

Mutasd a hangod – automatikus jeltolmács

Mátételki Péter, Pataki Máté, Turbucz Sándor, Kovács László

MTA SZTAKI Elosztott Rendszerek Osztály

1111 Budapest, Lágymányosi utca 11.

{peter.matetelki, mate.pataki, sandor.turbucz, laszlo.kovacs}@sztaki.mta.hu

Kivonat: *A Tolmácskesztyű projektben egy olyan segédeszközt alkotunk, mellyel a beszéd- és halláskárosult emberek kézmozgását, vagyis gesztusokat használva képesek a mindennapi életben kapcsolatot teremteni ép embertársaikkal. A kifejlesztett segédeszköz egy innovatív hardver-szoftver-rendszer, amely kézmozgást érzékelő kesztyűből valamint kézjeleket felismerő és nyelvi feldolgozást végző szoftverből áll. A Tolmácskesztyű eszkörendszer jelnyelvi szinkrontolmácsként működik, segítségével a fogyatékkal élők anyanyelvükön – vagyis jelnyelven – kommunikálhatnak az épekkel. A Tolmácskesztyű applikáció a jelelt szöveget hangosan felolvassa, így a sérültek és a (jelnyelvet nem ismerő) épek között folytonos kommunikáció jön létre.*

Bevezetés

A siket emberek jelenleg jelnyelvi tolmács segítségével kommunikálnak az épekkel, tolmács híján sokszor hosszás, olykor nehézkes magyarázkodással tudják csak megértetni magukat. Problémájuk hasonló – sőt teljesen analóg – azzal, amikor két, közös nyelvet nem beszélő ember próbál kommunikálni. A siketek anyanyelve a jelnyelv, melyet a Tolmácskesztyű szinkrontolmácsként képes beszélt nyelvre alakítani, így általa a fogyatékos csoport tagjai képesek magukat megértetni a jelnyelvet nem ismerőkkel. A Tolmácskesztyű ezzel jelentősen előmozdítja a siketek társadalmi integrációját, üzletben, piacon való vásárlástól kezdve a banki ügyintézésen át az oktatásban való részvételig számos kaput nyit meg számukra. Ezen felül az eszköz nagy mértékben segítheti a beszéd- és halláskárosult emberek munkavállalási lehetőségeit is.

Az általunk kifejlesztett rendszer két alkotórészből áll: egy kesztyűből, mely a rá erősített orientációérzékelő szenzorok segítségével gesztusleírók formájában megalkotja a kéz digitális mását, gesztusleírók formájában. Ezen leírókat folyamatosan sugározza a hozzá vezeték nélkül csatlakoztatott mobiltelefonos applikációnak, melyen a jelfeldolgozást, betű-, szó- és mondataalkotást, valamint a felolvasást végző algoritmus fut. A mobilapplikáció feldolgozza a

bejövő gesztusleírókat, intelligens szegmentáló algoritmusok és automatikus szövegjavítás során érthető szöveget hoz létre, melyet hangosan felolvas.

Az alábbiakban ismertetjük a rendszer magasszintű áttkintését és működését, illetve részleteket írunk le a mobilapplikációként működő feldolgozó algoritmusok működéséről. A hardvert és alacsony szintű hardvervezérlő szoftvereket nem részletezzük.

A „Jelnyelvi tolmácskesztyű fejlesztése” KMR_12-1-2012-0024 projekt az MTA SZTAKI és az Euronet Magyarország Zrt. konzorciális együttműködése során, a Magyar Kormány támogatásával, a Nemzeti Fejlesztési Ügynökség kezelésében, a Kutatási és Technológiai Innovációs Alap finanszírozásával valósul meg.

