

**UNIVERZA V LJUBLJANI
FAKULTETA ZA UPRAVO**

Diplomsko delo

**SISTEMI ZA PREPOZNAVANJE GOVORA –
Speech Recognition**

Miha Eržen

Ljubljana, september 2012

**UNIVERZA V LJUBLJANI
FAKULTETA ZA UPRAVO**

DIPLOMSKO DELO

SISTEMI ZA PREPOZNAVANJE GOVORA – Speech Recognition

Kandidat:	Miha Eržen
Vpisna številka:	04039417
Študijski program:	visokošolski študijski program uprava prve stopnje
Mentor:	viš. pred. dr. Bojan Peček

Ljubljana, september 2012

IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Podpisani Miha Eržen, študent visokošolskega študijskega programa uprava prve stopnje, z vpisno številko 04039417, sem avtor diplomskega dela z naslovom: Sistemi za prepoznavanje govora – Speech Recognition.

S svojim podpisom zagotavljam, da:

- je predloženo delo izključno rezultat mojega lastnega raziskovalnega dela,
- sem poskrbel, da so dela in mnenja drugih avtorjev oz. avtoric, ki jih uporabljam v predloženem delu, navedena oz. citirana v skladu s fakultetnimi navodili,
- sem pridobil vsa dovoljenja za uporabo avtorskih del, ki so v celoti prenesena v predloženo delo, in sem to tudi jasno zapisal v predloženem delu,
- se zavedam, da je plagiatorstvo – predstavljanje tujih del, bodisi v obliki citata bodisi v obliki skoraj dobesednega parafraziranja bodisi v grafični obliki, s katerim so tuje misli oz. ideje predstavljene kot moje lastne – kaznivo po zakonu (Zakon o avtorstvu in sorodnih pravicah, Ur. list RS, št. 21/95), kršitev pa se sankcionira tudi z ukrepi po pravilih Univerze v Ljubljani in Fakultete za upravo,
- se zavedam posledic, ki jih dokazano plagiatorstvo lahko predstavlja za predloženo delo in za moj status na Fakulteti za upravo,
- je elektronska oblika identična s tiskano obliko diplomskega dela ter soglašam z objavo dela v zbirki »Dela FU«.

Diplomsko delo je lektorirala: mag. Darja Rabzelj

Ljubljana, 4. 9. 2012

Podpis:

POVZETEK

Mnogo let je že minilo, odkar smo lahko zadnjič govorili, kako pomemben je in bo računalnik v življenjih povprečnih ljudi. Še več let je minilo od tedaj, ko smo zadnjič rekli, kakšno vlogo ima v gospodarskem, tehnološkem in civilno-razvojnem smislu. Brez računalnika dandanes posameznik ne najde službe, ne more opraviti izobraževanja, ne more odpreti podjetja ter opraviti še marsikatero druge, bolj malenkostne, stvari. Že kar nekaj časa pa poznamo tudi upravljanje računalnika glasovno, pri čemer nista potrebni miška ali tipkovnica.

V tej diplomski nalogi sem s pomočjo preverjanja zmogljivosti ter razumljivosti programa Speech Recognition lahko pokazal njegovo uspešnost ter s tem podal subjektivno oceno delovanja in možnosti uporabe tega programa.

Ključne besede: program, računalnik, programiranje, glasovni ukaz, prepoznavanje govora, mikrofona.

SUMMARY

SPEECH RECOGNITION SYSTEMS – SPEECH RECOGNITION

Many years have passed since we could talk about present and future importance of computer usage in average life of an individual, and many more years have passed since we could talk about its usage in economy, technology and civil development. Without using computers, individual is not able to find a job, get a degree, start a business and even do the most simple daily tasks anymore. For more than a decade, programs for speech recognition have been developed and launched for using computers.

In this thesis, a program called Speech Recognition has been tested in order to evaluate its performance and ability to understand human voice. I have made a subjective evaluation on results and presented a possibility of its usage.

Key words: program, computer, programming, voice command, speech recognition, microphone.

KAZALO

IZJAVA O AVTORSTVU DIPLOMSKEGA DELA	iii
POVZETEK	iv
SUMMARY	v
KAZALO	vi
KAZALO PONAZORITEV.....	vii
1 UVOD	1
2 TEORIJA PROGRAMA »SPEECH RECOGNITION«	5
2.1 Zgodovinski okvir programov za prepoznavanje govora.....	5
2.2 Osnove programov za prepoznavanje govora	8
2.3 Tehnične lastnosti programa speech recognition in njegovo delovanje.....	9
3 UPORABA PROGRAMA SPEECH RECOGNITION	11
3.1 Namestitev in zagon programa speech recognition	11
3.2 Nastavitev mikrofona	12
3.3 Vodič po programu speech recognition	13
3.4 Priprava računalnika za boljše razumevanje	16
3.5 Kartica za pomoč.....	18
4 POTREBNA OPREMA ZA UPORABO PROGRAMA SPEECH RECOGNITION	20
4.1 Mikrofon	20
4.1.1 Kratka zgodovina razvoja mikrofona	20
4.1.2 Teorija delovanja mikrofona.....	21
4.2 Računalnik z operacijskim sistemom.....	22
4.3 Zvočna kartica.....	22
5 ANALIZA REZULTATOV RAZISKOVALNEGA DELA	24
5.1 Narekovano besedilo in njegova analiza.....	24
5.2 Rezultati raziskovalnega dela ter analiza	26
6 ZAKLJUČEK	32
LITERATURA IN VIRI	34

PRILOGE.....	37
--------------	----

KAZALO PONAZORITEV

KAZALO GRAFIKONOV

Grafikon 1: Uspešnost prepoznavanja govora - črtni	28
Grafikon 2: Uspešnost prepoznavanja govora – stolpčni	28
Grafikon 3: Čas narekovanja izbranega besedila	30

KAZALO PRILOG

Priloga 1: Tabela podatkov testiranja	37
---	----

KAZALO SLIK

Slika 1: Prvi računalnik iz leta 1941	1
Slika 2: Logotip podjetja IBM.....	5
Slika 3: Logotip podjetja DARPA	6
Slika 4: Govoreča lutka Julie iz leta 1987.....	6
Slika 5: Prikazni meni vodiča po programu Speech Recognition.....	14
Slika 6: Končni povzetek najbolj pogosto uporabljenih ukazov pri narekovanju	15
Slika 7: Pozdravno besedilo čarovnika za uvajanje programa	17
Slika 8: Začetni meni kartice za pomoč.....	18
Slika 9: Narekovanje posamezne črke za sestavo besede	19
Slika 10: Logotip podjetja Bell Laboratories	21
Slika 11: Narekovano besedilo	24
Slika 12: Nastavitev jezika poslušanja	25

KAZALO TABEL

Tabela 1: Prikaz rezultatov uspešnosti prepoznavanja govora	29
---	----

Tabela 2: Čas, potreben za narekovanje danega besedila na določen dan	30
---	----

1 UVOD

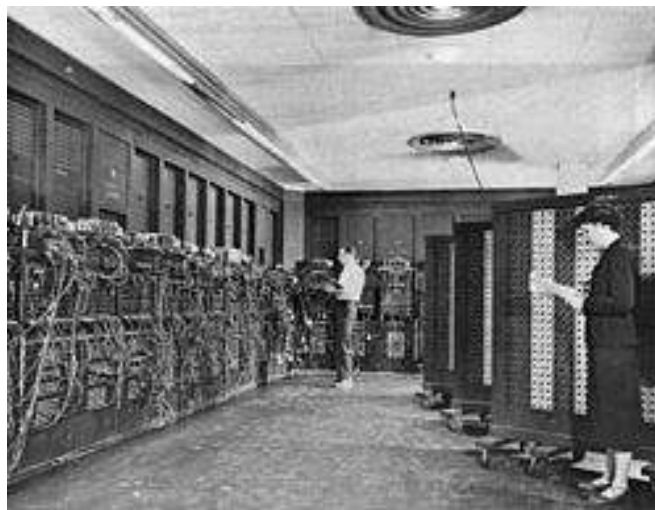
Delo z računalnikom je že pred mnogimi leti postalo vsakodnevno opravilo tako za starejše kot za mlade ljudi. Spremljamo lahko bliskovit razvoj računalniške tehnologije, v mehanskem in programskem smislu.

Če se za hip ozremo v zgodovino in ji posvetimo nekaj stavkov, moramo razvoj računalnika in njegovo starost najprej definirati z njegovim namenom. Nekateri avtorji med prve računalnike uvrščajo že prve mehanske pripomočke, poznane že iz stare Grčije in Rimskega imperija, kot so kotomeri in abaki, vendar pa sam le-teh ne bi povezoval s sodobnim pojmovanjem računalnika. Dandanes namreč računalnik pomeni napravo, s katero lahko ne samo računamo, rišemo, pišemo in slikamo, temveč tudi programiramo, si prirejamo njene funkcije tako, kot ustreza nam, posameznikom, ter seveda v znanstvene namene počnemo še ogromno drugih stvari, ki povprečnim uporabnikom sploh niso znane. (Wikipedia, 2012)

Menim pa, da vsi, ki uporabljamo osebni računalnik za take ali drugačne namene, že vemo, da verjetno ni stvari na svetu, ki ne bi bila v neki fazi izdelana, zapisana, zabeležena, shranjena, poslana, itd. s pomočjo sodobnega računalnika.

Prvi računalniki, ki jih poznamo, so se pojavili v velikosti sobe, velike kot šolska učilnica. Imeli so lastni klimatski sistem na strehi stavbe, za hlajenje, zmogli pa so obdelati manj podatkov, kot jih danes obdelata najmanjši premik miške na povprečnem osebem računalniku. To je bilo leta 1941 (Wikipedia, 2012). Primer računalnika te vrste je na sliki 1.

Slika 1: Prvi računalnik iz leta 1941



Vir: Wikipedija (2012)

Danes lahko najdemo mnogo naprav, ki vsebujejo znanje in tehnologijo, ki sta potrebna za glasovno ukazovanje. Že nekaj let poznamo na tržišču mobilne telefone, ki znajo poslušati

naš glas ter tako izpolnjujejo nekatere najenostavnejše ukaze, kot so klicanje vizitk v imeniku, shranjenim na telefonu. S podobnimi programi se lahko poigravamo tudi v naj sodobnejših avtomobilih, kjer lahko ne samo glasovno upravljamo z mobilnim aparatom, ki je brezžično povezan z vozilom, temveč tudi upravljamo s klimatsko napravo, nastavitvijo sedežev, itd. Kjer pa so se strokovnjaki srečali z problemom glasnosti vožnje, kot je šumenje zaradi upora vetra na avtomobil, zvok premikajočih se koles, itd. velik izziv je predstavljala tudi navigacija, saj je bilo težko vnesti vsa imena ulic, mest, krajev, katera so lastna, in tako težje izgovorljiva (Herbig in dr., 2012).

Še leta 1903 se je človeštvo posmehovalo bratoma Wright in dvomilo v njun polet na nebo z neko ljudem nepoznano čudno napravo, danes pa se z največjim udobjem vozimo v ogromnih letalih na dolge razdalje, na višinah in s hitrostmi, ki jih takrat še poznali niso. Podobno danes ne verjamemo tehnologiji za prepoznavanje govora, ki je razmeroma mlada, zato moramo dati tej tehnologiji, tako kot vsaki drugi, dovolj časa, da jo uporabniki sprejmejo in, seveda, da postane dovolj nezmotljiva, da bi lahko bila primerna za vsakodnevno uporabo.

Program, ki sem ga preizkusil za potrebe te diplomske naloge, se, kot sem že omenil, imenuje Prepoznavanje govora (v nadaljevanju uporabljam angleški izraz »Speech Recognition«) in ga najdemo v nadzorni plošči v vseh Microsoftovih operacijskih sistemih, starejših od XP različice. Operacijski sistem Windows XP je bil razvit 25. oktobra 2001 (Microsoft corporation, 2012).

Namen moje diplomske naloge je študija uporabe omenjenega programa ter oris njegovih značilnosti, prednosti in pomanjkljivosti. Opisati sem želel, kako poteka uporaba programa, in sicer vse od začetka, kjer program uporabnika sam vodi skozi vodič ter mu tako pokaže osnovne ukaze za njegovo razumevanje in uporabo. Hkrati pa sem hotel pokazati, kakšna naj bo potrebna mehanska oprema za uporabo programa.

