

ÉRZELEM FELISMERŐ INTERFÉSZ KONCEPCIONÁLIS TERVEZÉSE OKOSOTTHON RENDSZERHEZ

L. Kiss Márton

PhD hallgató, Miskolci Egyetem, Automatizálási és Infokommunikációs Intézet
3515 Miskolc, Miskolc-Egyetemváros, e-mail: autkiss@uni-miskolc.hu

Pintér Judit Mária

tudományos főmunkatárs, Miskolci Egyetem, Automatizálási és Infokommunikációs Intézet
3515 Miskolc, Miskolc-Egyetemváros, e-mail: pinterjm@uni-miskolc.hu

Absztrakt

A beszéd alapú navigáció térhódítása az okosotthon rendszereket sem kerülte el, egyre több olyan modulokat és komplett rendszereket találhatunk a piacon, amik segítségével szóbeli utasításokkal tudunk az otthonunkban lámpákat kapcsolni, hőmérsékletet állítani, kapcsolgatni a szellőztetést, vezérelni a multimédia eszközöket. Célunk egy olyan interfész megtervezése és fejlesztése, ami a hangvezérlésen túl az érzelmek felismerésére is alkalmas. Diszpécser központokban és call centerekben már használatos megoldás, de házautomatizálási rendszerek esetén sok kérdést felvet annak alkalmazhatósága.

Kulcsszavak: IoT, érzelemfelismerés, hangvezérlés, okosotthon

Abstract

The spread of voice-based navigation has not escaped smart home systems either, with more and more modules and complete systems on the market that can be used to turn on lights, set temperatures, turn on ventilation, and control multimedia devices in your home with verbal instructions. Our goal is to design and develop an interface that, in addition to voice control, is also capable of recognizing emotions. This solution is already in use in dispatcher centers and call centers, but in the case of home automation systems, its applicability raises many questions.

Keywords: IoT, emotion recognition, voice control, smart home

1. Bevezetés

A beszéd alapú érzelemfelismerés kevésbé kapott még helyet okosotthon modulokban. Az Amazon cég például 2014 végén jelentette be az első virtuális hangasszisztensét, amelyet Amazon Alexának nevezett el és az egyik legközkedveltebb termék a piacon. Az Amazon Alexa könnyen integrálható meglévő okosotthonok vezérlésébe. Különböző előre beprogramozott folyamatokat képes végrehajtani egyetlen szó hatására, amit a telefonos alkalmazásán belül lehet beállítani és lehetőséget kínál más gyártók eszközeinek vezérlésére is [1]. Az Amazon Alexa hasonlóan társaival, még nem alkalmas érzelmek felismerésére és reakcióképzésre, bár a fejlesztések és tesztelések kezdetét vették, de piacon még nem fellelhető.

Diszpécser központokban már használatos megoldás, de házautomatizálási rendszerek esetén sok kérdést felvet annak alkalmazhatósága, hiszen meg kell alkotnunk egy olyan szabályrendszert, ami alapján a felismert érzelmekre adott válaszreakciója a háznak ne legyen tola-
kodó és ne kapcsoljon be

minden egyes parancsunk után, illetve ne bírálja felül a már kiadott parancsot, ha az eltérő lenne. A cikkben részletesen bemutatásra kerülő rendszertervezetet megelőzte egy kísérleti jellegű fejlesztés, amely szintén bemutatásra kerül az alábbiakban.

2. Az érzelemfelismerés alkalmazhatósága okosotthon rendszerekben

A tervezett interfész kísérleti fejlesztését a Miskolci Egyetem vezette csapat Solar Decathlon Europe 2019 (SDE19) Nemzetközi Innovációs Passzívház építő versenyen megépített Magyar Fészek+ pályamunkájában végeztük el. Ebben a verzióban a házban elhelyezett nappalin és konyhán végigfutó LED szalag színét tudtuk változtatni és egy adott zeneszámot elindítani.

A kísérleti rendszer megalkotásának elsődleges célja az volt, hogy kézzelfogható eredményt kapjunk arról, hogy valóban megoldható a kommunikációs útvonal kialakítása, valamint a versenyen a szakmai zsűri is meg tudja osztani gondolatait az elképzeléssel kapcsolatban. Maga a tervezett funkció túlmutat az alapszükségleteken, azonban olyan mértékben hozzájárulhat a komfortérzethez, amelynek élettani hatásai is lehetnek, illetve a kutatási terület 3 szakterület határán helyezkedik el, ami szintén emeli a fejlesztésben rejlő potenciált.

