FIATAL MŰSZAKIAK TUDOMÁNYOS ÜLÉSSZAKA XVII.

Kolozsvár, 2012. március 22–23.

HANGVEZÉRELT OPERÁTORI KEZELŐFELÜLET FEJLESZTÉSE IRÁNYÍTÓ RENDSZEREKHEZ

PINTÉR Judit Mária, TROHÁK Attila

Abstract

In our paper we introduce a part of our research where we develop a voice-commanded operator surface which can help to improve the efficiency of the operator's work. We intend to integrate the voice recognition module to the control system via OPC connection. Now we will describe the speech recognition module in details.

Key words:

voice control, control systems, operator surface

Összefoglalás

A cikkünkben egy kutatásunk egyik részéről számolunk be. A célunk egy olyan hangvezérelt operátori kezelőfelület létrehozása, mely az operátorok munkáját képes lehet hatékonyabbá tenni. A hangfelismerő modul irányítórendszerhez történő integrációját OPC-n keresztül kívánjuk megvalósítani. Most a beszédfelismerő modult ismertetjük részletesebben.

Kulcsszavak:

hangvezérlés, irányító rendszer, operátori kezelőfelület

1. Bevezetés

A beszéd nem más, mint akusztikus hullámok keltése, azaz beszédhangok, fonémák (hangok olyan elemi, elvont egysége, amely szavakat különböztet meg egymástól, önálló jelentéssel nem rendelkezik) kibocsátása. A beszédfelismerés az a folyamat, melynek során a beszédfelismerő gép azonosítja a kiejtett beszédjeleket és átalakítja ezeket szöveggé, vagy más, számítógép által feldolgozható adattá. A beszédfelismerés meglehetősen tág témakör. Szűkebb értelemben a tartalom felismerését értjük alatta, tágabb értelemben azonban alkalmazások egész sora használ egészében vagy komponenseként beszédfelismerőt [1]. Több különböző beszédfelismerési feladatokat különböztetünk meg [2]:

- Beszéd-szöveg átalakítás (speech recognition);
- Beszédőfelismerés: Beszélő személy azonosítása (speaker verification);

Felhasználó felismerése, kiválasztása egy előre definiált halmazból
(*speaker identification*);

- Beszéddetekció;

- Kulcsszókeresés;
- Érzelmi töltet felismerése.

A különböző szempontokat figyelembe véve az alábbi osztályokba sorolhatók a beszédfelismerő szoftverek (a felsorolás nem teljes körű, de a főbb osztályokat tartalmazza) [2]:

- A szótár nagysága szerint: kis (kötött, zárt) szótáras, közép szótáras, nagy szótáras;
- A megvalósítás módja szerint: (milyen módon kezeli a bemeneti hangadatot, illetve, hogy milyen algoritmusokat használ) mintafelismerő, akusztikus-fonetikus;
- Beszédmód szerinti osztályozás: izolált szavas, kapcsolt szavas, folyamatos beszéd;
- Beszélőre adaptáltság szempontjából: beszélőfüggetlen, beszélőfüggő;
- Akusztikai környezet szempontjából való osztályozás: csendes, zajos, telefonbeszéd;
- Üzem mód szempontjából való osztályozás: parancsmód, diktáló üzemmód, más szempontból: online/offline.

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.

1. ábra. Statisztikai számítógépes beszédfelismerő működése

A példa alkalmas annak bemutatására, hogy milyen működési elven alapszik a vezérléshez felhasznált beszédfelismerő. A bemenetre kapott beszédjelet előzetesen egy akusztikai feldolgozásnak vetjük alá (lényegkiemelés). A célunk hogy olyan, az egyes beszédhangokat jól elkülönítő jellemzőket tartsunk meg, amelyek a beszédet a lehető legtömörebben reprezentálják lényegesebb információvesztés nélkül. -A tömörítés szükségessége a 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 szótárat és a nyelv szintaktikai viszonyait statisztikai alapokon leíró nyelvi modellt használhatjuk fel. A Markov-modellek, 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 beszédatadbázisnak 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 [3].

