



Taligenkänning - har den en framtid som nyttoverktyg?

*Examens arbete I, 10 p.
Handelshögskolan vid Göteborgs Universitet
Institutionen för Informatik – HT 2001
Handledare: Göran Walske
Författare: Zineta Komad*

Abstrakt

Informationsteknologi utvecklas allt snabbare och idag har vi möjlighet att utföra det mesta, även att kommunicera med datorer. För att vi skall kunna använda datorer på mest effektivt sätt, är det oerhört viktigt att vi kan kommunicera med datorer på mänskligt sätt – genom att tala. Taligenkännings-teknologi innebär att datorn förstår och känner igen det talade språket. Syftet med uppsatsen är att presentera området taligenkänning, att undersöka dess användningsområde samt att försöka ta reda på vilka är dess för- och nackdelar. Att undersöka hur taligenkänning skulle kunna utvecklas i framtiden var också av intressen. För att bilda en uppfattning om teorier och undersökningar inom området taligenkänning, har jag utfört en litteraturstudie. För att undersöka om det finns något allmänt intresse för taligenkänningsprogram ringde jag till ett flertal företag som säljer hård- och mjukvara för PC. Även personer på olika företag som utvecklar/använder tjänster har kontaktats. Vad jag har kommit fram till i undersökningen är att det finns ett brett användningsområde för taligenkänning, bl.a. i mobiltelefoner, ordbehandlingsprogram, datorstyrning, telefons-tjänster, m.m. Utveckling av hjälpmedel för funktionshindrade personer är ett stort område där det finns intresset för taligenkänningsprodukter. Att teknologin inte har kommit till större användning kan bero på att det inte har funnits speciellt många utvecklare av den samt att det saknas något större intresse för den. Undersökningen har dock visat att många tillfrågade har förhoppningar och stor tron på taligenkänning samt att det finns ett flertal projekt på gång. Med denna undersökning ville jag, framför allt uppmuntra forskare och studenter att undersöka taligenkänningsområdet bättre, för att kunna påpeka teknologins brister och på så sätt påverka utvecklingen av den.

TALIGENKÄNNING - HAR DEN EN FRAMTID SOM NYTTOVERKTYG?	1
VARFÖR TALIGENKÄNNING?	5
FRÅGESTÄLLNING	6
SYFTET OCH INTRESSEENTER.....	6
AVGRÄNSNING	6
DISPOSITION.....	7
METOD.....	8
LITTERATURSTUDIE.....	8
Om taligenkänning och dess utsträckning.....	8
Om språk	9
INTERVJUER	9
TESTNING AV OLIKA TJÄNSTER.....	10
TALIGENKÄNNING – BAKGRUND.....	11
NÄRLIGGANDE OMRÅDEN	11
Artificiell intelligens	11
Expertsystem	12
Artificiella neurala nätverk.....	12
Natural Language Processing.....	13
HISTORIK.....	14
SPRÅK	15
ALLA TALAR OLIKA.....	15
TALKOMMUNIKATION	16
VERBAL OCH ICKE-VERBAL KOMMUNIKATION	16
TALIGENKÄNNINGSTEKNOLOGI.....	17
HUR FUNGERAR TALIGENKÄNNING?.....	17
Akustik- och språkmodell	17
OLIKA TYPER AV TALIGENKÄNNING	18
Talarberoende och talaroberoende system	18
Diskreta och kontinuerliga system	18
Diktering och kommandostyrning.....	18
RESULTAT	19
TALIGENKÄNNINGSPRODUKTER – VAR FINNS DEM?	19
Mobil taligenkänning	19
Telefontjänster med taligenkänning.....	19
Taligenkänning och PC	20
Taligenkänning för personer med olika funktionshinder	21
FÖR- OCH NACKDELAR MED TALIGENKÄNNING?	22
FRAMTID	23
DISKUSSION	25
SLUTSATS	26
FORTSATT FORSKNING.....	26
REFERENSER.....	27
BIBLIOGRAFI.....	28

BILAGA: INTERVJUFRÅGOR..... 29

Människan har länge fascinerats av idén att kommunicera med maskiner på ett naturligt sätt. Vi har sett gamla science fiction filmer där man talade till sina hushållsapparater i önskan om att de skall utföra något, som t.ex. spela musik eller tända lyset. I andra fall har olika figurer, som reste genom tiden, haft små apparater som de kunnat tala till, inbyggda i sina kläder. I verkligheten har vi inte kunnat göra något utav detta fram tills nu.

Informationsteknologi utvecklas allt snabbare och idag har vi möjlighet att utföra det mesta, även att kommunicera med datorer.

Kommunikation är ”överförandet av information från en individ (organism, system) till en annan” (P. Linell, 1990). Människans naturliga sätt att kommunicera är genom att tala, genom att använda röst, vilket skiljer sig från den vanliga kommunikationen med datorer. Kommunikationen mellan människa och dator sker via datormusen och tangentbordet. För vana datoranvändare är det naturligt men de som inte är vana vid att använda tangentbordet, eller de som p.g.a. ett handikapp inte kan använda det, kan uppfatta det som en svårighet.

Alla ansträngningar att skapa bra gränssnitt och göra datorer mer användarvänliga, har som syfte att göra mötet mellan människa och maskin så naturligt som möjligt. Vi måste känna att vi lätt, och utan större kunskaper om datorer, kan använda dem. För att vi skall kunna använda datorer på mest effektivt sätt, och att kunna utnyttja alla fördelar med dem, är det oerhört viktigt att vi kan kommunicera med datorer på mänskligt sätt. En teknologi som möjliggör sådan kommunikation kallas för taligenkänning. Taligenkänning (eng. speech recognition) innebär att datorn förstår och känner igen det talade språket. System för automatisk taligenkänning används för att underlätta människa-dator-kommunikation och för att tillåta inmatning av styrkommandon, text och data via tal.

Forskningen inom taligenkänning har pågått i mer än 50 år. Första försök att skapa maskin som kan tala skedde redan på 1930-talet. Den första taligenkänningsprodukten (i dagens bemärkelse) lanserades av IBM, 1993. Det kallades för IBM Personal Dictation System for OS/2. Forskningen idag kretsar mest kring medicin, bl.a. huruvida läkare kan använda taligenkänningsprodukter för att underlätta dagliga aktiviteter (Wormek et al., 1997), och huruvida funktionshindrade människor kan använda taligenkänning som kompletterande styrsätt för dator (Buchholz & Havstam, 2000), även projekt ”Datorbaserad talträningstrustning för personer med stamning”(Forne-Wästlund, 1998)

Varför taligenkänning?

Första gången jag kom i kontakt med taligenkänning var när jag köpte en mobiltelefon med möjlighet till röstuppringning. Röstuppringning innebär att man kan ringa upp någon genom att säga personens namn, samt att svara och avvisa samtal med rösten när telefonen är ansluten till en s.k. *'hands free'* enhet. Innan man kan använda röstuppringning så måste man spela in röstmärken till poster i telefonboken.

Det mest intressanta var att jag inte alltid lyckades med röstuppringningen, och att andra personer som testade min telefon lyckades ringa ett samtal först efter att de hade härmat min röst under ett antal försök. Jag började undersöka ämnet för att få reda på hur röststyrningen fungerar och varför den misslyckas ibland.

Frågeställning

I mitt sökande hittade jag många intressanta användningsområden för taligenkänningsystem samt en hel del programvaror från olika tillverkare, vilket fick mig att ställa frågor som:

1. Till vad används taligenkänning idag?
2. Finns det intresse för taligenkänningsprogram?
3. Vilka för- och nackdelar finns?
4. Hur kan taligenkänning utvecklas i framtiden?

Syftet och intressenter

Under tiden jag letade efter information om taligenkänningsprogram, kom jag fram till att det finns väldigt lite grundläggande information som är tillgänglig utanför vetenskapskretsen. Det slog mig också att en lekman måste ha väldigt svårt att förstå den befintliga informationen.

Syftet med uppsatsen är att presentera området taligenkänning, att undersöka dess användningsområde samt att försöka svara på ovanstående frågor. Mina förhoppningar är att väcka intresse för taligenkänningsteknologi, främst hos studenter och blivande forskare. Även vanliga människor som är nyfikna på och villiga att använda en ny teknologi kan ha intresse av uppsatsen.