A következő fejlesztési lépésben célunk, hogy még több paramétert változtassunk a házon belül. A LED szalag színén kívül a többi lámpa vezérlése is integrálásra került az energiahatékonyságot szem előtt tartva. Tervezett fejlesztési lépés a zenei rész bővítése, a hőmérséklet, a páratartalom és a szellőztetés változtatása. További vizsgálati pont munkánkban, hogy az egyes érzelmekhez mely értékek az optimálisak. Itt már pszichológiai szakirodalomra van szükség, hogy az egyes érzelmi állapotok ellensúlyozására, vagy öröm és nyugalom esetén azok fenntartására milyen színek, fényerősség, zene készlet, hőmérséklet ajánlott.

A vizsgálatok megkezdéséhez az 1. táblázat tartalmazza azokat az értékeket, amiket meghatároztunk az egyes érzelmekhez. Az értékek megválasztásánál a szakirodalomban megfogalmazott komfortérzet értékeket is figyelembe vettük [2].

1. táblázat. Az egyes érzelmekhez rendelt paraméterértékek

Érzelmek	LED szalag szín	Egyéb lámpák színhőmérséklete	Hőmérséklet [C°]	Zene	Páratartalom	Szellőztetés
Szomorúság		meleg	22	Vidám zeneválogatás	60%	Nem releváns
Harag		meleg	24	Tenger morajlás, relaxációs zeneválogatás	50%	Bekapcsol
Semleges		hideg	24	Jazz zeneválogatás	40-60%	Nem releváns
Boldogság		meleg	23	Természetből felvett hangok	60%	Nem releváns
Nyugalom		meleg	23	Klasszikus zeneválogatás	40-60%	Nem releváns
Unalom		meleg	21	Elektronikus zeneválogatás	40-60%	Bekapcsol
Izgalom		semleges	21	Zongora zeneválogatás	50%	Bekapcsol
Félelem		semleges	23	Meditációs zeneválogatás	50%	Bekapcsol

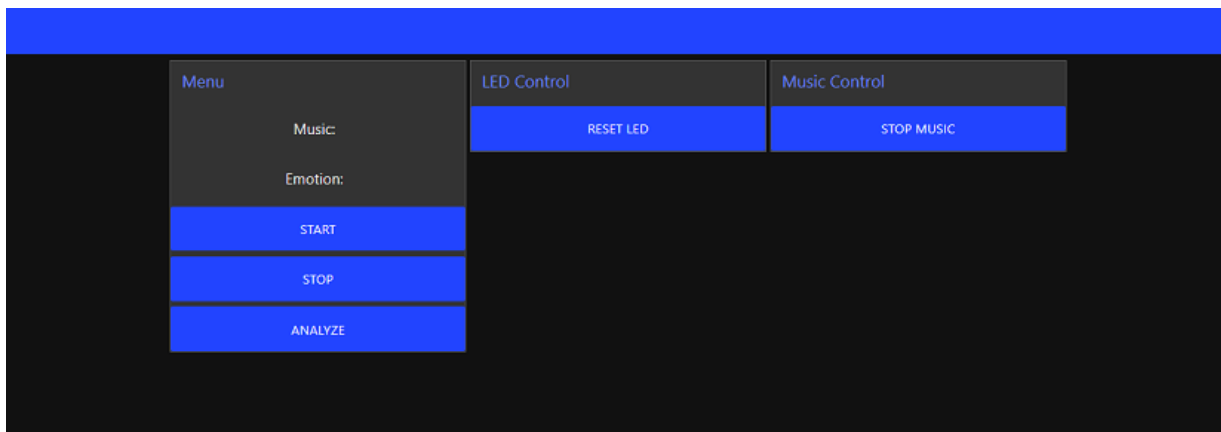
A táblázatban szereplő utolsó három sárgával jelölt sorban, olyan érzelmek szerepelnek, amiket az interfész kísérleti fejlesztésében mellőztünk, de a következőkben alkalmazni fogunk.

