

Óbudai Egyetem

Doktori (PhD) értekezés
tézisfüzete



**Eljárások, modellek és módszerek beszéd- és
mozgáskorlátozott személyek támogatására intelligens térben**

Simon-Nagy Gabriella

Témavezetők:

Dr. Várkonyiné Prof. Dr. Kóczy Annamária

Dr. Kutor László

**Alkalmazott Informatikai és Alkalmazott Matematikai
Doktori Iskola**

Budapest, 2021. április 12.

I. A kutatás előzményei

Kutatási témám a súlyosan mozgás- és beszéd fogyatékos személyek életvitelének támogatására szolgáló eljárások, modellek és módszerek fejlesztése. Eredményeim két fő területre bonthatók. Az egyik a publikus térben (elsősorban középületekben) való akadálymentes navigáció, a másik az okosotthon rendszereknek a speciális követelményekhez való adaptálásának aspektusai. Utóbbi területen belül elsősorban azzal foglalkoztam, hogy a progresszív neuromuszkuláris betegségeknek a beszédre gyakorolt hatása milyen módosításokat tesz szükségessé az okosotthon rendszer beszédvezérlésében.

A súlyosan mozgássérült személyek asszisztenciához való hozzáférése, a körülményekhez mérten önálló életvitel lehetőségei Magyarországon nem nevezhetők ideálisnak. Mind az utcák, mind a középületek akadálymentessége számos kívánnivalót hagy maga után. Bár a középületek akadálymentesítéséről törvény rendelkezik, ez azonban csak az újonnan épített vagy felújítandó épületekre vonatkozik. Számos régi épület csak részben akadálymentes, vagy a bejáratnál található lépcső miatt egyáltalán nem tekinthető kerekesszéssel járhatónak. Ugyanakkor az akadálymentesség törvényi kritériumai egyes mozgássérült személyek számára szükségtelenül szigorúak, tehát az épületen belül egy meghatározott útvonal bejárhatósága nem egyszerű kérdés.

Az otthon falai között sem egyszerű a helyzet, hiszen aki nagyrészt képtelen a (koordinált) mozgásra, annak a legegyszerűbb tevékenységekhez is segítségre lenne szüksége, de az otthonápolás áldozatkész hozzátartozók és megfelelő anyagi lehetőségek mellett is nagyon nehezen biztosítható a nap 24 órájában.

Mindezt tetézi, hogy több olyan betegség is van (például a szklerózis multiplex vagy az amiotrófiás laterálszklerózis), amely a mozgási képességek leépülése mellett a beszédminőséget is rontja (mind a beszédben résztvevő izmokra és idegekre, mind a légzésre való negatív hatás által). Ez nem csak a segítőkkal való kommunikációt nehezíti meg, de a beszéd felismeréssel vezérelhető segítő rendszerek használatát is ellehetetleníti hosszú távon. Az önálló(bb) életvitel informatikai támogatását szolgáló okosotthon rendszerek kisebb-nagyobb szolgáltatáskínálattal már régóta léteznek, ám ezek nincsenek felkészítve sem a torzult, nehezen kivehető beszéd felismerésére, sem pedig a megfelelően biztonságos, megbízható működésre. Egy utasítás félreértése esetén

ugyanis egy súlyosan mozgássérült felhasználó nem képes egy gombnyomással korrigálni a hibát, és ez akár veszélyes helyzetekhez is vezethet.

Beltéri navigációt támogató ontológiák a szakirodalomban

Az OntoNav [Anagnostopoulos et al, 2005] szemantikus beltéri navigációs rendszer és egy ontológiai keretrendszer útkeresési feladatok kezelésére. Képes épületen belüli navigációra folyosókon keresztül, de nem képes helyiségen belüli navigációra, amely szükséges lenne például egy több bejáratú rendelkező, nagy méretű terem vagy aula esetén.

Az ONALIN [Dudas et al, 2009] az útkeresés során képes figyelembe venni a különböző szükségletekkel és preferenciákkal rendelkező egyének igényeit; egyéb követelmények mellett tartalmazza az ADA (American Disability Act) által előírtakat is. Az épületeket folyosók hálózataként modellezi, és képes megadni a felhasználó által megadott specifikus megszorításoknak megfelelő útvonalakat. Ezt úgy éri el, hogy minden lehetséges fogyatékosághoz előre számított hálózatot használ, és mindig a megfelelő hálózaton futtatja a keresést, ami meglehetősen rugalmatlan.

Scholz és Schabus [Scholz et al, 2014] olyan beltéri navigációs ontológiát készített, ami támogatja a termelőeszközök autonóm navigációját termelési környezetben. Az eszközök egy workflow által meghatározott lépéseknek megfelelő sorrendben haladva, minden lépéshez megkeresik a megfelelő berendezést, és kiszámítják az adott berendezéshez a legmegfelelőbb útvonalat.

A Geodint [Matuszka et al, 2013] az általános legrövidebb út (shortest path) algoritmussal navigál egy gráf modellen. Jelenleg a fenti ontológiák egyike sem érhető el, azonban az ismertett szemantikus modell egyes részei adtak ötletet az osztályhierarchia felépítéséhez.

Benner és Karimi [Benner et al, 2014] megvizsgálta az elérhető, gyalogos navigációra szolgáló ontológiákat (beleértve az INO és ONALIN ontológiákat is), és nem találtak olyan megközelítést, amely minden fogyatékoságra általánosan illene, a létező megoldások csak a spektrum egy részét fedik le.