Avgränsning

Frågeställningarna avser användningen i Sverige mest därför att det är praktiskt omöjligt att få tillräckligt med aktuell information från andra länder (t.ex. USA) där systemen redan är välkända och används i stor omfattning. Detta gäller, främst, förfrågor om försäljning av olika produkter. När det gäller traditionell försäljning (affärer), av praktiska skäl, behandlas enbart en region. On-line försäljning avser hela landet.

Taligenkänning är ett brett och nästan outforskat område. Tillsammans med en del andra vetenskapsområde, utgör den idag grunden till informationssystem. I litteraturen behandlas taligenkänning oftast i samband med andra discipliner och teknologier, och gränsen mellan de är otydlig. En av teknologier, som förekommer ofta i samband med taligenkänning, är Natural Language Processing (NLP). Jag har inte för avsikt att gå djupt in på skillnader mellan röstigenkänning och NLP, eller att mer ingående förklara alla talteknologier som t.ex. talarverifiering.

I uppsatsen nämns olika delar inom talteknologi, talarverifiering, talsyntes och taligenkänning. Talsyntes och taligenkänning ofta behandlas under samma begrepp, alltså taligenkänning. Jag har valt att utesluta talarverifiering från studien.

Översättning av olika applikationer från språk till språk är en problematik som förekommer även inom taligenkänning. Uppsatsen tar inte upp problemområdet, eftersom det är komplext och behandlas inom många andra områden, som t.ex. i ordbehandling, textinläsning med läspenna, mm.

Disposition

Nästkommade kapitel handlar om metodval. På vilket sätt har informationen samlats, samt motivering för valda metoder förklaras. I kapitlet ”*Taligenkänning – bakgrund*” ges en kort introduktion till vetenskapliga discipliner (datalogistik, artificiell intelligens, expertsystem, mm), som har nära koppling till taligenkänning. Med kapitlet ”*Språk*” belyses språkets komplexitet samt språkvetenskapens betydelse för utveckling av taligenkänningsteknologin. ”*Taligenkänningsteknologi*” ger en sammanfattad beskrivning av hur tekniken fungerar. Även olika typer av taligenkänning tas upp. I nästkommade kapitel presenteras resultaten av studien. Resultatdelen är fördelat i olika underkategorier, såsom användning, för-/nackdelar, samt framtid, för att ge en bättre överblick. ”*Diskussion*” relateras till det som redovisats i Resultatkapitlet. Det diskuteras huruvida studien har gett svar till frågeställningarna samt ges förslag på en eventuellt bättre studie. I det avslutande kapitel beskrivs slutsatser och ges förslag till fortsatt forskning. Bilaga presenterar en samling av frågor som ställdes under intervjuerna.

Metod

Med uppsatsen har jag tänkt presentera området taligenkänning och bl.a. undersöka dess för- och nackdelar. För att få förståelse för problemområdet har jag inlett mitt arbete med att studera befintlig litteratur och studier inom ämnet.

Jag har haft för åsikt att upplysa om hur taligenkänning används idag. För att få veta mer om systemets utsträckning här i Sverige, har jag kontaktat ett flertal företag som tillhandahåller röststyrda tjänster samt företag som jobbar med utveckling av teknologin och produkter. Vår kommunikation utfördes via e-post och/eller telefonsamtal.

För att själv uppleva hur det är att använda en tjänst som baseras på taligenkänning, har jag ringt till ett fåtal tjänster. Om man har för åsikt att diskutera ett systems för- och nackdelar, anser jag att man själv måste testa ett sådant system, för att få bättre insyn. Eftersom jag inte har haft tillgång till taligenkänningsystem för PC, så har jag avgränsat testning till telefontjänster som baseras på taligenkänning.

Jag ville också veta om det finns något allmänt intresse för taligenkänningsprogram. Frågor om detta har jag ställt under intervjuer, via e-post. Tanken var att även ställa dessa frågor till kontaktade mjukvaruförsäljare. Jag prövade därför att ringa upp ett flertal olika företag i branschen. De flesta saluförde dessvärre inga taligenkänningsprodukter.

Litteraturstudie

Under studien fann jag att det krävs en viss kunskap om språk och kommunikation för att bättre förstå hur taligenkänning fungerar och, framför allt, för att inse att taligenkänningsutvecklingen måste klara av flertalet motgångar.

Om taligenkänning och dess utsträckning

För att bilda en uppfattning om teorier och undersökningar inom området taligenkänning, har jag utfört en litteraturstudie. Jag har eftersökt litteratur bl.a. i samråd med lärare och bibliotekarier. Det visade sig dock att det fanns bara ett fåtal böcker att tillgå, då området ännu är outforskat.

Mest information hittade jag på Nätet, i form av uppsatser och pressreleaser. Jag sökte aktivt bland kommersiella sidor, för att få en uppfattning om systemets utsträckning och om vilka produkter som finns på marknaden.

Det är viktigt att kritiskt granska all information då man utför en studie. Information som finns att tillgå på Nätet är inte alltid pålitlig. Därför har jag bestämt att rangordna dem på en egenhändigt konstruerad pålitlighetsskala.

Jag har valt att utvärdera texter från Nätet enligt deras källor. Avhandlingar och skrifter som har anknytning till olika universitet och forskningscentra har jag värderat högst. Därefter har artiklar från tekniska on-line tidningar hamnat. Övriga artiklar och pressreleaser har hamnat på tredje plats. Längst ner har jag placerat information från kommersiella webbsidor. Nämda rangordningen har jag hållit mig till främst när det gäller för- och nackdelar med taligenkänning samt framtidsvision. En verklig uppskattning av situationen skulle negativt

kunna påverka försäljningen av taligenkänningsprodukter. Därför har inte information från kommersiella webbsidor tagits med större betydelse.

När det gäller teknisk information om hur produkter fungerar samt information om den bakomliggande teknologin, har uppgifter från kommersiella webbsidor värderats högt.

Om språk

Eftersom utveckling av taligenkänningsteknologi är beroende av kunskaper om mänsklig kommunikation och språk, har jag valt att titta lite närmare på studier inom språkvetenskap. Jag har valt bort flertalet skrifter och böcker inom ämnet eftersom de mest riktade sig till människor som är välbekanta med området språkvetenskap. Till slut valde jag en bok som, enligt min mening, förklarade olika lingvistiska teorier på så sätt att även en lekman kan förstå dem.

Som hjälp till att förklara och definiera olika begrepp, har jag använt mig av olika typer av uppslagsböcker och lexikon, eftersom materialet är, för det mesta, skrivet på engelska.

Intervjuer

En av de viktigaste delarna av studien är att ta reda på vad andra människor, som använder eller utvecklar system, själva tycker om dem. Eftersom taligenkänning inte har haft något större genombrott i Sverige, har det inte funnits möjlighet att intervjua eller observera någon tillräcklig mängd av användare av taligenkänningssystem. Då återstod bara att kontakta människor som på något sätt jobbar med dessa produkter. Personer som intervjuades har valts ut någorlunda slumpmässigt.

Den första tanken som slog mig var att vända mig till olika institutioner på universitetet, för att eventuellt hitta personer som forskar inom området. För att få tag på rätt personer har jag sökt bland universitets webbsidor om forskning. Det visade sig att det inte fanns många publikationer som rör området taligenkänning. Till slut har jag kontaktat ett flertal personer på Institutionen för lingvistik/datalingvistik, för att på så sätt få tag på någon som besitter kunskaper inom området.

Enbart två av personer kontaktade på Institutionen för lingvistik svarade på mina brev. Svaret var att de inte hade tid att ställa upp. Övriga svarade inte ens, av till mig okänd anledning.

Från personer som jag har kontaktat vid Institutionen för informatik, har jag fått råd om olika företag som borde känna till problemområdet. Personerna i frågan kunde ej ställa upp och svara på frågor kring taligenkänning.

Kontaktade personer som svarade på mina frågor jobbar på olika företag inom olika område, bl. a. utveckling av plattformar för taligenkänningssystem inom olika miljöer, distribuering av taligenkänningsprodukter för PC samt företag som tillhandahåller röststyrda telefontjänster.

Eftersom deras bakgrund och ansvarsområde skiljer sig väldigt mycket åt, har det varit omöjligt att ställa samma frågor till dem. Därför har jag valt att ställa ett antal öppna frågor i anknytning till min problemställning.