3. Kísérleti rendszer bemutatása

Az SDE19-es versenyen integrált prototípus felismerő modulja nem a kimondott szavak értelmét vizsgálta, hanem a beszéd szupraszegmentális jellemzőit (hangmagasság, hangerő, hangsúly, beszéddal- lam, beszédtempó, beszédrítmus, beszédszünet, egyéni hangszínezet). A felismert érzelmi állapot alapján változtatjuk a LED megvilágítás színét és indítunk el zenét. Így kiépítésre került a kapcsolatot a felismerő és a ház automatizálási rendszere között, de ebben az esetben a szóbeli vezérlés teljesen függetlenül működött, míg az érzelemfelismerő, csak manuális indítással lehetett használni és az érzelmek által három paraméter került változtatásra [3].

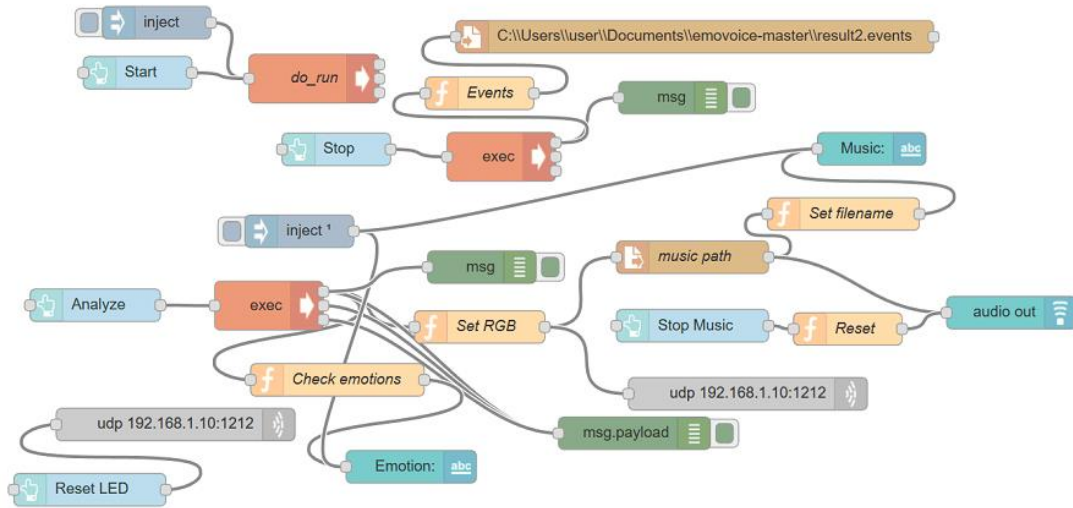
A fejlesztés megkezdésekor mindenképpen azt szerettük volna megvalósítani, hogy a tervezett interfész bemenete az Amazon Alexa legyen, mint mikrofon. Ez azonban meghíúsult, mivel jelenleg az Alexa néhány javítandó hibával küzd és ezért a PC-re telepíthető alkalmazást nem lehet elérni, így csak jobbára hangszóróként lehet közvetlenül számítógéphez csatlakoztatni.

A kezdeti verziónál így a WoMic alkalmazást használtuk –aminek segítségével az Andoridos mobil eszközök vezeték nélküli mikrofonként tudnak funkcionálni, így mobilt vagy tabletet tudtunk használni, mint mikrofon. Az indítást és az elemzést a Node RED fejlesztői környezetben létrehozott webes alkalmazással lehetett vezérelni. Az 1. – 2. ábrán látható a kísérleti fejlesztés eredménye, amellyel a fentebb említett folyamatot tudtuk megvalósítani.