Az ontológia segítségével megvalósított navigáció lehetőségeit Magyarországon is kutatják az MTA Számítástechnikai és Automatizálási Kutatóintézetében, a Debreceni

Egyetemen, valamint az Óbudai Egyetem Neumann János Informatikai Karán dolgozó kollégák. Ennek eredménye az iLOC ontológia [Szász et al, 2016], amelynek fejlesztésébe én is bekapcsolódtam. Ez egy általános épület belső szerkezetének leírását, az épületen belüli közlekedés lehetőségeit tartalmazta, akadálymentességi információkat viszont kezdetben még nem. Azonban az iLOC ontológiát úgy tervezték, hogy egyszerűen bővíthető legyen további ontológiákkal, amelyek akár környezet-specifikus adatok (például kórházakra vonatkozóan [15]), akár egyéb fontos tulajdonságok leírására szolgálnak. Ezt kihasználva alakítottam ki az iLOC akadálymentességi kiegészítéseit.

Karaktorsorozatok között értelmezett távolság metrikák

Számos metrikát ismerünk karaktorsorozatok távolságának meghatározására, ezeket tekintetem át, amikor a dizartria-specifikus távolság definiálásán kezdtem dolgozni [Bilenko et al, 2003] [Cohen et al, 2003] [Ukkonen, 1992]. Közülük a legegyszerűbb a Hamming távolság, amely a két karaktorsorozat megfelelő pozícióiban lévő karakterek közti eltérések száma. Definíciójának megfelelően csak egyenlő hosszúságú karaktorsorozatokra értelmezett, ezért ritkán használják általános szövegek közötti távolság megadására.

Két szöveg Levenshtein (szerkesztési) távolságát azon egykarakteres módosítások (csere, beszúrás, törlés) minimális száma adja, amely ahhoz szükséges, hogy az egyik szöveget a másikba átalakítsuk. Ezt a metrikát általában rövid szövegek illeszkedésének vizsgálatára használják olyan esetekben, amikor kevés eltérés várható (pl. keresés szótárban, helyesíráseellenőrzés), de hosszabb szövegek közötti távolság számítása ezzel a módszerrel meglehetősen költséges művelet.

A Damerau-Levenshtein távolság kiegészíti az eredeti Levenshtein távolságot a szomszédos karakterek cseréjének műveletével. A leghosszabb közös részsorozat (longest common subsequence - LCS) olyan távolságot ad meg, ahol csak a beszúrás/törlés művelet engedélyezett. A "szerkesztési távolság" megnevezést egyébként gyakran használják olyan általános string metrikák megnevezésére, ahol a megengedett módosító műveletek a hozzájuk kapcsolódó költség értékével a metrika paramétereként adóttak.

Léteznek egyéb megoldások is a szövegek közti távolság számítására, amelyek nem feltétlenül metrikák matematikai értelemben. Ilyen például a Jaro-Winkler távolság, amelyet duplikáció-detektálásra fejlesztettek ki. Ez az eljárás az egyező karakterek száma és a szükséges cserék száma alapján adja meg a hasonlóság mértékét. Emellett (paraméterezéstől függően) hasonlóbbnak tekintheti azokat a szövegeket, amelyeknek az eleje egyezik.

A fenti eljárások közös tulajdonsága, hogy nem képesek a karakterek "fontosságát" figyelembe venni, azaz bizonyos megadott karakterek egyezése nem produkál nagyobb hasonlósági értéket, mint más karaktereké. Ezért változtatás nélkül egyik sem alkalmas dizartria-specifikus távolság kifejezésére, a Jaro-Winkler távolság pozíció-alapú differenciálása azonban jó irányba mutat a probléma szempontjából. Ezt az ötletet felhasználva fejlesztettem a saját elgondolásomat.

Anytime algoritmusok a beszéd felismerésben

Kifejezetten az okosotthon rendszerek megbízható és biztonságos működésének kritériumairól kevés forrás található, ezért az "Információra és a kapcsolatos technológiára vonatkozó kontroll célkitűzések" (Control Objectives for Information and related Technology – COBIT®¹) keretrendszerben megfogalmazott kritériumokat vettem alapul, közülük is elsősorban a hatékonyság, rendelkezésre állás és megbízhatóság elveit. [Szenes, 2011] szerint ezen kritériumok általánosíthatók vállalati működési kritériumokká, ezekre alapoztam a saját kritériumaimat. Mivel az okosotthon megbízható működése nagy mértékben a megbízható beszéd felismerésen múlik, ezért a COBIT-ra alapozott elvek jelentősen meghatározták a beszéd felismerő modulom tervezésének és megvalósításának irányát és szempontjait.

Az anytime algoritmusok, modellek és rendszerek olyan speciális valós idejű rendszerek, amelyek képesek változó mennyiségű, vagy kevés erőforrással működni (ide értendő a futási idő vagy a rendelkezésre álló adatok is), a kimenet minőségének csökkenése árán. Ez azt jelenti, hogy tetszőleges idejű futtatást követően képesek kimenetet szolgáltatni, amelyhez hozzá tartozik a minőség illetve hiba mérőszáma is.

¹ COBIT 4.1 Framework, Management Guidelines, Maturity Models, Copyright © IT Governance Institute, 2007., <https://www.isaca.org/Knowledge-Center/cobit/Documents/COBIT4.pdf>

Anytime algoritmusokat ritkán használnak a beszédfelismerés területén annak ellenére, hogy általános jelfeldolgozási feladatokban használatos, és alkalmas lenne a real-time rendszerekben fennálló időkorlátnak való megfelelésre. Található néhány kivételes példa az anytime megközelítés alkalmazására beszédfelismerés jellegű projektben, például [George et al, 2004] és [Menzel, 1994] esetében természetes nyelvi mondatok szerkezetének valós idejű elemzésénél használják, a keresési tér méretének csökkentésére. Ez a megközelítés (a lehetséges találatok mennyiségének szisztematikus redukálása) adta az ötletet a saját megoldásomhoz, mivel a stratégia alkalmazható a fix utasításkészlettel rendelkező okosotthon rendszereknél az elhangzott utasítások osztályozásában is.

