Govorni uporabniški vmesnik v industrijskem okolju

Peter Rogelj, Simon Dobrišek, France Mihelič
Fakulteta za elektrotehniko
Univerza v Ljubljani
Tržaška 25, 1001 Ljubljana, Slovenija
peter.rogelj@fe.uni-lj.si, simon.dobrisek@fe.uni-lj.si, france.mihelic@fe.uni-lj.si

Abstract

We present a speech user interface as a part of a system for recognition of serial numbers of electricity meters, used at final testing in a production process. In case of successful serial number recognition, the system operates automatically, while in contrarily, user intervention is needed to manually adjust the serial number. The speech interface enables the remote adjustment of the unrecognized serial numbers on a distributed production line. A special care was taken on reliability, applicability, and user friendliness of the user interface to meet the requirements of the industrial environment.

1 Uvod

Z razvojem tehnologije strmo narašča tudi uporaba sistemov za razpoznavanje govora. Velike možnosti uporabe razpoznavalnikov govora se kažejo v izdelkih za dom in gospodinstvo, trenutno pa je njihova uporaba najbolj razširjena v mobilni telefoniji. Razpoznavalnike govora lahko koristno uporabimo tudi v bolj zahtevnih aplikacijah in s tem tudi v industriji. Tu se poleg problema razpoznavanja govora pojavi še drug problem, to je njegova učinkovita in za uporabnika prijazna uporaba. Ta problem rešuje govorni uporabniški vmesnik.

Ob načrtovanju govornega vmesnika se moramo najprej vprašati o njegovi upravičenosti. Dejstvo je, da ima govorni vmesnik poleg svojih prednosti tudi slabosti. Kot prednosti govornih vmesnikov bi lahko navedli naslednje njihove lastnosti:

- možnost komunikacije tudi v primeru zasedenosti oči,
- pri komunikaciji roke niso zasedene in jih je moč uporabiti v druge namene,
- manjša navezanost na prostor,
- hitro reagiranje v kritičnih situacijah,
- brezkontaktno ukazovanje, kar je prednost predvsem v onesnaženih okoljih.

Slabosti govornih vmesnikov pa izhajajo predvsem iz težav z zanesljivostjo delovanja in so sledeče:

- v splošnem manjša hitrost komunikacije kot pri grafičnem vmesniku,
- manjša zanesljivost, kar v primeru pomembnih podatkov zahteva dodatno preverjanje pravilnosti vnosa,
- občutljivost na zvoke iz okolice,
- možne so težave pri priučevanju uporabnikov.

V nadaljevanju je predstavljena izvedba govornega uporabniškega vmesnika, namenjenega popravljanju morebitnih napak pri avtomatskem odčitavanju serijskih številk električnih števcev v zadnji fazi proizvodnje, končni kontroli.

