

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA  
ZAVOD ZA ELEKTRONIČKE SUSTAVE I OBRADBU  
INFORMACIJA

Seminarski rad iz kolegija  
Podatkovni višemedijski prijenos i računalne  
mreže

## SPEECH RECOGNITION

Joško Rogulj, Joško Rokov, Nino Talian

Zagreb, siječanj 2006.

## Sadržaj

1. Sažetak.....	2
2. Uvod.....	3
3. Princip rada i primjene.....	4
4. Sustavi na tržištu.....	7
5. Zaključak.....	25
6. Literatura.....	26

## **1.Sažetak**

U ovom seminarskom radu dati ćemo temeljni opis tehnologije raspoznavanja govora, objasniti glavne pojmove vezane za temu, podjelu sustava te opis najkvalitetnijih komercijalnih i besplatnih alata za raspoznavanje govora trenutno. Poseban naglasak staviti ćemo na opis alata te njihovu kategorizaciju.

## **2. Uvod**

Prepoznavanje govora je proces kojim računalo (ili neki drugi stroj) identificira izgovorene riječi.

U osnovi to znači razgovor s računalom u kojem računalo ispravno prepoznaje naše riječi. Poznavanje slijedećih definicija je nužno za razumijevanje tehnologije prepoznavanja govora:

### ***Iskaz***

Iskaz je vokalizacija (izgovor) jedne riječi ili više riječi koji predstavlja neko jedinstveno značenje računalu. Iskaz može biti jedna riječ, nekoliko riječi, rečenica ili čak nekoliko rečenica.

### ***Ovisnost o govorniku***

Sustavi ovisni o govorniku su dizajnirani tako da se prilagođavaju određenom govorniku. Općenito su mnogo točniji za tog određenog govornika, ali zato puno manje precizni za ostale govornike. Pretpostavljaju da govornik govori istim tonom i tempom. Sustavi neovisni o govorniku su dizajnirani za veći broj raznovrsnih govornika. Prilagodivi sustavi obično počinju s radom kao sustavi neovisni o govorniku da bi kasnije tehnikama treniranja postali prilagođeni korisniku u svrhu povećanja preciznosti samog sustava.

### ***Rječnici***

Rječnici su liste riječi ili iskazi koji mogu biti prepoznati od strane sustava za prepoznavanje govora. Općenito, manji rječnici su lakši računalu za prepoznavanje, dok su veći rječnici teži. Za razliku od normalnih rječnika svaki ulaz ne mora nužno biti jedna riječ. Pojedini zapis u rječniku može biti u obliku jedne ili čak dvije rečenice. Manji rječnici mogu imati malo prepoznatih iskaza dok veoma veliki rječnici mogu imati do stotine tisuća i više!

### ***Točnost***

Sposobnost sustava za raspoznavanje može biti ispitana mjerenjem točnosti – odnosno kako dobro sustav prepoznaje iskaze. To uključuje ne samo točno prepoznate iskaze već i identificirane iskaza koji nisu sadržani u rječniku. Dobar sustav za raspoznavanje govora može imati točnost u iznosu od 98% i više! Prihvatljiva razina točnosti ustvari ovisi o specifičnoj primjeni sustava.

### ***Treniranje***

Neki sustavi za raspoznavanje govora imaju mogućnost prilagodbe govorniku. Ako sustav ima tu mogućnost tada nastupa treniranje. Sustav za raspoznavanje govora se trenira tako što traži od govornika da standardne izraze ili fraze čime podešava svoje algoritme usporedbe da što više odgovaraju aktivnom govorniku. Treniranje obično povećava točnost sustava. Treniranje se također može iskoristiti za govornike koji imaju problema s općenito govorom ili izgovorom nekih određenih riječi. Dokle god govornik može konzistentno ponavljati iskaze, sustav za raspoznavanje govora bi se trebao prilagođavati.

## **3. Princip rada i primjene**

### **Princip rada**

Sustavi za raspoznavanje govora mogu biti u dvije glavne skupine: *Sustavi za raspoznavanje uzoraka* uspoređuju ulazne uzorke s ugrađenim/naučenim uzorcima te nastoje pronaći odgovarajući par. *Zvučni fonetski sustavi* koriste znanje o čovjekovoj građi (prirodna sinteza govora i sluh) pri usporedbi karakteristika govora (fonema i glasova). Većina modernih sustava bazira se na pristupu baziranom na raspoznavanju uzoraka jer je on prilagođen modernim tehnikama procesiranja i teži ka većoj točnosti. Većina SRG-a može se rasčlaniti na slijedeće korake:

1. Snimanje govora i detekcija iskaza.
2. Predfiltriranje (pred-isticanje, normalizacija, izdvajanje pojasa, itd.)
3. Uokviravanje (sjeckanje podataka u uporabljiv format)
4. Filtriranje (filtriranje svakog okvira)
5. Usporedba (prepoznavanje iskaza)
6. Djelovanje (izvršavanje funkcije pridjeljene prepoznatom uzorku)

Iako se svaki korak čini jednostavnim, svaki može sadržavati više različitih (ponekad i potpuno različitih) metoda.

(1) *Snimanje govora i detekcija iskaza:* može biti postignuto na više načina. Početne točke moguće je odrediti usporedbom razine zvuka ambijenta (zvučne energije u nekim slučajevima) i tek snimljenog zapisa.

Detekcija završne točke je teža jer govornik uglavnom ostavi artefakte koji uključuju disanje, škr gutanje zubiju i jeku.

(2) *Predfiltriranje:* se postiže na razne načine ovisno o ostalim značajkama SRG-a.

Najčešća metoda je «Bank-of-Filters» metoda koja koristi cijeli niz audio filtera za pripremu zapisa, te metodu linearne predikcije koja računa razliku (pogrešku) predviđanja. Različiti oblici spektralne analize se također koriste (npr. kepstar).

(3) *Uokviravanje:* uključuje odvajanje uzorkovanih podataka u određenu veličinu. To se često ubraja pod korake 2 ili 4. Ovaj korak također uključuje pripremu granica za analizu (odvajanje rubnih pojava, itd.).

(4) *Filtriranje:* nije uvijek prisutno. To je posljednja priprema za svaki okvir prije usporedbe i slaganja. Često se ovaj korak sastoji od poredavanja u vremenu i normalizacije.