1. ábra. A kísérleti fejlesztés grafikus felülete

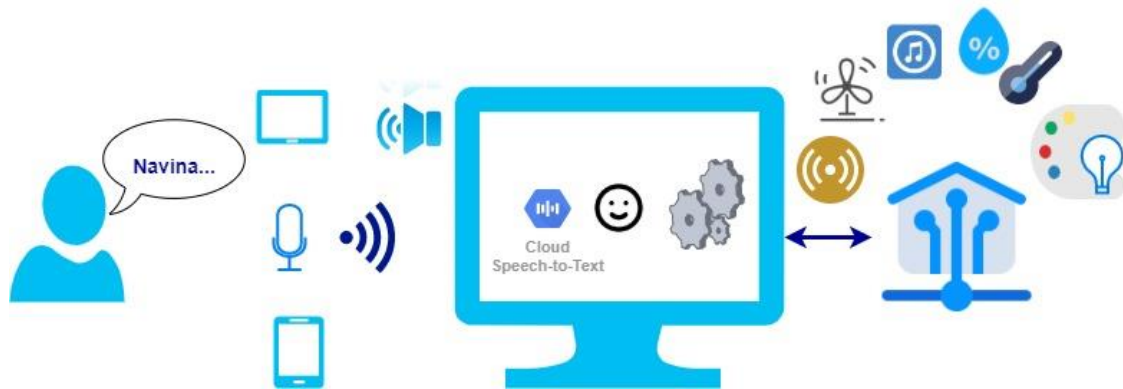
A 2. ábrán a Node-RED folyamat ábrája látható, ami egy nyílt forráskódú FBP (folyamat alapú programozás) grafikus fejlesztő eszköz, amely segít gyorsan létrehozni alkalmazásokat. Böngésző alapú szerkesztővel rendelkezik, ami megkönnyíti a folyamatok összekapcsolását. Széles választékú node-jaival, amelyek egyetlen kattintással hasznosíthatóak, könnyen és logikusan fejleszthetünk.



2. ábra. A kísérleti fejlesztés Node-RED folyamat ábrája

4. A tervezett rendszer felépítése

A szerzett tapasztalatok és véleményezések alapján a rendszer újbóli átgondolása a 3. ábrán szemléltetett felépítési elképzelést eredményezte.



3. ábra. A tervezett rendszer felépítése

Célunk egy olyan komplett vezérlés létrehozása lett, ami képes a parancsokat értelmezni és a mögöttes érzelmeket is detektálni, majd a megfelelő paramétereket a megszabott értékekre állítani, azaz közvetlenül hozzáfér a ház automatizálási rendszeréhez. Ezek alapján a komplett rendszer a központi szerveren fog futni, amihez vezeték nélküli mikrofont fogunk csatlakoztatni, illetve mobil eszközökről is elindítható lesz a hangvezérlés. Mobil eszközök esetén a program webes felületén keresztül nyomógomb segítségével lesz elérhető a hangvezérlési opció. Elsősorban a szerver géphez csatlakoztatott mikrofont tekintjük fő egységnek. A szerver géphez csatlakoztatott mikrofon folyamatosan figyelő állapotban lesz és csak egy adott kifejezés – esetünkben a Navina, ami a rendszer neve – felismerése aktivizálja a rendszert és dolgozza fel az utána következő kifejezést, azaz a parancsot. A parancsot

feldolgozva, illetve a parancs érzelmi töltetét analizálva végrehajtja az adott utasítást, majd visszajelez nyugtázva ezzel a feladatot, illetve indokolt esetben rákérdez az adott személy érzelmi állapotára, és ha a felhasználó valóban egyetért azzal, hogy a rendszer a hangulatának megfelelő paramétereket állítson be a házban, akkor a program azt is végrehajtja közvetlen beavatkozással a ház automatizálási rendszerébe. A Loxone mini szerver és a szerveren futó Node-RED tartalmaz egy UDP (User Datagram Protocol) szervert. Így lehetővé válik, hogy pont-pont kommunikációt folytasson UDP-n keresztül a két eszköz. A Node-RED UDP-n keresztül küldi el a LED kivilágítás színeit. A Loxone UDP-n keresztül küldi el a Philips Hue lámpák állapotát.

A rendszer nem csak szóbeli utasítások végrehajtására lesz alkalmas, hanem értékek lekérdezésére és megjelenítésére is. Összességében elmondhatjuk, hogy egy olyan kiegészítő interfészt hozunk létre a ház Loxone rendszeréhez csatolva, ami nem csak a kommunikációs csatornák számát bővíti, hanem a funkciók tárházát is.

4.1 A tervezett interfész felépítése

Az előzőleg ismertetett funkciók alapján a 4. ábrán szereplő elemekből épül fel a tervezett interfész, amely elemek különböző csoportosításából adódóan 4 fő modulját különböztetjük meg:

- érzelem felismerő modul;
- parancsértelmező és végrehajtó modul;
- zene lejátszó modul;
- megjelenítő modul.