Testning av olika tjänster

För att kunna bilda en uppfattning om användningen av taligenkänningsprodukter, har jag valt testningen av produkter som mest lämplig metod. Jag har inte haft tillgång till mjukvara för PC som jag skulle kunna testa. Det återstod då bara att testa befintliga telefontjänster som baseras på taligenkänning. Jag har valt att testa två tjänster: Telias röststyrda nummerupplysning och SAS Speech Line (biljett- och flyginformation).

Gemensamt för alla tester var att jag ställde ett antal vanliga, enkla frågor som man skulle ha gjort om man ringde till en vanlig kundtjänst. Jag har också försökt att ställa enkla frågor, men att göra det på mer informellt sätt, genom att t.ex. fnissa eller gäspa medan jag talade. Lite vikt har jag även lagt vid att ställa frågor med annorlunda uttal av vanliga svenska ord. Det var också viktigt att försöka avbryta pågående samtal när som helst (p.g.a. samtalskostnader).

Taligenkänning – bakgrund

Under litteraturstudien om taligenkänning kom jag i kontakt med en del vetenskapliga discipliner som har koppling till taligenkänning, bl.a. datalingvistik och datavetenskap.

Närliggande områden

En av vetenskapliga discipliner som har nära koppling till taligenkänning är datalingvistik. Datalingvistik är en disciplin mellan lingvistik och datavetenskap som behandlar tillämpningar och teori kring dator och mänskligt språk. Målet med datalingvistiken är att utveckla mjukvaror i syfte att förbättra människa-dator-kommunikation. (Nationalencyklopedin, 2000)

Inom datalingvistiken kan man finna även forskning kring artificiell intelligens, som oftast behandlas inom datavetenskap. Artificiell intelligens (AI) utgör den vetenskapliga grunden till taligenkänning. (Turban & Aronson, 2001)

Artificiell intelligens

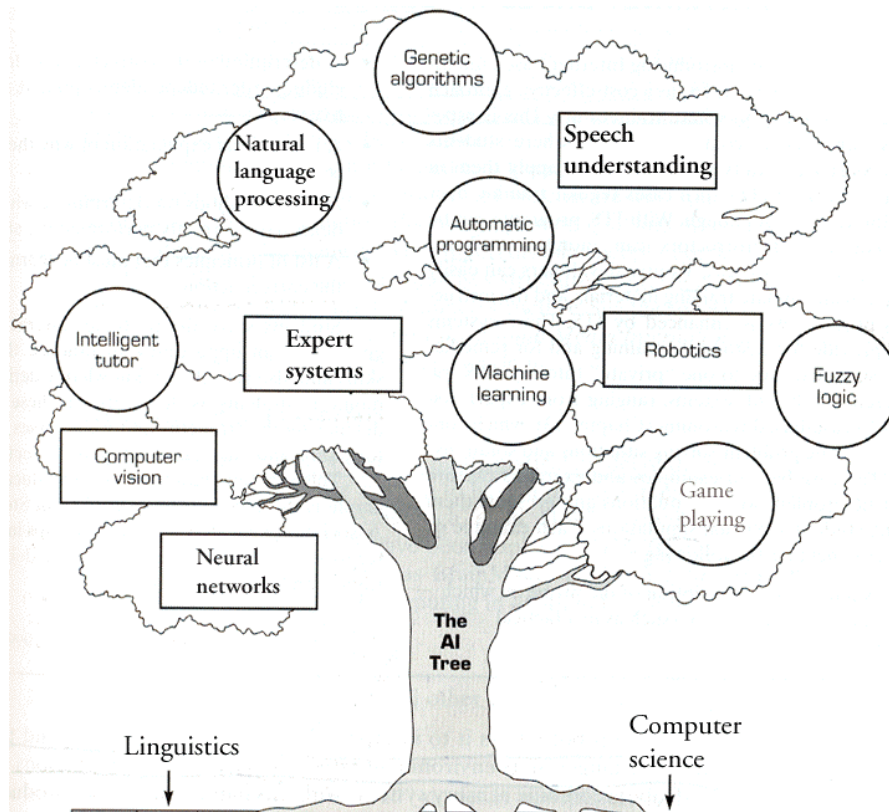
Talsyntes, förståelse och automatisk översättning av språk räknas ibland till AI. Det finns ett flertal definitioner av artificiell intelligens. Gemensamt för de alla, enligt Turban, är att AI innebär studier om processen som omfattar mänsklig tänkande, samt en maskinell representation av denna.

Man kan säga att ”AI är ett forskningsområde inriktat på att konstruera datorsystem som uppvisar intelligent beteende”. (Nationalencyklopedin)

Det som utmärker mänsklig intelligens är förmåga att bl.a.:

- förstå och lära sig från erfarenhet
- agera rätt och snabbt i nya situationer
- lösa problem
- klara av komplexa situationer
- använda kunskap
- tänka och resonera

Det intelligenta systemet bildar resonemangen genom att försöka dra slutsatser, utifrån fakta och bestämda regler. Detta sker med hjälp av mönsterigenkänning, vilket är unikt för AI. (Turban) Syftet med artificiell intelligens är att efterlikna mänsklig intelligens. Alan Turing utvecklade ett test (*Turing test*) för att avgöra om en dator uppvisar intelligent beteende. Datorn anses vara intelligent, enligt testet, ”enbart om en människa, som kommunicerar med en dator och en annan människa, inte kan avgöra vilken är vilken”. [*min översättning*]



Not. Från "Decision support systems and intelligent systems", av Turban & Aronson, 2001, Prentice Hall

Bilden representerar olika grenar inom artificiell intelligens. Expertsystem, neurala nätverk och NLP är bara några av användningsområden för AI, som har koppling till taligenkänning.

Expertsystem

System som klarar av att planera, dra slutsatser, lösa problem, etc. kallas expertsystem (ES). Expertsystem är ett "datorprogram som inom ett avgränsat kunskapsområde erbjuder ett stöd för problemlösning och beslutsfattande, jämförbart med förmågan hos mänskliga experter". (Nationalencyklopedin, 2000) Med andra ord, med hjälp av den kunskapen som lagrats i datorn, ES härmar experternas sätt att resonera.

Expertsystem erbjuder användarvänliga gränssnitt, genom vilka användaren kan ställa frågor och få svar från systemet. Svaret genereras med hjälp av en kunskapsbas som innehåller fakta och en samling regler om hur fakta kommer att hanteras. Förutom svaret genererar systemet en förklaring till varför just det svaret valdes ut. Ett intelligent program skall inte bara kunna dra slutsatser, utan skall även lära sig av erfarenhet från tidigare fall.

Artificiella neurala nätverk

I sin strävan att göra datorer så intelligenta att de kan ersätta och förstärka människans tänkande, har man skapat system som efterliknar hjärnans nätverk. Sådana system kallas ANN (eng. Artificial Neural Networks).

Ett artificiellt neuralt nätverk (neuronnät) består av ett flertal sammankopplade processenheter. Det tar emot en signal (input), behandlar den, och ger resultat i form av en ny signal (output). Behandlingen av data sker med hjälp av matematiska modeller och

algoritmer. En algoritm är ett stegvis process med förbestämda start- och slutpunkt. Målet, med att använda en algoritm, är att hitta en lösning för ett specifikt problem.

Det som utmärker ANN metodologin är bl.a. dess inlärningsförmåga, möjlighet att lösa nya sorts problem, robusthet, och parallellitet som medger hög beräkningshastighet, enligt Turban. Robustheten avser ofullständig och/eller felaktig data, vilket i taligenkänningsfall innebär brus i indata och i nätverket. Parallelliteten innebär att nätverkets processenheter, likt hjärnans neuroner, samtidigt bearbetar data.

Inlärningsförmåga hos datorer bygger på att ett expertsystem tränas upp med kända exempel. Därefter används matematiskt och statistiskt baserade metoder för att, med mönsterigenkänning, hitta den troligaste lösningen. (Nationalencyklopedin) En av grundtankarna bakom ANN är, att tränas genom exempel för att bli bättre och bättre vid varje utförande av en viss uppgift. Vissa ANN kan klassificera data utan att använda kända exempel. De kan hitta en struktur (ett mönster) i datamängden utan träning med hjälp av extern information om den. Neurala nätverk används oftast i dag för olika former av mönsterigenkänning; såsom automatisk signalbehandling, textläsning och taligenkänning.