(5) *Usporedba:* Postoji niz raznih tehnika usporedbe. Većina koristi usporedbu trenutnog okvira s poznatim uzorcima. Postoje metode koje koriste skrivene Markovljeve lance, frekvencijsku analizu, diferencijalnu analizu, tehnike linearne algebre, metode distorzije u spektralnoj i vremenskoj domeni. Sve ove metode koriste se za generiranje parova po vjerojatnosti i točnosti.

(6) *Djelovanje:* može biti bilo koja akcija koju razvojni programer zamisli.

## **Tipovi raspoznavanja govora**

Sustavi za raspoznavanje govora mogu se podijeliti u nekoliko različitih klasa obzirom na tipove iskaza koje imaju mogućnost raspoznati. Ove klase su bazirane na činjenici da je jedan od problema SRG-a sposobnost određivanja trenutaka početka i kraja govornikova iskaza. Većina paketa može pripadati više nego jednoj klasi ovisno o načinu rada koji se koristi.

### ***Izolirane riječi***

Sustavi bazirani na izoliranim riječima obično zahtijevaju da svaki iskaz ima tišinu (izostanak audio signala) s OBJE strane okvira uzorka. To ne znači da prihvaća isključivo riječi nego zahtijeva da iskaze idu jedan za drugim.

Često ovi sustavi imaju «Slušanje/Neslušanje» stanja s kojima zahtijevaju od govornika da pričeka između izgovaranja dvaju uzastopnih iskaza (obično sustav radi procesiranje upravo za vrijeme pauze). Prikladniji naziv za ovu klasu bi ustvari bio izolirani iskazi.

### ***Povezane riječi***

Sustav baziran na povezanim riječima (ili točnije: povezanim iskazima) je sličan prethodno navedenom ali s razlikom da dozvoljava da se različiti iskazi procesiraju zajedno uz minimalnu pauzu između njih.

### ***Kontinuirani govor***

Kontinuirani govor je slijedeći korak. Sustav tih mogućnosti je najteže stvoriti jer mora koristiti specijalne metode određivanja granica iskaza. Sustavi bazirani na kontinuiranom govoru dozvoljavaju korisnicima korištenje prirodnog govora dok računalo određuje sadržaj. Načelno to je diktiranje računalo.

### ***Spontani govor***

Izgleda da postoji više definicija za spontani govor. Na osnovnoj razini to može biti govor s prirodnim zvukom i neuvježban. SRG baziran na spontanom govoru morao bi imati mogućnost prepoznavanja značajki spontanog govora kao što su spojene riječi, «um», «am» i čak lagana zamuckivanja (zastajkivanja).

### ***Provjeravanje i identifikacija govora***

Neki sustavi za raspoznavanje govora imaju mogućnost prepoznavanja određenih korisnika što može biti korisno za sigurnosne svrhe ali još uvijek nedovoljno pouzdano.

## **Primjene**

Iako mnogi zadatci koji koriste sučelje s računalom mogu potencijalno koristiti SRG slijedeće primjene su najčešće trenutno:

### ***Diktiranje***

Diktiranje je danas najčešća upotreba sustava za raspoznavanje govora. Ono uključuje liječničke recepte, poslovne diktate, kao i općenitu obradu riječi. U nekim slučajevima koriste se specijalni rječnici u svrhu povećanja točnosti sustava.

### ***Naređivanje i kontrola***

Sustavi za raspoznavanje govora koji su dizajnirani da izvode funkcije na sustavu definirani su kao «Sustavi za naredbe i kontrolu». Iskazi kao «Open Mozilla» i «Pokreni terminal» će učiniti upravo to.

### ***Telefonija***

Neki PBX ili sustavi za glasovnu poštu omogućavaju korisnicima izgovaranje naredbi umjesto pritiskanja tipki za slanje određenih tonova.

### ***Prijenosnici***

Zbog ograničenih ulaznih dodataka za prijenosne uređaje (PDA, mobiteli) govor je otvorena mogućnost.

### ***Medicinski (nedostatci)***

Mnogi ljudi imaju problem s tipkanjem zbog određenih fizičkih ograničenja kao što su: učestale ozljede naprezanja, mišićna distrofija i mnoge druge. Uzmimo za primjer ljude s slušnim problemima kojima je znatna pomoć sustav koji njihov dolazni telefonski signal pretvara u tekst.

### ***Ugrađeni sustavi***

Većina novijih mobilnih telefona ima mogućnost C&C raspoznavanja govora koje dopušta iskaze tipa «Zovi doma». A sutra bi možda mogli glasovnim naredbama upravljati nekim kućanskim aparatima ili svim elektroničkim uređajima!

## **4. Sustavi na tržištu**

Neki od komercijalnih alata dostupnih na tržištu su:

- 1) Dragon Naturally speaking (Nuance) -- <http://www.nuance.com/naturallyspeaking>
- 2) IBM ViaVoice -- <http://www.scansoft.co.uk/viavoice/>
- 3) SpeakToText (CoolSoft) -- <http://www.coolsoftllc.com>
- 4) VR Commander (Interactive Voice Technologies) -- <http://www.vrcommander.com/>
- 5) Audiomatic (Wise Riddles Software) -- <http://www.wiseriddles.com/Audiomatic>
- 6) e-Speaking (e-speaking) -- <http://www.e-speaking.com/>
- 7) Realize Voice Lite (Realize Software) -- <http://www.realizesoftware.com>



- 8) Voice Studio (Ultimate Interactive Desktops) -- <http://www.voicestudio.us>
- 9) Abbot (Softsound) -- <http://www.softsound.com>
- 10) Dictation Buddy (High Criteria) -- <http://www.highcriteria.com>
- 11) MacSpeech iListen -- <http://www.macspeech.com/>

Programi za prepoznavanje govora relativno su slični po svojim funkcijama. Gotovo svi imaju svojstvo upravljanja glasom unutar osnovnih aplikacija i većina ih ima mogućnost diktiranja teksta u razne tekst editore. Za ozbiljnu upotrebu i kvalitetan rad dva su daleko najbolja programa na tržištu i oni će u daljnjem tekstu biti detaljnije opisani. To su Dragon Naturally Speaking i Via Voice.

