAKORLATI ALKALMAZÁSÁNAK ELŐKÉSZÍTÉSE

A beszédfelismerő rendszer gyakorlati alkalmazását gondosan elő kell készíteni és megfelelően szabályozni kell. A beszédfelismerő rendszert felhasználók, a funkciójukhoz mérten más-más felkészítést kapnak. A beszédfelismerő rendszer általános gyakorlati alkalmazásához lényegesen kevesebb ismeret kell, mint a rendszer üzemkészen tartásához, ezért el kell különítenünk a beszédfelismerő rendszer fenntartóját, a rendszer üzemeltetőjét és a felhasználókat.

A beszédfelismerő rendszer fenntartóra akkor van szükség, ha az üzemeltető szervezet állományában senki nem rendelkezik olyan digitális írástudással, hogy képes lenne ellátni a

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 4. szám

beszédfelismerő rendszer rendszeres felülvizsgálatát. Ilyen esetekben legcélszerűbb egy beszédfelismerő rendszert szállító és fenntartó vállalattal fenntartási szerződést kötni.

Abban az esetben, ha egy vállalatnál van olyan személy, aki rendelkezik olyan szintű digitális írástudással, hogy megfelelő előkészítés és dokumentumok rendelkezésre állás esetén képes ellátni a beszédfelismerő rendszer felügyeletét, fenntartását, akkor ezt a személyt rendszer üzemeltetőnek nevezzük.

Ennek megfelelően a beszédfelismerő rendszer kezelő személyzetét és felhasználóit az alábbi csoportokba oszthatjuk:

- beszédfelismerő rendszer létrehozója, szállítója;
- beszédfelismerő rendszer üzemeltetője (fenntartó, karbantartó, „rendszergazda”);
- beszédfelismerő rendszer felhasználója (kezelőszemélyzet).

A beszédfelismerő rendszer felhasználóinak a rendszer működését olyan mértékben kell ismerni, hogy képesek legyenek a szakmai szókészlet folyamatos bővítésére. A rendszer üzemeltetői segítséget tudjanak nyújtani az "egyszerű" felhasználóknak esetleges problémáik esetén. Ha a beszédfelismerő rendszert alkalmazók egyetlen egyszerű beszédfelismerő rendszer használóval kezdik a munkát, akkor elkerülhetetlen a folyamatos kapcsolat a beszédfelismerő rendszert szállítóval, amelynek kereteit szerződésben kell rögzíteni.

A beszédfelismerő rendszert a konkrét felhasználó hangjához adaptálni, beállítani kell. A megfelelően magas felismerési ráta érdekében a beszédfelismerő rendszert meg kell ismertetni a felhasználó jellegzetes személyi „hang-spektrumaival”. A hang-spektrum megismerése általában néhány percnyi előre elkészített szöveg felolvasásával történik. A gép, ha a spektrum felismerő folyamat során, egy megadott felhasználói névhez már elegendő információval rendelkezik, akkor ezt jelzi. (A felhasználói beszéd-spektrum felismerést célszerű megismételni, ha a felismerési ráta egy bizonyos eltűrt szint alá csökkenne.)

Mielőtt rátérnénk a konkrét megvalósítási lehetőségekre bemutatására tekintsük át, hogy melyek azok a problémák, amelyekre mindenképpen választ kell adnia az alkalmazandó beszédfelismerőnek:

- folyamatos hangnyomás változásból kvantált minőségi jellemzők elkülönítése és azonosítása;
- a hullámforma erősen változik az akusztikai környezet hatására – visszaverődések, zajok, interferencia, háttérbeszélgetés, zene stb.;
- egyazon mondanivaló végtelen sok akusztikai formában jelenhet meg;
- a beszédsebesség változik a beszélő személyétől függően, és egy beszélő esetében is. A fonémák⁴, sőt, fonémáreszek időtartama nem egységesen változik meg;
- a beszéd több, mint elemek egymásutánisága! Lényegesek a hangsúly, a dallam, a szünetek, a ritmus és a tempóváltások is, mert ezek a felismerendő üzenet lényegi elemei;

⁴ Fonéma: Jelentés megkülönböztető hang, melynek megváltoztatása az egész szó jelentését megváltoztatja.