Področje sem izbral, ker menim, da je omenjeni program za razpoznavanje govora, vgrajen v operacijski sistem Microsoft Windows, že dandanes dovolj dobro zasnovan za vsakdanjo uporabo.

Moj cilj je bil dokazati, da je mogoče z minimalno vajo

- **v primernem prostoru in s primerno opremo doseči natančnost programa, višjo od 95 %.**

Hipoteza se mi je zdela na začetku drzna, saj nisem imel popolnoma nobenih izkušenj z uporabo programa. Na predavanjih sem videl zgolj kratko, 10-minutno predstavitev predavatelja, ki me je navdušila do te mere, da sem julija in avgusta 2012 začel z vadbo programa. Tako sem se s pomočjo vsakodnevne vaje in testiranj, ki sem jih opravljal sproti in vsak dan pod istimi pogoji, dnevno izpopolnjeval.

Izdelava diplomske naloge je potekala predvsem doma, saj sem za testiranje programa potreboval vedno isti prostor, z enakimi zvočnimi motnjami, saj sem le tako lahko beležil rezultate, ki mi jih je testiranje dajalo. Za testiranje sem potreboval osebni računalnik in mikrofona, ki je bil najbolj primeren za to nalogo.

Testiranje sem opravljal 30 ur: programu sem narekoval besedila, nato pa opravil pregled ter rezultate med seboj primerjal. Ker znam slepo desetprstno tipkati, sem rezultate tipkanja primerjal z vnosom preko narekovanja z mikrofonom.

Ker je virov in literature s tega področja izjemno malo, sem veliko pisal na podlagi osebnega znanja in izkušenj s programom, saj sem le-tega temeljito preučil. Opisal sem postopke, od začetka do konca, ki so potrebni za uporabo programa. Preostanek znanja in literaturo sem črpal z internetnih strani, kjer sem prebiral forume in članke, in nato povzetke zapisal v diplomsko nalogo.

V diplomski nalogi sem strogo ločil program Speech Recognition od računalnika. Zato je opazna razlika, da nekateri avtorji, med drugimi celo podjetje Microsoft v svojih vodičih in orodjih, uporabljajo izraz računalnik tako, kot da uporabnik govori v računalnik. Verjetno z namenom zblížanja z uporabnikom, da bi ta lažje razumel situacijo in uporabo celote, tako strojne kot programske opreme. Dejstvo je, da za vsako aplikacijo, ki jo ima računalnik, stoji operacijski sistem. Sestavljen je iz mnogo programov, za nekatere od njih ni nujno, da so bili razviti v laboratorijih Microsofta. In ker vemo, da je program Speech Recognition končni izdelek več projektov in programov, sem se odločil, da te stvari med seboj ločim. Zato v nalogi, kadar omenjam program, mislim na program Speech Recognition.

V drugem poglavju sem opisal zgodovinski okvir programa Speech Recognition. Začel sem v letu 1950, ko sicer še ni bilo nič znanega ali omenjenega o programu Speech Recognition, vendar pa se prve zasnove tovrstnih programov začnejo razvijati v tem obdobju in želel sem opisati celotno zgodovino ideje o programu za prepoznavanje govora.

V tretjem poglavju sem dosledno opisal postopke uporabe programa Speech Recognition v operacijskem sistemu Windows 7. Šel sem po korakih, od začetka do konca, ter sproti opozarjal na morebitne pomanjkljivosti, na katere moramo biti pozorni pri prvi ter nadaljnji uporabi tega programa. Z nekaj slikami sem poskušal prikazati osnovne postopke pri uporabi programa.

V četrtem poglavju sem predstavil tehnično podporo, ki jo program potrebuje za svoje delovanje. Na kratko sem opisal zgodovino mikrofona, kot primarnega pripomočka za delo s programom. Opisal sem še delovanje mikrofona ter predstavil osnovne vrste mikrofona. Predstavitvi mikrofona sem v petem poglavju dodal tudi kratek opis zvočne kartice in operacijskega sistema, potrebnega za delovanje programa Speech Recognition.

V petem poglavju sem analiziral in primerjal rezultate raziskovalnega dela, ki sem jih zbiral skozi testiranje programa Speech Recognition. Podal sem besedilo, ki sem ga uporabljal za

testiranje programa, ter prikazal nekaj besed, ki so izstopale pri narekovanju in razumevanju. V tem poglavju je prikazano, koliko časa sem porabil za narekovanja teksta in kako se je leta sproti spreminjal ter kako se je odstotek uspešnosti prepoznavanja govora spreminjal sorazmerno s časom.

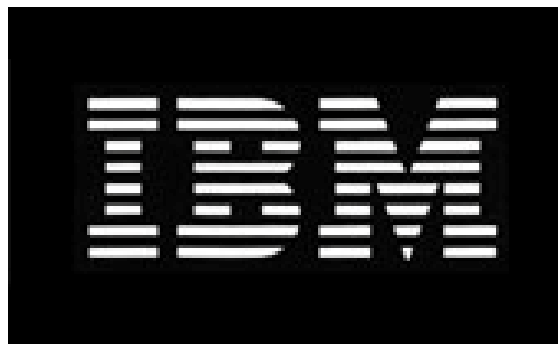
Šesto poglavje sem namenil zaključku diplomske naloge, v katerem sem podal svoje subjektivno mnenje o programu Speech Recognition. Ocenil sem njegovo uporabnost in delovanje. Dotaknil sem se vnaprej zastavljene hipoteze in se do nje opredelil.

2 TEORIJA PROGRAMA »SPEECH RECOGNITION«

2.1 ZGODOVINSKI OKVIR PROGRAMOV ZA PREPOZNAVANJE GOVORA

Začetki samega programa za prepoznavanje govora segajo v leto 1950, ko so znanstveniki in programerji s pomočjo tehnologije razvili program, ki je lahko razumel prve ukaze, ki so takrat bili še v obliki številčk. Kasnejši skok v razvoju je svetu predstavilo podjetje IBM s programom, ki je razumel zgolj 16 izgovorjenih besed, seveda v angleškem jeziku. To se je zgodilo leta 1962. Kot navaja avtorica članka na spletni strani PCworld, Melanie Pinola, je bil to velik napredek za tiste čase, sploh če pomislimo, kako osnovni in primitivni so takrat bili računalniki ter upoštevamo dejstvo, da je bil prvi računalnik izdelan leta 1941. To pomeni, da so bili prvi poskusi izdelave sistemov za prepoznavanja govora ustvarjeni zgolj 21 let po predstavitvi prvega računalnika (Pinola, 2011). Spodnja slika prikazuje logotip podjetja IBM, ki je veliko pripomoglo k razvoju sistemov za prepoznavanje govora.

Slika 2: Logotip podjetja IBM



Vir: (IBM, 2012)

Največji razvoj je program doživel v letih od 1971 do 1976, ko je podjetje DARPA razvilo program z imenom »Harpy«, ki je lahko prepoznal 1011 besed, zopet seveda le angleške. Za primerjavo je avtorica zapisala, da je to približno toliko, kot jih zna izgovoriti povprečni triletni otrok. V enakem obdobju so tudi druga podjetja bila med seboj bitko v hitrosti razvoja sistemov za prepoznavanje govora. Avtorica v članku omenja glavna tekmeca Bell Laboratories in Threshold Technology (Pinola, 2011). Naslednja slika prikazuje logotip podjetja Beel Laboratories.

Slika 3: Logotip podjetja DARPA



Vir: (Baichtal, 2011)

V začetku leta 1980 se prvič pojavijo naprave, ki so bile namenjene za komercialno uporabo, in ne samo za vojsko in znanost. Prvič zaznamo tudi željo po uporabi teh programov v medicinske namene. Tako je leta 1987 iz podjetja Worlds of Wonder na tržišče prišla lutka za otroke z imenom Julie, s katero si lahko spregovoril nekaj besed. Lutka je znala odgovarjati na ukaze, ki ji jih je uporabnik zadal. Preproste otroške pogovore je razumela, vendar se je pojavil problem izgovarjave, saj otroci kdaj pa kdaj kakšno črko pojedjo (Pinola, 2011). Na spodnji sliki je prikazana govoreča lutka Julie, kot prvi izdelek za komercialno prodajo, ki je imela vgrajen sistem za prepoznavanje govora.

Slika 4: Govoreča lutka Julie iz leta 1987



Vir: (Squidoo, 2012)

Za to obdobje je nekako značilno, da so začeli program razvijati v smeri, da program ne le prepozna mnogo besed (in jih nato ali izgovori naglas ali pa jih zapiše v nek dokument), ampak je bila ideja tega obdobja tudi predvidevanje. To pomeni, da te, ko v nareku poveš

več besed v istem stavku in brez večjih premorov, program posluša ter o tistih besedah, ki jih ne razloči povsem dobro, sklepa sam oziroma predvidi, katera beseda bi bila najbolj primerna in tako dokonča stavek. Tak način narekovanja je kot osnova uporabljen še danes. O tem bom pisal še kasneje, ko se bom osredotočil na program Speech Recognition v operacijskem sistemu Windows 7 (Pinola, 2011).

V tem obdobju je še vedno ostala neka osrednja težava teh programov. Programi so temeljili na diskretni obravnavi posameznih besed. To pomeni, da je uporabnik za izdajanje ukazov programu moral govoriti s premori. Program je vsako izgovorjeno besedo analiziral. Avtorica članka to opiše kot: Danes.....je.....lep.....sončen.....dan (Pinola, 2011).

Po desetletju boja programerjev s programi, njihovim razumevanjem ter sedaj še s predvidevanjem nastopi leto 1990, ko povprečni uporabniki po vsem svetu začnejo uporabljati zmogljivejši ter hitrejši osebni računalnik. Tukaj govorimo o sposobnosti in hitrosti notranjega spomina ali RAM ter o hitrostih procesorjev, vgrajenih na matično ploščo, ki zagotavljajo bistveno hitrejše delovanje samega računalnika.

V tem obdobju so se razvijali računalniški programi, ki so lahko poslušali po 100 besed na minuto ter jih na koncu zapisali ali izgovorili. Podjetje Dragon, ki se razvija in ustvarja še danes, je izdalo program, ki se je imenoval Dragon Dictate. Velika pomanjkljivost je bila njegova visoka cena, znašala je kar 9000 dolarjev. To je bilo tedaj veliko ne samo za posameznika, temveč tudi za marsikatero podjetje. Pa vendar je podjetje ustvarjalo naprej in sedem let kasneje izdalo nov program z imenom Dragon Naturally Speaking. Slednji je bil prvi program, ki je zagotavljal naravno govorjenje brez premorov in je lahko sledil 100 ukazom na minuto. Za njegovo uporabo je bilo potrebnih 45 minut uvajanja, cena pa je bila 695 dolarjev, za kar nekateri menijo, da je še vedno predrago (Pinola, 2011). Osebo se mi zdi ta cena bolj dostopna, sploh za tisto obdobje in v primerjavi s ceno predhodnika. Zadnja cena polnega paketa je približno 200 dolarjev (Voice-recognition-software-review.toptenreviews.com, 2012).

Avtor John P. Mello je na spletnem portalu PCworld zapisal: *»Program Dragon Naturally Speaking zahteva veliko konjskih moči. Računalniki z operacijskim sistemom Windows 95 bodo potrebovali 32 MB trenutnega spomina ali RAM-a, 60 MB prostora na trdem disku in procesor Pentium-133 (čeprav bi program deloval s procesorjem Pentium 90«* (Mello, 1997). Za še eno ponazoritev lahko povem, da bi danes, 15 let kasneje, z računalnikom opisanih karakteristik potrebovali verjetno več minut že za odprtje aplikacije v sedanji obliki.

Mnogi znanstveniki in programerji, so mnenja, da je nujno potreben prvi korak v razvoju sistemov za prepoznavanje govora, razviti programe ki delujejo kot »govor-v-besedilo«. To pomeni da program posluša govor uporabnika ter ga nato izpiše kot besedilo. Mnogi so mnenja, da je ta korak nujno potreben za razumljivost uporabnikovega glasu, kot tudi za pravilni odziv na govornikov glas (B. H. Juang in R. Rabiner Lawrence., 2011, str. 18).

