

FÜGGŐSÍNPÁLYA HANGVEZÉRLÉSÉNEK KIFEJLESZTÉSE

DEVELOPMENT OF A VOICE-COMMANDED INTERFACE FOR A MONORAIL

Pintér Judit Mária, Trohák Attila***

ABSTRACT

In our paper we introduce the steps and results of our research where our aim is to develop a system which is able to operate a monorail with voice commands.

1. BEVEZETÉS

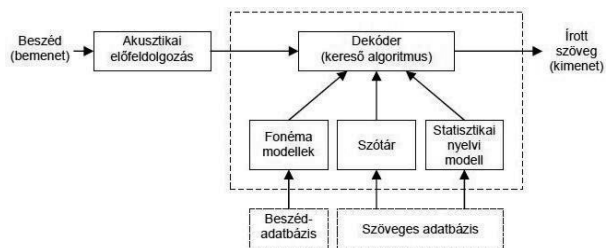
A cikkünkben bemutatásra kerülnek egy függősínpálya hangvezérlésének kifejlesztésére irányuló kutatás lépései, eredményei, melyben a célunk az, hogy szóbeli utasításokkal lehessen a rendszert működtetni. A cikkben bemutatásra kerülnek a számítógépes beszédfelismerők, azok általános felépítése és működése, majd a feladat ellátására alkalmasnak talált beszédfelismerő jellemzői, valamint a fejlesztés és a rendszerintegráció lépései.

2. A SZÁMÍTÓGÉPES BESZÉDFELISMERŐK ÁLTALÁNOS FELÉPÍTÉSE ÉS MŰKÖDÉSE

A számítógépes beszédfelismerők nem túl részletes működési vázlatát láthatjuk az 1. ábrán, amely fonémák modelljére épülő statisztikai alapú beszédfelismerő rendszert mutat be. A példa alkalmas annak bemutatására, hogy milyen működési elven alapszik a vezérléshez felhasznált beszédfelismerő. A bemenetre kapott beszédjelet előzetesen egy akusztikai feldolgozásnak vetjük alá, amelyet szokás lényegkiemelésnek is nevezni. A cél az, hogy olyan, az egyes beszédhangokat jól elkülönítő jellemzőket tartsunk meg, amelyek a beszédet a lehető legtömörebben reprezentálják lényegesebb információvesztés nélkül. A tömörítés szükségessége az eredendően hatalmas számítási igény kordában tartása végett is felmerül, a legtöbb alkalmazáshoz kívánatos ugyanis, hogy a beszédfelismerő valós időben is működőképes legyen (online üzemmód).

Ezután következhet a dekódolás folyamata. A művelet során a felismerés alapegységeinek modelljeit, a szótárt és a nyelv szintaktikai viszonyait statisztikai alapokon leíró nyelvi modellt használhatjuk fel. A Markov-modell, a szótár és a nyelvi modell együttesen egyfajta tudást visz a rendszerbe, melyet

számítógépes betanítás (training) során alapozhatunk meg. Ehhez beszéd, illetve szöveges adatbázis egyaránt szükséges. A beszédatadabázisnak a nyelvben előforduló valamennyi építőelemet és kapcsolatot is többszörösen tartalmaznia kell, hogy statisztikailag megfelelő lefedettséget adjon. Hasonlóképpen, a nyelvi modellnek is meg kell felelnie a beszédfelismerő használati területére jellemző szóhasználati szokásoknak.



1. ábra. Statisztikai számítógépes beszédfelismerő működése

3. A BESZÉDFELISMERŐ JELLEMZŐI

A függősínpálya kezelési feladatok megkönnyítésének és hatékonyabbá tételének egyik lehetősége a szóbeli utasításokkal való vezérlés. Bizonyos funkciók előhívása, vagy több ugyanazokból a lépésekből álló utasítások, folyamatok végrehajtása kiváltható lenne egészen rövid kifejezéseket tartalmazó szóbeli paranccsal. Ezeknek a kifejezéseknek természetesen egyértelműnek kell lenniük, azaz tartalmazniuk kell például egy rögzített kifejezést amivel "jelezzük" a rendszer felé, hogy egy utasítást fogunk közölni, nem pedig éppen csak beszélgetünk valaki mással. A "nyitó" kifejezést követően pedig következhet a fő utasítás, például adott funkció megnyitása, vagy értékek módosítása, lekérdezése stb., vagyis adottak lesznek kulcsszavak, amikkel egyértelműsíthető lesz a parancs. Bizonyos esetekben pedig biztonsági kérdésekkel (amiket a rendszer intéz a felhasználó felé pl. biztonsági kód kérése) láthatjuk el az egyes folyamatok elindítását vagy lezárását ezzel növelve a biztonságot.