Kapcsolódó kutatások

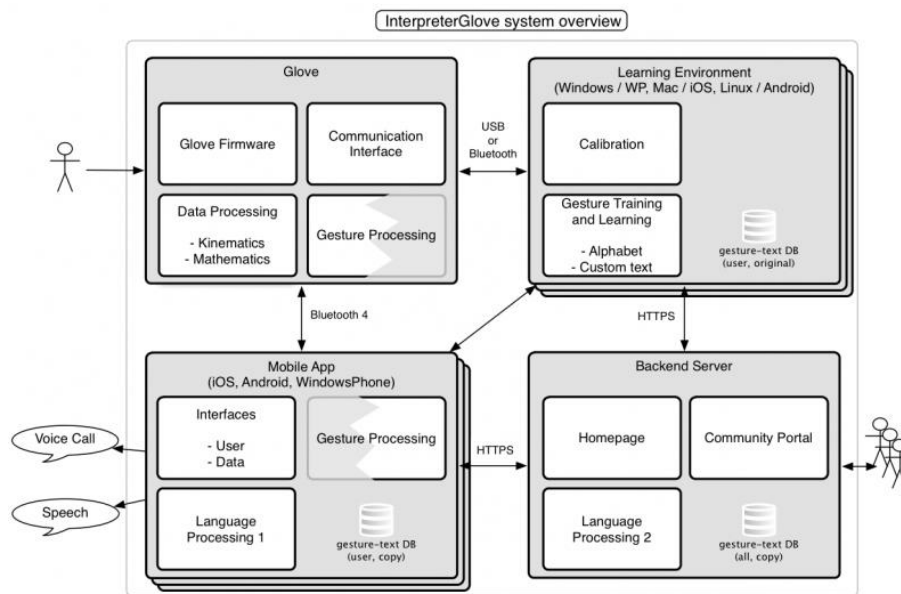
A tudományos szakirodalomban a gesztusfelismerésre két megközelítéssel találkozhatunk. Nagyobb múltra tekint vissza az optikai módszereken alapuló elemzés, mely esetben kamerák által rögzített képet analizálva rekonstruálják a kéz állapotát. Az elmúlt években egy másik módszer is napvilágot látott, mely orientációérzékelő szenzorokat vesz alapul. Ilyen szenzorokkal találkozhatunk a modern mobilkészülékekben is, melynek köszönhetően a telefonok a fizikai elfordításra reagálva képesek elfordítani a képernyőn megjelenő grafikát.

Taiwoo Park és társai [1] egy viselhető mozgásdetektáló szenzorhálózat energetikai analizálására fókuszáltak. In-Kwon Park [2] munkájában egy hasonló kesztyű-alapú rendszerről olvashatunk, ám ez a – tanítható – Tolmácskesztyűvel ellentétben hajlításerzékelő (flex) szenzorokat alkalmaz, és csak 17 különféle gesztus megkülönböztetésére alkalmas. Yun Li [3] egy accelerométer és EMG szenzorok [8] segítségével dolgozott ki kínai jelnyelvi felismerőt, mely főként a kéz pozícióját veszi alapul, az ujjak szerepe csak másodlagos – szemben a Tolmácskesztyűvel, ahol az ujjak döntő szerepet játszanak. Seungki Min [4] és Cemil Oz [5] egyaránt orientációszenzor alapú rendszerekkel kísérelték meg a gesztusfelismerést, de HMM [9] és ANN [10] módszereket alkalmaztak, melyek a Tolmácskesztyű esetében komoly korlátot jelentenének a betaníthatóságra nézve, ezért mi elvetettük ezen módszerek alkalmazását.

Farid Parvini [6] [7] CyberGloves [11] kesztyű felhasználásával készített gesztusfelismerő rendszert, munkáiban ismerteti a biomechanikai validáción alapuló felismerést, valamint összehasonlítja azt az ANN-alapú módszerekkel; rámutat előbbi hatékonyságára. A Tolmácskesztyű projektben mi is nagy hasznát vettük a biomechanikai megközelítésnek, melyet elsősorban a gesztusleíró megalkotásakor kamatoztattunk.