V začetku naslednjega obdobja, leta 2001, se je pojavil operacijski sistem, ki je vseboval program za prepoznavanje govora, Speech Recognition. Ta program je bil že močno dodelan in izboljšan. Prav tako v tem času je divje naraslo število osebnih računalnikov, močno se je razširila uporaba interneta. Tako je imel praktično vsak, ki je imel računalnik, tudi internet. Glede na tako veliko mrežo uporabnikov, ki so si med seboj sicer fizično oddaljeni več tisoč kilometrov, vendar zblížani v neposredno bližino s pomočjo interneta, se je preko pošiljanja podatkov, izmenjave informacij in z redno nadgradnjo tudi program za prepoznavanje govora lahko hitro razvijal. Že omenjena avtorica Pinola v svojem članku sicer omenja, da povprečni uporabniki verjetno sploh niso vedeli za takšen program, ki so ga imeli naloženega v kompletu z operacijskim sistemom, saj nekako niso imeli potrebe po zamenjavi tipkovnice in miške. Tukaj velja izpostaviti še, da je bil program brezplačen, kar je redkost v zgodovini računalništva, še posebej ob upoštevanju že prej omenjenih cen predhodnikov tega programa (Pinola, 2011).

V istem obdobju pa Google za podjetje Apple izdelala aplikacijo, ki je lahko prepoznavala govor uporabnika ter tako izpolnjevala njegove ukaze. Ti ukazi še danes predstavljajo klicanje vizitk v imeniku, iskanje zelene vsebine po internetu, brskanje po nastavitvah telefona, itd. Avtor članka na portalu PCworld Daniel Ionescu pravi, da aplikacija zazna, kdaj telefon primemo v roke in ga približamo obrazu, ter tako aktivira program za poslušanje, ki nato izvede operacije, ki smo mu jih glasovno ukazali. Potek delovanja avtor opiše tako: *»Prepoznavanje glasu ne poteka izključno na iPhone-u. Ko telefon nekaj vprašate, le-ta pretvori vaš glas v digitalni zapis in ga pošlje na Googlove strežnike, ki analizirajo zapis in ga pretvorijo v iskalni niz. Nato rezultate pošljejo nazaj na telefon. Celoten proces traja približno sekundo, odvisno seveda od hitrosti internetnega omrežja, ki ga uporabljate.«* (Ionescu, 2008) Avtor v nadaljevanju opisuje tudi postopek, ko uporabnik glasovno zaprosi mobilni telefon, naj mu pove, kje se nahaja najbližja restavracija, le-ta pa nato preko navigacijskega sistema, ki je vgrajen v mobilni telefon, locira napravo in s tem tudi uporabnika, nato na spletu preveri naslove vseh zelenih restavracij, izbere najbližjo ter nastavi in izriše navodila in pot do nje (Ionescu, 2008).

Velik razvoj je naredil Google leta 2010, ko je svoje predhodne programe za prepoznavo govora nadgradil s tako imenovano »osebno prepoznavo«. Pri tej program beleži glasovne ukaze uporabnika in jih dodaja na splet, tako pa ustvarja in nadgrajuje bazo besed, narečij, izgovorjav. S tem so dosegli, da program do sedaj obvlada 230 milijard besed (Ionescu, 2008).

2.2 OSNOVE PROGRAMOV ZA PREPOZNAVANJE GOVORA

Tehnologija programov za prepoznavanje govora se glede na namen v osnovi deli na dve, medsebojno malo povezani, skupini (Hou, 2012):

- Tehnologija z omejenim obsegom slovarja pomni nekaj besed oziroma ukazov. Deluje relativno zanesljivo. Dnevno se uporablja v prenosnih telefonih.

- Tehnologija z neomejenim slovarjem pa ima vnaprej pripravljene vse besede in izgovorjeno besedo išče v slovarju. To je mnogo bolj kompleksna rešitev, ki še vedno ne deluje popolnoma brezhibno. Še vedno je potrebno govoriti nekoliko počasneje, z mnogimi korekcijami.

V nadaljevanju svoje raziskave se osredotočam izključno na drugo tehnologijo. Zanima me uporabnost programa, ki omogoča narekovanje poljubnega besedila namesto njegovega tipkanja.

Glede na način delovanja pa ločimo diskretne in zvezne programe (Schalk in dr., 1993):

- Diskretni programi prepoznavajo vnos vsake posamezne besede. Uporabnik pove besedo in program poišče besedo, katere izgovor je najbolj podoben. Programi te vrste omogočajo vnos 40 do 60 besed na minuto. Tipični predstavnik programske opreme te vrste je Kurzweil Voice Pad (Voice Recognition.com, 2012).
- Poleg teh pa poznamo še programe, ki uporabljajo tehniko zveznega narekovanja. Uporabnik govori tekoče. Program si pomaga z medsebojnim prepoznavanjem besed. Kot primer naj navedem, da ko program 'ugotovi' v stavku besedo »are«, samodejno postavi pripadajoči samostalnik v množino. V praksi to pomeni, da čeprav program 'prepozna' vnos stavka »car are..«, ponudi rešitev »cars are«. Torej pravilno spremeni prepoznano besedno zvezo »vozila so« namesto »vozilo so«. Procesiranje zveznih programov je seveda mnogo bolj zahtevno in je mogoče šele z novejšo tehnologijo. Prav zaradi zahteve po tekočem narekovanju pa ta tehnologija zmore vnos tudi preko 100 besed na minuto (Voice Recognition.com, 2012). Na trgu je več tovrstnih produktov, kot so »Dragon Naturally Speaking« podjetja Dragon Systems, »viaVoice« podjetja IBM, »Voice Xpress« podjetja L&H in »FreeSpeech« podjetja Philips (Chunrong in dr., 2004, str. 1).

Program Microsoft Word 2007, s katerim sem delal za potrebe te diplomske naloge, podpira zvezno tehnologijo. Najbolj priljubljen program za prepoznavanje govora na spletu je sicer že omenjeni Dragon Naturally Speaking, a je njegova pomanjkljivost previsoka cena, prednost Wordovega programa pa je, da je implementiran v informacijski sistem Windows.

2.3 TEHNIČNE LASTNOSTI PROGRAMA SPEECH RECOGNITION IN NJEGOVO DELOVANJE

V osnovi program deluje tako:

1. Mikrofon beleži izgovorjene besede. Membrana niha in njeno nihanje beleži z električnimi signali.
2. Zvočna kartica, ki je pri novejših računalnikih del CPE, pretvori nihanje mikrofonske membrane v digitalni zapis.
3. Aplikacija oziroma program za prepoznavanje zvoka poišče najbližji digitalni zapis korena besede v svojem slovarju besed. Sledi druga faza, ki 'popravi' stavek ali del

stavka v dokončno obliko s pomočjo medsebojnega dopolnjevanja. Primer korekcije samostalnika v množinsko obliko glede na glagol je v nadaljevanju besedila raziskave opisan. Tako se na ekranu izpiše celoten stavek. Ukaze, ki mu jih zadamo, pa izvaja in beleži enako kot ukaze miške ali tipkovnice (Microsoft corporation, 2012).

Velika ovira pri nastanku in dopolnjevanju programa Speech Recognition je bila seveda izgovorjava besed, pa tudi narečja posameznika. Poleg tega so programerji in znanstveniki morali najti pot, kako ločiti narekovanje od ukazovanja. Do dvoumnosti pride, kadar uporabnik narekuje besedilo s tehnično vsebino o programu Speech Recognition. Program bo pravilno zaznal, da gre za besedilo in ne bo odpiral in zapiral programov, spreminjal velikosti črk ali pa celo izklopil računalnika in podobno. To oviro so znanstveniki in programerji rešili tako, da program loči modus poslušanja od ukazovana.

3 UPORABA PROGRAMA SPEECH RECOGNITION

Uporaba programa Speech Recognition je zelo enostavna in je približana uporabniku, kolikor je mogoče. Prepričan sem, da bi vsak s svojim osnovnim znanjem lahko osvojil in učinkovito uporabljal omenjeni program.

Prav zato menim, da je uporaba programa slabo dokumentirana. V operacijskem sistemu Windows 7 je pomoč o uporabi programa za prepoznavanje govora kompleksno zaobsežena. Besedilno raztolmači, kaj lahko uporabnik s programom naredi, kako ga uporabi, kaj potrebuje. Površno opiše tudi optimalno okolje, ki omogoča čim boljše prepoznavnost govorečega uporabnika. Prav prepoznavnost pa je ključnega pomena za uporabnost tega programa.

3.1 NAMESTITEV IN ZAGON PROGRAMA SPEECH RECOGNITION

Na začetku je potrebno namestiti operacijski sistem, ki je novejši od Windows XP, ki je izšel leta 2001. To so operacijski sistemi Windows XP, Windows Vista, Windows 7 ter v zadnjem času tudi že Windows 8. Seveda mora biti strojna oprema računalnika dovolj zmogljiva za delovanje nameščenega operacijskega sistema. Program nima posebnih zahtev za dodatno opremo.

Najprej je potrebno program aktivirati. Program se zažene s klikom na ikono »Start« v spodnjem levem kotu namizja operacijskega sistema. Lahko tudi s pomočjo tipke za leteče okno, ki odpre vrstico Start. V odprti vrstici Start mora uporabnik poiskati ikono »Control Panel« oziroma, v slovenščini, Nadzorna plošča. V njej je potrebno aktivirati bližnjico do programa, ki se imenuje Speech Recognition oziroma, v slovenščini, Sistem za prepoznavanje govora. Uporabniku se nato odpre novo okno, ki prikaže pet možnosti za naslednje povezave.

Prva možnost se imenuje »Zaženite Sistem za prepoznavanje govora«, druga možnost je ukaz »Nastavitev mikrofona«, kot tretjo možnost računalnik ponudi »Vodič po programu Speech Recognition«, četrta povezava ponuja možnost »Izurite računalnik za boljše razumevanje«, kot zadnjo, peto možnost, pa računalnik poda možnost »Odpri kartico za pomoč«.

Tukaj moram omeniti, da je postopek pri prvi uporabi vse do klika na ikono Speech Recognition v nadzorni plošči nekoliko drugačen od omenjenega, saj pri prvi uporabi program samostojno popelje uporabnika skozi uvod, pripravo mikrofona in ostale postopke. Bi pa v nadaljevanju vseeno rad malce podrobneje opisal vsako od petih možnosti, saj jih vsak uporabnik programa lahko kar nekajkrat uporabi, vsaj v začetnem obdobju uporabe.

Prvo možnost »Zaženi Speech Recognition« bom izpustil, saj se praviloma uporablja samo za ročni zagon programa, razen če ni nastavljena možnost, da se program zažene samodejno ob zagonu operacijskega sistema.

3.2 NASTAVITEV MIKROFONA

Izbor namestitve mikrofona sta izredno pomembna. Mikrofon so ob nepravilni namestitvi podvrženi 'sikanju'. To je zapis 'razmazanega' glasu črke »s«. Zato je nujno potrebno izbrati drugo možnost »Nastavitev mikrofona«. Klik z miško odpre novo podokno, kjer program vpraša po izbiri vrste uporabljenega mikrofona.

Prva vrsta je tako imenovani »Headset microphone«, kar v prevodu pomeni mikrofona, ki je po navadi pritrjen na okvir slušalk, ki jih imamo na glavi. Menim, da je ta vrsta mikrofona najboljša, saj program zahteva, da je oddaljenost mikrofonske glave oddaljena od ust točno in konstantno razdaljo. To lahko zagotovi le ta vrsta mikrofona, saj ostala tipa te možnosti nimata.

Druga možnost izbire mikrofona se imenuje »Desktop microphone«. Gre za mikrofona, ki so na mizi in naj bi bili nepremično postavljeni na trdo podlago. Takšne mikrofona uporabljajo na primer fakultete v večjih učilnicah, kjer profesor ali profesorica stoji za pultom in govori v mikrofona, ta pa s pomočjo ojačevalnika pošilja zvok v zvočnike, nameščene po celotni predavalnici.

Tretjo možnost za mikrofona program ponudi kot »Other«, kar pomeni vse druge možnosti. To so na primer mikrofona, vgrajeni v prenosne računalnike, ali pa mikrofona, ki se jih uporablja za petje.

Kot je mogoče razbrati iz opisa vseh možnosti, je prva izbira vrste mikrofona najboljša. Opis mikrofona bo predstavljen v nadaljevanju.