* MSc. mérnök informatikus hallgató, Miskolci Egyetem, Automatizálási és Kommunikáció-technológiai Tanszék

** Egyetemi tanársegéd, Miskolci Egyetem, Automatizálási és Kommunikáció-technológiai Tanszék

Figyelembe kell venni azt is, hogy a felületet nem csak egy adott személy fogja nagy valószínűséggel használni, ezért a hangvezérlésnek ugyanolyan jó hatékonysággal kell majd működnie, minden egyes ember esetén. Ez két féle módon valósítható meg. Létrehozhatunk egy beszélőfüggetlen felismerőt, aminek minden személy esetén ugyanolyan magas hatékonysággal kell működnie, vagy minden felhasználó esetén külön tanítási folyamatot hajtunk végre és profilszerűen az adott személyhez kapcsolódó beszéd-modelleket (rejtett Markov-modellek) alkalmazunk a beszédfelismerőben és attól függően, hogy ki az aktuális felhasználó, úgy töltődik be az megfelelő konfiguráció. A beszélőfüggetlen megoldáshoz képest nagyobb biztonságot és könnyebb megvalósíthatóságot eredményez a felhasználónként történő betanítás. A munka során elsősorban beszélőfüggetlen rendszert szeretnénk létrehozni. Ehhez nagy számú tanítóanyagra van szükség.

Triviálisnak tűnő elvárás, de mégis meg kell jegyeznünk, hogy a hangvezérlés lehetőségének folyamatosan adottnak kell lennie és alkalmazkodnia kell a folyamatos emberi beszédhez.

Az igények és a kritériumok áttekintése után összegezhethetjük, hogy a navigáció megvalósításához pontosan milyen beszédfelismerőre is van szükség:

- kötött szótáras (kulcsszó alapú felismerés);
- mintafelismerő (rejtett Markov-modellek alkalmazása);
- kapcsolt szavak felismerésére alkalmas;
- beszélőfüggetlő;
- zajos környezetben alkalmazható;
- parancs üzemmódú;
- online üzemmódú.

4. A HANGVEZÉRLÉS FEJLESZTÉSÉNEK LÉPÉSEI

A projekt munka során, - amit az Anyagmozgatási és Logisztikai tanszékkel együtt valósítunk meg – az első lépés a kulcsszavak meghatározása volt. Mivel a függősín-pályák olyan fejeletti, kötöttpályás szakaszos működésű anyagmozgató rendszerek, amelyek általában több, tipizált, önálló villamos hajtással rendelkező szállítóegységből (kocsiból), függősín-ekből és tetszőleges mértékig automatizált vezérlésből állnak, ezért azok koordinálása is nagy mértékben korlátozott. Használhatjuk például az előre/hátra és fel/le parancsokat kiegészítve, hogy milyen tempóban mozogjon, mivel a függősín-pályán a kocsi két eltérő sebességgel is tud haladni. A pályán a kocsi két helyen tudjuk pozicionálni, a váltóban és az állomásnál, ezért ezeket a szavakat is belevettük a szókészletbe. Az Anyagmozgatási és Logisztikai Tanszék jövőbeli tervei közt szerepel, hogy koordináták megadásával is tudják pozicionálni a rendszert, ezért a számjegyek is a szókészlet részét képezik. Külön kérésre a vészleállítás

parancsot is felvettük a kulcsszavak közé. Így összesen 26 db kulcsszót határoztunk meg.