## **Dragon naturally speaking (Nuance)**

### *Glavne značajke*

- namijenjen za Windows OS
- visoka točnost prepoznavanja govora (preko 98% za engleski jezik)
- jednostavna instalacija te mogućnost upravljanja govorom gotovo svim aplikacijama pod Windowsima (uključujući MS Word, MS Excel, prijenosna računala i Tablet PC)
- jednostavno i brzo podešavanje (upoznavanje aplikacije sa korisnikovim glasom)
- diktiranje teksta u bilo koji tekst editor (Word, elektronička pošta)
- lako provjeravanje lekture (korisnik kaže: «Play that back!» pa mu program pročita tekst)
- ugrađena NBS tehnologija (Nothing But Speech) za ignoriranje svega osim čistog govora
- mogućnost odabira posebnog načina rada za slovanje, numeriranje, zadavanje naredbi ili diktiranje
- kontroliranje miša glasom
- web pretraživanje glasom
- formatiranje i preuređivanje teksta glasom
- jednostavne glasovne naredbe za inače složenije akcije
- automatsko umetanje interpunkcijskih znakova
- mogućnost slušanja teksta «najljudskijim» glasom dostupnim na tržištu
- mogućnost spremanja svog profila i njegove upotrebe na drugom računalu

- cijena: Preferred verzija oko 200 \$, Professional verzija oko 500 \$

### ***Instalacija***

Program je jednostavan za instalaciju na sustavima sa Windows 98 ili novijim OS. Ovisno o brzini računala instalacija traje do 20 min. Kod nadograđivanja, primjerice, verzije 7 na 8, instalacija briše prethodnu verziju i, ukoliko postoje veliki sačuvani korisnički profili, proces može potrajati i do 30 min. Verzija 8 zahtjeva online aktivaciju.

### ***Vježbanje i potrebno vrijeme***

Program zahtjeva vježbanje prije upotrebe, da bi točno prepoznao korisnikov govor. Vježba se sastoji od toga da korisnik pročita jedan ili dva teksta računalu da bi ono analiziralo korisnikov glas. Prije vježbe, program radi optimalnog rada vrši samopodešavajuće testove za mikrofoni i zvučni sustav koji traju oko 20 sekundi. Korisniku se pruža mogućnost ponavljanja dijela teksta kojim program nije zadovoljan, radi stvaranja boljeg glasovnog modela.

Za korisnike koji imaju poteškoća sa čitanjem postoje lakše skripte za vježbu, a postoje i skripte za djecu. Iako je font zadebljan i lak za čitanje, ne postoji mogućnost njegove promjene.

### ***Lakoća korištenja***

Najčešće se softver za prepoznavanje govora koristi za pisanje teksta neke vrste. U tu svrhu ovaj je program relativno lako koristiti nakon dobrog uvježbavanja. Neke glasovne naredbe manje su formalne nego kod drugih sličnih programa. Umjesto "move to end of line", može se reći "go to end of line" itd.

### ***Točnost prepoznavanja govora***

Program pruža visoku točnost prepoznavanja govora nakon dobrog treninga, ali uz dobar i ujednačen izgovor engleskog jezika (>98%). Stranicu teksta moguće je tako izdiktirati za nekoliko minuta, a što se više koristi, program točnije razumije korisnikov tekst.

### ***Integracija s ostalim aplikacijama***

Dragon je kompatibilan s većinom Microsoft programa, te je u njih tako moguće, primjerice, diktirati. Najbolje funkcionira sa MS Word-om, dok više poteškoća ima s Excel-om ili Access-om, budući da oni koriste znatno više resursa računala. To, dakle nije do softvera za prepoznavanje govora. Preporuča se izgovaranje u sam Dragon Naturally Speaking, to jest

u za to predviđeni Dragon pad, ta da se potom tekst kopira u željeni tekst editor. Što je jači računalni sustav, to se manje pojavljuje ovaj problem.

Program omogućuje diktiranje, ispravljanje, formatiranje, navigiranje i u drugim Windows aplikacijama, no često prije korištenja u nekim aplikacijama treba podesiti glasovne komande u programu za prepoznavanje. Inače može doći do aktiviranja neželjenih akcija unutar aplikacije. Moguće je zadati da se isključivo diktira, da ne bi program shvatio dio teksta kao neku naredbu.

### ***Zahtjevi za računalo i kompatibilnost s ostalim hardverom***

Za dobar rad računalo treba imati brz procesor i novije verzije softvera (verzija suvremena verziji programa za prepoznavanje). Vrlo je važna kvaliteta zvučne kartice i, naravno, mikrofona, a ukoliko postoji mnogo buke i smetnji unutar kućišta računala, preporuča se korištenje USB mikrofona koji ima sustav za obradu signala izvan kućišta računala pa mu ono ne smeta.

Programi za prepoznavanje govora koriste RAM memoriju u procesu raspoznavanja izgovorenih riječi, pa što je više ima, to bolje.

Minimalni zahtjevi: Intel® Pentium® III / 500 MHz procesor (ili ekvivalentni AMD® procesor), 256 MB RAM (512 MB RAM preporučeno), 500 MB slobodnog prostora na tvrdom disku, Microsoft® Windows® XP (SP1 ili više), 2000 (SP4 ili više) Creative® SoundBlaster ili ekvivalentna zvučna kartica, Internet Explorer 5 ili noviji, veza s internetom potrebna za aktivaciju.

Navedene specifikacije su potrebne da bi program uopće radio, pa se preporuča znatno jača konfiguracija.

## **Via voice (IBM)**

### ***Glavne značajke***

- namijenjen za Windows OS, Mac OS, Linux
- vokabular od 300 000 riječi (moguće ubaciti nove riječi, komande, imena, adrese...)
- visoka točnost prepoznavanja govora
- mogućnost odabira UK engleskog ili US engleskog

- jednostavna instalacija te mogućnost upravljanja govorom gotovo svim aplikacijama pod Windowsima (uključujući MS Word, MS Excel, Outlook, Internet Explorer...)
- diktiranje teksta u bilo koji tekst editor (Word, elektronička pošta)
- mogućnost slušanja teksta
- dodatni vokabular za provjeru gramatike novih unesenih riječi
- vodi računa o kontekstu teksta pa razlikuje vrlo slične riječi koje se izgovaraju
- prepoznaje i filtrira nepoželjne zvukove (dah, mljackanje...)
- konstantno prilagođavanje korisnikovom glasu (što se više koristi, to bolje radi)
- brzina i preciznost programa automatski se prilagođavaju mogućnostima računala prilikom instalacije
- mogućnost korištenja glasovnih kratica za veće izraze (npr. adrese)
- upravljanje i kontroliranje aplikacijama pomoću glasa
- kreiranje, uređivanje i slanje elektroničke pošte glasom
- mogućnost dobivanja glasovne potvrde nakon izvršene naredbe
- korištenje jedne riječi umjesto više tipki
- cijena oko 30 \$

### ***Instalacija***

Program je jednostavan za instalaciju na sustavima sa Linux, Mac OS X, Windows 98 ili novijim OS. Ovisno o brzini računala instalacija traje do 20 min.