Natural Language Processing

Taligenkänning innebär automatisk identifiering av ord uttalade av en mänsklig talare. Det innebär enbart att datorn kan ta emot diktering, inte att datorn förstår sig på vad man har sagt. Att förstå innebörden faller under en annan datorvetenskap som kallas för NLP.

Även om den nämns ofta i samband med taligenkänning så finns det stora skillnader mellan dem. NLP förstår verkligen vad som sägs. Detta sker genom en process (parsning), varigenom orden i en sats och deras grammatiska kategorier och relationer identifieras. Ordformer och deras kategorier, i form av lexikon, lagras i dator. Med hjälp av grammatiska regler avgörs vilka ord eller fraser som kan kombineras och vilka typer av fraser som finns. (NE)

Målet med NLP är att åstadkomma en effektiv mänskliga-dator-kommunikation genom användning av naturligt språk. Om vi talar idag om t.ex. Sveriges statsminister och Göran Persson, måste systemet veta vilket årtal det handlar om, för att förstå att det handlar om en och samma person. För att underlätta förståelse och undvika fel, krävs det en stor kunskapsbas.

NLP handlar om hur vi förstår betydelse av en mening eller av en text. Ett av stora problem för utvecklingen av NLP är ords och frasers flertydighet. Trots brister, används NLP i system för automatisk textförståelse, grammatikanalys och översättning. (NE)

Historik

Att ha maskiner som talar, har varit människans dröm ganska länge. Redan på 30-talet gjorde man försök och det var inte helt omöjligt. Man förstod redan då hur den mänskliga talapparaten fungerar, och man lyckades skapa en maskin som kunde tala. AT & T's Bell Labs utvecklade 1936 den första elektroniska språksynthesizern, **VODER** (voice coder). Voder bestod av en person (operator) som jobbade med ett slags tangentbord och fotpedaler för att omvandla ljud till förståeligt tal.



Not. Från "Speech Analysis, Synthesis and Perception", Flanagan, 1965, Springer-Verlag

Bilden visar en demonstration av Voder, som utfördes av en utbildad "Voder-användare", på Världsmässa 1939 i New York, enligt Flanagan (1965).

Första försöket att utveckla system som kan förstå och känna igen tal var inte så lyckat. Det var under det sena 40-talet, då USA:s försvarsmakt finansierade utvecklingen av automatiska språköversättare. Man har sedan dess försökt att bygga ett bra taligenkänningsprogram genom att använda olika tekniker.

I början av 80-talet kunde program identifiera en liten vokabulär uttalade i isolation av en enskild talare. Utvecklingen har sedan inriktats på att få systemet att klara löpande tal och inte enbart isolerade ord. En ökning av vokabulären och en övergång från talarberoende till talarberoende var lika viktigt att uppnå.

Det var först i mitten av 90-talet som man började närma sig en fungerande lösning. De tidiga systemen, från början av 90-talet, krävde mycket datorkraft och man fick tala ord för ord.

Genom att datorernas prestanda har ökat, har tillverkarna av taligenkänningsprogram fått en chans att förbättra sina system. Idag kan systemen klara av löpande tal och nu fungerar de med mer än 90 % rätt.

Men varför tog det då så lång tid att utveckla ett godtyckligt taligenkänningsprogram? Problemet ligger i det mänskliga språket. Det som människan förstår lätt, är helt obegripligt för en dator.

Människans sätt att prata bygger mycket på antagande om att kontexten är självklar och inte behöver uttryckas. En exakt taligenkänning är en mycket komplex process. Även om datorn skulle kunna förstå alla ljud rätt, så är det inte säkert att den skulle förstå innehållet då talet enbart är en presentation av språket.(Linell)

Språk

Människor använder språket hela tiden utan att fundera på vad det är som gör att vi förstår vad en annan person säger, eller skriver. Men om vi hör ett främmande språk för första gången, hör vi bara en ström av ljud. Det som gör att vi hör olika ord i ett språk vi talar, är att vi har lärt oss att se mönster (olika bitar) i språket. Detta gör vi enligt regler eller vanor vi har lärt oss.

Mönsterigenkänning inom AI baseras på kunskaper om mänsklig mönsterigenkänning, sådan som vi använder t.ex. för att förstå ett språk. Att känna igen olika ord i ett språk är inte tillräckligt för att förstå ett språk, för att uppfatta betydelse av de uttalade orden.

Språk består av lexikon, dvs. ordförråd och olika regler eller konventioner för hur dessa ord skall hanteras. (Linell) Dessa regler är följande:

Fonologiska regler - reglerar ordets uttal

Morfologiska regler - reglerar hur orden böjes och hur nya ord bildas. Detta innebär användning av singular/plural form, prefix och suffix, osv.

Syntaktiska regler - reglerar hur orden får sammanställas till fraser, satser och meningar.

Semantiska regler - reglerar ords och ordkombinationens betydelse. Det är med hjälp av dessa regler som vi förstår (utifrån kontexten) skillnader mellan två ord som låter likadant.

Textgrammatiska regler - reglerar hur meningar kan bygga upp texter. Många dokument har en förbestämd struktur. T.ex. en vetenskaplig uppsats består av abstrakt, introduktion, metod osv. En menings betydelse avgörs enligt dess placering i uppsatsen.

Pragmatiska regler - reglerar hur ord och meningar användes för olika kommunikativa syften i olika situationer. Det innebär den allmänna kunskapen om verkliga världen.

Taligenkänningsprogram klarar av språkanalysen enligt fonologiska, morfologiska och även syntaktiska regler. Det är inte svårt att t.ex. tillämpa regler om ordföljd. Men vi börjar närma oss den stora problematiken, redan på semantiska nivå. Det är på den nivå som det börjar uppstå skillnader mellan taligenkänning och NLP.

Alla talar olika

Även om det kan kännas ganska enkelt för oss att avgöra vad en person säger, så är inte fallet med datorer. Det som gör det enkelt för oss, är att vi känner till de olika reglerna. Utan den kunskapen skulle det vara svårt för oss att skilja åt t.ex. orden ”par” och ”bar”.

Det är många faktorer som försvårar utvecklingen av taligenkänning, bl.a. olika ljudegenskaper hos olika individer. Generellt gäller att språk är beroende av talarnas biologiska begränsningar i talorgan, hjärna och nervsystem, vilka sätter vissa gränser för vilka ljud som kan uttalas, på vilket sätt man uttrycker sina tankar, etc. (Linell) Ljudegenskaperna hos orden varierar också beroende på talarens kön, ålder, dialekt, sinnesstämning och talstil. Talstilen i sin tur kan variera beroende på talarens bakgrund, utbildning, osv. Ännu en faktor som påverkar ljudegenskaper hos ord, är talarens modersmål.

Problemet med förståelse av tal ligger inte bara i talvariationer hos olika människor, utan även i att en person uttalar samma ord på olika sätt.

Talkommunikation

Talkommunikationen innebär ett invecklat socialt samspel mellan den som talar och den som lyssnar. ”Talaren måste ta hänsyn till lyssnarens situation när han formulerar sitt budskap, och på liknande sätt måste lyssnaren sätta sig in i talarens roll när han försöker förstå budskapet [...]” (Linell)

Vårt sätt att formulera budskapet på, varierar beroende på vad vi vet om lyssnaren. Som talare, väljer man medvetet på vilket sätt man skall framföra budskapet. Lyssnaren i sin tur försöker förstå budskapet. Talaren och lyssnaren är medvetna om att de kommunicerar med varandra. De är också medvetna varför de kommunicerar, hur och vad de kommunicerar om. Den insikten är väsentlig för förståelsen av språklig kommunikation. (Linell) I mänsklig kommunikation är det inte nödvändigt att säga allt. Där finns det underförstådda som en dator inte kan förstå.

Människor kommunicerar med varandra på olika sätt. Hur det uttalade budskapet utformas beror bl.a. på lyssnaren, situationen, miljön, etc. Män även sättet man framför det underförstådda på skiljer sig. Det skrivs om två sorts kommunikation, verbal och icke-verbal kommunikation. Den sist nämnda delas även upp i paralingvistisk och extralingvistisk kommunikation.

Verbal och icke-verbal kommunikation

Med verbal kommunikation menas språklig kommunikation, kommunikation med ord och kombinationer av ord. Förutom budskapet, signalerar vi med rösten dessutom kön, ålder, socialt och geografiskt ursprung.