Több modulból álló és adaptív rendszerek a beszédfelismerésben

A változó, idővel romló minőségű beszéd problémájára hosszú távon a legkényelmesebb megoldást a működés közbeni tanítás szolgáltatná.

A beszédfelismerésben széles körben használatosak különféle adaptív eljárások különféle célokra, ilyenek például a zaj kompenzáció [Kalinli et al, 2010] [Juang et al, 2001], egy adott akcentushoz való igazítás [Wang et al, 2014], vagy hasznos tanítóminták kiválasztása címkézésre egy nagyobb címkézetlen korpuszból [Riccardi et al, 2005]. Maga az adaptálódó rendszer jellemzően valamilyen jól ismert lágy számítási módszer, mint a neurális háló [Xue et al, 2014] vagy fuzzy filter [Juang et al, 2001].

A folyamatos, használat közbeni adaptáció azonban veszélyeztetheti a felismerés pontosságát, különösen mivel az átlagos élettérben megtalálható háttérzajok emberi beszédet is tartalmazhatnak, akár más személyek beszéde, akár a televízió vagy rádió műsorainak formájában.

Egy több modulból álló rendszert megfelelően felépítve és működtetve kiküszöbölheti ezt a veszélyt, ha garantáljuk azt, hogy a használat közbeni tanítás ne tehesse tönkre a meglévő, jól működő modult, valamint ha ezt az elsődleges modult csak szigorú feltételek teljesülése esetén cseréljük le.

Számos szabadalom és néhány tudományos cikk is található olyan beszédfelismerő rendszerekről, amelyek több felismerő modult tartalmaznak, azonban ezeknek a célja a felismerés pontosságának a növelése. Az egyes felismerő modulok azonos típusúak, csak működési paramétereikben különböznek. Ugyanazon beszédmintát megkapva

bemenetként, az egyedi kimeneteik egy kombinált kimenetet adnak valamilyen döntési folyamat eredményeképp [Baker, 2000] [Endo et al, 2007].

Olyan példákat is találhatunk a szakirodalomban, ahol az osztályozás felgyorsítása a cél, és ehhez különböző típusú neurális hálókat alkalmaznak (például egyet az előfeldolgozásra, egy másikat a pontos felismerésre) [Murveit et al, 2006]. A kombinált felismerő működhet fonéma, szótag vagy akár teljes szavak szintjén is [Schwenk et al, 2000] [Natori et al, 2010] [Natori et al, 2013].

II. Célkitűzések

A célkitűzésem olyan eljárások, modellek és módszerek tervezése volt, amelyek mind a nyilvános térben, mind otthon segítenek áthidalni a fenti problémákat a meglévő megoldások és eszközök fejlesztésével, kiegészítésével.

A részben akadálymentes épületek problémájának áthidalására olyan beltéri akadálymentes útvonaltervező megoldást kerestem, amely speciális hardver telepítése nélkül lehetővé teszi az egyén igényeire szabott, általa (a lehető legkevesebb segítséggel) bejárható útvonal megtalálását.

A létező mérnöki megoldások sorában az "akadálymentes beltéri navigáció" a legtöbb esetben a látássérült felhasználók audio alapú navigálására vonatkozik. Ha képesek is a navigációs algoritmusok akadálymentes utakat megkülönböztetni, az egyenértékű a "kerekeszkék számára akadálymentes" kategória törvényi definíciójával (például [Dudas et al, 2009]). Mivel a való életben ez rugalmatlan és rosszul kihasználható, ezért a saját munkámban arra koncentráltam, hogy az útvonaltervezés maximálisan testreszabható legyen, és lehessen értelmezni rajta a "segítséggel bejárható útvonal" kategóriát. Ehhez szükség van az épület részeinek sok különféle adatára, valamint arra, hogy az adatokat gyorsan és praktikus kereshető formában és rendszerben tároljuk, ezeket valósítottam meg a doktori képzésem keretében.

A súlyosan mozgás- és beszéd fogyatékos felhasználók okosotthon rendszereinek megfelelő felépítésére pedig olyan irányelveket terveztem, amelyek a speciális igények mellett is megbízható használatot tesznek lehetővé, valamint a beszéd felismerés ismert módszereinek olyan jellegű továbbfejlesztését adtam meg, amely kifejezetten a torzult, és a betegség következtében lassan változó beszéd megbízható felismerését teszi lehetővé.

A jelenleg létező okosotthon rendszereknek több olyan hiányossága van, amelyek miatt nem alkalmasak súlyosan mozgás- és beszéd fogyatékos személyek számára. A súlyosan mozgássérült személyek ugyanis nem biztos, hogy képesek néhány lépéssel és mozdulattal korrigálni az esetleges hibás parancsvégrehajtásból származó kényelmetlen vagy esetleg kifejezetten veszélyes helyzeteket. Egy ideális rendszer úgy kellene működjön, hogy az utasításokat csak akkor hajtja végre, ha biztos abban, hogy melyik utasítás hangzott el. (A gyakorlatban ez persze egy kellően magas elfogadási küszöböt

jelent, nem teljes bizonyosságot.) Ez alól azonban kivétel a vészjelzés és a segítségkérés funkció, amelyek esetében viszont sokkal nagyobb gond, ha nem hajtja végre a rendszer, pedig kellene. Az utasításoknak ez a fajta megkülönböztetése a jelenlegi rendszerekből hiányzik.