2 Predstavitev problema

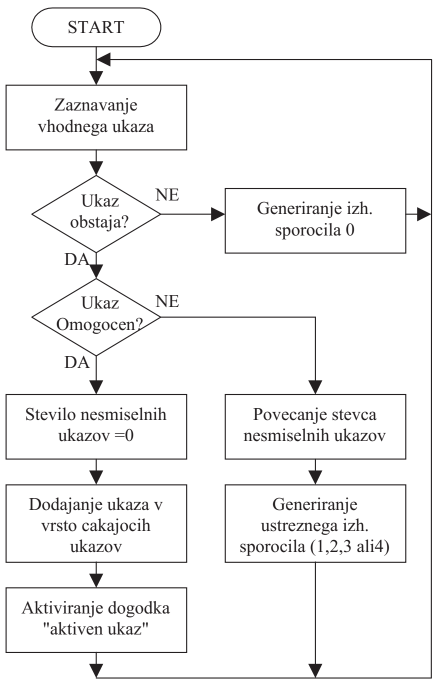
Proizvodni proces električnih števcev se zaključi s fazo končne kontrole kvalitete. V tej fazi se večje število števcev (po 40) postavi na nosilce in zapelje v prostor, kjer se preizkus izvrši. Med procesom preizkušanja, je potrebno prebrati tudi serijske številke. To se izvrši tako, da sistem za industrijsko optično razpoznavanje znakov (industrijski OCR) s kamero, ki se premika od števca do števca, razpozna serijske številke eno za drugo [2], [3]. Če razpoznavanje ni uspešno, mora serijsko številko prebrati operater. Ker za celotno kontrolo električnih števcev skrbi en računalnik, je le ta odmaknjen od večine preizkušanih števcev in vnos preko tipkovnice je otežen. Tu se kot rešitev ponuja uporaba govornega uporabniškega vmesnika. S tem odpravimo problem zaradi odmaknjenosti procesa in operater lahko popravke vnaša kar z mesta procesa (z uporabo daljinskega mikrofona). Dobra lastnost govornega vmesnika je v tem primeru tudi ta, da operater pri komunikaciji nima zasedenih oči, kajti potrebuje jih za branje serijskih številk. Slabost pa je gotovo zanesljivost. Da to pomanjkljivost odpravimo, je uporabniški vmesnik načrtan tako, da vnosu serijske številke vedno sledi še njeno potrjevanje.

3 Izvedba govornega vmesnika

Govorni uporabniški vmesnik zahteva v sistemu za industrijski OCR dva dodatna sklopa. Prvi je dodatna nit v programu OCR-a, tako imenovana *nit govornega vmesnika*, drugi sklop pa je *govorni dialog*. Poleg tega govorni vmesnik zahteva tudi dve dodatni podatkovni strukturi, ki opisujeta možne vhodne ukaze in možna izhodna sporočila.

3.1 Nit govornega vmesnika

Nit govornega vmesnika je del programa, ki deluje neodvisno od sklopov za obdelavo slike, govornega dialoga ter grafičnega uporabniškega vesnika. Namenjena je komunikaciji z razpoznavalnikom govora, preverjanju pravilnosti in omogočenosti vhodnih ukazov ter generiranju izhodnih sporočil, ki se nanašajo na (ne)pravilnost vhodnih ukazov. To so funkcije, ki zahtevajo hitro delovanje, saj je to zelo pomembno za ustrezno uporabniško prijaznost govornega vmesnika. Hitrost je zagotovljena s tem, da nit govornega vmesnika ne vsebuje nobenih časovno zahtevnih funkcij. Njeno delovanje je z ostalimi sklopi programa povezano le preko tabele vhodnih ukazov oziroma v njej vsebovanih podatkov o omogočenosti ukazov. Diagram poteka niti govornega vmesnika je prikazan na sliki 1.



Slika 1: Diagram poteka niti govornega vmesnika.