Bland de vanligaste funktionerna hos den icke-verbala kommunikationen, finner vi, enligt Linell, är att uttrycka känslor (t.ex. glädje, ilska) och attityder till samtalspartner (t.ex. överlägsenhet och underlägsenhet, ironi) och ämne (t.ex. tvivel).

Paralingvistisk kommunikation omfattar ljud, ansiktsuttryck, gester, kroppsspråk mm som ackompanjerar och kompletterar den verbala kommunikationen. Extralingvistisk kommunikation uppträder oberoende av den verbala och omfattar både ljud (t.ex. nysningar, rapningar, stånkanden) och andra medel (t.ex. ansiktsuttryck, klädedräkt). (Linell) Människor kan inte utesluta den icke-verbala kommunikationen. Den icke-verbala kommunikationen kompletterar det vi språkligt uttrycker (med den verbala kommunikationen). Datorer saknar, än så länge, möjlighet att tolka den icke-verbala kommunikationen, vilket gör det svårt för människor att tala till datorer och för datorer att förstå tal.

Taligenkänningsteknologi

När man talar om talteknologi bör man skilja på följande: talarverifiering, talsyntes och taligenkänning.

Talarverifiering är en teknik som modellerar en talares röst/talapparat. Talarens identitet kan sedan verifieras genom att ett uttalande jämförs med den kända modellen.

Talsyntes, som också kallas **TTS** (text-to-speech), innebär att en given text omvandlas till en utgående akustisk signal. Med andra ord: datorn kan läsa upp ur en text, t.ex. e-post eller nyheter.

Taligenkänning är datorns uppgift att omvandla inkommande akustiska signaler till en text (transkription).

Hur fungerar taligenkänning?

För att datorn skall kunna förstå vad vi säger, måste det översättas. Taligenkänning är en teknologi som omvandlar uttalade ord till alfanumerisk text och styrkommando förståeliga för en dator.

Olika tillverkare av taligenkänningssystem använder sig av olika metoder för att göra den omvandlingen. Gemensamt för alla dem är hur omvandlingen går till, dvs. två grundläggande steg som följer olika modeller.

Akustik- och språkmodell

Den ena är en akustikmodell, då inkommande ljud analyseras och omvandlas till fonem - det grundläggande elementet i ett språk.

Omvandlingen sker genom att brus och all onödig information (t.ex. volymändringar) tas bort. Sedan, genom användning av automatiska beräkningar, reduceras data till en mängd av frekvenser. Den analyseras och omvandlas från ord till en digital representation av fonemen.

Den andra är en språkmodell som analyserar innehåll och jämför fonemkombinationer med ord som finns i ett inbyggt lexikon, för att bestämma vilka ord som uttalades. (Miastkowski, 2000)

Processen används av en parser, ett automatiskt system som analyserar satsen enligt en viss grammatik. Den avgör vilka ordklasser orden tillhör och hur orden hör ihop.

I alla språk finns det ord som låter likadant men de betyder olika saker. För att systemet skall känna av vilket ord det handlar om, används så kallade trigrams (eng.), vilket innebär att systemet analyserar kontexten som ordet i fråga befinner sig i. I många fall kan systemet känna igen ordet genom att titta på de två orden som har sagts innan.

NLP, till skillnad från taligenkänning, kan avgöra ordets betydelse och hitta dess kontext, genom att titta på en hel mening.

Olika typer av taligenkänning

I vissa fall anses taligenkänning vara mycket bra, då de bästa system fungerar till 95 % rätt. I andra fall anses det att taligenkänning inte duger p.g.a. felaktiga tolkningar. Detta bero på att det finns olika typer av taligenkänning, som fungerar olika. Man skiljer, först och främst, mellan talarberoende och talaroberoende system.

Talarberoende och talaroberoende system

Talarberoende systemen är de som idag fungerar allra bäst. De är mer eller mindre beroende av att användaren är alltid densamme. Sådana system finns idag bl.a. i mobiltelefoner och bilar. Först efter att användaren har lagrat in röstmärken kan denne använda taligenkänningsfunktioner. Processen går till så att systemet jämför det talade ordet med tidigare lagrade röstmärken.

Talaroberoende systemen innebär att olika användare kan använda en och samma produkt. För att kunna använda produkten krävs det en träning av systemet. Träningstid varierar beroende på produkten. Träningen innebär att användaren läser upp olika texter. Uppläsning analyseras och ord jämförs med det inbyggda lexikonet. Med träningen lär man systemet att lyssna på olika användare. Talaroberoende systemen används bland annat i olika applikationer för PC.

Diskreta och kontinuerliga system

Det finns ännu en uppdelning av taligenkänning: diskreta system och kontinuerliga system. Diskreta system presenterar en äldre teknik och kräver att användaren talar tydligt, långsamt och att den säger ett ord i taget, med paus mellan varje ord. Även då diskreta system inte är särskilt användarvänliga, så finns de kvar. Kontinuerliga system tillåter användaren att tala naturligt.

Diktering och kommandostyrning

Taligenkänningsprogram möjliggör att två uppgifter kan utföras med hjälp av datorn. Dessa är diktering och kommandostyrning.

Diktering innebär inläsning av text till datorn, dvs. till ordbehandling, e-post, fax, mm. Kommandostyrning innebär styrning av datorn genom talkommandon. I stället för att använda sig av kortkommandon eller att klicka med musen för att aktivera olika funktioner, som t.ex. öppna en fil eller skriva ut filen, ger man talkommandon i som t.ex. ”Öppna fil A”, ”Skriv ut fil A”, mm. Många system på marknaden klarar av både kommandostyrning och diktering.

Vi är fortfarande långt ifrån ett felfritt taligenkänningsystem och det beror på en del hinder, bl.a. störande ljud i bullriga miljöer.

På grund av systemens begränsningar och höga priser, så har system för taligenkänning hittills bara använts vid speciella situationer, t.ex. för användare som p.g.a. ett handikapp inte kan använda tangentbordet. Tack vare större satsning på utvecklingen och prissänkning finns det idag ett flertal väl fungerande kommersiella system för taligenkänning. Idag finns det många produkter och tjänster som använder taligenkänningstekniken, alltifrån röststyrda mobiltelefoner, automatiska telefoncentraler till biljettbokningssystem. De bästa systemen kan känna igen tusentals ord. Sådana system behöver träning så att de t.ex. kan känna av en viss dialekt. Oftast är sådana system anpassade till en användare.

Resultat

Taligenkänningsprodukter – var finns dem?

I undersökningen har jag stött på många användningsområden för taligenkänning, bl.a. i mobiltelefoner, ordbehandlingsprogram, webbapplikationer, datorstyrning, telefontjänster, m.m.

Taligenkänning används inte bara för att underlätta arbetet med PC utan även i andra miljöer. Företaget Voxi AB har utvecklat en ny typ av lösning för talstyrning av olika elektroniska apparater och informationssystem. Dessa produkter, enligt D. Adler (personlig kommunikation, december 2001), används och efterfrågas i inbyggda miljöer, handdatorer, desktop-datorer, servrar, m.m.

Ett av mer kända användningsområden för taligenkänning är mobiltelefoner.

Mobil taligenkänning

I mobiltelefoner används det talarberoende system, vilket innebär att användaren måste spela in röstmärken till poster i telefonboken. Sedan, när användaren säger namnet till personen han/hon vill ringa till, jämförs det med de inspelade röstmärkena.

En av de första mobiltelefonstilverkarna som använde taligenkänningsteknologin var Ericsson. År 1999 lanserade den första mobiltelefonen, med möjlighet till röstuppringning och styrning. (S. Dobler, Ericsson Review No. 3, 2000) Det största problematiken på området är, enligt J. Rasmusson (personlig kommunikation, december 2001), att det inte finns något större intresse (bland köpare) för produkter som de är idag. När de första produkterna kom ut på marknaden, för 3-4 år sedan, har de marknadsförts på så sätt att vanliga människor har förväntat sig mycket mer av röstuppringningen än vad den kunde åstadkomma.

”Mobiltelefoner med röstuppringning var inte så användbara som man trodde, vilket förklarar det svaga intresset”, säger Rasmusson (Ericsson Research, Lund).

Telefontjänster med taligenkänning

En större del företag med telefontjänster såsom kundsupport, biljettbokning, tidtabeller, nummerupplysning, etc. får dagligen samma frågor från många kunder. Att standardisera sådana frågor och införa en automatisk röststyrd tjänst, är ett sätt att utnyttja personalens kunskaper på bättre sätt. Många företag har infört röststyrda telefontjänster, bl.a. SJ, SMHI (Vilket Väder), Vägverket (Bilregistret), Telia, SAS, mm.