A pontos felismerést azonban megnehezíti a beszéd torzulása, azaz a dizartria. A publikusan elérhető beszélőfüggetlen beszédfelismerő rendszerek ezzel nem tudnak megküzdeni (néha az ép beszéddel sem), a személyre szabott megoldások is idővel veszítenek hatékonyságukból a dizartria súlyosbodása miatt. Emellett ezek a megoldások nem használják ki azt, hogy az érintettek beszédének bizonyos elemei tovább maradnak azonosíthatók, mint más elemek. A működés közbeni adaptáció nem ismeretlen a szakirodalomban, azonban esetünkben gondoskodni kell arról, hogy a felismerés pontossága ne csökkenjen a tipikusan zajos, más emberek beszédét is tartalmazó tanítóminták miatt. Különböző szűrési módszerek léteznek ilyen célra, de a lehető legbiztonságosabb megoldásra kell törekedni. Bár minden személy betegsége egyedi lefolyású, a szakorvosok képesek megkülönböztetni kategóriákat már a kezdeti tünetek alapján, amelyek előrevetítik a beszéd romlásának jellegét is. Ez az előzetes tudás is felhasználható a beszédfelismerő tervezésénél, főképp a parancsszavak megválasztásánál.

III. Vizsgálati módszerek

Akadálymentes beltéri navigáció

Az ontológia fejlesztése során a diagram ábrázolás jelöléseit a Protégé² ontológia szerkesztő szoftver által alkalmazott jelölések alapján választottam meg, tehát ellipszisek jelölik az osztályokat, és nyilak a közöttük lévő kapcsolatokat. A kapcsolatok lehetnek a saját ontológia által meghatározott kapcsolatok, vagy publikus kapcsolat típusok, például az `rdfs:subClassOf`, amely az osztályhierarchia ábrázolására hivatott. Az ontológia leírása során az RDF³ (Resource Description Framework), RDFS⁴ (RDF Schema) és OWL⁵ (W3C Web Ontology Language) által meghatározott szintaktikát követtem.

Az iLOC számos publikus ontológiához kapcsolódik, és az akadálymentességi ontológia kiegészítem is alkalmaz ilyen. Ezek közül a leglényegesebbek:

- a FOAF⁶ (friend of a friend), amelyből az Agent és az Organization osztályokra hivatkozik az iLOC,
- a World Wide Web Consortium által létrehozott Geo szótár⁷ (SpatialThing osztály) és vCard⁸ ontológia (Location osztály)
- a QUDT⁹ (Quantities, Units, Dimensions and Types) séma, amelyből a `qudt:value-t` és a `qudt:unit-ot` használtam fel az objektumok tulajdonságainak leírására.

Az ontológiákra való hivatkozás úgy történik, hogy az ontológia névtér azonosítójához (IRI – Internationalized Resource Identifier) egy prefixet rendelünk, és a továbbiakban ezzel a prefixszel hivatkozunk rá. Például egy ontológiát leíró dokumentum elején szereplő