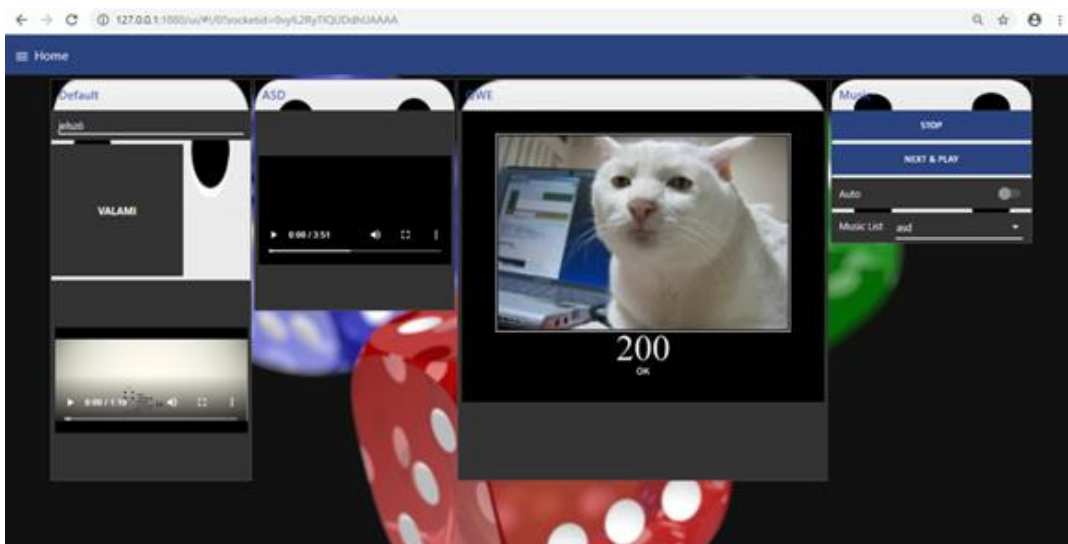


4. ábra. Az interfész elemei

4.2 Fejlesztéshez alkalmazott rendszerek

A fejlesztés elvégzéséhez meg kellett ismerkednünk a Node-RED fejlesztői környezettel, a különböző integrált modulokkal és az érzelemfelismerő keretrendszerrel is. Az alábbi pontok ezen elemek rövid ismertetőjét tartalmazzák.

- **Node-RED:** egy nyílt forráskódú FBP grafikus fejlesztő eszköz, amely segít gyorsan létrehozni alkalmazásokat, különösen esemény vezérelteket, mint például az IoT alkalmazások. Bárhol telepíthető ahol node.js futtatható, beleértve az IoT eszközöket, például Raspberry Pi, BeagleBone vagy Intel Edison. Az IBM fejlesztette ki a hardver eszközök, API-k és online szolgáltatások összekapcsolásához új és érdekes módon a javascript programozási nyelv alkalmazásával. A számítógépes programozásban a folyamat alapú programozás (FBP) egy olyan programozási paradigma, amely az alkalmazásokat „fekete doboz” folyamatok hálózataként határozza meg, amelyek üzenetek átadásakor előre meghatározott kapcsolatok között cserélnek adatokat, ahol a kapcsolatok a folyamatokhoz külsőleg vannak megadva. Ezek a fekete dobozos folyamatok végtelenül újracsatlakoztathatóak különböző alkalmazások létrehozásához anélkül, hogy azokat belsőleg meg kellene változtatni. Az FBP tehát természetesen komponens-orientált [4]. Bőngésző alapú szerkesztővel rendelkezik, ami megkönnyíti a folyamatok összekapcsolását a paletta széles választékú node-jaival, amelyek egyetlen kattintással hasznosíthatóak. Továbbá, ha nincs számunkra megfelelő node, akkor van paletta menedzselési opció, ahol tölthetünk le közösség által létrehozott node-okat vagy saját magunk is készíthetünk. Ilyenre példa a zene lejátszásra alkalmazott node, ami nem tartozik az alapvető node-ok közé, de már többet is megírtak és könnyen letölthetünk egy nekünk megfelelőt. Grafikus felület is könnyen készíthető a Node-Red Dashboard kiegészítő modul segítségével. Ez egy opcionális modul, nem tartozik az alapértelmezett verzióhoz. A Dashboard rendkívül széles lehetőséget kínál testreszabás szempontjából, ahogy azt a 5. ábrán is láthatjuk. Például mérő eszközök által kapott adatok kiírása egy diagramra. Ráadásul azt megoldani, hogy egy programot elég legyen csak grafikusan kezelni se túl nehéz. A Node-RED-ben létrehozott folyamatok JSON fájlokba kerülnek tárolásra.