Ko je izbrana možnost mikrofona, nas program popelje skozi vodič nastavitve mikrofona. V prvem koraku pokaže, kakšna je pravilna lega mikrofona za uporabo, in sicer v treh alinejah:

- Postavite mikrofona približno 2,5 cm od ust, na eno stran od obraza.
- Ne dihajte neposredno v mikrofona.
- Poskrbite, da stikalo za glasnost ni nastavljeno na nemost.

Ko je mikrofona pravilno nameščen, se glas uporabnika in program prvič srečata. Program zahteva branje stavka, ki ga izpiše na ekran. Stavek se glasi: »Peter dictates to his computer. He prefers it to typing, and particulary prefers it to pen and paper.« Prevod tega stavka ni potreben, saj program zazna izgovarjavo samo v angleškem jeziku. Pod stavkom se pokaže parameter, ki sledi glasu.

To prvo srečanje programa z uporabnikovim glasom še ni navajanje programa na uporabnika in se lahko poljubno krat ponovi. V tej fazi program samo preverja sposobnosti in delovanje mikrofona. To pomeni, da program opravi hitri pregled delovanja mikrofona.

3.3 VODIČ PO PROGRAMU SPEECH RECOGNITION

Po priključitvi mikrofona sledi faza prepoznavanja uporabnikovega govora. Nekateri ljudje govorijo hitro, odrezavo, drugi lenobno, počasi. Program se priuči tudi logopedskih napak, kot so sesljanje, napačna izgovarjava črke »r« in podobno. Program se mora 'navaditi na uporabnika'. To program opravi v vodiču po programu, sočasno se uporabnik nauči uporabljati program. Tako sta ubiti dve muhi na en mah.

S klikom na vodiča po programu se odpre program. Osebni vtis je izredno pozitiven. Menim, da je vodič odlično izdelan. Uporabnika dobro seznanj s programom, z njegovimi zmogljivostmi, pomanjkljivostmi ter ga nauči osnovnih ukazov, ki jih potrebuje za uspešno uporabo programa. Kot prikazuje spodnja slika, program na desni strani ekrana izpiše dobrodošlico ter kratek povzetek, kaj vse se bo uporabnik naučil skozi vodič.

Slika 5: Prikazni meni vodiča po programu Speech Recognition



Vir: lasten

V primeru želje o posameznih možnostih dajanja ukazov programu je potrebno izbrati eno izmed možnosti, ki so izpisane na spodnjem delu ekrana. To so:

1. »Basics/Osnove«: V tem delu program popelje skozi osnove, ki uporabniku stalno pomagajo pri uporabi programa. Program najprej nazorno opiše in pokaže zagon programa. Nato raztolmači znake. V tem delu program prikaže pregled in spreminjanje možnosti nastavitvev programa, spremembo iz angleškega jezika v ameriško angleščino in podobno. Angleščina in ameriška angleščina sta v osnovi isti jezik, vendar se nekatere besede drugače izgovarjajo oziroma naglašujejo, nekaj skoraj neopaznih, vendar za ta program zelo pomembnih, razlik je v slovnici, nekaj pa je tudi takih, zaradi katerih se spremenijo celo ukazi. Kot primer naj navedem, da je pri angleškem jeziku ukaz za piko ob koncu stavka beseda »full stop«, pri ameriški različici angleščine pa je ta beseda »period«. Poleg tega program omogoča odličen ukaz, ki ga aktiviramo z izgovorom besedne zveze »What can I say«. Program to zazna kot prošnjo, da naj izpiše možne ukaze, ki bi bili v tisti situaciji najbolj primerni. To je resnično dobra pomoč, saj program premore mnogo možnosti in s tem tudi ukazov. Tako se ni potrebno 'na pamet' naučiti vseh ukazov, ampak lahko uporabnik možnosti prikliče s prej omenjeno funkcijo.
2. »Dictation/Narekovanje«: Morda bi lahko rekli, da je osnovna ideja programa Speech Recognition prav narekovanje; seveda poleg mnogih ostalih funkcij, ki jih program ponuja. V zavihku »Dictation« program najprej pove, da mora biti izgovarjava

razločna in dovolj glasna. To sta seveda najpomembnejši zadevi, ki jih je potrebno upoštevati med uporabo programa. Opazil sem, da glede na dejstvo, da nisem rojeni govorec angleškega jezika, program kaj hitro zameša besede in jih ne izpiše pravilno. Tako je pogosto bolje, da uporabnik besede ne izgovori pravilno, pač pa tako, da jo program najbolje prepozna. O tem več kasneje, ko bom komentiral in primerjal rezultate raziskovalnega dela. Program izpiše samo ukaze, ki jih potrebujemo pri nareku besedila. Taki ukazi so enostavno narekovanje poljubnega besedila in posebni ukazi, kot so končanje stavkov s pikami, vejicami, vprašaji, klicaji ter zahteva za izpis narekovajev, oklepajev in zaklepajev. Med narekovanje štejejo tudi ukazi, ki nadomestijo komande z miško, kot so »scroll up«, »scroll down«, »go to«, itd. Na sliki 6 je predstavljen povzetek najbolj pogosto uporabljenih ukazov pri narekovanju.

Slika 6: Končni povzetek najbolj pogosto uporabljenih ukazov pri narekovanju

Summary

There are really just a few commands you need to memorize for dictation.

They are:

- NEW PARAGRAPH, NEW LINE
- Correct WORD
- Select WORD
- Go to WORD, Go after WORD, Go to the end of the document
- Delete WORD, Delete that
- Press KEY

As you use dictation more and more, you'll learn more advanced dictation commands. These commands are listed on your Speech Reference Card.

You can always open the Speech Reference Card by saying "What can I say?"

Vir: lasten

Kot je razvidno na sliki 6, program skrči najnujnejše ukaze in za lažjo uporabo svetuje pomnjenje naslednjih ukazov: »New paragraph«, »New line«, »Correct word«, »Select word«, »Go to word«, »Go after word«, »Go to the end of the document«, »Delete word«, »Delete that«, »Press key« (slovenski prevod: »nov odstavek«, »nova vrstica«, »popravi besedo«, »izberi besedo«, »pojdi na besedo«, »pojdi za besedo«, »pojdi na konec besedila«, »izbriši besedo«, »izbriši«, »izberi gumb«).

Na koncu je prikazano še opozorilo na možnost »What can I say«, ki omogoča izbor najprimernejših ukazov za tisti trenutek.

3. »Commanding/Ukazovanje«: Ta zavihek poda informacijo o upravljanju računalnika z glasovnimi ukazi. Pove načelo: 'recite, kar vidite'. Tako v mapi Moj računalnik, z ukazom »Click pictures« program odpre mapo z slikami. Pokaže osnovne prijeme, ki lajšajo uporabo programa, in sicer tako pri tipkanju v programih, kot so Word, Wordpad, Excel, kot tudi pri uporabi samega programa v operacijskem sistemu Windows.
4. »Working with Windows/Delo z Windowsi«: Program v tem koraku nauči izvajati preproste operacije v operacijskem sistemu Windows. Pokaže vzorec zapiranja in odpiranja programov, njihovo minimaliziranje, maksimiziranje in podobno. Mnogi od teh ukazov veljajo za vse programe in aplikacije v sistemu Windows.
5. »Conclucion/Zaključek«: V zaključku pa program nagovori uporabnika, da naj bo pri uporabi strpen in naj ne prehitava z narekom, saj se program še v veliki meri nadgrajuje in dopolnjuje, zato so še vedno področja, ki jih ne obvlada povsem. Uporabniku na kratko razloži, kaj se je naučil skozi vodič in ga tako pripravi za nadaljnjo uporabo programa.

3.4 PRIPRAVA RAČUNALNIKA ZA BOLJŠE RAZUMEVANJE

Že v kratkem opisu program pove, da je potrebno opraviti to urjenje za izboljšanje razumevanja uporabnikovega glasu. To urjenje sicer ni obvezno, vendar so rezultati prepoznavnosti bistveno boljši s pomočjo urjenja.

V tem delu program začne pomniti glas uporabnika in njegove posebnosti. S pomočjo vaje in uporabe program postaja vse boljši in boljši. Vsak od nas ima svoj ton glasu, celo neko svojo izgovorjavo besed, zato je zelo pomembno, da program na enem računalniku uporablja samo ena oseba. Če bi na primer na mojem računalniku skozi to urjenje šel nekdo drug, bi zagotovo pri testiranju mojega glasu kasneje zaznali nekaj več napak.

Kot prvo predstavljam funkcijo »Train your computer to better understand you«, ki se, kot sem že omenil, nahaja v izhodiščnem meniju programa Speech Recognition v nadzorni plošči operacijskega sistema. Slika 7 prikazuje prva navodila, ki jih uporabnik prejme s strani programa.

Slika 7: Pozdravno besedilo čarovnika za uvajanje programa

Welcome To Speech Recognition Voice Training

By listening to you read aloud to the computer, Speech Recognition learns how you speak.

On the next page some training text will be displayed. When the training text appears, read it aloud in a natural and even tone.

The system will display one line at a time. When you have finished reading a line, the system will, after a short pause, automatically present the next line.

If the system doesn't understand what you read, it will skip to the next line. This is normal, and not a cause for concern.

Click Next to begin training.

Note: We recommend that you take the Speech Recognition tutorial first. In addition to improving the recognition of your voice, the tutorial also gives you an opportunity to learn the most important Speech Recognition commands.

Vir: lasten

V navodilih med drugim izvemo, da se program s poslušanjem uporabnikovega govorjenja sproti uči načina našega govora; da mora uporabnik stavke, ki jih program prikaže na zaslonu, brati glasno in v naravnem in enakomernem tonu ter da je pred pričetkom uporabe programa priporočljivo iti skozi vodič, saj se bo uporabnik tako naučil najpomembnejših ukazov (Microsoft corporation, 2012).

Program približno 7-10 minut prikazuje stavke, ki jih je potrebno diktirati v mikrofonski napravi čim bolj natančno, razločno in dovolj glasno. Program ne navaja povsem nesmiselnih stavkov, temveč preko njih sporoča glavne lastnosti njega samega, kot na primer način uporabe programa in izgovorjave besed. Program velikokrat omeni, da se s pogostostjo njegove uporabe večja njegova zanesljivost in natančnost.

Tako tudi pove, da je izgovorjeno besedo, ki je program ne razume pravilno, najbolje popraviti z ukazom »correct that« in ne z ukazom »delete that«. V drugem primeru (»delete that«) program besedo izbriše in nadaljuje s poslušanjem uporabnika. V prvem primeru (»correct that«) pa program odpre okence, kjer so nanizane besede, ki bi morda ustrezale izgovorjavi. Ob tem si program zapomni izbiro, jo shrani in naslednjič ob isti besedi poskusi uporabiti besedo, ki je bila naknadno izbrana iz seznama. Tako se program uči uporabnikovega glasu in postaja vedno bolj nezmotljiv.

Priporočljivo je opraviti čim več podobnih uvajanj.

3.5 KARTICA ZA POMOČ

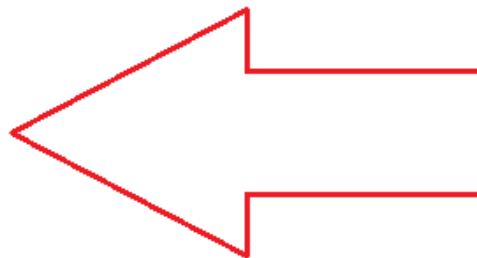
Kot zadnjo možnost programa v osnovnem meniju v nadzorni plošči program ponudi kartico za pomoč, v kateri so zgoščeni vsi ukazi in vsi pripomočki, ki jih uporabnik potrebuje. Prikazana je na sliki 8.

Slika 8: Začetni meni kartice za pomoč

Common commands in Speech Recognition

You can print this topic for quick reference while you're using Windows Speech Recognition.

- ▶ [How do I use Speech Recognition?](#)
- ▶ [Common Speech Recognition commands](#)
- ▶ [Dictation](#)
- ▶ [Keyboard keys](#)
- ▶ [Punctuation marks and special characters](#)
- ▶ [Controls](#)
- ▶ [Windows](#)
- ▶ [Click anywhere on the screen](#)



Was this helpful?



Vir: lasten

V kartici se nahaja tudi možnost za »on line« pomoč, ki uporabniku omogoča postavljanje vprašanj, na katera nato odgovorijo zaposleni v podjetju Microsoft. Hkrati program sprašuje uporabnika, ali je pomoč, ki je bila samodejno dana s strani programa, kaj pomagala pri reševanju problema. Pri tej možnosti je dobro, da si uporabnik zanjo vzame čas in Microsoftu posreduje potrebne informacije, saj danes nadgradnja programov deluje prav s pomočjo tovrstnih povratnih informacij s strani uporabnikov.