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 4. szám

- a beszédjel a beszéd tartalmi információt részben akusztikus, részben nyelvi szinten hordozza. (Jelentést nem hordozó szavak érthetősége 70-75%.) A beszédfeldolgozás hierarchiaszintjei (pragmatikai, megértési szint, szemantikai jelentéstani szint, szintaktikai, mondattani szint, lexikai szint, fonetikai, fonológiai szint, akusztikai szint) egyidejűleg aktívak!

MŰKÖDŐ BESZÉDFELISMERŐ RENDSZEREK

A beszéd felismerő rendszerrel foglalkozó vállalatok közül az első a Philips volt, amely 50 éve fejlesztette ki az első diktáló eszközt. Az analóg és a digitális diktafonjaiból 1990-re már 10 millió eladás történt. A Philips fejlesztő csapata már 1994 óta foglalkozik beszéd felismeréssel, melynek eredményeként 2001-ben készült el az intelligens beszéd interpretáció az „ISI”. A Philips piacvezető terméke a SpeechMagic™ beszéd felismerő rendszer, amely már magyarul is elérhető volt, ezért ennek a rövid bemutatásával kezdjük a működő rendszerek bemutatását. A másik ok, hogy miért ezt a szoftvert választottuk az volt, hogy ma is jelen van a piacon, továbbfejlesztett változata már huszonegy nyelven elérhető. Katonai jelentősége pedig abban áll, hogy két jelentős területen (gyógyítás, jog) gyakorlatilag azonnal alkalmazható lehet különösebb „katonásítás” nélkül is.

A Philips felismerte, hogy a szakmai beszéd felismerő rendszerek egyik legjelentősebb alkalmazási területe a jogi és az orvosi dokumentáció készítés lehet. Ezen szakterületek jól behatárolható szóincsel rendelkeztek, amelyek lehetővé tették a 90-97%-os beszéd felismerési arány elérését. Ezért a Philips erre a két szakterületre fókuszálva fejlesztette ki a beszéd felismerő szoftvert és az annak működését biztosító különböző nyelvű szakmai szakszótárakat. A működést biztosító szakszótárakat a RECO SYST Kft.⁵ készítette el, és informatikusának a közreműködésével Kelet-Közép Európában Magyarországon elsőként 2004-ben elkészült a szoftver háttérbázisát képező magyar nyelvű radiológiai és 2006-ban a magyar nyelvű jogi szakszótár. A magyar nyelvű radiológiai szakszótár 100 millió szóból álló, 500-600 ezer darab, a radiológia minden modalitását, teljes szakterületét felölelő leletállományból, továbbá magyar nyelvű férfi és női hang-anyagból készült 9 havi fejlesztői munka eredményeként.

Nézzük át röviden mit tud a SpeechMagic™ beszéd felismerő rendszer, amely egy szövegértő, beszélőfüggő rendszer.

Ez a rendszer nem szavakat, hanem szöveget értelmezett, a beszéd felismerés pedig akkor érte el a maximális 90-97%-os arányt, ha a felhasználó személye huzamosabb ideig állandó volt, a rendszer ugyanis azonosította a beszélőt annak hangszíne, kiejtése, tájshozlása alapján. Ez a függés a felhasználó személyétől természetesen hátrányos is lehet, amely azonban kiküszöbölhető azzal, hogy a szoftver rendszerek biztosítják a beszéd felismerő szoftver folyamatosan „öntanulását”. Az utólagos korrekciókat a szoftver tárolta és a számított ismétlések után önmagát megbízhatóan korrigálta. A rendszer alkalmazásakor

⁵ <http://www.recosyst.hu/>

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 4. szám

eredményként a diktálással közel egyidejűleg az írott szöveg is előállításra kerül, ami esetünkben gyakorlatilag kötelező.