5. ábra. Példa kezelőfelület létrehozására

- **Watson Speech to Text:** egy olyan szolgáltatás, ami az emberi hangot átalakítja át írott szóvá. Bárhol használható, ahol szükséges lehet a beszéd és az írás áthidalására, ideértve a beágyazott rendszerek hangvezérlését, az ülések és konferenciahívások hangfelvételét, valamint az e-mailek és a jegyzetek diktálását. Ez a könnyen használható szolgáltatás mesterséges intelligenciát használ a nyelvtan és a nyelv felépítésére vonatkozó információk összekapcsolásához az audio jel összetételével kapcsolatos ismeretekkel, hogy pontos hangfelvételt kapjunk. A Watson Speech to Text egy felhő-natív megoldás, amely mélyen tanuló mesterséges intelligencia algoritmusokat alkalmaz [5]. A piacon több megoldás is létezik a beszéd írott szöveggé való átalakításához, melyeket előszeretettel alkalmaznak különböző applikációkban (Google Cloud Speech-to-Text, Amazon AWS Text-to-Speech Service, Speechmatics, Microsoft Cognitive Service).
- **EmoVoice:** egy olyan eszközkészlet, amely lehetővé teszi a valós idejű érzelmi felismerők felépítését a beszéd akusztikus tulajdonságai alapján (nem szóinformációk felhasználásával). Azon akusztikai paraméterek megfelelő megválasztása, amelyek az emberi beszéd érzelmi tartalmát jól jellemzik, szintén kritikus fontosságúak [6,7]. A prozódia elemeinek megvalósulása inkább akaratfüggő, kultúránként eltérő. Ezekkel színesítjük a hangzást, érzékeltetjük mondanivalónk lényegi elemeit. A hangsúlyváltással, dallamvezetéssel, tempóváltással megkönnyítjük az értelmezést. Az EmoVoice-t kezdetben Thurid Vogt keltette életre, és a "Valós idejű érzelmek automatikus felismerése a beszédből" című disszertációja részletesen leírja a technológiát. A jelenlegi megvalósítás továbbra is támogatja az értekezésben ismertetett eredeti szolgáltatáskészletet, bár most már át lehet váltani más szolgáltatáskészletre is. Különösen az openSMILE által nyújtott szolgáltatás használható. Hasonlóképpen, a még széles körben alkalmazott LibSVM osztályozó mellett további osztályozási modelleket is lehet tesztelni az eszközkészlettel [8,9,10].

5. Tervezett vizsgálatok és fejlesztési lehetőségek

Összegezve a tervezett munkálatokat, az alábbi fejlesztések és vizsgálatok kerültek meghatározásra a rendszer fejlesztése során. A rendszer folyamatosan naplózni fogja az eseményeket és a parancsokat, így válik visszakövethetővé, hogy az előzetesen definiált értékek és reakciók mennyire voltak jól időzítve és jól megválasztva. Az alábbi lista tételesen tartalmazza a jelenleg folyó fejlesztéseket.

- Hatékonyságnövelés az érzelemfelismerő modulon: paraméterek és adatbázisok tesztelése.
- A felismerhető érzelmek számának bővítése.
- Az érzelemfelismerés feltételrendszerének kidolgozása és tesztelése.
- A megválasztott paraméterek validálása egyes érzelmeknél.
- A grafikusan megjelenítendő paraméterek bővítése.
- Zajcsökkentő eljárások szükség esetén való integrálása és hatékonyságvizsgálata.
- A rendszer integrálhatóságának vizsgálata.
- A rendszer által értelmezhető parancsok bővítése és rugalmasabbá tétele.