A következő lépés a hangminták begyűjtése volt. A hallgatók segítőkészségében bízva megpróbáltunk a Sphinx beszédfelismerő szoftver - amit majd a hangvezérlés megvalósításához fogunk használni – honlapján ajánlott betanító adatbázisához a szükséges mennyiségű hangmintára szert tenni. Az ajánlás szerint beszélőfüggetlen, vezérlő és irányító (command and control), kis szótáras beszédfelismerőkhöz összesen 5 órányi hanganyagot ír elő 200 beszélőtől lehetőleg egyenlő eloszlásban kell származnia nemekre és korcsoportokra nézve. Eddig összesen 56 embertől sikerült hangmintát összegyűjtenünk, durván a negyede az ajánlásnak de a tesztelesek megkezdéséhez elegendőnek bizonyult. A hangminták 22050 Hz-en lettek WAV formátumban (előjeles 16 bites PCM). rögzítve. (A hallás frekvenciatartományát 20 Hz és 20 kHz közé teszik a technológiai tervezés során. Például a legtöbb jó minőségű hangszóró ebben a sávban sugároz, a Hi-Fi minőségű hanganyagot 44100 Hz-en mintavételezik, ez a Shannon-féle mintavételezési törvény szerint 22050 Hz-es sáv szélességet jelent. 20 kHz-nél nagyobb sáv szélességet – pl. 96 kHz-es mintavételezés – általában akkor használnak, ha további jelfeldolgozási feladatokat kívánnak elvégezni.) [2] A hangmintákat a betanítás előtt rendszerezni és ellenőrizni kellett és a rögzítés során használt kérdőívek alapján pedig megállapítható, hogy a minták nemekre vonatkozó eloszlása megfelelő, korcsoport tekintetében (mivel hallgatóktól származnak a minták többségében) pedig szinte az összes minta 20 és 25 év közötti személyektől származik. A betanítás megkezdése előtt a hangmintákon lényegkiemelést (MFCC konverziót) kellett végrehajtani, amit a HTK toolkit szoftvercsomag HCopy programjával végeztük el.

A betanításhoz szükséges még:

- a szótár, ami a kulcsszavakat tartalmazza némileg módosult formában, melynek oka, hogy a HTK toolkit szoftver, (amin a betanítást és a teszteleket a jövőben végezzük) nem értelmezi az ékezetes karaktereket, ezért némely esetben a SAMPA* szabványt követtem,
- nyelvtan fájl, ami a parancsok szintaktikáját tartalmazza. A jellemzőkben megállapítottuk, hogy a biztonságosabb működéshez szükséges lenne egy "kezdő" kifejezés használata a parancsoknál. Így a nyelvtan fájlban meghatároztuk, hogy minden kifejezésnek a start kulcsszóval kell kezdődnie (kivéve stop parancs),
- a betanítandó rejtett Markov- modell prototípusa. A modellek különböző állapotszámúak lehetnek, hogy melyik lesz a leoptimálisabb, a tesztelesek során fog kiderülni.

A felsorolás csak a legfontosabb tanító elemeket tartalmazza, ezen felül számos más összetevő is szükséges.

* a nemzetközi fonetikai ábécé ASCII átírása a SAMPA

Miután ezek a tanításhoz szükséges fájlok rendelkezésre állnak, a továbbiakban elvégezzük a betanítást, tesztlesekkel megpróbáljuk a lehető legjobb hatékonyságot elérni, majd a „végleges” konfiguráción zajteszteket végrehajtani, hogy a rendszer mennyire érzékeny a különböző zajokra. Mivel logisztikai rendszerhez fejlesztünk hangvezérlést, ezért annak ipari környezetben is – ahol általában egyszerre több zajforrás is fellelhető - magas hatékonysággal kell működnie.

5. A FELISMERŐ INTEGRÁCIÓJA

A vezérlés megvalósításához a beszéd felismerő által felismert utasításokat valamilyen módon továbbítani kell a függősín-pálya vezérlőegysége felé, ahol adott folyamatok végrehajtását eredményezi. Ennek megfelelően egyértelműen meg kell határozni, hogy melyik kulcsszó melyik folyamat végrehajtásának felel meg és biztosítani kell a valós idejű végrehajtást.

Első lépésként a függősín-pálya jelenlegi vezérlését mértük fel. Megállapítottuk, hogy a jelenlegi vezérlés egy nagy, összetett rendszer vezérlő PLC-jén kapott helyet. Ez egy Mitsubishi PLC QX81 típusú digitális bemeneti és QY81P típusú digitális kimeneti modulokkal. Ezek a modulok 24 VDC jelszinten működnek. Mivel a beszéd felismerő egy számítógépen futó szoftver, ezért a felismerő által detektált parancsot be kell juttatni valamilyen módon a függősín-pálya vezérlő rendszerébe. A komplex irányító rendszernek ha a digitális bemeneti csatornáira juttatjuk a jelet, akkor a PLC programjában is módosításokat kell végrehajtani, ami egy több száz jelet kezelő, rendszeresen használt berendezés esetében nem kockázatmentes, ezért a vezérlő rendszer ideiglenes helyettesítése mellett döntöttünk.