A SpeechMagic™ egy alaptechnológiai rendszer. Szoftverfejlesztői környezet segítségével illeszthető volt az egészségügyi szolgáltatóknál már működő szakmai informatikai rendszerhez (HIS-hez), radiológiai informatikai rendszerhez (RIS-hez); digitális képparchiváló rendszerekhez (PACS-hoz), vagy akár egyéni PC-hez. Az illesztést a RECOsYST Kft., illetve az egészségügyi informatikai rendszert szállító és követő informatikai cég közösen végezte. Az első hazai SpeechMagic™ rendszerillesztést az ISH Kft. végezte el, aki a MedSolution kórházi informatikai, illetve RIS rendszerét illesztette az AGFA PACS rendszeréhez, ezáltal megvalósítva az első magyar nyelven is működő beszédfelismerő és HIS-RIS-PACS komplett informatikai rendszert. Egy év elteltével elkészült a GlobeNet Rt. közreműködésével a Medworks kórházi informatikai rendszerbe történő illesztés, valamint ezzel egyidejűleg a Meditcom Kft. kórházi informatikai rendszerének radiológiai moduljához történő illesztés és a Medimon Kft. radiológiai és képparchiváló rendszeréhez történő illesztés.

A SpeechMagic™ működtethető szóló személyi számítógépen, vagy számítógépek lokális hálózatán, LAN, WAN, vagy internet közbeiktatásával egyaránt. Szóló számítógép esetében ez egy akkor használatos személyi számítógép, (Intel pentium 4), 512 MB RAM és 1 GB winchester kapacitással. Hálózatos alkalmazás a gyakoribb, de a beszédfelismerő szerver és az alkalmazás szerver is Intel Pentium 4 számítógép, minimum 2GHz processzorral, 512MB RAM, 1 GB winchester kapacitással. Az operációs rendszer Windows NT legújabb verziói voltak. Kiegészítő tartozék volt a számítógépekben lévő kiváló minőségű hangkártya. A beszédfelismerő rendszerhez tartozott a kézben fogható speciális diktafon (SpeechMike), mely a diktálás során a hanganyagot továbbította a számítógép szoftvere felé. További tartozék volt a visszahallgatás alkalmával az elmondott szöveg meghallgatást megszakító lábpedál. A SpeechMagic™ használata egyszerű, gyorsan megtanulható volt. A SpeechMagic™ felhasználó szám függő licence formájában került forgalmazásra, melyet a folyamatos karbantartás, rendszerkövetés havi díja egészített ki. Az orvosi dokumentáció minden esetben számítógépes alkalmazások támogatásával készült, ahol a beszédfelismerő rendszer segített a hatékonyabb munkavégzésben és a munkafolyamatok optimalizálásban.

Ezek után tekintsük át általánosságban, hogy milyen előnyei lehetnek a SpeechMagic™ használatának, hiszen ezek az előnyök realizálhatóak a katonai feladatok végrehajtása során is:

- megújíthatja, korszerűsítheti az orvosi dokumentációs tevékenységet;
- nem szükséges az orvosi dokumentáció előállításához egyidejűleg két munkaerő lefoglalása: az orvos és az írnok;
- nem kell a diagnosztikai leletezéshez a (szak)asszisztens írnoki tevékenysége;
- elmarad a hanganyag leírása (a leírás szükségessége esetén megoldható);
- nem szükséges analóg hangkazetták alkalmazása, nincs kazetta csere (a tárolás a számítógépen történik);
- nincs rossz hangfelvétel miatti leletezés ismétlés (a rossz minőségű hangfelvétel más alkalmazások segítségével javítható, ismételhető);

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 4. szám

- javítható a szöveges orvosi dokumentum minősége;
- gyorsabban készülnek az orvosi dokumentációk;
- idő takarítható meg, mert azonos időben több képalkotó diagnosztikai/terápiás vizsgálat készíthető, ezáltal csökkenthető a betegek előjegyzési és várakozási ideje a vizsgálati leletre (a jelentés azonnal elkészülhet írásos formában is);
- fokozható a betegelégedettség (gyors reagálási lehetőség a betegség felismerése utáni beavatkozáshoz);
- munkaerő szabadítható fel a betegekkel való foglalkozásra (a labornak a kezelőszemélyzete minimális);
- a rendszer a PC szervereken kívül általában nem igényel újabb számítógépes munkaállomások beszerzését, csak a meglévő leletező munkaállomások kiegészítő kellekkel való ellátását.