### ***Vježbanje i potrebno vrijeme***

Program zahtjeva vježbanje prije upotrebe, da bi točno prepoznao korisnikov govor. Vježba se sastoji od toga da korisnik pročita jedan ili dva teksta računalu da bi ono analiziralo korisnikov glas. Prije vježbe, program radi optimalnog rada vrši samopodešavajuće testove za mikrofoni i zvučni sustav koji traju oko 20 sekundi. Najveći nedostatak vježbe je u tome što se ne mogu preskakati riječi, tj. Ako program ne shvati riječ, korisnik ju mora ponavljati sve dok ne shvati.

Za korisnike koji imaju poteškoća sa čitanjem postoje lakše skripte za vježbu. Font teksta za vježbu može se mijenjati. Općenito, da bi se postigao visok stupanj točnosti, potrebno je dodatno vježbanje i uhođavanje programa nego kod Dragon Naturally Speaking.

### ***Lakoća korištenja***

Najčešće se softver za prepoznavanje govora koristi za pisanje teksta neke vrste. U tu svrhu ovaj je program relativno lako koristiti nakon dobrog uvježbavanja. U odnosu na Dragon, neke komande moraju biti izrečene formalnije.

### ***Točnost prepoznavanja govora***

Program pruža visoku točnost prepoznavanja govora nakon dobrog treninga, ali uz dobar i ujednačen izgovor engleskog jezika (>96%). Stranicu teksta moguće je tako izdiktirati za nekoliko minuta, a što se više koristi, program točnije razumije korisnikov tekst.

### ***Integracija s ostalim aplikacijama***

Via Voice je kompatibilan s većinom Microsoft programa, te je u njih tako moguće, primjerice, diktirati. Najbolje funkcionira sa MS Word-om, dok više poteškoća ima s Excel-om ili Access-om, budući da oni koriste znatno više resursa računala. To, dakle nije do softvera za prepoznavanje govora. Preporuča se izgovaranje u sam Via Voice, to jest u za to predviđeni Speak pad, ta da se potom tekst kopira u željeni tekst editor. Što je jači računalni sustav, to se manje pojavljuje ovaj problem.

Program omogućuje diktiranje, ispravljanje, formatiranje, navigiranje i u drugim Windows aplikacijama, no često prije korištenja u nekim aplikacijama treba podesiti glasovne komande u programu za prepoznavanje. Inače može doći do aktiviranja neželjenih akcija unutar aplikacije. Moguće je zadati da se isključivo diktira, da ne bi program shvatio dio teksta kao neku naredbu.

I na relativno brzom računalu može proći dugo vremena prije nego što se Via Voice programom ostvari neka akcija. Računalo se tako može potpuno zaglaviti čestim ponavljanjem takve problematične naredbe.

### ***Zahtjevi za računalo i kompatibilnost s ostalim hardverom***

Za dobar rad računalo treba imati brz procesor i novije verzije softvera (verzija suvremena verziji programa za prepoznavanje). Vrlo je važna kvaliteta zvučne kartice i, naravno, mikrofona, a ukoliko postoji mnogo buke i smetnji unutar kućišta računala, preporuča se

korištenje USB mikrofona koji ima sustav za obradu signala izvan kućišta računala pa mu ono ne smeta.

Programi za prepoznavanje govora koriste RAM memoriju u procesu raspoznavanja izgovorenih riječi, pa što je više ima, to bolje.

Minimalni zahtjevi: Pentium 600MHz procesor ili ekvivalentni AMD, 192 MB RAM, 510MB slobodnog prostora na tvrdom disku.

Navedene specifikacije su potrebne da bi program uopće radio, pa se preporuča znatno jača konfiguracija.

U slijedećoj (opsežnoj) tablici navodimo razlike ovih dvaju, na tržištu najkvalitetnijih paketa:

Speech Recognition Feature Comparison Matrix				
FEATURE	Dragon Naturally Speaking 8 PROFESSIONAL	Dragon Naturally Speaking 8 PREFERRED	IBM Via Voice Standard V10	IBM Via Voice Professional USB V10
Microphone included	Yes	Yes	Yes	Yes
Short set-up and user enrollment	Yes	Yes	Yes (But not as fast as Dragon)	Yes (But not as fast as Dragon)
Dictate into most Microsoft® Windows®-based applications	Yes	Yes	Yes	Yes
Control menus and dialog boxes in most Microsoft® Windows® XP & Windows® 2000-based applications by voice	Yes	Yes	No (only MS Word & Internet Explorer)	Yes
Simultaneous dictation and command modes	Yes	Yes	Yes	Yes
Format and edit by voice	Yes	Yes	Yes	Yes
Mouse control by voice	Yes	Yes	No	Yes
Nothing But Speech (NBS) <sup>TM</sup>	Yes	Yes	No	No
USB audio support	Yes	Yes	Yes	Yes
Performance Optimizer	Yes	Yes	Yes	Yes
Natural Punctuation	Yes	Yes	No	No
Say Web and browser links by voice	Yes	Yes	No	Yes

Award-Winning RealSpeak™ 2 Text-to-Speech	Yes	Yes	No	No
Dictation Playback	Yes	Yes	Yes	Yes
Use with handheld digital recorder	Yes	Yes	No	Yes
Dictate into Pocket PC or Palm Tungsten	Yes	Yes	No	No
Support for cordless or array microphones	Yes	No	No	No
Save audio with text dictation	Yes	No	Yes	Yes
Third-Party Correction	Yes	No	No	No
Roaming User	Yes	No	No	No
Smart Formatting	Yes	No	No	No
Smart Commands	Yes	No	No	No
Create multiple custom vocabularies	Yes	No	No	Yes
Macro Creation Tools				
FEATURE	PROFESSIONAL	PREFERRED	IBM Via Voice Standard V10	IBM Via Voice Professional USB V10
Text and Graphics Dictation shortcuts	Yes	Yes	No	Yes
Complex macro support	Yes	No	No	No
Macro recorder	Yes	No	No	No
Basic scripting commands	Yes	No	No	No
Advanced Microsoft® VBA-compatible scripting	Yes	No	No	No
Specialized Application Support				
FEATURE	PROFESSIONAL	PREFERRED	IBM Via Voice Standard V10	IBM Via Voice Professional USB V10