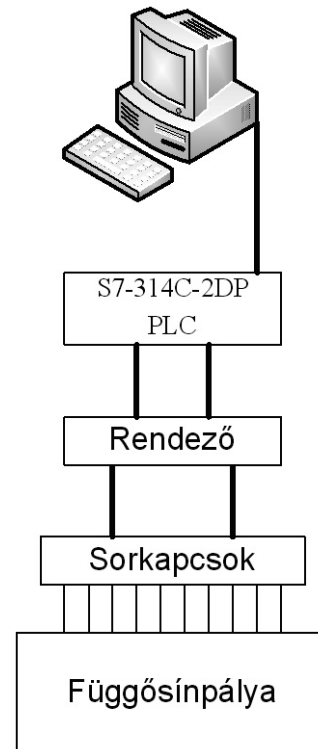
A technológia jelei egy sorkapcsos rendszert követően gyűjtőkábelek segítségével csatlakoznak a vezérlő rendszerre. Az IO modulok „apa” csatlakozóval vannak ellátva, míg a gyűjtőkábel végén „anya” csatlakozó található. A megoldásunk az, hogy egy új PLC-t csatlakoztatunk a technológiára úgy, hogy a PLC IO jeleinek vezetékét „apa” csatlakozó végével látjuk el és így a technológia jeleivel a gyűjtőkábel „anya” csatlakozójával szembe tudjuk kötni. Így a fejlesztés során a komplex technológiai rendszerről másodpercek alatt le lehet választani a függősín-pályát.

Az új PLC egy Siemens S7-314C-2DP típusú kompakt vezérlőberendezés, melyre az alábbi IO jelek kerülnek csatlakoztatásra:

- Bemenetek:
 - futómacska működik;
 - forgató végállás kapcsoló keresztben;
 - forgató végállás kapcsoló hosszában;
 - forgató, macska nincs bent;
 - forgató, ready lámpa;

- forgató, vészstop gomb benyomva.
- Kimenetek:
 - keresztbe forgatás Kone forgató;
 - hosszába forgatás Kone forgató;
 - fel;
 - le;
 - emelés gyors;
 - macska jobbra/előre;
 - macska balra/vissza;
 - futómacska gyors;
 - kürt.

A számítógépes szoftver kimenetének és a PLC-nek az összekötési problémája továbbra is megmaradt, de ezzel a rendszer kialakítással ez már könnyedén áthidalhatóvá válik. Amennyiben a PLC digitális bemeneteire szeretnénk csatlakozni egy számítógépről, akkor a számítógépet el kell látni egy olyan modullal, ami képes 24 VDC jelet küldeni a PLC irányába. Ezt a fajta bővítést a költséghatékony rendszer kifejlesztése érdekében elvetettük. Egy másik lehetséges megoldás a PLC bővítése kommunikációs kártyával, például egy soros kártyával és ekkor a számítógép soros portján küldjük ki a detektálás után az utasításokat. Valós üzemi körülmények között ez a megoldás is igen költséges lehet az irányítórendszerek bővítő moduljainak magas ára miatt.



2. ábra. A vezérlés kialakítása

A választott megoldásunk végül az lett, hogy kihasználjuk azt, hogy az ilyen irányítórendszerek operátori kezelőfelülettel ellátottak, azaz egy

számítógépen futó SCADA/HMI alkalmazás az operátoroknak lehetővé teszi a technológia állapotainak megjelenítését és beavatkozási, irányítási lehetőséget is biztosít számukra.

A választott PLC-hez egy PC alapú HMI felületet készítettünk WinCC Flexible Runtime segítségével és a klasszikus megjelenítő felületről működtethetővé tesszük a függősínpálya rendszert. Ebben az esetben a HMI felület képes a PLC változóit írni és olvasni, tehát valamilyen módon azt kell megoldanunk, hogy a PLC változójának írása ne egy kezelőfelületen elhelyezett gomb megnyomására történjen meg, hanem egy szóbeli utasítás hatására. A megoldás kulcsa az irányítástechnikai informatikai rendszerek esetében igen elterjedt OPC technológia lehet. Az OLE for Process Control (OPC) a windows alkalmazások és a termelési folyamatok irányítását végző hardvereszközök közti kapcsolatteremtés eszköze szerver/kliens formában. [5.] Nyitott szabvány, amely a termelés folyamatát irányító eszközök adatainak állandó, torzítatlan elérését célozza meg. Módszere független az adat típusától, illetve az adat forrásától. A HMI felületen nem csak a PLC irányába lehet adatcserére irányuló kapcsolatot konfigurálni, hanem egy OPC szerver felé is.