Az általános előnyök után nézzük meg azt is, hogy milyen kiemelt előnyöket jelenthet a rendszer különböző szintű vezetők (parancsnokok) és a konkrét feladat-végrehajtók számára:

- jelentős a hatékonyság növekedés;
- átszervezi, racionalizálja a munkafolyamatokat (szigorúan leszabályozott feladatok, hatáskörök);
- munkaidőt, helyet szabadít fel;
- munkaerőhiányt pótol;
- munkabért takarít meg;
- térben korlátlanul alkalmazható és jótékonyan hat a munkahelyi légkörre;
- nem ért félre, „jobb, mint az emberi fül”;
- rövidítés nélkül ír, értelmesen rövidít (a szakszótár alapján);
- az öntanulás képessége, a jelentést végrehajtó személye előzetesen meghatározható;
- önfejlesztő, önellenőrző;
- a tudományos munka írását is jól támogatja, latinul is tud;
- nem vét helyesírási hibát.

A KATONAI ALKALMAZÁS LEHETŐSÉGEI

Először foglaljuk össze SpeechMagic™ beszéd felismerő rendszer alkalmazási lehetőségeit, amelyből egyértelműen látható a katonai alkalmazás lehetősége is. A SpeechMagic™ egy szövegértő, beszélőfüggő rendszer. Ez azt jelenti, hogy a rendszer nem szavakat, hanem szöveget értelmez, a beszéd felismerés pedig akkor éri el a maximális 90-97%-os arányt, ha a felhasználó személye huzamosabb ideig állandó, a rendszer ugyanis azonosítja a beszélőt annak hangszíne, kiejtése, tájékozása alapján. Szoftver rendszerek biztosítják, hogy a beszéd felismerő rendszer folyamatosan „öntanul”. Az utólagos korrekciókat tárolja, s számított ismétlések után önmagát megbízhatóan korrigálja. A rendszer alkalmazásának eredményeként a diktálással közel egyidejű az írott szöveg előállítás.

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 4. szám

Segítségével lehetséges volt a munkaállomásról (saját PC) történő diktálás, s külön munkafázisban a beszédfelismerés, majd egy másik (adminisztrátori) munkaállomásról, más személy által történő szöveg korrektúra, javítás elvégzése, majd a dokumentum nyomtatása. Lehetséges volt olyan alkalmazás, ahol azonos munkaállomásról történik a diktálás, s ezt követően azonos személy által a hanganyag lehallgatásával, a szöveg visszaolvasásával korrektúra elvégzése, majd a nyomtatás. További lehetőség volt a diktálással egyidejűleg történő szövegolvasás és javítás is. Lehetséges volt az interaktív diktálás, amikor egy már korábban létrehozott dokumentumban, iratmintában parancsszóval a kipontozott szövegrészhez lép a program, s oda diktálható a kívánt szöveg. A diktálás történhetett a számítógéphez kapcsolt vezetékű diktafonon, vagy mobil diktafon használatával is és létezik „stand-alone” alkalmazási módszer vagyis a diktálás történhetett üres, új dokumentumba, vagy már meglévő iratmintába is.