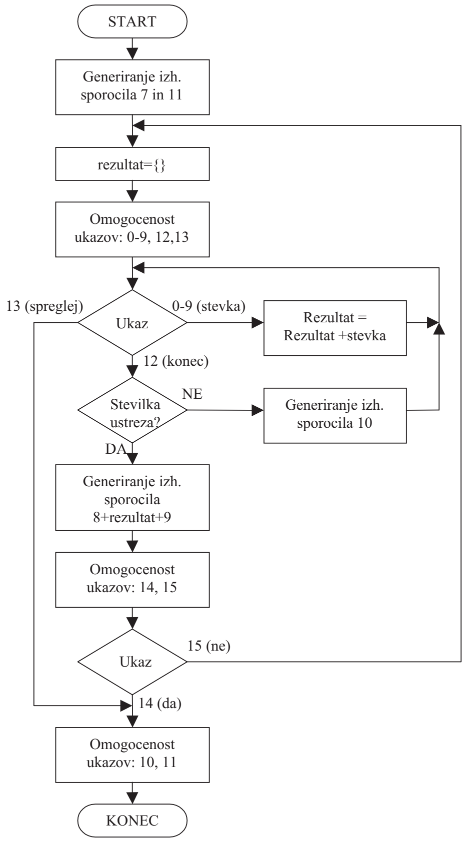
Komunikacija z razpoznavalnikom govora poteka preko TCP/IP protokola. Program za industrijski OCR je TCP/IP strežnik, razpoznavalnik govora pa mora biti realiziran kot ločen program, ki se nanj priključi kot TCP/IP odjemalec, kot na primer v [4]. Zvokovni vhod iz mikrofona razpoznavalnik zvoka razpoznava in rezultate v obliki tekstovnih nizov pošila programu za OCR. Tu nit govornega vmesnika vsak vhodni niz razdeli na posamezne ukaze (besede), ter glede na njihov pomen izvede ustrezno akcijo. Če ukaz nima pomena in ga torej ni v tabeli vhodnih ukazov (tabela 1), se takoj generira zvočno izhodno sporočilo, s katerim se na to opzori. Podobno se opozori, če ukaz v tabeli sicer obstaja, ampak v danem trenutku ni omogočen. V tem primeru se izhodno sporočilo izbere glede na število zaporednih neomogočenih ukazov, njihovo ponavljanje in glede na čas, ki je minil od zadnjega takega ukaza. Vsa možna izhodna so navedena v tabeli izhodnih sporočil (tabela 2). Če je vhodni ukaz omogočen, se iz tabele vhodnih ukazov razbere njegova identifikacijska številka, s katero je ukaz opisan, ter se ga doda v vrsto ukazov, ki čakajo na izvršitev. Poleg tega se aktivira dogodek *aktiven ukaz*, s katerim se od niti razpoznavalnika slike oziroma od govornega dialoga zahteva, da v prvem možnem trenutku strežeta čakajočim ukazom.

3.2 Govorni dialog

Za razliko od niti govornega vmesnika, je govorni dialog tesno povezan s problemom in načrtovanim potekom komunikacije. Za načrovanje govornega dialoga si je najprej potrebno zamisliti potek komunikacije z vsemi možnimi ukazi in odzivi. Definirati je potrebno nabor možnih vhodnih ukazov (tabela 1), ter predvideti možne zvokovne odzive (tabela 2). Za vsako stanje govornega vmesnika moramo predvideti odzive na vse možne ukaze, ki jih v danem trenutku nit govornega vmesnika lahko posreduje govornemu dialogu. Nabor teh ukazov izmed vseh možnih ukazov določimo z ustrezno nastavitvijo parametov omogočenosti ukazov. S tem dosežemo, da v vsakem trenutku govorni dialog lahko sprejme le ukaz, ki ima v danem trenutku tudi logičen pomen. To zelo olajša načrtovanje govornega dialoga, saj namesto določanja odziviv na vse možne ukaze za vsako stanje določimo le nabor možnih ukazov, odzive pa le za omogočene ukaze. Za ostale ukaze poskrbi nit govornega vmesnika.

Za rešitev našega problema smo izbrali skupno 17 ukazov (tabela 1) ter predvideli 13 možnih zvočnih odzivov (tabela 2). Povezave med ukazi, in akcijami, ki jih ti ukazi izzovejo so prikazani na sliki 2.

Izvajanje govornega dialoga se začne po ugotovitvi neuspešnosti razpoznavanja serijske številke z metodami industrijskega OCR-a. Govorni dialog se zato začne z opozorilom na nastalo situacijo. Nato je omogočeno narekovanje serijske številke, kar je v



Slika 2: Diagram poteka govornega dialoga.

bistvu zaporedje ukazov od 0 do 9 (števke) in se zaključi z ukazom 12 (konec). Če vnesena številka ustreza pogojem za serijsko številko, se zaradi potrebe po veliki zanesljivosti preveri pravilnost njenega vnosa. V ta namen se generira zvokovni odziv s katerim sistem ponovi vneseno številko ter se zaprosi za potrditev pravilnosti (da/ne). Po uspešnem vnosu serijske številke se govorni dialog konča, sistem pa nadaljuje z novim razpoznavanjem slike. Če se pri preverjanju pravilnosti vnosa ugotovi, da vnesena številka ne ustreza zahtevam, ki so v naprej podane za serijsko številko, se operaterja zaprosi za ponoven vnos. V vsakem trenutku je omogočen tudi preklic vnašanja serijske številke z ukazom 13 (spreglej), kar povzroči, da sistem razpoznavano serijsko številko označi kot nerazpoznano.