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
```

kódsor azt határozza meg, hogy a dokumentum további részében az `rdfs:` rövidítés az RDF Schema 1.1 verzióját jelenti.

² <https://protege.stanford.edu/>

³ <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

⁴ <http://www.w3.org/2000/01/rdf-schema#>

⁵ <http://www.w3.org/2002/07/owl#>

⁶ <http://xmlns.com/foaf/0.1/>

⁷ http://www.w3.org/2003/01/geo/wgs84_pos#

⁸ <http://www.w3.org/2006/vcard/ns>

⁹ <http://qudt.org/schema/qudt#>

Egy ontológiának megfelelően leírt adathalmazon SPARQL nyelvű lekérdezések futtathatók, a tesztek során használt Linked Data Platform (más néven triplestore, vagy subject-predicate object database) szoftvereknek van is ilyen képessége. Az Apache Marmotta¹⁰ és az OpenLink Virtuoso¹¹ szerverek rendkívül különbözően teljesítettek ugyanazon az adathalmazon ugyanazokat a lekérdezéseket futtatva. A Virtuoso azért volt jelentősen gyorsabb a kettő közül, mert nem szabványos megoldásokat is használ.

A gráfadatbázison végzett tesztekhez a Neo4j¹² szoftver 3.4 Community verzióját használtam. A Neo4j hasznos tulajdonsága, hogy számos fájlformátumból képes adatokat importálni (elsősorban a JSON és CSV támogatott, de XML vagy XLS beolvasására is van lehetőség), így az OWL-ban meglévő adatok a triplestore-ból exportálva betölthetők címkézett tulajdonsággráf adatszerkezetbe. A Neo4j képes futtatni Cypher és Gremlin nyelvű lekérdezéseket is, az előbbi deklaratív, míg az utóbbi elsősorban procedurális szemléletű (bár deklaratív lekérdezések alkotására is van lehetőség). A tesztjeimhez Cyphert használtam, mivel egyrészt ez a Neo4j hivatalosan támogatott lekérdezőnyelve, másrészt pedig felfogásában jobban hasonlít a SPARQL-hez, ezért könnyebben összehasonlítható azzal.

Beszédvezérlés dizartriás felhasználók esetén

A beszédfelismerés alapelve, hogy kinyerjük a rögzített hangminta releváns jellemzőit, amelyekből fonéma szintű modellt alkotunk. A fonéma a nyelv legkisebb egysége, ami önálló jelentéssel vagy jelentés-megkülönböztető szereppel bír (nem mindig egyenértékű a beszédhanggal, mivel előfordul, hogy ugyanannak a fonémának pozíciótól függően más-más kiejtése lesznek), ezért szokás a beszédfelismerés feladatát olyan osztályozási feladatnak tekinteni, ahol a rögzített hangot egymást átfedő keretekre bontjuk, keretenként kiszámítjuk a spektrumát, az így összeálló spektrogramon jellemzően további feldolgozást végzünk (például az emberi hallás változó frekvencia-felbontását modellező mel-skálának megfelelően átszámítjuk), meghatározzuk a hangok határait, majd ezeket az egységeket kíséreljük meg fonéma osztályokba sorolni valamilyen módszer, modell segítségével. A kapott sorozatról pedig eldöntjük, hogy a szótárunk melyik elemére illeszkedik legjobban. Ezt az egyszerű lépéssorozatot

¹⁰ <https://marmotta.apache.org/>

¹¹ <https://virtuoso.openlinksw.com>

¹² <https://neo4j.com/>

természetesen kiegészítheti számos előfeldolgozó eljárás vagy speciális megoldás attól függően, hogy milyen környezetben, milyen célra készül a beszédfelismerő.

Esetemben a klasszikus kötött szavas, beszélőfüggő felismerési feladat kiegészült a dizartria tünetei által okozott problémákkal. A tervezés során abból kellett kiindulnom, hogy a légzés és az artikuláció sérülése miatt a beszélő hangerő-kontrollja bizonytalan, a beszéd szaggatott, bizonyos hangok (jellemzően a mássalhangzók) elmosódnak, a szótagok hossza eltérhet a normálistól. Ráadásul a betegség progresszív természete miatt a beszéd minősége is romlik idővel, tehát nem elég a pillanatnyi kiejtésre betanítani egy modellt úgy-ahogy. Hosszú távon leginkább a magánhangzók maradnak felismerhetők, tehát elsősorban ezekre lehet alapozni a felismerést, ám a magánhangzók esetén is számítani kell némi torzulásra: például a rövid-hosszú magánhangzó párok megkülönböztetése nehézkessé válhat egy idő után, vagy a normális esetben különböző nyelvvállásokkal (alsó-középső-felső) képezett megfelelő magánhangzók kiejtése hasonlóvá válik.

Az átlagos magnitudo-különbség függvény (Average Magnitude Difference Function - AMDF) használható a beszéd alapfrekvenciájának megállapítására [Suma et al, 2010], a beszéd határainak meghatározására a felvett hang és annak eltolt változata közötti távolság megadásával. Az AMDF hasonló az autokorreláció függvényéhez, az alapfrekvencia mellett megmutatja a beszéd zöngés fonémáit (azaz a magánhangzókat és a zöngés mássalhangzókat), ezért felhasználtam a mintában előforduló magánhangzók célzott vizsgálatára.

A magánhangzók olyan, önmagukban kiejthető beszédhangok, amikor a levegő nagyrészt akadálytalanul távozik a tüdőből, megrezegtetve a hangszálakat. Ezen fonémák esetén megállapíthatunk rezonancia frekvenciákat, amelyeket a vokális traktus határoz meg, ezeket formánsoknak nevezzük. A formánsok csúcsocként jelennek meg a hang spektrumában az egyes magánhangzókra jellemző, jól azonosítható módon, így az első három formáns (F1, F2 és F3) kiszámításával lehetővé teszik a magánhangzók formáns alapú osztályozását [Abari et al, 2011]. Ezt a megközelítést használtam a beszédfelismeréssel kapcsolatos tesztjeimben.

A teszteléshez használt hangminták rögzítését, szerkesztését, vizualizálását és elemzését is a Matlab R2011a verziójában futtattam, a Signal Processing Toolbox függvényeinek felhasználásával. A teszteket relatíve csendes környezetben futtattam (a háttérzaj 50 dB-