Végezetül nézzünk egy mai példát az amerikai szárazföldi erőktől. A bemutatásra kerülő MFLTS (Machine Foreign Language Translation System) idegen nyelvi fordítási rendszer Michael Doney, az MFLTS termékigazgatója, szerint a katonák nyelvi fordítási képességének biztosításával áttöri a nyelvi korlátokat. Az MFLTS egy olyan szoftveralkalmazás, amely megfelel a hadsereg kiemelt nyelvi fordítási követelményeinek is. Az MFLTS segíti a katonákat abban, hogy hatékonyan kommunikáljanak a külföldi, nem angolul beszélő populációkkal automatizált külföldi beszéd- és szövegnyelv-fordítási képességgel. Bár nem helyettesíti az emberi nyelvi támogatást, mint a tolmácsok és fordítók, az MFLTS sikeresen kiegészíti és kiegészíti ezt a támogatást. Ennek eredményeképpen a katonák teljes mértékben kihasználhatják a fordítói és tolmácsolási erőforrásokat a küldetésük végrehajtása érdekében. Ezenkívül az MFLTS-t „átvizsgálásként” lehet használni a kínált fordítás pontosságára vonatkozóan. Az MFLTS rendszer elsődleges összetevője a gépi fordítás, az automatikus beszédfelismerés és az optikai karakterfelismerés.

A szoftveralkalmazást úgy tervezték, hogy mind a kereskedelmi forgalomban kapható hardveres, mind a kormányzati automata rendszereken működjön. A különböző warfighter igények kielégítésére három MFLTS szoftver konfiguráció áll rendelkezésre:

- Web-Enabled (elosztott közös földi rendszer, szöveg-szöveg);
- Mobil (szöveg-szöveg, beszéd-beszéd);
- Hordozható (Nett Warrior, beszéd-beszéd).

Milyen további erőfeszítéseket terveznek a jövőre nézve? Az MFLTS nyílt rendszer architektúrája lehetővé teszi a további nyelvi összetevők folyamatos integrációját, hogy megfeleljen a hadsereg egyre növekvő nyelvi fordítási követelményeinek. Mint ilyen, közvetlenül hozzájárul a hadsereg felkészültségének elsőbbségi állományának vezetőjéhez, lehetővé téve a warfighterek számára, hogy kommunikáljanak, és viszont hatékonyan működjenek a nem angol nyelvű lakosság körében világszerte. Hagyományosan egy hadsereg-nyelvszaki, a katonai műveleti nyelv specialitásai miatt hat-tizenöt hónapig kerülnek külön képzésre ahhoz, hogy megszerezzék az adott nyelvhez szükséges jártasságot. Azonban a szakképzett nyelvészek, az anyanyelvűek, akik a MOS 09L-hez tartozó hadsereg tolmácsaként szolgálnak,

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 4. szám

és még a nyelvi fordítói vállalkozók is mindig hiányoznak és rendszerint túlterheltek, miután telepítették őket.

Azzal, hogy MFLTS-t rendszeresítik adott lesz egy szoftver alkalmazás, amely nyelvi fordítási képességet biztosít az egyes katonáknak. Mivel a mai hadsereg egyre inkább globálisan működik, fontosabb, mint valaha, hogy a katonák képesek legyenek hatékonyan kommunikálni a nem angol nyelvű népeiségekkel. Például, ha egy katona alapképzést folytat, vagy az őshonos lakosság körében dolgozik, hogy taktikai kihallgatással összegyűjtse az információkat, akkor képesnek kell lennie arra, hogy kommunikáljon ezekkel a populációkkal a misszió hatékony végrehajtása érdekében.

Jelenleg az előzőekben felsorolt három lehetőség közül két MFLTS alkalmazás van használatban. Az egyik kétirányú, valós idejű speech-to-speech fordítást biztosít, míg a másik szöveges szöveges fordítást biztosít elektronikus dokumentumok, weboldalak és közösségi média számára. Mindkét verzió a legmodernebb gépi nyelvfordítási technológiát alkalmazza, és lehetővé teszi a felhasználók számára, hogy több nyelvi "csomagból" választhassanak, attól függően, hogy milyen nyelven beszélnek a művelet területén vagy más konkrét felhasználói követelményeknek megfelelően. A misszió szemszögéből az MFLTS programhivatal egy webalapú portált is kifejleszt, amely lehetővé teszi a felhasználók számára az MFLTS alkalmazás letöltését és telepítését, az alkalmazás frissítését vagy a nyelvi csomagok szükség szerinti letöltését. A jövőben a felhasználó képes lesz saját MFLTS alkalmazásának testreszabására és több mint 65 különálló nyelvi csomaghoz fér hozzá.