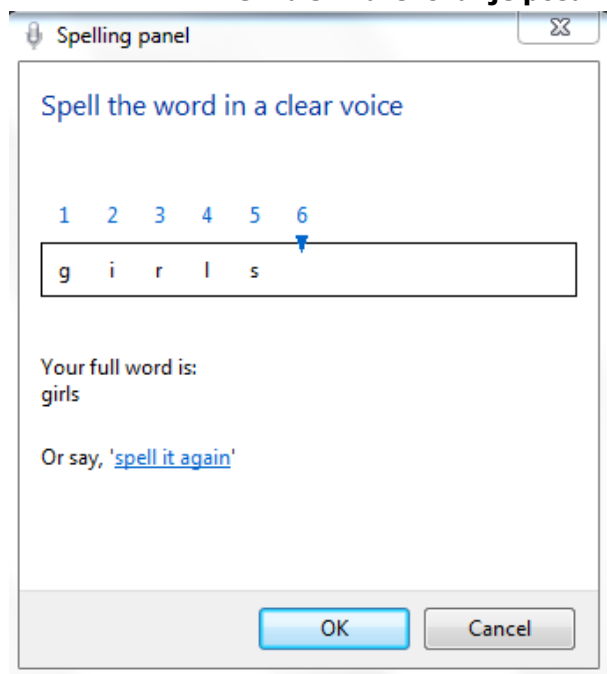
Uporabnik med uporabo dokaj hitro ugotovi pomanjkljivosti, ki bi jih programerji zelo težko odkrili. Kasneje poroča, kako uspešno je bila napaka odpravljena, kar je v veliko pomoč vsem programerjem.

Sam sem pri uporabi programa nekajkrat uporabil kartico za pomoč. Z njo sem si zelo veliko pomagal, zato menim, da je izdelana kakovostno, uporabno in prijazno do uporabnika.

Druga zanimiva možnost, ki jo program Speech Recognition ponuja, je črkovanje. Možnost je uporabna, ko uporabnik pride do besede, ki je navkljub večkratnemu ponavljanju, program nikakor ne razpozna. Še posebno se to izkaže za koristno pri nareku lastnih imen in tujk, ki nimajo prevoda v angleški jezik. Takrat ima uporabnik na voljo funkcijo »Spell it«. Ta funkcija v posebnem oknu omogoči narekovanje posameznih črk, ki sestavljajo besedo.

Spodnja slika 9 prikazuje primer. Po končani narekovani besedi, sestavljeni iz posameznih črk, lahko program to besedo doda v svoj slovar; pri tem pa nam program ponudi glasovni izrek besede, ki ga sam zabeleži, in je tako naslednjič beseda lažje prepoznana.

Slika 9: Narekovanje posamezne črke za sestavo besede



Vir: lasten

4 POTREBNA OPREMA ZA UPORABO PROGRAMA SPEECH RECOGNITION

V osnovi uporabniki programa za prepoznavanje govora s pomočjo tehničnih pripomočkov, ki pretvarjajo akustično energijo v prepoznavni digitalni zapis, upravljajo z računalnikom tako, kot bi uporabljali naprave, prirejene za roke. Program realizira vnos preko izvršenih ukazov ali izdelanega besedila, ki ga izpiše na zaslon. Pri tem pa so potrebne tri stvari:

- mikrofoni,
- računalnik z operacijskim sistemom,
- zvočna kartica.

Računalnik, ki sem ga uporabljal ob izdelavi diplomske naloge, je znamke Samsung, model R540. Premore 2.53 GHz procesne moči Intelovega procesorja i3, ki poganja in vzdržuje operacijski sistem Windows 7. Pri tem mu pomaga 3 GB trenutnega spomina in 280 GB spomina na trdem disku. Računalnik je bil kupljen v ZDA leta 2011.

Mikrofon, ki sem ga uporabil, je nameščen na okvir slušalk znamke Canyon, model CNR-HS7. Vrsta mikrofona je opredeljena kot zunanji, delovanje Mono, občutljivost mikrofona pa je zabeležena kot -58 dB. Mikrofon je bil kupljen v Sloveniji leta 2012.

Kljub temu, da sem za uporabo programa in testiranje uporabljal prenosni računalnik in slušalke z mikrofonom nižjega razreda, sem testiranje vseeno opravil brez večjih zapletov. Verjetno bi bila ob uporabi boljšega mikrofona uspešnost večja in bi tako prej zaznal napredek pri prepoznavanju besed, saj boljši mikrofoni bolje odstranjujejo šume iz okolice, ki so ves čas prisotni. Verjamem pa, da višja kakovost računalnika na rezultate testiranja ne bi imela posebnega vpliva, saj so danes vsi računalniki že toliko zmogljivi, da program, kakršen je Speech Recognition, obvladujejo brez težav.

4.1 MIKROFON

4.1.1 KRATKA ZGODOVINA RAZVOJA MIKROFONA

Prvi mikrofoni se je pojavil leta 1876, njegov avtor pa je bil Emile Berliner. Takratni mikrofoni seveda ni bil takšen, kot pa poznamo danes, vendar je predstavljal neko osnovo. Vgrajen je bil v takratne telefonske pretvornike glasu (Bellis, 2012).

Kasneje, leta 1878, pa je David Edward Hughes izumil mikrofoni, kakršen se uporablja še danes. Večji vzpon razvoja mikrofona je sledil po letu 1942, ko so izumili radijske postaje in prek njih pošiljali zvočne zapise (Bellis, 2012).

V letu 1964 je podjetje Bell Laboratories s pomočjo znanstvenikov Jamesa Westa in Gerharda Sesslerja prejelo patent za elektroakustični pretvornik, ki je ponujal večjo zanesljivost in natančnost. Omenjena znanstvenika sta za tiste čase ponudila nizko ceno, velikost samega mikrofona pa je bila precej manjša kot dotlej. To je bila mala revolucija v razvoju mikrofona in zaslediti je podatek, da je bilo prodanih skoraj milijarda teh mikrofonov na leto (Bellis, 2012).

Slika 10: Logotip podjetja Bell Laboratories



Vir: (Bell Laboratories , 2012)

Dinamični in kondenzatorski mikrofoni so bili razviti po letu 1970. Dopuščali so občutljivost na nižje tone in čistejšo elektronsko zapisovanje zvoka.

Razvoj mikrofonov se je nato po letu 1990 samo še stopnjeval, tako danes poznamo dinamične, kondenzatorske, keramične, ogljene, kristalne ter laserske mikrofone (Jež in Fašalek, 2011, str. 4).

4.1.2 TEORIJA DELOVANJA MIKROFONA

»Mikrofon je naprava oziroma pretvornik, ki spreminja nihanje zračnega tlaka v izmenično električno napetost. Tlak zračnega valovanja, ki deluje na membrano mikrofona, povzroča, da membrana niha v ritmu zračnega valovanja« (Jež in Fašalek, 2011, str. 4).

Sklepamo lahko, da je mikrofon pretvornik energije, ki akustično energijo ali zvok pretvarja v mehanično in na koncu v električno. Delovanje lahko primerjamo s človeškim ušesom, vendar pa so zmogljivosti slednjega bistveno manjše in slabše. Kot sta zapisala avtorja Pelicon in Doplihar v svojem raziskovalnem delu, uho zazna zvok od 20 Hz do 20.000 Hz, medtem ko mikrofon zazna tudi zvok, ki presega 100.000 Hz. To pomeni, da lahko mikrofon zazna in identificira tako glasen zvok, ki bi močno poškodoval ali celo trajno uničil človekovo uho (Doplihar in Pelicon, 2011, str 2).

Avtorja Peter Jež in Jernej Fašalek v svojem raziskovalnem delu delita mikrofone na šest skupin, in sicer glede na postavitev v prostoru, namembnost, impedanco, način dela, usmerjenost in zgradbo (Jež in Fašalek, 2011, str. 4).

Glede na zgradbo pa lahko mikrofone delimo na naslednje skupine: dinamične, kondenzatorske, keramične, ogljene, kristalne in laserske (Jež in Fašalek, 2011, str. 4).

»Če potrebujemo mikrofon, s katerim bomo snemali izključno govor, bi nam zadostovala karakteristika, ki je vodoravna med 300 in 3000 Hz, za osnovni prenos glasbe pa vsaj od 100

do 5000 Hz. Za kakovostno studijsko snemanje glasbe je potrebna karakteristika vsaj med 30 in 16.000 Hz«. (Doplihar in Pelicon, 2011, str. 3)

Iz zgornjega citata je razvidno, da za uporabo programa Speech Recognition potrebujemo mikrofona, ki ustreza vsaj minimalnim standardom kakovosti. Vendar je v tem razredu mikrofonov prisoten velik cenovni razpon. Sam sem za uporabo programa uporabljal star mikrofonski komplet s slušalkami, kupljen v trgovini z mešanimi izdelki. Cena mojega mikrofona je bila med 15 in 20 evrov. V nadaljevanju natančneje opišem mikrofona in računalnik, katerega sem sam uporabljal.

Za mikrofona, ki se uporablja pri programu Speech Recognition, je pomembna njegova sposobnost izločanja šumov iz okolice. Šum je zvočna motnja, ki ga proizvaja okolica, in je po navadi ljudem zaradi pozornosti neslišna. Mikrofona je izdelan tako, da preusmeri pozornost zgolj na zvok, ki prihaja neposredno v njegovo smer. Seveda pa se dostikrat zgodi, da ob večjih ali manjših zvočnih motnjah mikrofona preusmeri pozornost na zunanji dejavnik in tako narobe ali popačeno razume ukaz ali besedo.

4.2 RAČUNALNIK Z OPERACIJSKIM SISTEMOM

Med osnovno opremo sodi tudi računalnik, na katerem je nameščen operacijski sistem. Ker je program Speech Recognition last podjetja Microsoft Corporation, bom v diplomski nalogi pisal samo o operacijskih sistemih Windows podjetja Microsoft Corporation. Trenutno je na trgu več programov, s katerimi lahko uporabniki glasovno ukazujejo ali narekujejo besedila računalniku, ki jih ta nato izpisuje na zaslon. Kot sem v uvodu že napisal, je za dobro delovanje programa, ki sem ga preučeval, potreben najmanj operacijski sistem različice Windows XP, ki je izšel leta 2001 (Microsoft corporation, 2012). To pomeni, da je mogoče uporabljati tudi starejše računalnike, saj računalniku iz leta 2001 dodeljemo visoko starost zaradi sorazmerno hitrega razvoja tehnologije (Voicerecognition.com, 2012).

4.3 ZVOČNA KARTICA

Zvočna kartica je danes osnovna oprema vsakega računalnika. Standardni del matične plošče je postala z nastankom procesorja Pentium.

Zvočna kartica je sekundarni sprejemnik zvoka in primarni pošiljatelj zvoka. To pomeni, da računalnik zvok prejme preko mikrofona, ki je zato primarni prejemnik. Pošlje ga v zvočno kartico, ki se nahaja na matični plošči v računalniku. Mikrofona je zato primarni, zvočna kartica pa sekundarni prejemnik zvoka. Primarni pošiljatelj zvoka pa je zato, ker digitalni zapis, ki ga prejme iz računalnika, spremeni v zapis, ki ga nato predvajajo zvočniki. Zvočne kartice Sound Blaster so bile prve, ki jih je podjetje IBM vstavljalo v svoje osebne računalnike. To se je dogajalo vse dokler na trg niso prišli računalniki, ki so že imeli vgrajene zvočne kartice v matične plošče. Tako so predstavljali nov standard opreme osebnih računalnikov. Sprva so računalniki proizvajali samo kratka, tiha zvočna opozorila, ki jim ni

bilo mogoče spreminjati frekvence, glasnosti in trajanja zvoka. Večji razvojni napredki so se pričeli konec 20. stoletja, ko računalniki zmorejo proizvajati vse vrste možnih glasov, tudi v 3D obliki.

Če citiram avtorico prispevka na spletni strani Wikipedije, je zvočna kartica posrednik med računalnikom, ki komunicira digitalno, in svetom okoli njega, ki komunicira analogno (Babic, 2006).