nél kisebb). A tesztkörnyezethez a mintát egy éles üzemben lévő okosotthon rendszer adta, amelynek későbbi továbbfejlesztése, illetve újratervezése az itt ismertetett eredmények alapján egy jövőbeli célom. A rendszer egy súlyosan mozgássérült, enyhe mértékben beszéd fogyatékos, szklerózis multiplexben szenvedő hölgy lakásában üzemelt sok éven át, egészen a közelmúltig. A tesztekhez az eredeti utasításkészletet használtam, de nyilvánvalóan nem az éles üzemben lévő gépen futtattam őket.

IV. Új tudományos eredmények

1. téziscsoport: akadálymentes beltéri navigáció

1.1 tézis

Egy meglévő, épületek belső szerkezetét leíró ontológiához kidolgoztam egy olyan kiterjesztést, amely lehetővé teszi az adott ontológia szerint leírt épületekben a beltéri akadálymentes navigáció implementálását különféle fokú mozgássérültség szintek speciális igényeihez igazodva [1, 2].

1.2 tézis

Kidolgoztam egy megfeleltetést, amely lehetővé teszi a fenti ontológia szerint leírt adatok transzformálását gráfadatbázisban való tároláshoz, és a gráfadatbázison olyan lekérdezések futtatását, amely mozgássérült felhasználók részletesen megadott igényeinek megfelelő útvonalat talál az épület két pontja között, amennyiben ilyen létezik [3, 4].

2. téziscsoport: beszédfelismerés robusztussága intelligens otthon rendszerekben

2.1 tézis

Dizartriában szenvedő felhasználók számára készült beszédvezérléshez megalkottam egy szövegekre értelmezett távolság mértéket, amely dizartria-specifikus távolságot állapít meg két utasítás között beszédvezérelt intelligens otthon rendszerekben való használat szempontjából [5, 6].

2.2 tézis

Kidolgoztam egy anytime megközelítést a beszédvezérelt intelligens otthon rendszerek beszédfelismerő algoritmusának felépítésére a robusztus működés elősegítése céljából [7, 8, 9, 10, 11].

2.3 tézis

Megfogalmaztam egy olyan, több beszédfelismerő modulból álló keretrendszernek az elveit, amelynek célja beszédvezérelt intelligens otthon rendszerek beszédmodelljének működés közbeni folyamatos adaptációja olyan módon, hogy a felismerési pontosság eközben ne csökkenjen [12, 13, 14].

V. Az eredmények hasznosítási lehetősége

A közvetlen hasznosíthatóság érdekében a munkám során igyekeztem mindig az aktuálisan fennálló körülményekből és lehetőségekből kiindulni, nem pedig egy ideális, illetve jövőbeli helyzethez tervezni. Ez számos korlátozó tényezőt jelentett, de ezek néha éppenséggel ötletet adtak a megoldáshoz, és rávezetnek a gyakorlati felhasználás módjára.

Beltéri akadálymentes navigáció implementálása valós környezetben

A beltéri akadálymentes navigáció gráfadatbázisra épülő megvalósítása (ugyanúgy, mint a korábbi, RDF store-ban tárolt verzió) megköveteli az épület szerkezetének, helyiségeinek, valamint azok akadálymentességi tulajdonságainak leírását, és a későbbiekben természetesen ezen adathalmaznak a karbantartását, aktualizálását is. Ez a feladat emberi erőforrást igényel, a teljes automatizálás lehetősége ugyanis egyelőre nem látszik reálisnak.

Az adatok beviteléhez mindenképpen érdemes egy grafikus felülettel rendelkező programot adni, amely megkönnyíti a bevitt végző személy dolgát. Ez a program, azon túl, hogy képes az egyszerű és áttekinthető felületen bevitt adatokból a megfelelő formátumú fájlokat legenerálni, szolgálhatna egyéb kényelmi funkciókkal is, amelyek az épületek elrendezésének a tipikusan ismétlődő jellegére alapozhatnak (például többemeletes épületen belül a mosdók elhelyezkedése, az ajtók szélessége, a villanykapcsolók magassága, stb.).

Az útvonaltervező hasznosságának jelentős komponensét adja a felhasználói felület használhatósága. Tipikusan mozgássérült felhasználók esetében ez nem csak a program képernyőjének és az adatbevitelnek a logikus kialakítását, és a minél gyorsabb és intuitívabb használatra való törekedést jelenti. Az általános ergonómiai megfontolások mellett legalább ilyen fontos a felhasználói felületnek a mozgássérült felhasználók számára akadálymentes kialakítása is. A Web Content Accessibility Guidelines (WCAG) jelenleg a globálisan elfogadott szabvány az akadálymentes weboldalak és szoftverek tervezésére és implementálására. Léteznek emellett egyéb szempontrendszerek és segédanyagok is, de ezek értelemszerűen mind ugyanazokat az

alapelveket fogalmazzák meg. A mozgássérült felhasználók szempontjából ezek az alapelvek a következők:

- A felhasználói felület működtethető legyen speciális beviteli módszerekkel
- A felhasználói felület biztosítson elegendő időt a bevitel elvégzésére, tehát automatikus váltások vagy (látszólagos) inaktivitás miatti kiléptetés ne, vagy csak hosszabbítható időkorláttal forduljon elő.
- A felhasználói felület minél kevesebb interakciót igényeljen a keresés adatainak megadásához.

Fontos megjegyzés a fentiekhez, hogy az akadálymentes szoftvertervezés egyik alapvető feltevése, hogy a felhasználónak van valamilyen módszere arra, hogy a számítógép vagy táblagép vagy mobiltelefon bemeneti perifériáit kezelni és kimeneti perifériáit érzéklni tudja. A mozgássérült felhasználók esetén ez nyilván a bemenet működtetését jelenti. A gépi beszédfelismeréssel működő diktálás abban az esetben lehet hasznos segédeszköz a gépelés kiváltására, ha a felhasználónak nincs vagy enyhe mértékű a beszéd fogyatéka, ezért erre ne alapozzunk. A kiindulási hely bevitel megoldható az épület egyes részeiben elhelyezett rádiófrekvenciás címkék leolvasásával NFC-képes mobiltelefonok esetén [16].

Okosotthon kialakítása beszéd- és mozgássérült személyek számára

Az általam alkotott módszerek leginkább az okosotthon rendszerek tervezésében játszanak szerepet, már üzemben lévő megoldás megbízhatóbbá alakítására korlátozottan, vagy legalábbis jelentős átalakítások mellett alkalmasak. Jól kiegészítik a jelenleg bevett, megszokott megoldásokat, esetleg egybe is vágnak ezekkel. Ez utóbbira jó példa az okosotthon moduljainak a szabványos felületeken keresztül való csatlakoztatására vonatkozó megbízhatósági kritérium.