Bár az MFLTS a nyílt rendszerek architektúráját és a magánszektor által inspirált, fejlett gépi tanulási technológiákat alkalmazza, azt kifejezetten arra tervezték, hogy megfeleljen a rendkívül különleges katonai követelményeknek. A katonák tipikusan olyan környezetben működnek, ahol a hálózatokhoz és a távoli szerverekhez való csatlakozás nem garantált, ezért az MFLTS-t helyileg kell elhelyezni a felhasználók okostelefonjain és számítógépein. A hadsereg olyan katonai sajátosságú nyelvi tartalmat is alkalmaz, amelyet a kereskedelmi nyelvű fordítási termékek nem értenek meg, és azokat a teljes spektrumú katonai műveletek támogatására fejleszteni és integrálni kell az MFLTS-be.

Az MFLTS tehát támogatja a hadsereg első számú prioritását (készenlét) egy olyan automatizált nyelvi fordítási képesség biztosításával, amelyet a bevetett katonák használhatnak, akiknek szükségük van a helyi idegen nyelvű emberekkel való kommunikációra, amikor egy tolmács sem áll rendelkezésre. A bevetett katonák az MFLTS-t is használhatják az idegen nyelvű dokumentumok és a közösségi média, például weboldalak és blogoldalak fordítása révén a szituációs tudatosságuk és megértésük javítása érdekében. Az MFLTS program továbbra is támogatni fogja a hadsereget azért, hogy új nyelvi csomagokat fejlessz és elérhetővé teszi őket az MFLTS Language Portal segítségével. Ez az új, alkalmazható és alkalmazkodó nyelvi fordítási képesség közvetlenül lehetővé teszi a hadsereg számára, hogy „győzzön egy komplex világban”, hatékonyan áthágva azokat a nyelvi akadályokat, amelyeket a hadsereg ma és holnap találni fog.



2. sz. ábra

(Forrás: <https://peoiews.army.mil/2017/army-language-translation-system-assists-soldiers-readiness>)

ÖSSZEFOGLALÁS, KÖVETKEZTETÉSEK

Ez a publikáció a mesterséges intelligenciák alkalmazási területein belül, a különböző típusú munkák hatékonyságát növelő lehetőségek közül, a beszédfelismerő rendszerek alkalmazásának lehetőségeit vizsgálta. A beszédfelismerő rendszer alkalmazásának alapelveit egy előző cikkemben már tisztáztam, ezért most a hangsúlyt áthelyeztem a gyakorlati alkalmazás alapjainak vizsgálatára. Miután így teljessé vált az alapelvek bemutatása, bemutatásra került egy működő beszédfelismerő rendszer, amely jelentős múltra tekint vissza és amelynek használata katonai környezetben is indokolt lehet. A következtetés egyértelműen levonható, mely szerint a mesterséges intelligencia alkalmazása a beszédfelismerésben a katonai feladatok végrehajtásánál is jelentős távlatokat jelenthet. Ennek a ténynek az alátámasztására bemutatásra került egy – az amerikai hadseregben – működő rendszer, bizonyítva ezzel is a cikk aktualitását és kapcsolódását a folyamatban levő katonai kutatásokhoz is.