»Najbolj osnovna zvočna kartica je tiskano vezje, ki vsebuje:

- analogno-digitalni pretvornik (ADC),
- digitalno-analogni pretvornik (DAC),
- ISA ali PCI vmesnik, ki povezuje kartico z matično ploščo,
- vhodno-izhodne povezave za mikrofona in zvočnike.

Če želi uporabnik posneti svoj glas, mora mikrofona priključiti na računalnik (torej na zvočno kartico) in vanj govoriti. ADC pretvori analogni zapis glasu v digitalnega. To naredi tako, da zvočni zapis razseka na kratke intervale. Število meritev na sekundo se imenuje frekvenca vzorčenja. Večja ko je slednja, več podatkov dobi računalnik in bolj natančno je zvočni val opisan« (Babic, 2006).

5 ANALIZA REZULTATOV RAZISKOVALNEGA DELA

5.1 NAREKOVANO BESEDILO IN NJEGOVA ANALIZA

Spodaj je citirano besedilo, ki sem ga vsak dan narekoval programu. Besedilo vsebuje 309 besed in je sestavljeno iz enostavnih, osnovnih angleških besed. Knjigo Slovenian Folk Tales sem v ta namen uporabil zato, ker sem menil, da bo lažje, če za prve besede uporabim preproste angleške besede Tako sem si izbral pravljice za otroke. Čeprav je besedilo zelo enostavno, sem imel pri nekaterih besedah veliko problemov. Zato se je vsaj na začetku, čas narekovanja močno zavlekel. Kot bom v naslednjem poglavju ponazoril s podatki v grafu in tabeli, se je zaradi zatikanja pri nekaterih besedah čas narekovanja iz začetnih dobrih 16 minut zmanjšal zgolj na slabih 6 minut narekovanja. Tukaj moram poudariti, da pred 27. julijem 2012, ko sem začel narekovati in testirati program, na svojem računalniku nikoli nisem uporabljal programa Speech Recognition. Zato sem sem za uvajanje in navajanje programa potreboval določen čas.

Slika 11: Narekovano besedilo

»Once upon a time, in Istria near the sea, a young boy was tending some cows and some sheep and goats.

It was afternoon, and the heat of the sun was burning. Suddenly, the boy saw three beautiful girls asleep on the soft grass. They were fairies. They were incredibly pretty, and looked very much like each other, in fact they looked almost the same.

The boy never thought they might be fairies. He thought they were just girls who had tired themselves in the sun and fallen asleep.

But the sun will burn them, he taught to himself. Such pretty faces, I must help them.

He climbed the nearby lime-tree, broke off some thick branches and made of shelter from the sun for the girls.

Soon the fairies awoke and got up. They were surprised and asked each other who could have been so merciful to shield them from the sun.

But they knew very well what had happened, as fairies never sleep but only pretend to be asleep. They asked questions only to see whether the boy would speak up or not.

But the shepherd didn't say a word, he even tried to run away for he couldn't stand looking at the fairies, their hair gleamed too much. It glittered like pure gold.

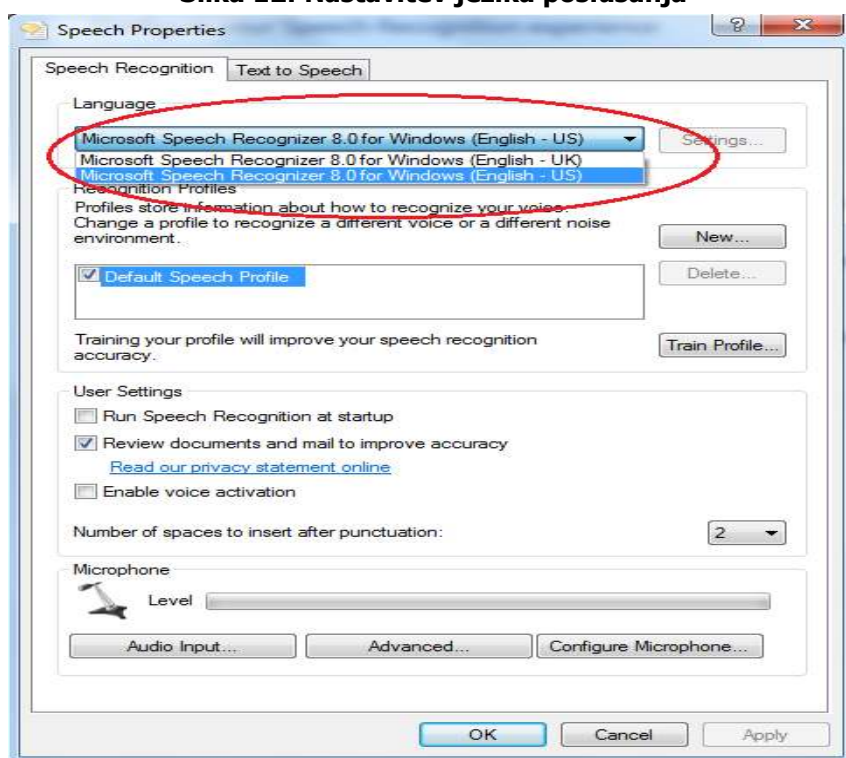
In an instant all three were by his side. He couldn't run away. They asked him what he wanted for having sheltered them from the burning sun. But the boy dared ask for nothing. They offered him a miraculous purse, which would forever stay for of gold coins. But the shepherd didn't want it for he didn't know the value of money. He didn't know what to make of it, to play with it he didn't want, for money is a dead thing, and he had live cows and sheep which he loved more than anything.«

V besedilu so podčrtane besede, s katerimi sem imel največ težav. Pri podčrtanih besedah sem moral uporabiti funkcijo črkovanja (*angl. »Spell it«*), ki jo program ponuja, kadar po večkratni ponovitvi izgovarjave besede program še vedno ni prepoznal prave.

Skozi čas uporabe programa se je prepoznavnost vidno izboljševala in tako tudi nisem imel več veliko problemov z besedami, ki sem jih prej moral črkovati. Besede, ki so mi delale največ težav, so bile: 'gleamed', 'purse', 'Istria', 'lime-tree', 'glittered'. Samo besedi 'purse' in 'gleamed' sem na koncu osvojil tako, da sem ju lahko izgovoril in ju je program nemudoma zaznal in izpisal v besedilo. Ostale tri pa program do konca raziskave ni prepoznal in sem jih moral vnašati v besedilo s pomočjo možnosti črkovanja. Zelo sem se trudil za pravilno izgovarjavo teh besed, vendar mi to ni nikakor uspevalo. Poizkušal sem vse vrste narečij, menjal poudarke na črkah, itd., a program vseeno ni in ni zaznal teh besed.

Dokler sem programu narekoval v angleški različici, sem imel veliko težav tudi pri besedi 'girls'. Nisem je mogel izreči tako, da bi me program razumel. Nekajkrat sem bil primoran uporabiti funkcijo črkovanja. Ko uporabnik večkrat zapored poskuša izgovoriti določeno besedo, se mu jezik še dodatno zapleta. Zato je razumljivo, da kakšna beseda programu ni prepoznavna. V trenutku, ko sem v jezikovnih nastavitvah programa spremenil jezik iz angleščine v ameriško angleščino, se napaka ni več pojavila. Izgled okna, kjer sem spremenil jezik, je prikazan na sliki 12.

Slika 12: Nastavitev jezika poslušanja



Vir: lasten

Sprememba nastavitve jezika iz angleške angleščine v ameriško angleščino je bistveno izboljšala razpoznavnost in s tem natančnost prepoznavanja besed. Morda zato, ker nas dnevno obdaja ameriška angleščina. Čeprav na internetu, televiziji in v knjigah prebiramo besedila v angleškem jeziku se ne zavedamo, da je to pravzaprav ameriška angleščina. Predvsem se nanjo navajamo ob gledanju filmov, ki v veliki meri prihajajo iz Združenih držav Amerike. Še bolj kot v pisanju angleških besedil se to opazi pri izgovarjavi ali pri poslušanju. Tako lahko opazimo razliko in ločimo med govorom Angleža in Američana, medtem ko pri prebiranju besedila to razliko težje zaznamo. Večinoma gre pri slednjem za skoraj neopazne razlike, kot so npr. vidne pri besedah 'organisation' (angleška verzija) in 'organization' (ameriška). Govorim seveda iz izkušenj občasnega uporabnika angleškega jezika in ne nekoga, ki se z njim aktivno ali profesionalno ukvarja.

Poleg citiranega besedila sem programu narekoval druga besedila, ki so bila bolj ali manj zapletena. Le tako sem lahko zagotovil dovolj vaje in navajanja s strani programa na moj glas in izgovarjavo. Sočasno sem se učil in privajal na program in njegove zmogljivosti. Veliko sem se delal s programom na uporabi operacijskega sistema kot takega. Program sem poizkusil tudi na aplikaciji spletnega brskalnika ter prebiral spletne strani in se sproti učil ukazov, ki so potrebni za upravljanje drugih aplikacij na svojem računalniku. Ugotovil sem tudi, da je ukazovanje programu za delo z računalnikom povsem nova tema, o kateri se v tej diplomski nalogi ne bi opredeljeval niti je resneje utemeljeval.

5.2 REZULTATI RAZISKOVALNEGA DELA TER ANALIZA

V spodnjem grafu (Graf 1) je prikazana krivulja rasti uspešnosti prepoznavanja besedila. Meril sem besedilo, ki sem ga prikazal na sliki 11. Za prikaz učinkovitosti programa sem uporabil zelo preprosto formulo izračuna prepoznavnosti, čeprav je v literaturi zaslediti tudi bolj sofisticirane metode: (Gaikwad in dr., 2010)

$$\text{uspešnost} = \frac{\text{število pravilno prepoznanih besed}}{\text{vseh besed}} * 100$$

Podobno meri tudi Peček v svoji raziskavi (Peček, 2004).

Krivulja se začne z datumom 27. 7. 2012 pri rezultatu 83,5 %, kar pomeni, da je bilo na ta dan brez napak prepoznanih 83,5 % besed, kar znaša 258 besed od 309. V celotnem besedilu je bilo največ 9 in najmanj 4 neprepoznanih besed. To pomeni, da kljub uporabi možnosti »correct that« program ni prepoznal besede, zato sem moral uporabiti funkcijo »spell it«, kjer sem nato narekoval vsako črko posebej, da sem sestavil končno besedo.

Prvi rezultat je slabši od raziskave, ki jo je objavil Peček (2004). Tukaj moram izpostaviti, da sem se takrat trudil z angleško različico angleščine.

Krivulja je v splošnem pogledu rasla. Vendar, kot je razvidno iz njenega gibanja, je na nekaterih mestih močno upadla. V celotni raziskavi sta bila dva trenutka, ki sta močnejše

spremenila gibanje krivulje. Prvi dogodek se je zgodil 1. 8. 2012, ko sem imel zaradi prehlada drugačen glas in me program ni razumel najbolje. Tako me program ni prepoznal kot tisti glas, ki sem ga uvajal skozi čas. V tej točki se vidi odvisnost osebne vezanosti uporabnika na program. Menim tudi, da je program potreboval nekaj časa, tudi ko sem dobil pravi glas nazaj, da se je zopet navadil na moj način izgovorjave. To dokazuje tudi krivulja, saj je potrebovala kar nekaj dni, da je prišla zopet na raven pred 1. 8. 2012.

Nato se krivulja počasi vzpenja in raste, nato pa, kot je razvidno, zopet pride dogodek, ki je še bolj izrazito kot prvi spremenil smer krivulje navzgor.

Drugi dogodek se je zgodil 13. 8. 2012, ko sem v nastavitvah programa Speech Recognition nastavil jezik poslušanja iz angleščine v ameriško angleščino. Sprva sem želel imeti nastavljen na angleško angleščino. Mentor mi je že na začetku svetoval, naj imam nastavljen program na ameriško angleščino (glej sliko 12), vendar sem bil mnenja, do bodo rezultati boljši na angleški verziji, ker je knjiga napisana v tej jezikovni različici. Izkazalo se je, da me program veliko bolje razume z ameriško angleščino.