Amiről mindenképp szót kell ejtenem az üzemeltetői tapasztalataim kapcsán, az a terméktámogatás és karbantartás kérdése. Egy súlyosan mozgássérült felhasználó otthonában ez egy 24 órában működő rendszer, amely ha megáll, akkor megáll az élet is. A felhasználó semmit sem fog tudni csinálni, amíg a hibát el nem hárítja valaki. Ezért ezeknek a rendszereknek a tervezésénél két kritikus üzemeltetési szempont van, amit mindenféleképpen figyelembe kell venni. Az egyik, hogy a hardver jellegű problémákat (amennyire ez lehetséges) a segítők a helyszínen el tudják végezni. Tehát a

rendszer összeszerelésénél figyelni kell arra, hogy a hozzátartozó vagy a nővér is könnyedén tudjon elemet cserélni, esetleg egy kihúzódt kábelt újra összedugni.

A másik pedig, hogy a szoftver jellegű problémák orvoslása viszont csak a legkritkább esetben várható el a helyszínen lévő személyektől, ezért (nyilván megfelelő autentikációval működő, védett csatornán keresztül) távoli felügyeleti lehetőséget kell biztosítani a vezérlő számítógéphez. Ha a felhasználó állapota lehetővé teszi, akkor a biztonság kedvéért adjunk alternatív segítségkérő csatornát az üzemeltető felé.

Fejlesztői szemlélet tekintetében fel kell készülni arra, hogy a felhasználó egyéni képességei, állapota miatt személyre szabott megoldásokra lehet szükség a rendszerrel való interakció kialakításában. Ebbe az is beleértendő, hogy előfordulhat, hogy nem lehetséges minden szempontból tökéletes és kényelmes módokat találni erre. Arra azonban többnyire lehet számítani, hogy a felhasználó és a segítői is maximálisan együttműködők lesznek a rendszer használatának elsajátításában.

VI. Irodalmi hivatkozások listája

[Anagnostopoulos et al, 2005] Anagnostopoulos, C., et al., "OntoNav: A Semantic Indoor Navigation System", In Proc. 1st Workshop on Semantics in Mobile Environments in conjunction with 6th ACM SIGMOBILE / SIGMOD International Conference on Mobile Data Management, Agua Napa, Cyprus, May 9-13, 2005.

[Dudas et al, 2009] Dudas, P. M., Ghafourian, M., Karimi, H., "ONALIN: Ontology and Algorithm for Indoor Routing", In Proc. Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, MDM'09, Taipei, Taiwan, May 18-21, 2009, pp. 720-725.

[Scholz et al, 2014] Scholz, J., Schabus, S. "An Indoor Navigation Ontology for Production Assets in a Production Environment", In Proc. International Conference on Geographic Information Science, Vienna, Austria, Sept. 24-26, 2014, pp. 204-220.

[Matuszka et al, 2013] Matuszka, T., Gombos, G., Kiss, A., "A New Approach for Indoor Navigation Using Semantic Webtechnologies and Augmented Reality", In Proc. Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments, Las Vegas, NV, USA, July 21-26, 2013, pp. 202-210.

[Benner et al, 2014] Benner, J. G., Karimi, H. A., "Accessible Wayfinding Ontologies for People with Disabilities", Extended Abstract for the RDWG Online Symposium on Accessible Way-Finding Using Web Technologies, Dec. 3, 2014.

[Szász et al, 2016] Szász, B., Fleiner, R., Micsik, A., Simon-Nagy, G., "iLOC – An Indoor Ontology", 2016, <http://lod.nik.uni-obuda.hu/iloc/iloc-20161107.owl>

[Szenes, 2011] Szenes, K., "Enterprise Governance Against Hacking", In Proc. 3rd IEEE International Symposium on Logistics and Industrial Informatics, LINDI 2011, Aug. 25-27, 2011, Budapest, Hungary, pp. 229-233.

[Bilenko et al, 2003] Bilenko M. et al., "Adaptive name matching in information integration", IEEE Intell. Syst. 18 (5), 2003, pp. 16-23.

[Cohen et al, 2003] Cohen, W.W., Ravikumar, P., Fienberg, S.E., "A Comparison of String Distance Metrics for Name-Matching Tasks", In Proc. IJCAI-03 Workshop on Information Integration, Acapulco, Mexico, August 9-10, 2003, pp. 73-78.

- [Ukkonen, 1992] Ukkonen, E., "Approximate string-matching with q-grams and maximal matches", *Theor. Comput. Sci.* 92 (1), 1992, pp. 191-211.
- [George et al, 2004] George, S., Zukerman, I., Niemann, M., "An Anytime Algorithm for Interpreting Arguments," *PRICAI 2004: Trends in Artificial Intelligence Lecture Notes in Computer Science*, Vol. 3157, 2004, pp. 311-321.
- [Menzel, 1994] Menzel, W. "Parsing of Spoken Language under Time Constraints", In *Proc. 11th European Conference on Artificial Intelligence*, Amsterdam, The Netherlands, August 1994, pp. 560–564.
- [Kalinli et al, 2010] Kalinli, O. et al., "Noise Adaptive Training for Robust Automatic Speech Recognition", *IEEE Trans. Audio, Speech, and Language Processing*, Vol. 18, Issue 8, 2010, pp. 1889-1901.
- [Juang et al, 2001] Juang, C., Lin, C., "Noisy speech processing by recurrently adaptive fuzzy filters" *IEEE Trans. Fuzzy Systems*, Vol. 9, Issue 1, 2001, pp. 139-152.
- [Wang et al, 2014] Wang, H., Wang, L., Liu, X., "Multi-level adaptive network for accented Mandarin speech recognition", In *Proc. 4th IEEE Int. Conf. on Information Science and Technology (ICIST)*, Shenzhen, China, Apr. 26-28, 2014, pp. 602-605.
- [Riccardi et al, 2005] Riccardi, G., Hakkani-Tur, D., "Active learning: theory and applications to automatic speech recognition", *IEEE Trans. Speech and Audio Processing*, Vol. 13, Issue 4, 2005, pp. 504-511.
- [Xue et al, 2014] Xue, S. et al., "Fast Adaptation of Deep Neural Network Based on Discriminant Codes for Speech Recognition", *IEEE/ACM Trans. Audio, Speech, and Language Processing*, Vol. 22, Issue 12, 2014, pp. 1713-1725.
- [Baker, 2000] Baker, J. K., "Speech recognition using multiple recognizers (selectively) applied to the same input sample", U.S. Patent 6 122 613, Sept. 19, 2000.