FELHASZNÁLT IRODALOM

1. Vicsi Klára: A beszédfelismerés fejlődése, a mai beszédfelismerési módszerek ismertetése, alpha.tmit.bme.hu/speech/docs/education/beszedkomm_felism3.PPT, 2006, Letöltve: 2017.09.01.
2. Vicsi Klára, Gordos Géza, Naszodi Mátyás, Tatai Péter: Magyar nyelvű kötött, közép szótáros, folyamatos beszédfelismerő rendszer megvalósítási megoldásainak kutatása (Research on the

HADTUDOMÁNYI SZEMLE

2017. X. évfolyam 4. szám

- construction of continuous speech recognizer for a Hungarian middle sized vocabulary) <http://real.mtak.hu/1452/46487-ZJ1.pdf>, Munkabeszámoló, 2007, Letöltve: 2017.09.01.
3. Wilcoxon, F.: "Individual Comparisons by Ranking Methods." *Biometrics* 1, 80-83, 1945.
 4. L. R. Bahl, P. F. Brown, P. V. de Souza, R. L. Mercer: Maximum mutual information estimation of hidden Markov model parameters for speech recognition. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vol. 1, pp. 49–52, Tokyo, Japan, April 1986.
 5. L. E. Baum, J. A. Eagon: An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model of ecology. *Amer. Math. Soc. Bull.*, Vol. 73, pp. 360–362, 1967.
 6. M. H. Cohen: Phonological structures for speech recognition. Ph.D. dissertation, University of California, Berkeley, USA, 1989.
 7. Creutz, M. and Lagus, K.: "Unsupervised Morpheme Segmentation and Morphology Induction from Text Corpora Using Morfessor 1.0.", *Publications in Computer and Information Science, Report A81*, Helsinki University of Technology, March, (2005)
 8. Velkei Szabolcs, Vicsi Klára: Beszédfelismerő modellépítési kísérletek akusztikai, fonetikai szinten, kórházi leletező beszédfelismerő kifejlesztése céljából, MSZNY 2004.
 9. Vicsi, K. - Velkei Sz. - Szaszák Gy. - Borostyán G. –Gordos Géza: Folyamatos középszótáras, beszédfelismerő rendszer fejlesztési tapasztalatai, kórházi leletező beszédfelismerő. *Hiradástechnika LXI. évf. (p.14-21)*, 2006.
 10. Mihajlik Dénes: Spontán magyar nyelvű beszéd gépi felismerése nyelvspecifikus szabályok nélkül. Doktori értekezés, BME-VIK Villamosmérnöki Doktori Iskola, 2010
 11. Czap László: Audiovizuális beszédfelismerés és szintézis, PhD értekezés, BME, Budapest, 2005.
 12. Gordos G., Takács Gy.: Digitális beszédfeldolgozás, Műszaki Könyvkiadó, Budapest, 1983.
 13. Gósy Mária: A zöngésségi hasonulás a (spontán) beszédben. *Beszédkutató 1998*, Akadémiai kiadó, Budapest, pp. 1-20, 1998
 14. Németh B., Mihajlik P., Tikk D., Trón V.: Statisztikai és szabály alapú morfológiai elemzők kombinációja beszédfelismerő alkalmazáshoz. MSZNY 2007: V. Magyar Számítógépes Nyelvészeti Konferencia, pp. 95-105, Szeged, 2007.
 15. Siptár Péter: *Fonológia (Egyetemi jegyzet)*. Budapest: MTA Nyelvtudományi Intézet, 1995.
 16. Vicsi, K. - Vig, A.: Az első magyar nyelvű beszédadatbázis, *Beszédkutató '98*, MTA Nyelvtudományi Intézete, Budapest 1998, pp. 163-177
 17. Vicsi K. Velkei Sz., Szaszák Gy., Borostyán G., Teleki Cs., Tóth Sz. L., Gordos G.: Középszótáras, folyamatos beszédfelismerő-rendszer fejlesztési tapasztalatai, *Proc. of MSZNY 2005*, pp. 348 – 360.
 18. Vicsi K., Tóth L. Kocsor A., Gordos G. Csirik J.: MTBA – Magyar nyelvű telefonbeszéd adatbázis. *Hiradástechnika 2002/8. sz. pp. 35-39.*
 19. <https://www.army.mil>