Jag har testat två tjänster: SAS SpeechLine och Telias röststyrda nummerupplysning (Telia autosvar).

SAS SpeechLine

SAS SpeechLine, som utvecklades av Scandinavian IT Group, är en av de tidiga tillämpningarna av taligenkänning i Sverige (SJ var först). Tjänsten finns nu i drift på flera språk: svenska, norska, danska och engelska. Den innehåller tre delar, SAS Globala tidtabell, SAS Aktuella ankomst- och avgångstider (dvs. förseningar i avgångar eller ankomster) samt Information om *standby* eller Bokningsbara ungdomsbiljetter.

Enligt O. Lundvall (personlig kommunikation, januari 2002), finns det i systemet en

grundläggande språkmodell som beskriver och tolkar naturligt tal. För att skapa en applikation som SpeechLine, måste man skapa en grammatik som innehåller alla varianter på orter och kommandon som kunden kan säga, exempelvis har det lagts in Götet som alias på Göteborg.

Tjänsten anses vara framgångsrik eftersom den tar emot cirka 3-5000 samtal per dag, även då ingen reklam har gjorts för tjänsten. Klagomålen på tjänsten har varit väldigt få. Tjänstens succé beror på, enligt Lundvall, faktum att företaget kontinuerligt följer upp hur systemet förstår olika människor och hur de uppfattar tjänsten.

Min uppfattning av tjänsten var att den var enkelt att använda. Den tiden jag väntade på att få välja en fråga, uppfattade jag som lång (ibland upp till en minut). Tjänsten för övrigt verkar vara ”väl uppfostrad”. Även om jag sade några ord med t.ex. fel betoning, uppfattades de rätt. Att jag talar svenska med brytning påverkade inte resultatet. Det var mycket enkelt att avbryta pågående fråga/samtal.

Telia autosvar

Telias röststyrda nummerupplysning ägs av Respons AB. Företaget använder sig av Philips taligenkänning som vidare utvecklats, av ett antal företag, till en fullständig tjänst. Bland företag som jobbade med utvecklingen finner man Philips, olika bolag inom Telia koncernen samt ett litet konsultbolag i Uppsala. R. Döös (personlig kommunikation, 7 januari 2002). De svåraste delarna att utveckla har varit, enligt Döös, ”[...] hanteringen av den stora databasen och transkriberingen av alla namn och adresser” samt en dialog som är enkel att förstå.

Uppringning till tjänsten visade följande:

Den tiden jag väntade för att få rätt svar, uppfattade jag som lång. Till skillnad från en vanlig nummerupplysning, för att kunna få telefonnummer till en person, behövde man även ange stadsdel eller adressen där personen bor. Man fick ingen möjlighet att bokstavera ett namn i fall systemet inte uppfattar rätt. För att avbryta samtalet fick jag upprepa ”Avbryt” flera gånger.

Taligenkänning och PC

Studier av information på Nätet visar att de mest omtalade tillverkare av taligenkänningsprodukter är L & H (Lernout & Hauspie, Belgien), Dragon Systems (USA), IBM, och Philips. Produkter från Dragon Systems numera saluförs av L & H.

De mest kända produkter inom taligenkänning visade sig vara Dragon Dictate (Dragon Systems), Free Speech (Philips) och Real Speak (L & H). Dessa program används för ordbehandling och de är ett slags mellanting mellan de talarberoende och talaroberoende systemen. Det innebär att användaren måste göra en hel del inställningar, samt utföra en lång träning av systemet, för att kunna använda det. Ju mer man använder (tränar) systemet, desto bättre blir det.

För många år sedan gjordes, med stöd från Hjälpmedelsinstitutet, en svensk översättning av DD. Dess senare version, Natural Speaking, översattes aldrig till svenska. Free Speech är utvecklad för flera språk, bl.a. svenska, men den blev aldrig kvalitetsmässigt lika bra som DD, enligt E. Bergek. (personlig kommunikation, januari 2002) Det är dock lättare att lära in så att det går vidare.

HT Speech är en svensktillverkad produkt som har samma grundprogram för taligenkänning som Dragon Dictate. Det är mer avsett för att styra datorn än för att mata in text. (Bergek) Till

skillnad från Dragon Dictate och Free Speech är HT ett diskret system, vilket innebär att man måste bokstavera text.

Det är inte bara program till ordbehandling som använder taligenkänning. L & H har utvecklat ett flertal produkter för arbetet inom hela Microsoft Office 2000/97 paketet. Det finns ett flertal produkter som ger möjlighet att, bara genom att tala, surfa på Nätet eller styra datorns vanliga funktioner, såsom att öppna filer eller skriva ut dem.

IVAN, som står för *Intelligent Voice Animated Navigator*, är en animerad guide som underlättar sökning på webben. Applikationen visas på dataskärmen som en liten klot med ansiktet, armar och ben. Den kan tala, lyssna och förflytta en genom olika webbplatser. Man använder den genom att ställa enkla frågor, som t.ex. ”Show me a weather rapport for tomorrow”. Det som sägs, visas på skärmen i en textruta. En ny applikation, som finns att ladda ner gratis från Nätet, heter myIVAN. (<http://www.myivan.com/index.htm>)

Marknadsundersökning

Mina försök att ta reda på mera om den allmänna intresse för taligenkänningsprodukter, genom att ringa till ett flertal företag som säljer hård- och mjukvara för PC resulterade i att jag möttes av en tystnad då jag nämde taligenkänningsprogram. Reaktionen blev inte mycket annorlunda efter att jag har förklarat vad jag menar med taligenkänning, och nämnt ett antal ”kända” produkter såsom Dragon Dictate och Freedom of Speech. Detta gäller ett större antal uppringda personer.

Några personer hänvisade till en av de största mjukvaruförsäljare på Nätet, företaget Dustin AB. Det visade sig att inte ens de säljer några taligenkänningsprodukter.

Även då jag har ringt till 20-tals slumpmässigt utvalda företag, var det bara två av dem som säljer taligenkänningsprodukter (Freedom of Speech). Dessa är ONOFF och Computer City i Göteborg.

På Computer City kunde man inte samtala med någon som jobbar med försäljning. Alla samtal kopplas direkt till kundtjänsten. Jag fick inget svar angående försäljning förutom att kundtjänsten har fått några förfrågor om produkterna.

Intervjuade försäljare på ONOFF, S. Nilsson (personlig kommunikation, december 2001), svarade att det finns väldigt svagt intresse för taligenkänningsprodukter. Försäljningen ligger långt under förväntningarna, vilket kan bero på den dåliga marknadsföringen av produkterna. Den intervjuade hänvisade till några större försäljare av mjukvara.

Produkter fanns dock hos vissa företag som säljer hjälpmedel till funktionshindrade personer. Utveckling av hjälpmedel för funktionshindrade personer är ett stort område där det finns intresset för taligenkänningsprodukter.

Taligenkänning för personer med olika funktionshinder

Enligt M. Lundman på Hjälpmedelsinstitutet (personlig kommunikation, januari 2002), finns det flera grupper, bland funktionshindrade människor, som har nytta av denna teknik. Det gäller bland annat synskadade, rörelsehindrade, talhandikappade och personer med läs- och skrivsvårigheter. Studien har visat att flertalet företag, som arbetar med utveckling och/eller distribution av hjälpmedel för funktionshindrade, är väl bekanta med taligenkänningsteknologin. Några av de är Hjälpmedelsinstitutet, DART, Frölunda Data, mm.

DART är Västra Sveriges kommunikations- och dataresurscenter för barn, ungdomar och vuxna med funktionshinder. Enligt M. Lundälv (personlig kommunikation, januari 2002), har det funnits och finns ett stort intresse för taligenkänningsprodukter. ” Ibland är dock förväntningarna orealistiskt höga. Vår roll blir ganska ofta att ta ner förväntningarna på tillgänglig teknik, på en realistisk nivå”, skrev Lundälv.

De som intresserar sig för taligenkänningsprodukter, hos Dart, är främst personer med (läs- och) skrivproblem, samt rörelsehindrade personer utan, eller med lindriga, talhandikapp. Det handlar om en ganska speciell användning av talstyrning - ett mycket litet antal styrkommandon i kombination med något annat alternativt styrsätt.