Internet Explorer 5, 6	Yes	Yes	Yes	Yes
AOL® 6, 7, 8	Yes	Yes	Yes	Yes
Word 2000, XP, 2003	Yes	Yes	Yes	Yes
WordPerfect® 11 (Service Pack 1 & 2), 12	Yes	Yes	Limited	Limited
Outlook® Express 5, 6	Yes	Yes	Yes	Yes
Excel 2000, XP, 2003	Yes	Yes	Yes	Yes
Outlook® 2000, XP, 2003	Yes	No	Yes	Yes
PowerPoint® 2000, XP, 2003	Yes	No	Yes	Yes
Lotus Notes® 5, 6	Yes	No	Yes	Yes
InfoPath®	Yes	No	No	No
Enterprise Features				
FEATURE	PROFESSIONAL	PREFERRED	IBM Via Voice Standard V10	IBM Via Voice Professional USB V10
MSI Installer	Yes	Yes	No	No
Import/export user files	Yes	Yes	Yes	Yes
Roaming User	Yes	No	No	No
Import/export macros	Yes	No	No	No
Import/export vocabularies	Yes	No	No	Yes
Specialized vocabularies available for purchase	Yes	No	No	Yes
Network tools for centralized vocabulary management	Yes	No	No	No
Section 508 Certified (US ACCESSIBILITY LAW)	Yes	No	No	No

Macro Creation Tools				
FEATURE	PROFESSIONAL	PREFERRED	IBM Via Voice Standard V10	IBM Via Voice Professional USB V10
Text and Graphics Dictation shortcuts	Yes	Yes	No	Yes
Complex macro support	Yes	No	No	No



Macro recorder	Yes	No	No	No
Basic scripting commands	Yes	No	No	No
Advanced Microsoft® VBA-compatible scripting	Yes	No	No	No
Specialized Application Support				
FEATURE	PROFESSIONAL	PREFERRED	IBM Via Voice Standard V10	IBM Via Voice Professional USB V10
Internet Explorer 5, 6	Yes	Yes	Yes	Yes
AOL® 6, 7, 8	Yes	Yes	Yes	Yes
Word 2000, XP, 2003	Yes	Yes	Yes	Yes
WordPerfect® 11 (Service Pack 1 & 2), 12	Yes	Yes	Limited	Limited
Outlook® Express 5, 6	Yes	Yes	Yes	Yes
Excel 2000, XP, 2003	Yes	Yes	Yes	Yes
Outlook® 2000, XP, 2003	Yes	No	Yes	Yes
PowerPoint® 2000, XP, 2003	Yes	No	Yes	Yes
Lotus Notes® 5, 6	Yes	No	Yes	Yes
InfoPath®	Yes	No	No	No
Enterprise Features				
FEATURE	PROFESSIONAL	PREFERRED	IBM Via Voice Standard V10	IBM Via Voice Professional USB V10
MSI Installer	Yes	Yes	No	No
Import/export user files	Yes	Yes	Yes	Yes
Roaming User	Yes	No	No	No
Import/export macros	Yes	No	No	No
Import/export vocabularies	Yes	No	No	Yes
Specialized vocabularies available for purchase	Yes	No	No	Yes
Network tools for centralized vocabulary management	Yes	No	No	No
Section 508 Certified (US ACCESSIBILITY LAW)	Yes	No	No	No

## SpeakToText (CoolSoft)

### *Glavne značajke*

- namijenjen za Windows OS
- diktiranje teksta u razne tekst editore
- prepoznavanje snimljenih audio materijala
- čitanje odabranih tekstualnih datoteka
- stvaranje novih glasovnih naredbi
- diktiranje teksta i kopiranje istog u Word dokument jednom naredbom
- pisanje i slanje e-mailova pomoću glasovnih naredaba

### *Instalacija*

Program je vrlo jednostavan za instalaciju na Windows sustavima. Ovisno o konfiguraciji računala instalacija ne traje više od 30 min. Sama instalacija zauzima 58 MB koja se može preuzeti na službenoj stranici proizvođača uz cijenu od \$39.95, ili se može naručiti i CD za \$15 više. Uz kupnju softwera, dostupan je i besplatni *upgrade*.

### *Vježbanje i potrebno vrijeme*

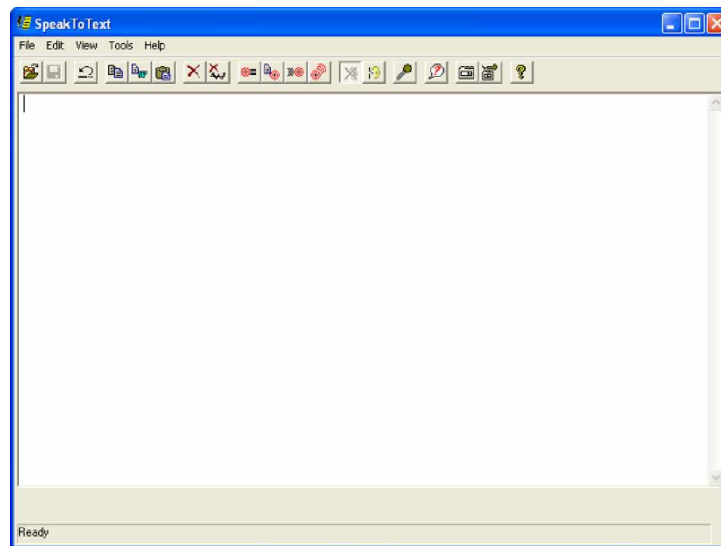
Kao i sa svakim drugim alatom, potrebno je neko vrijeme da se upozna sa svim mogućnostima koje alat posjeduje. Dok se u potpunosti ne upoznamo sa njima, programeri su integrirali opciju izlistavanja svih naredbi koje alat može prepoznati, a pozivaju se jednostavnim izgovorom “What Can I Say?”. Ili na hrvatskom jeziku : “Što smijem reći?”. Naravno, ukoliko korisnik ugradi i svoje naredbe, i njegove će biti izlistane.