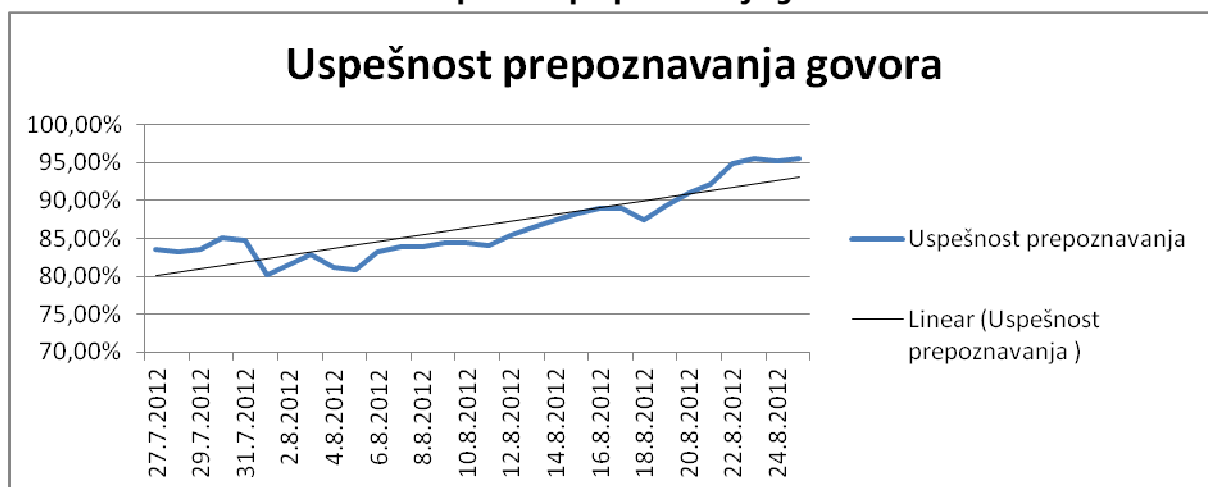
Kot sem že omenil v diplomski nalogi, sem sam uporabljal mikrofonski set, ki vsebuje mikrofona, ki je pritrjen z premično ročico na okvir slušalk. Mikrofona spada v kategorijo slabših in cenejših proizvodov in je tako dosegljiv za praktično vsakogar.

Pri začetku raziskave, po še zelo slabih izkušnjah, ki sem jih imel z namestitvijo, in pred spremembo jezika na ameriško različico, sva z mentorjem narekovala isto besedilo vsak v svoj računalnik in svojo opremo. Ugotovila sva, da je imel mentor boljši mikrofona, vendar je bila v prepoznavnost obeh podobna. Mentor je takrat besedilo videl in narekoval računalniku prvič, tako da je njegov slab rezultat razumljiv. Menim, da boljša in kakovostna oprema, kot je mikrofona ali sam računalnik, zanesljivo izboljša uspeh.

Vendarle pa sem v diplomski nalogi želel prikazati, da je lahko uporaba programa Speech Recognition prav tako uspešna, če uporabljamo osnovno opremo oziroma z nižjega kakovostnega razreda.

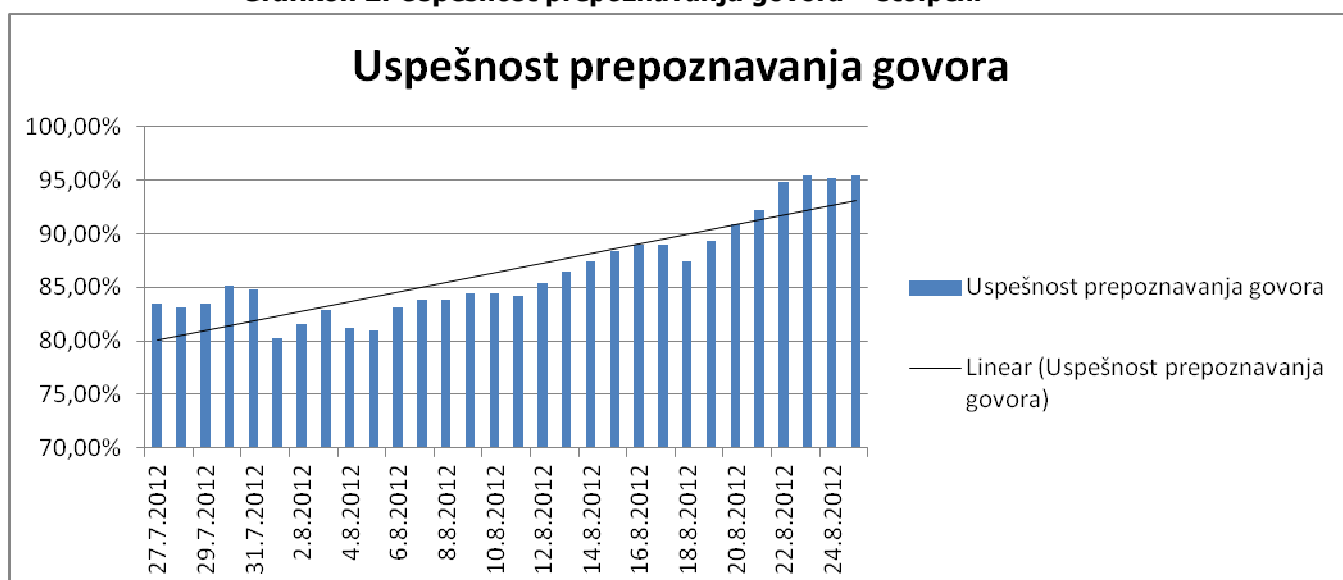
Za zanimivost sem omenjeno besedilo poizkusil narekovati s pomočjo primarno vgrajenega mikrofona, ki se nahaja na ohišju prenosnega računalnika. Ta mikrofona običajno uporabljamo za internetno video klicanje in ni namenjen strokovnemu in resnejšemu delu. Rezultati so bili porazni. Če bi ocenil uspešnost programa, je bila verjetno okoli 10-odstotna, ali pa še manj. Tako sem se prepričal, da je za uporabo programa za prepoznavanje govora resnično potrebujemo mikrofona, ki je nameščen 2,5 cm od ust. Tako, kot to priporoča vodič po programu.

Grafikon 1: Uspešnost prepoznavanja govora - črtni



Vir: lasten

Grafikon 2: Uspešnost prepoznavanja govora – stolpčni



Vir: lasten

V graf sem dodal tudi trendno črto, ki prikazuje povprečno gibanje osnovne krivulje ter tako kaže smer in naklon gibanja primarne krivulje.

V nadaljevanju predstavljam tabelo, v kateri sem datumsko prikazal uspešnost prepoznavanja govora med svojim testiranjem (Tabela 1).

Tabela 1: Prikaz rezultatov uspešnosti prepoznavanja govora

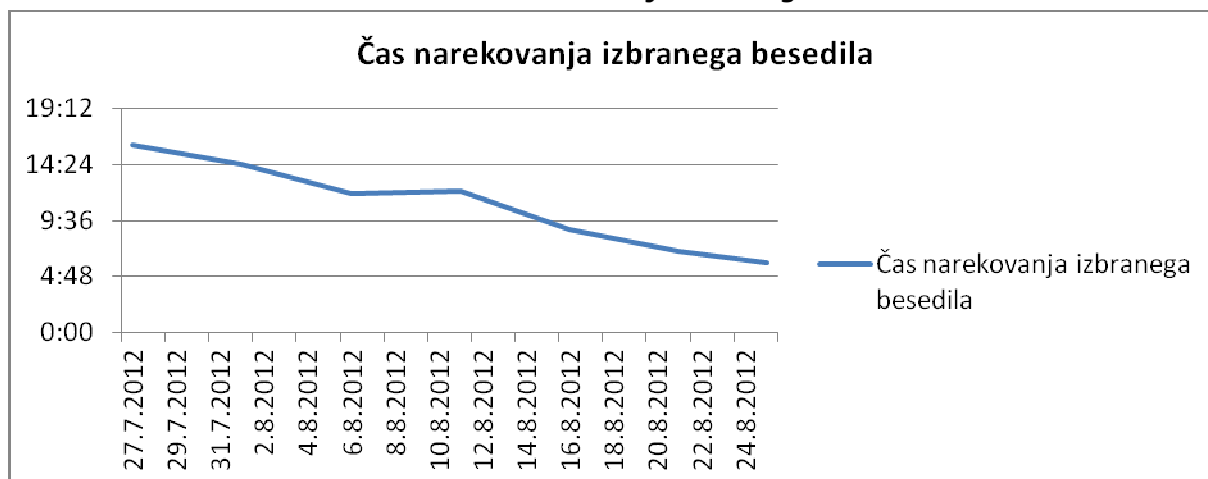
27.7.2012	83,50%	11.8.2012	84,14%
28.7.2012	83,17%	12.8.2012	85,44%
29.7.2012	83,50%	13.8.2012	86,41%
30.7.2012	85,11%	14.8.2012	87,38%
31.7.2012	84,79%	15.8.2012	88,35%
1.8.2012	80,26%	16.8.2012	89,00%
2.8.2012	81,55%	17.8.2012	89,00%
3.8.2012	82,85%	18.8.2012	87,38%
4.8.2012	81,23%	19.8.2012	89,32%
5.8.2012	80,91%	20.8.2012	90,94%
6.8.2012	83,17%	21.8.2012	92,23%
7.8.2012	83,82%	22.8.2012	94,82%
8.8.2012	83,82%	23.8.2012	95,47%
9.8.2012	84,47%	24.8.2012	95,15%
10.8.2012	84,47%	25.8.2012	95,47%

Vir: lasten

Med prilogami, je prikazana tabela, ki sem jo uporabljal za testiranje in prikazuje podrobnejše številke merjenj.

Spodnji graf (Graf 3) in tabela (Tabela 2) prikazujeta čas narekovanja danega besedila. Z narekovanji sem začel 27. 7. 2012 in končal 25. 8. 2012. Kot lahko vidimo, je gibanje krivulje pozitivno in se s časom vadbe manjša. Tako je iz tabele razvidno, da sem iz časa šestnajst minut (16) in sedem (7) sekund, ki sem ga zabeležil na začetku testiranja, skozi sedem vmesnih merjenj prišel na čas pet minut (5) in sedeminpetdeset (57) sekund. To pomeni, da se je skozi trideset dni testiran čas nareka bistveno izboljšal.

Grafikon 3: Čas narekovanja izbranega besedila



Vir: lasten

Menim, da je bilo sedem vmesnih merjenj dovolj, da se prikaže dejanski rezultat trajanja narekovanja besedila. Zato sem vsaki peti dan narekovanja meril in na koncu zabeležil čas. Rezultati merjenj so prikazani v spodnji tabeli.

Tabela 2: Čas, potreben za narekovanje danega besedila na določen dan

Datum narekovanja	Čas narekovanja
27.7.2012	16.07 minut
1.8.2012	14.27 minut
6.8.2012	11.55 minut
11.8.2012	12.08 minut
16.8.2012	8.47 minut
21.8.2012	6.56 minut
25.8.2012	5.57 minut

Vir: lasten

Verjetno največje presenečenje testiranja programa sem doživel na koncu, ko sem izbrano besedilo natipkal ročno. V srednji šoli smo v učnem načrtu imeli predmet, pri katerem smo se učili desetprstno slepo tipkati. Predmet je bil obvezni predmet učnega procesa in je bil na

koncu šolskega leta tudi ocenjen, tako kot vsak drug. Skozi uporabo računalnika, interneta ter nadaljnega šolanja sem hitrost slepega tipkanja še močno izboljšal. Sedaj skoraj ne delam več napak. Tako pri slepem tipkanju izgubljam izjemno malo časa za popravljanje morebitnih napak, lahko bi rekel, da ta čas ni vreden omembe.

Na začetku testiranja sem bil zato prepričan, da tipkam mnogo hitreje, kot lahko program posluša in zapisuje moje izrečene besede in ukaze za korekcijo. Na začetku sem namreč porabil veliko časa za odpravljanje napak, ki jih je naredil program, kadar ni 'razumel' besede. To se vidi tudi iz rezultatov, saj sem prvič porabil kar 16 minut za narekovanje celotnega besedila programu. To je bilo veliko več, kot ga potrebujem za ročno tipkanje besedila.

Za tipkanje celotnega besedila sem potreboval 8 minut in 47 sekund. Prepričan sem bil, da mi v 30 dneh ne bo uspelo osvojiti programa za prepoznavanje govora do stopnje, ko bi prekosil čas tipkanja. Predpostavka se je izkazala za napačno, saj sem sodeč po podatkih ta čas osvojil 16. 8. 2012, kar predstavlja 21 dni od prvega testiranja. V zvezi s časom 8 minut in 47 sekund naj omenim, da sem za dosego tega časa s tipkanjem hitel, zaradi česar so me po končanih 309 besedah oziroma 1625 natipkanih znakih močno boleli prsti in zapestje.