- [Endo et al, 2007] Endo, N. et al., "Speech recognition system having multiple speech recognizers", U.S. Patent 7 228 275, June 5, 2007.
- [Murveit et al, 2006] Murveit, H. et al., "Speech recognition system to selectively utilize different speech recognition techniques over multiple speech recognition passes", U.S. Patent 7 058 573, June 6, 2006.

[Schwenk et al, 2000] Schwenk, H., Gauvain, J., "Combining Multiple Speech Recognizers using Voting and Language Model Information", In Proc. IEEE International Conference on Speech and Language Processing, ICSLP 2000, Beijing, China, Oct. 16-20, 2000, pp. 915-918.

[Natori et al, 2010] Natori, S., Nishizaki, H., Sekiguchi, Y., "Japanese Spoken Term Detection Using Syllable Transition Network Derived from Multiple Speech Recognizers' Outputs", In Proc. 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010, Makuhari, Chiba, Japan, Sept. 26-30, 2010, pp. 681-684.

[Natori et al, 2013] Natori, S. et al., "Spoken Term Detection Using Phoneme Transition Network from Multiple Speech Recognizers' Outputs", Information and Media Technologies 8.2, 2013, pp. 457-466.

[Suma et al, 2010] Suma, S.A., Gurumurthy, K.S., "Novel pitch extraction methods using average magnitude difference function (AMDF) for LPC speech coders in noisy environments", In Proc. of the 2nd International Conference on Signal Processing Systems, ICSPS 2010, Vol. 1, Dalian, China, July 5-7, 2010, pp. 636-640.

[Abari et al, 2011] Abari, K., Rácz, Zs. Zs., Olaszy, G., "Formant maps in Hungarian vowels - online data inventory for research, and education", In Proc. of the Int. Conf. on Speech and Technology, Interspeech 2011, Florence, Italy, August 27-31, 2011, pp. 1609-1612.

VII. A tézispontokhoz kapcsolódó tudományos közlemények

- [1] Fleiner, R., Szász, B., Simon-Nagy, G., Micsik, M., "Indoor navigation for motion disabled persons in medical facilities", *Acta Polytechnica Hungarica*, Vol. 14, No. 1, 2017.
- [2] Simon-Nagy, G., Fleiner, R., "Ontology Extension for Personalized Accessible Indoor Navigation", *Advances in Intelligent Systems and Computing* 660, 2017, pp. 281-288.
- [3] Chalhoub, N., Simon-Nagy, G., "Indoor Navigation based on Linked Data at Honvéd Hospital, Budapest" In Proc. IEEE 12th International Symposium on Applied Computational Intelligence and Informatics, SACI 2018, Temesvár, Románia: IEEE Romania Section, IEEE Hungary Section, May 17-19, 2018, pp. 485-490.
- [4] Simon-Nagy, G., Chalhoub, N., Fleiner, R., "Graph-based Data for Accessible Indoor Navigation" In Proc. IEEE 23th International Conference on Intelligent Engineering Systems, INES 2019, Gödöllő, Hungary, Apr. 25-27, 2019.
- [5] Simon-Nagy, G., Várkonyi-Kóczy, A.R., "Distance metric for speech commands in smart home systems of dysarthric users", *Advances in Intelligent Systems and Computing* 519, *Recent Global Research and Education: Technological Challenges*. Berlin; Heidelberg: Springer Verlag, 2017, pp. 325-330.
- [6] Simon-Nagy, G., Várkonyi-Kóczy, A. R., "Distance metric for speech commands in smart home systems of dysarthric users", In Proc: 15th Int. Conf. on Global Research and Education in Intelligent Systems, Inter-academia'2016, Warsaw, Poland, Sept. 26-28, pp. PS31-1-PS31-6.
- [7] Nagy, G., "An interpretation of the COBIT information criteria to operational criteria of voice controlled Ambient Assisted Living systems", In Proc. 5th IEEE International Symposium on Logistics and Industrial Informatics, LINDI 2013, Wildau, Germany, Sept. 5–7, 2013, pp. 49-53.
- [8] Nagy, G., "Construction of anytime algorithms for robust speech recognition", *Advanced Materials Research* 1117, 2015, pp. 265-268.

- [9] Nagy, G., "Construction of Anytime Algorithms for Robust Speech Recognition", In Proc. 13th Int. Conf. on Global Research and Education in Intelligent Systems, Interacademia'2014, Riga, Lettország, 2014, pp. 170-171.
- [10] Nagy, G., Várkonyi-Kóczy, A.R., Tóth, J.T., "An Anytime Voice Controlled Ambient Assisted Living System for Motion Disabled Persons", In Proc. IEEE International Symposium on Medical Measurements and Applications, MeMeA'2015, Torino, Italy, May 7-9, 2015, pp. 163-168.
- [11] Várkonyi-Kóczy, A. R., Kiss, D., Nagy, G., Tóth, J., "Anytime Classification", In Proc. 7th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, IEEE HNICEM - ISCIII 2014, Puerto Princesa, Philippines, Nov. 12-14, 2014. paper 162.1-6.
- [12] Nagy, G., Kutor, L., "Multiple ANN Recognizers for Adaptive Recognition of the Speech of Dysarthric Patients in AAL Systems", In Proc. Advancing Assistive Technology and eAccessibility for People with Disabilities and the Aging Population, AAATE 2015, Budapest, Hungary, Sept. 9-12, 2015, pp. 1013-1016.
- [13] Simon-Nagy, G., Várkonyi-Kóczy, A.R., "Adaptive Speech Recognition Framework for Dysarthric Patients", JJAP Conf. Proc. Vol. 4, 2016, Cikk azonosítója: 011614.
- [14] Nagy, G., Várkonyi-Kóczy, A. R., "Adaptive Speech Recognition Framework for Dysarthric Patients", In Proc. 14th Int. Conf. on Global Research and Education in Intelligent Systems, Interacademia'2015, Hamamatsu, Japán, 2015, pp. 130-131.

VIII. További tudományos közlemények

[15] Fleiner, R., Simon-Nagy, G., Szász, B., "Accessible Indoor Navigation based on Linked Data in Hospitals", In Proc. IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016, Budapest, Hungary, Oct. 9-12, 2016, pp. 002425-002430.

[16] Kutor, L., Domozi, Zs., Nagy, G., Peller, Cs., Véső, T., "Object Identification and Local Information Services Using Near Field Communication", In Proc. Regional Conference on Embedded and Ambient Systems, RCEAS 2007, Budapest, Hungary, Nov. 22-24, 2007, 91 p.