3. Az operátori kezelőfelület vezérléshez szükséges beszédfelismerő

A navigáció megkönnyítésének és hatékonyabbá tételének egyik lehetősége a szóbeli utasításokkal való vezérlés. Bizonyos funkciók előhívása, vagy több ugyanabból a lépésekből álló utasítások, kiértékelési 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, képernyőváltás, berendezés kiválasztása, 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. Figyelembe kell venni azt is, hogy a felületet nem csak egy adott személy fogja nagy valószínűséggel használni, ezért a szóbeli navigációnak ugyanolyan jó hatékonysággal kell majd működnie, minden egyes ember esetén. Ezt két féle módon valósítható meg. Létrehozhatunk egy beszélőfüggetlen felismerőt, ami 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) alkalmazzuk 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. Triviálisnak tűnő elvárás, de megjegyzendő, hogy a szóbeli navigáció 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);
- folyamatos beszéd felismerésére alkalmas;
- beszélőfüggetlő;
- zajos környezetben alkalmazható;

- parancs üzemmódú;
- online üzemmódú.

4. A felismerő integrációja

A vezérlés megvalósításához a felismert utasításokat valamilyen módon továbbítani kell az operátori felület felé, ahol adott folyamatok végrehajtását eredményezi valós időben. Ennek megfelelően egyértelműen meg kell határoznunk, hogy melyik kulcsszó melyik folyamat végrehajtásának felel meg. Ezeket a követelményeket a legegyszerűbben úgy valósíthatjuk meg, ha veszünk egy adott változót és annak értékét a kulcsszavaknak megfelelően változtatjuk (kevés kulcsszó -> kevés érték). Így az egyes kulcsszavakhoz egy adott értéket rendelünk, amit a változó vesz fel, ha elhangzik az adott utasítás. A rendszer folyamatosan figyelve ezt az értéket változás esetén végrehajtja a kiváltott eseményt. Az integráció megvalósítása megoldható OPC kapcsolat segítségével. Az OLE for Process Control (OPC) a windows alkalmazások és a termelési folyamat irányítását végző hardvereszközök közti kapcsolatteremtés eszköze. Nyitott szabvány, amely a termelés folyamatát irányító eszközök adatainak állandó, torzítatlan elérését célozza. Módszere független az adat típusától, illetve az adat forrásától.

5. Összefoglalás

A cikkünkben bemutatásra kerültek a beszédfelismerő modullal szemben támasztott követelmények. A kutatásunk eredményeit a gyakorlatban egy daru vezérléséhez kívánjuk illeszteni és ott kívánjuk tesztelni a kidolgozott módszerünk ipari környezetben történő alkalmazhatóságát.

6. Köszönetnyilvánítás

A bemutatott kutató munka a TÁMOP-4.2.1.B-10/2/KONV-2010-0001 jelű projekt részeként az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.

Irodalomjegyzék

- [1] Pintér Judit Mária: Beszédfelismerés hatékonyságának vizsgálata különböző nyelvtanokkal, TDK dolgozat, 2011, 4;6;
- [2] Németh Géza- Olasz Gábor: A Magyar Beszéd, Akadémiai Kiadó, Budapest, 2010
- [3] Pintér Judit Mária: Akusztikus beszédfelismerés rejtett Markov modellek alkalmazásával, Szakdolgozat, 2010, 12. oldal

Pintér Judit Mária, hallgató

Trohák Attila, egyetemi tanársegéd

Miskolci Egyetem, Gépészmérnöki és Informatikai Kar,
Automatizálási és Kommunikáció-technológiai Tanszék
3513 Hungary, Miskolc, Egyetemváros

Telefon: +36-46-565111

E-mail: pinter.judit.maria@gmail.com, trohak.attila@uni-miskolc.hu