Darts erfarenheter av användning av taligenkänning gäller mest flerhandikappade personer med gravare funktionshinder. De begränsade erfarenheter av taligenkänning/diktering för personer med grava dyslexiproblem i kombination med språkstörningar har inte varit särskilt positiva. Men, det har funnits några delvis positiva resultat - mest för kommandostyrning i kombination med andra styrsätt, enligt Lundälv.

Frölunda Data är ett företag som distribuerar utrustning till personer med funktionshinder. Av taligenkänningsprodukter har de, för närvarande, endast Free Speech till försäljning. Förfrågningar om taligenkänningsprodukter kommer, i första hand, från personer med funktionshinder, som inte klarar att skriva på vanligt tangentbord. I andra hand, kommer de från personer som har behov av att mata in texter snabbt, som t.ex. läkare och advokater. Även personer med förslitningsskador (musarm) har frågat efter taligenkänningsprodukter. (E. Bergek, personlig kommunikation, januari 2002)

Den gruppen som säkert har störst nytta, är gravt rörelsehindrade med normalbegåvning. Där kan programmen användas, både för att mata in text, och för att styra datorns olika funktioner. När det gäller dyslektiker som användargrupp, är Bergek tveksam. ”Det råder stor diskussion om huruvida det är bra eller inte för personer som inte direkt ser när det blir fel.” Upplärning av programmet kan vara svår eftersom man bör läsa in många texter/ord som i sig kan vara svåra att läsa för dyslektikern. Det rekommenderas alltid, att man kompletterar programmet med bra stavningskontroll och talsyntes, som kan läsa upp den inmatade texten så att man hör eventuella fel, skrev Bergek.

För- och nackdelar med taligenkänning?

Undersökningen har visat att det finns många för- och nackdelar med taligenkänningen.

Bland de största fördelarna är användning av olika taligenkänningsprodukter i bilen. Genom att använda en mobiltelefon utrustad med *headset* och röstuppringning slipper man trycka på knappar samtidigt som man kan koncentrera sig på trafiken. (Rasmusson)

Ännu en fördel med taligenkänning, speciellt talarberoende system som används i mobiltelefoner, är att det kan användas för alla språk. Detta p.g.a. att programmet består av ”*ett-ord-modeller*”, dvs. det uttalade ordet matchas med röstmärke som är lagrad i telefonen.

Vad det gäller taligenkänning för PC, är fördelen, framför allt, möjlighet att snabbt få en större textmängd, utan att skriva på tangentbordet. Detta är speciellt viktigt för personer som jobbar mycket med dator. Användning av taligenkänning som komplement till andra styrsätt

kan effektivisera och avlasta andra funktioner. ”För en läkare, t.ex. kan taligenkänning ersätta bandspelaren, som ju lyssnas av en läkarsekreterare som skriver in texten.”, skrev Bergek.

Nackdelen här är att taligenkänningsteknologi är tekniskt komplicerad och sårbar ännu så länge. Den är också ”för krånglig för vanliga människor, speciellt talarberoende produkter”, anser Bergek. Det krävs att man lär sig programmets funktioner och lägger tid på att ’lära’ programmet att förstå vad man säger.

Produkterna kräver inte bara mycket träning under kompetent handledning för ett rimligt gott resultat, utan även noggrant underhåll. (Lundälv)

Fackuttryck och konstiga ord kan vara svårt att diktera in. Men även variationen i rösten på verkar resultatet. För att programmet skall fungera som bäst, måste man ha lugnt och tyst omkring så att inte inmatningen störs av andra ljud. (Bergek)

En tidigare nämnd fördel, möjligheten att tala till datorn för att underlätta och effektivisera arbetet, kan ses som en nackdel också. Speciellt för personer som jobbar på ett modernt kontor med öppet landskap, där användning av taligenkänningsprodukter inte bara utgör en sekretessrisk, utan även försvårar arbete för kringplacerade kollegor som behöver jobba ostört. Ännu ett problem som kan uppstå här är att alla ljudförändringar och brus, på ett sådant kontor, skulle försvåra användningen av produkterna.

En av stor fördelar, när det gäller företag som erbjuder sina kunder en automatiserad service, är att med röststyrning kan dessa tjänster bli mycket mer kundvänliga och lättanvända. ”Det är klart att vissa människor alltid vill tala med en annan människa, även när de ringer till kundtjänst, men väldigt många uppskattar möjligheten att snabbt få svar utan att behöva vänta i kö.” (Lundvall)

Att ersätta en del av kundtjänst med ett röststyrt autosvar kan ses som ett hot till många jobb, som t.ex. läkarsekreterare. Även då den stora fördelen kan, i vissa fall, vara snarare en nackdel som innebär att personalen friställs p.g.a. införandet av tjänsten, så är inte fallet med SAS. Lundvall skrev: ”Självklart är det så att röststyrda tjänster är till för att avlasta personal, eftersom SAS konstant har för lite personal i sina kundservicefunktioner så har de nya tjänsterna hittills inte medfört att någon behöver sluta.”

När det gäller SpeechLine, tycker de flest människor att tjänsten är lätt att använda, det finns dock alltid några röster som systemet har svårt att förstå. Även då systemet förstår dialekter mycket bra, ibland kan det vara en speciell röst som av olika orsaker inte passar. (Lundvall) Detta är en av de största nackdelarna med taligenkänning, att den inte är felfri.

Framtid

Undersökningen har visat att många tillfrågade har förhoppningar och stor tro på taligenkänning.

Många investerare ser talteknologi som en framtidsbransch och har redan investerat i företag inom branschen. Marknaden idag är mest inriktad på olika tjänster över telefon (såsom beställning, information, bokning etc.). ”Framöver tror vi att marknaden för företagsspecifika applikationer på handdatorer blir en stor marknad”, skrev Adler.

Framför allt, kommer taligenkänningsteknologi att vara mycket viktig för mobilteknologi. Eftersom mobiltelefoner blir allt mindre, medan deras display blir större, utrymme för tangenterna minskas. Som alternativ till de smala knapparna kommer man att använda talstyrning. (Dobler, 2000)

Det råder även uppfattning att talstyrningstjänster kommer att få sitt stora genombrott inom "call-center" och andra telefontjänster. (Taltekniken erövrar ny mark. 2001, oktober 12. Dagens IT)

Frågan är bara när det stora genombrottet kommer att ske. En del tillfrågade anser att det kommer att ta ganska lång tid. "[...] särskilt för små språk som svenska, kanske 5 - 10 år eller mer, och det kommer inte att fungera för allt och alla.", svarade Lundälv.

Det visade sig också att en del tillfrågade personer tror på att det kommer att ske inom snarare framtid, 2-4 år, när produkterna utvecklas så pass att de blir lättare att använda.

Talsvarstjänster på telefon, där man ger sina svar med ord/meningar, förekommer ju redan och blir säkert fler lösningar där. På sikt tror Bergek att många fler grupper kan ha intresse av att kunna "läsa in text" i datorn.

Taligenkänning kommer att få ett genombrott, men det kommer att ta längre tid än vad de flesta föreställt och föreställer sig. (Lundälv) Detta problem illustreras av den kaotiska situationen på talteknologimarknaden under senaste tiden, vilket syftar på att Dragon systems blev uppköpt av H & L, vilket i sin tur gick i konkurs (Lotsson A., 2001).

Vad det gäller framtiden, studien visade att det har startats flertal projekt för utveckling av talstyrda tjänster, bl.a. på Sveriges Television (SVT), som arbetar med att utveckla ett talgrensnitt mot Text-TV-databasen och diverse program.

Vad det gäller Scandinavian IT Group, så arbetar de nu med att utöka funktionaliteten i deras röststyrda tjänster. Det väntas komma en mängd nya tjänster under det närmaste året, ex bokning, ombokning, olika informationstjänster etc. Många företag är intresserade av röststyrda tjänster och Scandinavian IT Group erbjuder sina tjänster även utanför SAS och har därför etablerat ett samarbete med ett företag som heter Voice Provider. De har idag flera uppdrag som de arbetar med, enligt Lundvall.

På frågan om taligenkänning kommer att få genombrott, svarade Lundvall så här: "Röststyrda tjänster är ännu så länge i sin barndom, men om du tänker på att rösten faktiskt är ett av de mest naturliga medel att kommunicera med så kanske du förstår att detta område kommer att växa. Varför skall man ha ett nummertangentbord på telefonen? Det räcker ju att säga namnet på den man söker."