### *Točnost prepoznavanja govora*

Alat vrlo dobro prepoznaje korisnikov govor uz pravilno izgovaranje engleskih riječi. Treba se napomenuti da postoji i opcija snimanja , a kasnije i prepoznavanja, osobnih stavki na jeziku koji korisnik želi ili pak riječi koje su samo njemu poznate. Time su korisniku otvorene i puno veće mogućnosti pri automatiziranju svoga operacijskog sustava. Napomenimo da SpeakToText dolazi integriran sa Microsoftovim mehanizmom za prepoznavanje govora SAPI 5.1, koji se pokazao vrlo dobro.

### ***Integracija s ostalim aplikacijama***

Što se tiče integracije s ostalim aplikacijama, ovaj alat pruža mnoštvo mogućnosti. Ugrađena su sučelja za programe koji su namjenjeni razgovaranje preko Interneta, kao što su MSN Messenger, AIM Instant Messenger, Yahoo Messenger, ICQ i PalTalk. Sposoban je, kao i većina alata iste namjene, ispisivati diktirani tekst u Word ili neki drugi tekstualni editor, kopirati izrečeni tekst u *clipboard* itd. Proizvođač navodi da u tzv. naprednom načinu rada sposoban pisati i slati elektronsku poštu u Outlook-u, Outlook Express-u and Netscape-u. Na slici 2. se vidi izgled glavnog prozora SpeakToText alata.



Slika 2. Izgled glavnog prozora SpeakToText alata

### ***Zahtjevi za računalo i kompatibilnost s ostalim hardverom***

Gledajući prosjek, nije potrebno previše moćno računalo za pokretanje i normalan rad ovog alata. Proizvođač zahtjeva instaliranu verziju Windows XP Home/Professional ili Windows-a 2000, te preporuča 500MHz ili jači procesor, 256 megabajta RAM memorije, 100 megabajta slobodnog prostora na disku. Što se tiče perifernih jedinica, kvalitetan mikrofoni ne bi bio na odmet.

## **VR Commander**

### *Glavne značajke*

- namjenjen za Windows OS
- upravlja igricama, simulatorima i CAD aplikacijama
- intuitivan
- poziva kompleksne naredbe
- simulira tipkovnicu
- unosi tekst u razne aplikacije
- integracija sa VoIP programima
- brzo reagira
- podržava bluetooth mikrofone
- potiskivanje šuma

### *Instalacija*

Instalacija programa je vrlo intuitivna i za nju je potrebno samo nekoliko minuta jer minimalna instalacija zauzima do 4 MB, što i nije puno obzirom na današnju cijenu tvrdih diskova velikih kapaciteta.

### *Vježbanje i potrebno vrijeme*

VR Commander je koncipiran na način “instaliraj i koristi”(eng. load and go). Sučelje je vrlo jednostavno i intuitivno što odmah u startu daje veliku prednost korisnicima kojima je čitanje upustava za korištenje iziskuje napor. Vrlo brzo reagira na izrečene naredbe. Kako se navodi u specifikacijama alata, odmah nakon izgovorenog poslijednjeg sloga riječi, VR Commander reagira. Slušanje i brzo prepoznavanje riječi prilikom kontinuiranog govora daje ovaj program nalazi primjenu u raznim igricama, simulatorima, te naravno, svim ostalim aplikacijama.

Uz jednostavno korištenje predložaka definiranih u interakciji sa korisnikom, svako otvaranje audio ili tekstualnog dokumenta, ili otvranje i rukavanje nekom aplikacijom postaje vrlo jednostavno. Svakim predloškom može biti definirana jedna ili više naredbi, što ovisi o željama korisnika koliko stvari želi kontrolirati samo jednom izrečenom naredbom.

Definirane naredbe se mogu dijeliti sa drugim korisnicima istog računala ili ih poslati elektronskom poštom prijatelju.



Slika 3. VR Commaderov control panel

### ***Točnost prepoznavanja govora***

Ukoliko je korisnik vrlo dobar poznavatelj engleskog jezika, ne treba prolaziti fazu treninga pravilnog izričaja engleskih riječi. Većina alata iste i slične namjene danas na tržištu zahtjeva i nalaže da se prije korištenja prođe trening, ali u ovom slučaju to nije potrebno upravo zbog visokog stupnja prepoznavanja korisnikovog, ponekad i ne potpuno točnog, izričaja. Alat posjeduje tehnologiju kontinuiranog prepoznavanja izrečenih riječi, oslanjajući se na novi mehanizam (eng. engine) prepoznavanja. Prije su se riječi morale izgovarati sa kratima pauzama između, što više nije slučaj. To svojstvo omogućava korisniku korištenje ovog alata u realnom vremenu, te govoreći srednje brzim tempom, vrlo brzi odziv alata na izrečene riječi. Bilo da se radi o ispisivanju teksta u neki tekstualni editor ili zadavanje naredbi računalu.

### ***Integracija s ostalim aplikacijama***

VR Commander je kompatibilan s većinom Microsoft programa poput tekstualnih editora (Word, Excel...), mail klijenata (Outlook, Outlook Express...) i mnogih drugih. Ovaj se

alat najviše razlikuje od svojih klonova po tome što ima mogućnosti korištenja u igricama i raznim simulatorima letenja i vožnje. Također se preporučuje zaljubljenicima u igrice u kojima sudjeluje veći broj igrača spojenih preko Interneta. Naime, VR Commander ima integrirano sučelje za njihovo međusobno razgovaranje. Naravno, svaki igrač mora imati instaliranu verziju ovog alata.

### ***Zahtjevi za računalo i kompatibilnost s ostalim hardverom***

Nije potrebno previše moćno računalo za pokretanje i normalan rad ovog alata. Proizvođač zahtjeva instaliranu verziju Windows XP Home/Professional ili Windows-a 2000, te preporuča 450MHz ili jači Pentium 2 procesor, minimalno 128 megabajta RAM memorije i barem 4 megabajta slobodnog prostora na disku. Što se tiče perifernih jedinica, kvalitetan mikروفon koji poništava šum bi bio najoptimalniji izbor. Također je potrebna i tzv. “full duplex” zvučna kartica. Ovaj tip kartice podržava slanje i primanje zvučnih signala istovremeno. Ukoliko to nije slučaj, od korisnika se zahtjeva da instalira mikروفon koji je spojen na USB priključak.