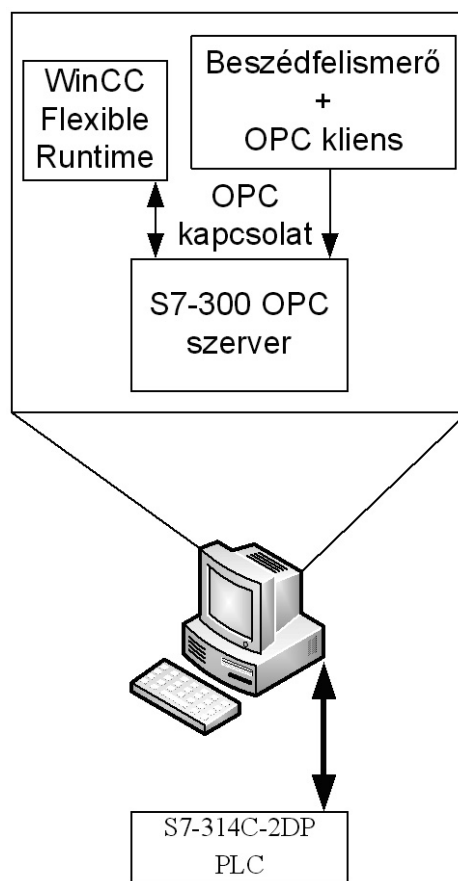
Tehát a cél az, hogy a beszéd felismerő szoftvert a HMI felületet futató PC-n futtatva egy OPC szerverbe juttatott adat segítségével a HMI felületen át hanggal vezérelhessük a függősínpályát. Ezt a legegyszerűbben úgy valósíthatjuk meg, ha veszünk egy adott változót és annak értékét annak megfelelően módosítjuk, hogy milyen kulcsszó hangzik el. Ez persze nem jelenthet sok értéket, de a mi esetünkben nincs is szükség rá (kötött szótár felismerő). Így az egyes kulcsszavakhoz egy adott értéket rendelünk, amit a változó vesz fel, ha elhangzik az adott utasítás. A HMI rendszernek ezt az értéket folyamatosan figyelnie kell és változása esetén az adott értéknek megfelelő kiváltott eseményt végrehajtani.

6. ÖSSZEFOGLALÁS

A cikkünkben bemutatásra kerültek egy függősínpálya hangvezérlésének kifejlesztésére irányuló kutatás lépései, eredményei, melyben a célunk az volt, hogy szóbeli utasításokkal lehessen a rendszert működtetni. A kutatás jelenleg a beszéd felismerő modul betanításánál tart és tesztelésénél tart, valamint az új PLC vezérlőprogram és a HMI felület készítése zajlik. Az OPC alapú egyedi kliensek és kapcsolatok kialakítása a korábbi tapasztalataink alapján nem fog problémát okozni. [3., 4.]

7. KÖSZÖNETNYILVÁNÍTÁS

A bemutatott kutató munka a TÁMOP-4.2.1.B-10/2/KONV-2010-0001 jelű projekt részeként az



3. ábra. Szoftveres integráció

Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.

8. IRODALOM

- [1.] VICSI K., NÉMETH G., SZASZÁK GY. A beszéd számítógépes feldolgozása
- [2.] NÉMETH G., OLASZY G.: A Magyar Beszéd, Akadémiai Kiadó, Budapest, 2010
- [3.] TROHÁK A.: Erőművi terhelés-elosztási stratégiák és megoldások kutatása és modellezése a hatásfoknövelés és a rugalmas terhelésváltoztatás céljából, Doktoranduszok Fóruma 2009, Gépészmérnöki és Informatikai Kar szekciókiadványa, pp. 176-181
- [4.] AJTONYI I., TROHÁK A.: Intelligens irányítórendszer tervezése és modellezése golyós őrlőmalomhoz, Doktoranduszok Fóruma 2008, Gépészmérnöki és Informatikai Kar szekciókiadványa, pp. 93-99
- [5.] IWANITZ F., LANGE J.: OPC Fundamentals, Implementation and Application. Heidelberg, Hüthig GmbH, 2005.