4 uporabniška prijaznost govornega vmesnika

Govorni dialog je načrtan tako, da ne zahteva posebnega priučevanja in z njim lahko upravljajo tudi neizkušeni uporabniki. Večjo enostavnost uporabe ter uporabniško prijaznost daje programu

Id. ukaza	Tekst	Omogočenost
0	„0“	0/1
1	„1“	0/1
2	„2“	0/1
3	„3“	0/1
4	„4“	0/1
5	„5“	0/1
6	„6“	0/1
7	„7“	0/1
8	„8“	0/1
9	„9“	0/1
10	„start“	0/1
11	„stop“	0/1
12	„konec“	0/1
13	„spreglej“	0/1
14	„da“	0/1
15	„ne“	0/1
16	„pomoč“	1

Tabela 1: Tabela vhodnih ukazov

Št.	Besedilo
0	„Ukaza ne razumem.“
1	„Ukaz trenutno ni možen.“
2	„Ukaz še vedno ni možen.“
3	„Tudi ta ukaz ni možen.“
4	„Ukaz ni možen ,prosim pristopi k računalniku.“
5	„Trenutno so omogočeni naslednji ukazi:“
6	„Vse števke od 0 do 9.“
7	„Napaka pri razpoznavanju.“
8	„Vnesena številka je:“
9	„Je to pravilno?“
10	„Vnesena številka ne ustreza zahtevam. Prosim, ponovi vnos.“
11	„Prosim, preberi serijsko številko.“
12	„Sprejeto.“

Tabela 2: Izhodna sporočila

tudi poseben ukaz „pomoč“, s katerim je možno kadarkoli ugotoviti kateri so v danem trenutku možni ukazi. Ta ukaz se izvede v niti govornega vmesnika, s čimer mu je zagotovljen hiter odziv.

Za doseganje uporabniške prijaznosti govornega vmesnika je nujno, da so njegovi ukazi logično izbrani, zvokivni odzivi (sporočila) pa razumjivi. Vendar pa se v industrijkem okolju pojavi dvom o upravičenosti takšne uporabniške prijaznosti, saj to hkrati pomeni, da je komunikacija počasnejša. Razumljiva sporočila so namreč tudi relativno dolga in tudi operater, ki bi nekaj časa delal s takim uporabniškim vmesnikom, bi se dolgih, vedno enakih sporočil kmalu naveličal. Želeli bi si torej, da so sporočila kratka in komunikacija tako hitrejša. V ta namen bi bilo potrebno dolga razumljiva sporočila zamenjati s kratkimi, neizkušenemu uporabniku sicer manj razumljivimi.

Operater pa bi njihov pomen moral poznati v naprej. Podobno bi lahko veljalo tudi za ukaze, vendar je tu problem veliko manj izrazit, saj so ukazi že v normalnem dialogu kratki.

Dober primer za tako izdelano komunikacijo je ameriška policija [5]. Sporočila so nadomeščena s kodami in kot taka neveščemu poslušalcu nerazumljiva. Za izurjene uporabnike pa je to hiter in učinkovit način komunikacije.