Működés

A rendszer fő alkotórészei a kesztyű – ideértve a hardvert és alacsony szintű vezérlőszoftvert egyaránt –, a mobilapplikáció, melyeket kiegészít a betanító környezet és a szerver. A logikai architektúrát az alábbi – angol nyelvű – ábra szemlélteti:



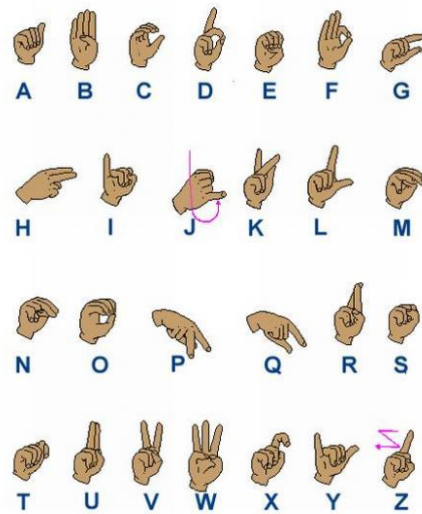
Ábra 1: Funkcionális architektúra

A kesztyű prototípusa lélegző elasztikus anyagból készül, ergonomikus tervezésének köszönhetően alkalmas mindennapi használatra. A kesztyűre integrálva orientációérzékelő szenzorok találhatóak, melyek adatait felhasználva szenzorfüziós, illetve térgeometriai műveletek segítségével megalkotjuk a kézfej digitális mását. Ezt speciális gesztusleírókkal reprezentáljuk. A gesztusleírókat a kesztyű vezérlőegységéhez illesztett vezeték nélküli kommunikációs egysége segítségével folyamatos jelfolyamként továbbítjuk a felhasználó mobiltelefonjára, másodpercenként 30 darabot.

Komoly kihívást jelentett a megfelelő gesztusdeszkriptor megalkotása. Miután analizáltuk az emberi kéz anatómiai struktúráját, bio-mechanikai és kinematikai jellemzőit [12], megalkottunk egy struktúrát, mely lehetőséget teremt tetszőleges előforduló kéz- és ujjállapot, valamint a szignifikáns kézmozgások kódolására. 20 helyiértéken tároljuk az ujjaknak a kézfej síkjához viszonyított állapotát, az ujjaknak a kézfejre merőleges síkhoz viszonyított állapotát, a hüvelykujj befördítöttségét, az ujjak szétnyitöttségét, a csukló és a kézfej helyzetét.

A jelelést megelőzően a kesztyűt a felhasználónak testre kell szabni, be kell tanítani a viselőjére jellemző nyújtott és hajlított, összezárt és nyitott ujjállásokat; többek közt erre szolgál a betanító környezet, mely vizuális visszacsatolást és beállítási lehetőségeket kínál egy platformfüggetlen GUI-n keresztül. Szintén itt készítjük el a felhasználó saját „szótárát”, mely a felismerendő kézjelekhez rendel tetszőleges szöveget.

Mivel a projektben egykezes kesztyűt készítünk, a nemzetközi Daktil ujjábécé [13] jelnyelvet választottuk ki tesztelésre. Ez minden egyes betűhöz egy gesztust rendel, az ábécé segítségével tehát tetszőleges szót le lehet betűzni.



Ábra 2: Daktil ABC

Annak érdekében, hogy a használatot megkönnyítsük, a gesztusokhoz nem csak betű rendelhető, hanem tetszőleges szöveg is. Felhasználóink így fel tudnak készülni az adott napon, adott szituációban várható tevékenységekre, és az általuk választott gesztusokhoz akár kifejezéseket, mondatokat, vagy hosszabb szövegeket is hozzárendelhetnek, pl. „Köszönöm!”, „Legyen szíves!”, „Szeretnék felvenni a bankszámlámról 25000 forintot.”.

A mobiltelefon a bejövő jeleket feldolgozza és értelmezi annak érdekében, hogy a felhasználó által mutatott gesztussorozathoz tartozó szöveget rekonstruálja. A feldolgozás két lépésből áll; az első lépésben az adatfolyam szegmentációja történik meg, ezt követően pedig kontextusérzékeny javítást alkalmazunk annak érdekében, hogy a végeredményként előálló szöveg felolvasva értelmes, lehetőleg hibamentes legyen.