Poleg časovne pridobitve, ki jo nudi program Speech Recognition v narekovanju besedila, sem ugotovil tudi prednost udobja takšne uporabe računalnika. Zgolj v slabih devetih minutah ročnega tipkanja danega besedila sem zasledil bolečine v rokah, ki jih pri narekovanju v program ne bi. Tako je prednost programa tudi s stališča ohranjanja zdravja izrazita in dobrodošla. Predvidevam, da profesionalcem, ki veliko tipkajo, na primer kakšno knjigo, bolečine v prstih, zapestju, komolcu in ramenih lahko predstavljajo veliko trpljenje.

6 ZAKLJUČEK

Kot sem v uvodu začel pisati o tem, kako je informacijska tehnologija dandanes nepogrešljiva, bi v zaključku rad to tudi prikazal in zapisano ovrednotil z rezultati svoje raziskave.

Skozi izdelavo diplomske naloge in testiranje programa Speech Recognition sem ugotovil, da znanje in uporaba računalnika nista zgolj nujna potreba za delo in vsakdanje življenje posameznika, podjetja, države, društva, fakultete, itd. Prišel sem do spoznanja, da si lahko vsakdanje delo in opravke z informacijsko tehnologijo v mnogih pogledih in na mnogih področjih bistveno skrajšamo in poenostavimo.

Program Speech Recognition sprva ni bil zasnovan za uporabnika v gospodinjstvu, na fakulteti ali morda v nekem podjetju. Na začetku je bila to le znanstvena fantastika, kot so danes mnoge druge stvari, ki jih vidimo v filmih. Toda zgodovina nas med drugim uči, da je veliko idej za tehnološki razvoj strokovna znanost dobila prav v filmih in drugih zvrsteh umetnosti, kjer domišljija nima meja. Tudi začetne zasnove o sistemih za prepoznavanje govora so znanstveniki s pomočjo programerjev prejeli na podoben način. Tako so se začeli za današnje čase majhni premiki v smer sistemov za prepoznavanje govora. Danes, nekaj desetletij po prvih primitivnih poizkusih izdelovanja takih sistemov, imamo na osebem računalniku, ki stane nekaj sto evrov, dovršen sistem prepoznavanja govora, ki se dnevno izboljšuje in samostojno dopolnjuje. Prvi začetki tovrstnega programa so bili namenjeni za vojsko in medicino. Prav programerji programa Speech Recognition pa so velik pomen dajali ljudem z nezmožnostjo uporabe tipkovnice ali miške. Ta program mnogim bolnikom po svetu, ki ne morejo uporabljati miške in tipkovnice, vsakodnevno lajša ali v celoti omogoča delo z računalnikom.

Tudi sam sem ob pisanju diplomske naloge spoznal, kako boleče je daljše tipkanje besedila brez prestanka, brez sproščanja in počivanja rok in prstov, ki so prikovani na tipkovnico ali miško. Tako si lahko samo predstavljam, kako težavno je to za nekoga, ki več ur dnevno neprekinjeno uporablja računalnik.

Hipotezo, ki sem si jo zadal pred pisanjem naloge, in sicer, da je lahko prepoznavnost programa Speech Recognition po 30 dneh uporabe, kar predstavlja približno 30 ur narekovanja, višja od 95 %, sem s pomočjo testiranja popolnoma potrdil. Sprva sem v to močno dvomil, saj so bili začetni rezultati porazni; še posebej ob upoštevanju podatkov o času, potrebnem za narekovanje izbranega besedila. Z vsakdanjo vajo pa sta se tako čas narekovanja, kot tudi uspešnost prepoznavanja besed tako izboljšala, da sta prehitela celo slepo desetprstno tipkanje.

Menim, da je program za končnega uporabnika dovolj prijazen ter tako enostavno dostopen, da ga lahko uporablja vsak izmed nas. V zakup je potrebno vzeti tudi nekaj potrpljenja, saj, kot pri vsaki stvari, tudi tukaj velja, da je vsak začetek težak. Tudi sam sem imel sprva močne dvome, ki so se na koncu izkazali za neupravičene.

Objektivno je potrebno priznati, da rezultati kažejo, da sem zastavljene ciljne rezultate težko in šele povsem na koncu osvojil. Delno je rezultat posledica pozne spremembe vrste angleščine iz angleške različice v ameriško. Zato menim, da bi bili rezultati ob stalni uporabi ameriške angleščine bistveno boljši, hkrati pa bi se hitreje kot sicer pokazal rezultat rasti uspešnosti prepoznavanja besed.

Poleg omenjene funkcije bi rad omenil, da sem uporabljal mikrofona najnižje cenovne vrednosti, ki ga trenutno najdemo na tržišču strojne računalniške opreme. Z gotovostjo trdim, da bi bili rezultati boljši in hitrejši ob uporabi novejših in tehnološko naprednejših mikrofonov.

Končni pregled rezultatov testiranja kaže, da mi je uspelo program navaditi na moj način narekovanja do 95,47 % prepoznavnosti. Ob upoštevanju napačnega izbora jezika poslušanja ter izbire mikrofona menim, da bi bil lahko rezultat dva do tri odstotke višji in bi na koncu prepoznavnost znašala morda celo 98 %.

Zato menim, da je bila v začetku raziskave zastavljena hipoteza, da je »v primernem prostoru in s primerno opremo doseči natančnost programa višjo od 95 %« v celoti dokazana in potrjena.

LITERATURA IN VIRI

1. B. H. Juang in R. Rabiner Lawrence, (2004). *Automatic Speech Recognition – A Brief History of the Technology Development*. Dostopno 22. 8. 2012 na: http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf.
2. Babic Lidija, (2006). *Zvočna kartica*. Dostopno 1. 9. 2012 na: http://wiki.fmf.uni-lj.si/wiki/Zvo%C4%8Dna_kartica.
3. Baichtal John, (2011). *Darpa to Fund Hackerspaces*. Dostopno 24. 08. 2012 na: <http://blog.makezine.com/2011/08/05/darpa-to-fund->.
4. Bell Laboratories, (2012). *Bell Laboratories*. Dostopno 27. 8. 2012 na: <http://www.belllabs.com/>.
5. Bellis Mary, (2012). *The history of Microphones – Microphones convert sound waves into electrical voltages*. Dostopno 27. 8. 2012 na: <http://inventors.about.com/od/mstartinventions/a/microphone.htm>.
6. BENIGAR Jože, ZOKŠEVA Eva, VALJAVEC Matija, ŠAVLI Andrej, MATIČETOV Milko, MILČINSKI Fran, (2002). *Slovenian Folk Tales*. Mladinska knjiga, Ljubljana.
7. CHUNRONG Lai, Shih-Lien Lu, Qingwei Zhao, 2004. *Performance Analysis of Speech Recognition Software*. V: . Russell Clapp, Kimberly Keeton (ur.), Fifth Workshop on Computer Architecture Evaluation using Commercial Workloads, Cambridge, Massachusetts.
8. GAIKWAD K. Santosh, (2010). *A Review on Speech Recognition Technique*. »V:« BHARTI W. Gawali (ur.): International Journal of Computer Application. University Aurangabad, str. 16-24.
9. HERBIG Tobias, GERL Franz, MINKER Wolfgang, (2011). *Self-Learning Speaker Identification (A System for Enhanced Speech Recognition)*. Springer, Berlin.
10. Hou Yunzhong, (2012). *Speech Recognition*. Dostopno 31. 8. 2012 na: http://future.wikia.com/wiki/Speech_recognition.

11. IBM, (2012). *IBM international recognition*. Dostopno 27. 8. 2012 na: http://www-03.ibm.com/ibm/history/exhibits/logo/logo_8.html.
12. Ionescu Daniel, (2008). *iPhone Gets Google Search By Voice*. Dostopno 22. 8. 2012 na: http://www.pcworld.com/article/153871/iphone_gets_google_search_by_voice.html.
13. Jež Peter in Fašalek Jernej, (2011). *Akustični senzorji*. Dostopno 22. 8. 2012 na: http://lirtme.fe.uni-lj.si/lirtme/slo/UNIVSS/daja_pret/seminarji_2011/Akusticni%20senzorji.pdf.
14. Mello P. John Jr., (1997). *NaturallySpeaking: Voice Recognition Leaps Forward*. Dostopno 22. 8. 2012 na: http://www.pcworld.com/article/4823/naturallyspeaking_voice_recognition_leaps_forward.html.
15. Microsoft corporation, (2012). *Zanimivosti iz prvih 25 let*. Dostopno 21. 8. 2012 na: <http://windows.microsoft.com/sl-SI/windows/history>.
16. PEČEK Bojan, 2004. *Use of discrete simulations as a tool for processes documentation*. V: Čičin-Šain Marina, Dragojlović Pavle, Turčić Prstačić Ivana (ur.), *Computers in education*. Hrvatska udruga za mikroprocesore, procesne i informacijske sustave, mikroelektroniku i elektroniku MIPRO HU, Rijeka, str. 70-75.
17. Pelicon Aljaž in Doplihar Matej, (2011). *Akustični senzorji (mikrofoni)*. Dostopno 25. 8. 2012 na: http://lirtme.fe.uni-lj.si/lirtme/slo/UNIVSS/meri_pret/seminar%202011/Porocilo%20Merlini%20pre tvorniki%20%20akusticni%20senzorji%20Doplihar%20Matej%20in%20Aljaz%20Pelicon.pdf.
18. Pinola Melanie, (2011). *Speech Recognition Through the Decades: How We Ended Up With Siri*. Dostopno 22. 8. 2012 na: http://www.pcworld.com/article/243060/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html.
19. SCHALK Peter in FOSTER Thomas, (1993). *Speech recognition*. CMP Books, California.
20. Squidoo, (2012). *Worlds of Wonder, Julie Talking Doll*. Dostopno 25. 8. 2012 na: <http://www.squidoo.com/julie-talking-doll>.

21. Voicerecognition.com, (2012). *What are the minimum hardware requirements that will allow me to use L&H Voice Xpress for Medicine 4.0?*. dostopno 31. 8. 2012 na: http://www.voicerecognition.com/products/lernout_hauspie/faq.html.
22. Voice-recognition-software-review.toptenreviews.com, (2012). *2012 Compare Best Voice Recognition Software*. Dostopno: 31. 8. 2012 na: <http://voice-recognition-software-review.toptenreviews.com/>
23. Wikipedia, (2012). *History of computing hardware*. Dostopno 1. 9. 2012 na: http://en.wikipedia.org/wiki/First_computer

PRILOGE

Priloga 1: Tabela podatkov testiranja

Datum testiranja	Neprepoznane besede	Popravljene besede	Prepoznane besede	Skupaj besede	Odstotek prepoznavnosti
27.7.2012	8	43	258	309	83,50%
28.7.2012	7	45	257	309	83,17%
29.7.2012	7	44	258	309	83,50%
30.7.2012	6	40	263	309	85,11%
31.7.2012	6	41	262	309	84,79%
1.8.2012	9	52	248	309	80,26%
2.8.2012	8	49	252	309	81,55%
3.8.2012	8	45	256	309	82,85%
4.8.2012	7	51	251	309	81,23%
5.8.2012	6	53	250	309	80,91%
6.8.2012	7	45	257	309	83,17%
7.8.2012	7	43	259	309	83,82%
8.8.2012	8	42	259	309	83,82%
9.8.2012	6	42	261	309	84,47%
10.8.2012	7	41	261	309	84,47%
11.8.2012	5	44	260	309	84,14%
12.8.2012	5	40	264	309	85,44%
13.8.2012	5	37	267	309	86,41%
14.8.2012	5	34	270	309	87,38%
15.8.2012	5	31	273	309	88,35%
16.8.2012	5	29	275	309	89,00%
17.8.2012	5	29	275	309	89,00%
18.8.2012	6	33	270	309	87,38%
19.8.2012	6	27	276	309	89,32%
20.8.2012	5	23	281	309	90,94%
21.8.2012	5	19	285	309	92,23%
22.8.2012	4	12	293	309	94,82%
23.8.2012	4	10	295	309	95,47%
24.8.2012	4	11	294	309	95,15%
25.8.2012	4	10	295	309	95,47%

Vir: lasten