## **Realize® Voice 4.1**

Realize® Voice 4.1 je nešto jednostavnija verzija alata koja omogućava kontrolu nad računalom koristeći korisnikove glasovne naredbe. Ovaj Microsoftov proizvod sjedinjuje posljednja saznanja na području prepoznavanja govora i reproduciranja govora. Reklamiraju ga naglašavajući da onu potpuno zamjenjuje tipkovnicu i miša, što je i dijelom točno, ali ne u tolikom obujmu kao što to čine Dragon Naturally speaking i IBM-ov ViaVoice.

### ***Glavne značajke***

- *Glasovne naredbe*

Alat ima mogućnost programiranja glasovnih naredbi za većinu poslova za što nam je inače potrebna tipkovnica i miš. Uz to, već u instaliranoj verziji alata postoji set naredbi koje su na engleskom jeziku. Ovim programom je moguće kontrolirati sučelje Windows operativnog sustava ali i definirati posebnu listu naredbi koji se odnose samo na određenu aplikaciju.

- *Personificiranje računala*

Na korisniku je da odluči hoće li koristiti ovu «ne baš svakidašnju» i «ne toliko korisnu» mogućnost ovog alata.

Korištenjem ovog alata se komunicira sa računalom davajući mu razne naredbe uvijek drugačijim redoslijedom.

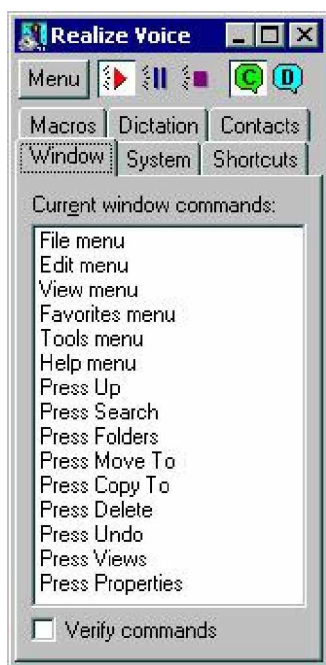
Uključivanjem ove opcije računalu se pokušava dati osobnost na način da ono svaki put na drugi način odgovori na korisnikovu naredbu. Npr. kada mu se kaže «Hvala», on jednom odgovori sa «Molim!», drugi put sa «Nema problema!». Iako je beskorisnost ove opcije očita ipak daje ovom alatu dozu spontanosti, ma koju ne nailazimo baš često.

- *Navigacija*

Program se može iskortistit za navigaciju kroz internet preglednike Internet Explorer, Mozillu Firefox i mnoge druge. Alat se također može intergrirati u mnoštvo drugih aplikacija, te ih učiniti mnogo lakšim za korištenje.

- *Diktiranje*

Kao i većina sličnih alata, ovaj program je sposoban prepoznati izrečene korisnikove riječi te ih simultano upisivati u odabrani tekstualni editor, radeći nad njima željene transformacije kao : mijenjanje fonta, zakošavanje i podebljavanje teksta, riječi ili samo pojedinih slova.



- *Diktiranje «slovo po slovo»*

Ovaj način rada omogućava unos teksta slovo po slovo, koristeći fonetičku abecedu. Npr. za upis slova «A» se koristi izraz Alpha, slova «B» Bravo itd.

- *Integracija sa adresarom*

Ovo svojstvo omogućava glasovno biranje brojeva ili pozivanje već postojećeg kontakta iz instaliranog adresara. Također se mogu mijenjati sve stavke vezane za taj kontakt, kao odlazak na njihovu web stranicu, dohvaćanje opisa puta koje vodi do njegovog mjesta stanovanja, i mnoštvo drugih opcija vezanih za elektronski adresar.

- *Glasovni prečaci*

Moguće je otvoriti bilo koju web stranicu, program, audio ili tekstualnu datoteku preko izgovora jedne riječi kojoj se definira željena akcija.

- *Tekst u govor*

Opcija “Tekst u govor” omogućava korisniku da čuje izdiktirani tekst od strane računala. Otvorena je mogućnost da korisnik sam bira što će računalo “izgovoriti” za svaku riječ posebno. Mijenjanje tog parametra daje osjećaj velike konfigurabilnosti ovog alata.

- *Instalacija i vježbanje*

Programje vrlo lako instalirati, bilo preko naručenog CD ili plaćene verzije koja se preuzela na službeneoj stranici Realize Softwarea. Preporučljivo je proći čarobnjak za testiranje i konfiguraciju mikrofona, te glasovne vježbe preko čarobnjaka(“wizard”) koji bi trebao pri korištenju alata uštedjeti mnogo vremena pri provjeravanju načina kako se koja riječ izgovara a da je alat prepozna kao takvu.

## **Besplatni alati za raspoznavanje govora**

Uz komercijalne, naravno, postoje i besplatne verzije alata za prepoznavanje govora. Komunikacija sa računalom, pri čemu se misli na govornu komunikaciju, dugo je bila tema koja se nerijetko svrstavala među znanstvenu fantastiku. Kroz posljednjih desetak godina tehnologija koja omogućava prepoznavanje govora je izuzetno napredovala. Mnoštvo takvih alata se proizvelo ne samo za *windows*, već i za druge operacijske sustave. Onima koji ti programi osiguravaju egzistenciju, poput ljudi sa posebnim potrebama, ili pak ljudima koji su dobro novčano potkovani na tržištu postoji cijela paleta proizvoda koji se međusobno razlikuju cijenom, kao i mogućnostima. No, postoji i uvijek prisutno *ali*. Naime, određene skupine programera su razvile alate koji po mogućnostima puno ne zaostaju za poviše spomenutim alatima ali se razlikuju u jednom bitnom faktoru : cijenom. Oni su besplatni. Od nekoliko desetaka besplatnih proizvoda, spomenuti ćemo one najbolje :

Ø Xvoice -- <http://xvoice.sourceforge.net/>

Ø Open Mind Speech -- <http://freespeech.sourceforge.net/>

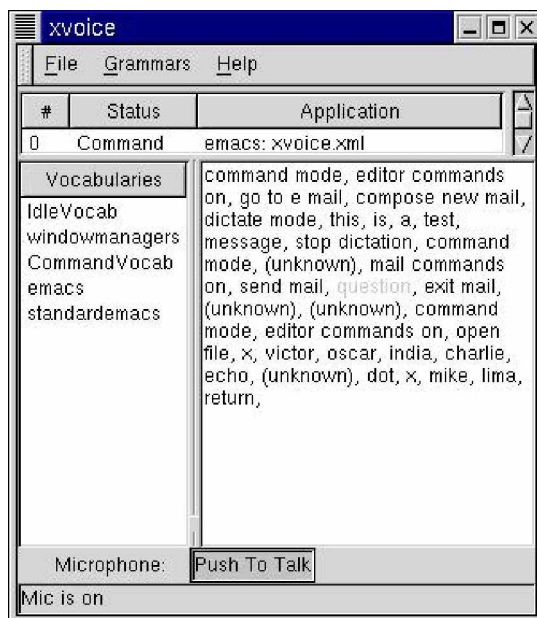


- Ø CMU Sphinx -- <http://cmusphinx.org>
- Ø NICO toolkit -- <http://nico.sourceforge.net/>