V našem primeru bi bilo mogoče predvideti obe vrsti sporočil. Klasična, uporabniško prijazna sporočila, so zelo primerna za delo neveščih uporabnikov ter za njihovo uvajanje. Ko je operater že dobro seznanjen s potekom komunikacije pa je izbira sporočil odvisna od pogostosti uporabe govornega vmesnika. V našem primeru pričakujemo veliko zanesljivost optičnega razpoznavanja številk in popravljanje s pomočjo govornega vmesnika tako ne bi bilo zelo pogosto. V tem primeru so morda še vedno bolj primerna klasična sporočila. Vendar pa bi bilo kljub temu operaterju dobro dati možnost, da jih zamenja s kratkimi, enostavnimi.

Govornemu vmesniku mora biti za zanesljivo delovanje v podporo tudi grafični v mesnik. Zelo pomembno je, da je s pomočjo grafičnega vmesnika možen izhod iz vsakega stanja, v katerega pridemo z govornim vmesnikom. V našem primeru vsebuje grafični uporabniški vmesnik tudi vse funkcije govornega vmesnika, izvzeto je le preverjanje pravilnost vnosa serijske številke, saj to pri grafičnem vmesniku ni potrebno.

5 Rezultati testiranja uporabniškega vmesnika

Govorni uporabniški vmesnik je bil testiran tako, da so prototipno verzijo uporabljali nevešči uporabniki. Pri testu je sodelovalo pet uporabnikov, ki niso bili deležni posebnega učenja, dana jim je bila le informacija o obsoju ukaza „pomoč“. Uporabljeno je bilo takšno testno zaporedje čelnih plošč števcev, da so bila prisotna tudi neuspešna avtomatska razpoznavanja. S tem je bila zagotovljena potreba po uporabi govornega vmesnika, tako da je vsak testni uporabnik izvedel vsaj tri vnose serijske številke. Pri delu porabnikov smo bili pozorni na pojavljanje nejasnosti in morebitnih napak, ki bi iz njih sledile. Opazili smo sledeča problema:

- nejasnosti se pojavljajo pri ukazu „pomoč“, kjer ni razvidno kdaj se naštevanje možnih ukazov konča,
- hitro vnašanje serijskih številk je zelo omejeno zaradi dolžine izhodnih sporočil. V ta namen je bila predlagana zamenjava sporočila „Napaka pri razpoznavanju.“ z enostavnim sporočilom „Napaka!“, spororčila „Vnesena številka ne ustreza zahtevam. Prosim, ponovi vnos.“ z „Neuspešen vnos!“ ter opustitev sporočila „Prosim, preberi serijsko številko.“.

Pokazali pa sta se tudi naslednji dobri lastnosti vmesnika:

- ukazi govornega vmesnika so dobro izbrani in se logično navezujejo na zahtevano akcijo,
- komunikacija je dovolj enostavna za učinkovito praktično uporabo.

6 Zaključek

Prikazali smo izvedbo govornega uporabniškega vmesnika, namenjenega uporabi v industrijskem okolju. Pri njegovem načrtovanju smo upoštevali zahteve po uporabniški prijaznosti, zanesljivosti in seveda uporabnosti. Glede rezultate, dobljene pri testiranju vmesnika, lahko sklepamo, da bi bil realiziran govorni vmesnik primeren za praktično uporabo v industriji.

Literatura

[1] B. Vatovec, Načrtovanje interaktivnih sistemov, zaključna naloga, Univerza v Ljubljani, Fakulteta za elektrotehniko, Maj 1996

[2] P. Rogelj, S. Kovačič, Odčitavanje serijske številke električnih števcev, diplomska naloga, Univerza v Ljubljani, Fakulteta za elektrotehniko, July 1998

[3] P. Rogelj, S. Kovačič, Odčitavanje serijske številke električnih števcev, ERK ’98, September 1998.

[4] S. Dobrišek, J. Gros, F. Mihelic, N. Pavešič, HOMER: a Voice Driven Text-to-Speech System for the Blind, Proceedings of the 1999 IEEE International Symposium on Industrial Electronics, str. 205-208, Bled, Slovenia, 12-16 July, 1999.

[5] Police Radio Codes, Internet stran: *http://cobras.org/police.htm*, 15.7.1999.