A felhasználó egymás után mutatja a korábban betanított jeleket. Siket és jeltolmács tesztelőink segítségével azt állapítottuk meg, hogy átlagosan azonos ideig formál a kezük egy érvényes gesztust, mint amennyi ideig a kéz átmeneti állapotban van két jel között. A kesztyű egyenlő időközönként, másodpercenként 30-szor mintavételezi a kéz állapotot, és küldi át a telefonnak. A szegmentálással ebből a folyamatos jelfolyamból kiválasztjuk az értékes gesztusleírókat, vagyis megállapítjuk minden egyes leíróról, hogy az egy definiált jelhez, vagy egy átmeneti állapothoz tartozik. Erre többféle algoritmust terveztünk és teszteltünk: a Hagdil hasonlóság alapú, ablakos simítást használó, ismeretlen karakterek mentén szegmentáló, betűismétlés alapján szegmentáló, kézdinamikai vizsgálaton alapuló, valamint átmeneti jelek mentén szegmentáló algoritmusok közül az utolsó kivételével mindegyiket implementáltuk. Az evaluáció során azt vizsgáltuk, hogy az előállított szöveg mennyire hasonló a jelelt szöveghez. Mérési eredményeink azt mutatják, hogy az (erőforrásigény szempontjából) legolcsóbb és legdrágább algoritmusok teljesítenek a legjobban, további vizsgálatainkhoz ezért a csúszóablakos simítást és a kézdinamikai vizsgálaton alapuló szegmentációk során előállított szöveget vettük alapul. A két algoritmus működésében

egy jelentős eltérést azonosítottunk, miszerint az ablakozó algoritmus állandó, kiszámítható körülmények között a lehető legjobb eredményt produkálja (pl. gesztusok precíz mutatása, állandó jelelési sebesség), míg a kézdinamikai vizsgálat a kontextusra érzéketlenül képes jó eredménnyel szegmentálni. Fontos továbbá megjegyezni, hogy bár az algoritmusok által előállított szöveg egyszerű távolságmétrikai mérése alapján hasonló minőséget látunk, de az eltérő működésük miatt a generált szöveg jelentősen eltér egyik és másik esetben, melynek a kontextusérzékeny javításnál később még szerepe lesz.

A kontextusérzékeny javítási algoritmus valójában egy természetes nyelvi feldolgozást (Natural Language Processing, NLP) foglal magában. A szegmentáció által előállított nyers szöveg tartalmazhat hibákat, melyeket a kontextusérzékeny javító hivatott értelmes szöveggé alakítani. Erre a célra a létező helyesírásellenőrzőket sajnos nem lehet alkalmazni, mert a kesztyű és a szegmentáló általi hibák egészen más jellegűek, mint a gépelés közben előforduló elírások, ezért ezek az algoritmusok nem képesek megbirkózni ezzel a speciális feladattal. A Tolmácskesztyű esetében figyelembe kell venni a kesztyű által készített gesztusleíró szemantikájából adódó speciális tulajdonságokat, a felhasználás tipikus jellemzőit, valamint a beszélt nyelv sajátosságait is. Az általunk készített komplex algoritmusban ennek megfelelően módosított Levenshtein távolságmétrikát, n-gram adatbázis alapú keresést és confusion-mátrixon alapuló súlyozási módszereket alkalmazunk. E három módszer kombinációjával létrehozott algoritmus segítségével javítunk a szegmentáció által előállított szövegen. Az evaluáció során azt tapasztaltuk, hogy a legjobb eredményeket közepes (kb. 100.000 szavas) szótárak esetén kaptuk, valamint megállapítottuk azt is, hogy a kézdinamikai szegmentáló által létrehozott szöveget nagyobb mértékben voltunk képesek javítani a fenti módszerrel, mint az ablakozó szegmentálóét. Mind a szegmentáció, mind pedig a kontextusérzékeny feldolgozás offline módon, az applikációba integráltan működik.

Az architektúraábrán látható backend szerver kiegészítő szerepet tölt be. Segítségével a gesztusok közösségi megosztására van lehetőség, illetve a felhasználói szótárak online felületen történő szerkesztésére.

Evaluáció és jövőbeli tervek

A projekt elejétől fogva szoros együttműködésben dolgozunk a célcsoport képviselőivel, siket emberekkel és jeltolmácsokkal. A tervezés során kritikáikat és észrevételeiket figyelembe vettük, és a lehető leghamarabb bevontuk őket a tesztelési folyamatba. Jeleléseik kielemezéséből adódó konklúziók és értékes visszajelzéseik a prototípus számos tulajdonságát segítették tökéletesíteni, többek közt az optimális jeltovábbítási frekvencia meghatározását, a Hagdil szemantika kifejlesztését, a megfelelő értékészletek és legjobban használható kézgesztusok meghatározását.