U slijedećem tekstu ćemo opisati najpraktičnijega, *XVoice*, jer je on jedini u potpunosti razvijen za krajnjeg korisnika tj. sadrži mehanizam prepoznavanja govora i odgovarajuće sučelje. Svrha ostalih besplatnih alata je pomoć pri usavršavanju samog mehanizma za prepoznavanje govora. Krajnji korisnici od ovih alata nemaju baš nikakve koristi ukoliko ih to područje ne zanima, dok za one zainteresirane postoje otvoreni kodovi za proučavanje i daljnji razvoj.

## Xvoice

Ovaj alat omogućava kontinuirano izricanje napisanog teksta te obavljanje naredbi koje su zadane govorom korisnika za većinu X aplikacija. Za prevođenje korisnikovog govora koristi se IBM-ov *ViaVoice* mehanizam za prepoznavanje govora koji se distribuira odvojeno od samog *XVoice*-a. IBM nudi *ViaVoice* u Americi i Kanadi za oko 40 dolara, što uključuje naglavne slušalice, no poštarina nije uključena u cijenu. Jeftinije se može proći ako se program plati te se zatim preuzme datoteka sa IBM-ove web stranice. Besplatna verzija *XVoice*-a se može preuzeti sa stranice [xvoice.sourceforge.net](http://xvoice.sourceforge.net).



Slika 4. *XVoice* korisničko sučelje

*XVoice* posjeduje četiri načina rada. Dok je naredbovnom načinu rada, alat povezuje govor sa već predefiniranim naredbama ili skupom naredbi. Na primjer, definirajmo riječ «list» koja označava izdavanje naredbe konzoli «ls -l». Kada korisnik u mikrofon kaže «list», naredba «ls -l» se šalje na konzolu kao da je otipkana. Kad je postavljen za diktiranje alat ispisuje samo ono što je uspio prepoznati od izgovorenih riječi, gdje se pritom eventualno izgovorene naredbe ne izvršavaju. U «idle» načinu rada se samo glavne predefinirane naredbe prepoznavaju i izvršavaju. Četvrti način rada je kombinacija prva dva, alat sluša i ispisuje sve prepoznate riječi ali i uzima u obzir prepoznate naredbe te ih izvršava. Prilikom prvog fokusiranja na određenu aplikaciju, alat se automatski postavlja u naredbovni način rada. Na slici 1. je prikazan tipičan izgled *XVoice* prozora. Prepoznate, i neke odbačene, riječi se vide na desnoj, dok se trenutno aktivne naredbe vide na lijevoj strani prozora. Ime aplikacije na koju se odnose ove naredbe se nalazi na vrhu prozora.

*XVoice* koristi *ViaVoice* datoteke za prepoznavanje predefiniranih naredbi ili običnog teksta. Osnovne naredbe su definirane u konfiguracijskoj datoteci *xvoice.xml*, koja se može nadopunjavati i mijenjati od strane korisnika.

### § Popis ugrađenih naredbi :

- "Microphone off" – gasi mikrofon; Za ponovno uključivanje potrebno je pritisnuti "Push To Talk" tipku
- "Command mode" – uključuje naredbovni način rada za fokusiranu aplikaciju, te učitava niz definiranih naredbi koji se odnose samo na tu aplikaciju
- "Stop command" – isključuje predefinirane naredbe za fokusiranu aplikaciju
- "Dictate mode" – uključuje funkciju diktiranja (napomena : prepoznate naredbe se ne izvršavaju).
- "Stop dictation" – isključuje funkciju diktiranja
- "Idle mode" – isto kao "stop dictation" zajedno sa "stop command"
- "Correction" – u fazi diktiranja briše najčešće spomenutu riječ
- "Build grammar files" – ponovno učitavanje predefiniranih naredbi

Neke aplikacije, poput onih kojima je miš nužan za navigaciju (Netscape) je teško kontrolirati pomoću govora. Alati poput ovoga, općenito, rade odlično za prepoznavanje naredbi dok za prepoznavanje običnog teksta nisu baš toliko praktični. Za ovo potonje razlog je ili mala greška u izricanju riječi ili pak sitne mutacije u glasu koje program ne može prepoznati. Iako

*XVoice* i *ViaVoice* omogućavaju mnogo toga, nije moguće kontrolirati cijeli Linux desktop u potpunosti. Slaba karika u lancu je IBM. Iako je uloženo dosta novca, unaprijeđivanje alata koji omogućavaju govorno upravljanje za *Linux* aplikacije nema svjetlu budućnost, iako su izašle neke poboljšane verzije *ViaVoice*-a ali za Windows operativne sustave.

## **5. Zaključak**

U ovom seminarskom radu dan je temeljni opis tehnologije raspoznavanja govora, objašnjeni su glavni pojmovi vezani za temu, podjela sustava te opis najkvalitetnijih komercijalnih i besplatnih alata za raspoznavanje govora trenutno.

U narednim godinama možemo očekivati još veći broj komercijalnih primjena, niže cijene postojećih rješenja te unaprijeđenja samih algoritama i prateće programske podrške.

Iako se većina sustava gradi isključivo za Microsoft Windows platforme pozitivan je i trend rasta OpenSource rješenja.

## **6. Literatura**

1. <http://www.tldp.org/HOWTO/Speech-Recognition-HOWTO/>
2. <http://cslu.cse.ogi.edu/HLTsurvey/ch1node4.html>
3. <http://www.ewh.ieee.org/r10/bombay/news6/AutoSpeechRecog/ASR.htm>
4. <http://www.nuance.com/naturallyspeaking>
5. <http://www.io.com/~hcexres/tcm1603/achtml/recomx7c.html>
6. <http://www.scansoft.co.uk/viaoice/>
7. <http://sourceforge.net>

...