További hosszútávú terveink között szerepel, hogy a rendszert egy kamera modullal is kiegészítsük, ami nem csak az adott személy felismerésére lesz alkalmas, hanem ugyanúgy az érzelmek felismerésére, ezzel alkotva egy összetettebb érzelemfelismerő rendszert. Másik fejlesztési elképzelésünk, hogy a rendszer grafikus megjelenítő részét egy alaprajz alapú megjelenítéssel egészítsük ki, amin a házhoz kapcsolódó összes információ megjelenik nem csak numerikusan, hanem grafikusan is, valamint szöveges üzenetekkel is tudjon a felhasználónak jelezni, ha szokatlan értékeket mér a rendszer vagy tanácsadó jelleggel energiahatékonyságot szem előtt tartó tippeket adjon a felhasználónak. Nagyon távoli célként pedig a magyar nyelvvel való bővítést tűztük ki célul.

6. Összefoglalás

Az ismertetett eddig elért eredmények és tervezett fejlesztések és vizsgálatok távlati célja, hogy a különböző fejlesztéshez alkalmazott keretrendszerekkel egy angol nyelven alkalmazható robusztus rendszert hozzunk létre, aminek segítségével nem csak infokommunikációs akadálymentesítés valósítható meg, hanem olyan extra kényelmi funkciók is, amiket a Loxone okos otthon automatizálási rendszer önmagában nem tud biztosítani. A rendszer másik nagy előnye, hogy webes grafikus felülettel rendelkezik, és nem igényli speciális eszközök beszerzését. Tervezett munkánk szorosan összefügg egyéb K+F tevékenységekkel az Automatizálási és Infokommunikációs Intézetben belül, ezért szélesebb látószögben tekintve a fejlesztésünk, egy összetett szenzoros vezérlő és tanácsadó rendszer részét is fogja képezni, amik szintén kényelmi funkciókat is be fognak tölteni az energiahatékonyság növelése mellett.

7. Köszönetnyilvánítás

A cikkben ismertetett kutató munka az EFOP-3.6.1-16-2016-00011 jelű „Fiatalodó és Megújuló Egyetem – Innovatív Tudásváros – a Miskolci Egyetem intelligens szakosodást szolgáló intézményi fejlesztése” projekt részeként – a Széchenyi 2020 keretében – az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Irodalom

- [1] https://en.wikipedia.org/wiki/Amazon_Alexa
- [2] Varga, Á., Pintér, J. M.: *Intelligens vezérlő rendszer fejlesztése komfortérzet konfigurálására*, Villamosmérnöki Tudományok 1 : 2 pp. 166-170., 5 p. (2018)
- [3] Pintér, J. M., Czap, L., L. Kiss, M.: *Development of speech-based interface for smart home systems*, Proceedings of the 19th International Carpathian Control Conference (ICCC 2018) IEEE, (2018) pp. 348-352. <https://doi.org/10.1109/CarpathianCC.2018.8399654>
- [4] <https://nodered.org/>
- [5] <https://www.ibm.com/cloud/watson-speech-to-text>
- [6] Álvarez, A., Cearreta, I., López, J. M., Arruti, A., Lazkano, E., Sierra, B., Garay, N.: *Application of feature subset selection based on evolutionary algorithms for automatic emotion recognition in speech*, ISCA Archive <http://www.iscaspeech.org/archive>, 2000.
- [7] Seppänen, T., Vayrynen, E., Toivanen, J.: *Rosody-based classification of emotions in spoken Finnish*, in, Eurospeech-2003, 2003.
- [8] <https://rawgit.com/hcmlab/emovoice/master/docs/index.html>
- [9] Wagner, J., Lingenfeller, F., Baur, T., Damian, I., Kistler, F., André, E.: 2013. *The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time*, In Proceedings of the 21st ACM international conference on Multimedia (MM '13). ACM, New York, NY, USA, pp. 831-834 <https://doi.org/10.1145/2502081.2502223>
- [10] Scherer, K. R., Banse, R., Wallbott, H. G.: *Emotion inferences from vocal expression correlate across languages and cultures*, Journal of Cross-Cultural Psychology, 32(1), pp. 76-92, 2001. <https://doi.org/10.1177/0022022101032001009>