A projekt célkitűzéseként megvalósuló jeltolmácskesztyű eredetileg a siketek támogatására készült. Fogyatékosügyi szervezetekkel történt egyeztetéseink rávilágítottak arra, hogy a Tolmácskesztyű a megfelelő változtatásokkal alkalmassá tehető számos másik célcsoport részére. Nagyothallók, beszéd fogyatékosok, mozgássérült és halmozottan sérült emberek, vagy

akár átmeneti beszéd fogyatékoságban szenvedő emberek a rehabilitációjuk során mind kiválóan tudnák hasznosítani a Tolmácskesztyű számukra testreszabott változatát.

Tervezzük továbbá a rendszer kiterjesztését több módon is. Egyrészt a meglévő rendszerbe – elsősorban – szoftveres változtatásoknak köszönhetően tervezzük integrálni a kézmozgás dinamikáját, melyre a jelenleg alkalmazott kesztyű-telefon közötti kommunikációs protokollt fel is készítettük. Másrészt tervezzük a rendszer kiterjesztését a test további részeire, elsősorban a karra, illetve a karokra és másikké kézzre, így létrehozva a kétkezes kesztyűt, melyet felhasználva megnyílik a lehetőség a teljes jelnyelvi felismerésre és számos más alkalmazási terület kiaknázására.

Referenciák

- [1] Taiwoo Park, Jinwon Lee, Inseok Hwang, Chungkuk Yoo, Lama Nachman, Junehwa Song. 2011. E-Gesture: a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems (SenSys '11)*. ACM, New York, NY, USA, 260-273.
- [2] In-Kwon Park, Jung-Hyun Kim, Kwang-Seok Hong. 2008. An implementation of an FPGA-based embedded gesture recognizer using a data glove. In *Proceedings of the 2nd international conference on Ubiquitous information management and communication (ICUIMC '08)*. ACM, New York, NY, USA, 496-500.
- [3] Yun Li, Xiang Chen, Jianxun Tian, Xu Zhang, Kongqiao Wang, Jihai Yang. 2010. Automatic recognition of sign language subwords based on portable accelerometer and EMG sensors. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI '10)*. ACM, New York, NY, USA, Article 17, 7 pages.
- [4] Seungki Min et al., "Simple Glove-Based Korean Finger Spelling Recognition System", Lecture Notes in Computer Science Volume 4705, 2007, 1063-1073.
- [5] Cemil Oz, Ming C. Leu. 2005. Recognition of finger spelling of American sign language with artificial neural network using position/orientation sensors and data glove. In *Proceedings of the Second international conference on Advances in neural networks - Volume Part II (ISNN'05)*, Berlin, Heidelberg, 157-164.
- [6] Parvini, F.; Shahabi, C., "Utilizing Bio-Mechanical Characteristics For User-Independent Gesture Recognition," *Data Engineering Workshops, 2005. 21st International Conference*, pp.1170,1170, 05-08 April 2005
- [7] Farid Parvini, Dennis Mcleod, Cyrus Shahabi, Bahareh Navai, Baharak Zali, Shahram Ghandeharizadeh. 2009. An Approach to Glove-Based Gesture Recognition. In *Proceedings of the 13th International Conference on Human-Computer Interaction. Part II: Novel Interaction Methods and Techniques*, Julie A. Jacko (Ed.). Springer-Verlag, Berlin, Heidelberg, 236-245.
- [8] <http://en.wikipedia.org/wiki/Electromyography>
- [9] http://en.wikipedia.org/wiki/Hidden_Markov_model
- [10] http://en.wikipedia.org/wiki/Artificial_neural_network
- [11] <http://www.cyberglovesystems.com/>
- [12] <http://www.eatonhand.com/nor/nor002.htm>
- [13] <http://byvonn.5mp.eu/web.php?a=byvonn&o=rMReKCyrMd>