Diskussion

Med denna undersökning ville jag, framför allt uppmuntra forskare och studenter att undersöka taligenkänningsområdet bättre, för att kunna påpeka teknologins brister och på så sätt påverka utvecklingen av den. Jag anser att jag har, med denna undersökning, besvarat mina frågeställningar.

Vad jag har kommit fram till i undersökningen är att det finns ett brett användningsområde för taligenkänning. Att teknologin inte har kommit till full användning kan bero på att det inte har funnits speciellt många utvecklare av den samt att det saknas något större intresse för den, inom vissa områden.

Taligenkänning har visat sig vara ett nyttverktyg och inte enbart hjälpmedel till den lata människan. Den används både för att förbättra säkerheten i trafiken, där förare kan uppmärksamma trafiken i stället för tangenterna på telefonen, och för att underlätta kommunikation med dator, speciellt för personer som av fysiska skäl har svårt att använda tangentbordet. Taligenkänning används även för att underlätta arbetet inom olika typer av företag.

Resultatet, framför allt när det gäller användningen, skulle ha kunnat vara annorlunda om jag utförde studien bl.a. under en längre (och en annan) tidsperiod. I så fall tror jag att jag skulle kunna intervjua några användare, eventuellt utföra bredare testning av t.ex. telefontjänster men även få bättre respons från de kontaktade personerna, med tanken på alla helger. Många av kontaktade personer svarade mycket kortfattat, vissa med förklaring att de inte hade mycket tid för att svara på frågor, och en del hade för mycket att göra för att kunna ställa upp.

Att det har visat sig att det finns väldigt svagt intresse för taligenkänningsprodukter för PC behöver inte nödvändigtvis vara så. I min studie har jag inte haft möjlighet att få information om alla produkter som köps t.ex. från utlandet, via Nätet. Det var inte heller möjligt att undersöka marknaden i hela Sverige, utan den baserades på en region. Dock har intresset bland företag och personer med funktionshinder visat sig vara stort oavsett geografisk område.

Under mitt arbete blev jag mest förvånad över att taligenkänningsprodukter inte marknadsförs bättre, speciellt produkter för användning på PC. Lika förvånande är att människor som jobbar inom försäljning av mjukvaror inte känner till produkterna, vilket inte stämmer på de som jobbar med produkter för funktionshindrade människor.

Vad det beträffar testning av röststyrda telefontjänster, vill jag påpeka att testningen utfördes enbart för att jag skall kunna få en uppfattning om användningen av taligenkänningsprodukter. Den skall inte representera en jämförelse mellan olika tjänster, eftersom den är avsedd att representera min egen (subjektiv) uppfattning av tjänsterna.

Studien har visat att taligenkänning har lika många nack- som fördelar, om ännu inte flera. En observation av användare skulle kunna visa även andra för- och nackdelar. Det visade sig att en av de stora nackdelarna med taligenkänningsprodukter är att de inte alltid är effektiva, vilket kan upplevas inom en del olika sammanhang. Om man skall lyckas med en

taligenkänningsprodukt bero det helt på vad som skall göras. Jag anser att de nämnda för- och nackdelar är tillräckliga för att visa att det finns en framgång i utvecklingen.

Det är ju helt klart att taligenkänning inte har börjat blomma i Sverige än. Detta kan ha många anledningar. Att det inte utvecklas några produkter här beror på, bland annat, att alla produkter måste översättas. Översättning av olika applikationer från språk till språk är en problematik för sig, som inte har nämnts i denna uppsats.

Vad det gäller utvecklingen i framtiden, anser jag att jag har fått svar på min fråga. Undersökningen har visat att det finns ett flertal projekt på gång, men också stora förhoppningar om att taligenkänning kommer att få ett genombrott snart. Detta trots alla nedgångar som, taligenkänningsjättar (Dragon, L & H) möttes med under den senaste tiden.

Slutsats

Jag har i studien kommit fram till att taligenkänning har ett flertal brister, vilket tyder på att det finns en hel del kvar att utveckla. Men det har också visats att teknologin är användbar och har mycket att tillföra både vår vardag i hemmet, och på jobbet. Att taligenkänningsapplikationer för PC inte har slagit igenom i Sverige, kan bero mer på dålig marknadsföring av produkter än på deras brister. Det kan också bero på att människor, överhuvudtaget, är försiktiga med att använda allt som är nytt och okänt, och att de håller sig gärna till gamla, väl invanda rutiner. En användargrupp som har insett styrkan och fördelarna med taligenkänning, och där det finns stort intresse för produkter, är personer med olika fysiska begränsningar. Taligenkänning har en framtid som ett nytt verktyg, både för företag och privata personer. Genom att intensivifiera forskningen kring taligenkänning, även för svenska språket, skulle man kunna minska brister som dagens produkter har, och erbjuda välfungerande produkter till människor som efterlyser dem. Säkerheten i bilen kommer inte att öka med användningen av taligenkänningsprodukter, så länge systemen ger felaktiga resultat. Det är snarare tvärtom, att man får farliga irriterade förare som har ett flertal gånger försökt ringa till någon, utan att lyckats. Läkarna har inte tid att sitta och diktera genom ett gränssnitt som genererar fel hela tiden, när det går snabbare att diktera till en sekreterare. För att taligenkänning skall förbättras och kunna användas krävs det mer forskningsarbete inom området. Så småningom, tror jag, kommer den att utgöra en allt vanligare del utav vår vardag.

Fortsatt forskning

Även då det redan pågår forskning kring taligenkänning inom medicin, tycker jag att taligenkänning för rörelsehindrade bör uppmärksammas ännu mer, allt eftersom där kan den vara till stor nytta.

Det skulle även vara intressant att se hur taligenkänning kan utvecklas som hjälpmedel till gamla människor. Jag tror att det skulle underlätta deras vardag om de kunde tala till sina hushållsapparater, men även styra alla telefonsamtal med rösten i stället för att använda telefonknappar.

Referenser

1. Buchholz M. & Havstam C. (2000). *Taligenkänning och dysartri (Delrapport i projektet: Talstyrning för personer med talstörning)*. Handikappkommittén Västra Götaland
<http://www.dart-gbg.org/dokument/rapporter/TOD-rapport.pdf>
2. Dobler S. (2000). *Speech recognition technology for mobile phones*. Ericsson Review No. 3
http://www.ericsson.com/about/publications/review/2000_03/article113.shtml
3. Flanagan J.L. (1965). *Speech Analysis, Synthesis and Perception*, Springer-Verlag.
<http://www.haskins.yale.edu/haskins/HEADS/SIMULACRA/voder.html>
4. Forne-Wästlund H. (projektledare). 1998. Datorbaserad talträningssutrustning för personer med stamning. Stamningsrehab AB, 703 78 Örebro.
<http://www.hi.se/it/NyTeknik/Projekt/0379.shtml>
5. Linell P., *Människans språk* (Halmstad: Läsprodukter AB, 1990).
6. Lottson A. (2001). *Belgisk språkteknik under kluban*. Computer Sweden
7. *Nationalencyklopedin Multimedia 2000 på CD-ROM*. (2000). Höganäs: Bra Böcker.
8. Miastkowski S. (2000). *How It Works: Speech Recognition*. PCWorld.
<http://www.pcworld.com/resource/printable/article/1,aid,16276,00.asp>
9. *Taltekniken erövrar ny mark*. (2001, 12 oktober). Dagens IT
http://www.dagensit.se/pub/ipsart.asp?art_id=15145
10. Turban E. & Aronson J.E. (2001). *Decision Support Systems and Intelligent Systems*. Prentice Hall International.
11. Wormek A.K., Ingenerf J. & Orthner H.F. (1997). *SAM: Speech-Aware Applications in Medicine to Support Structred Data Entry*. Institute of Medical Informatics and Health Service Research, Neuherberg, Germany
<http://medicine.ucsd.edu/f97/D004213.htm>

Bibliografi

1. Allwood J. *Några perspektiv på mänsklig kommunikation*. Institution för lingvistik, Göteborgsuniversitet.
<http://www.ling.gu.se/~jens/publications/docs001-050/044.pdf>
2. DART, Regional habilitering, Sahlgrenska Universitetssjukhuset, Göteborg
www.dart-gbg.org
3. Feldman S. (1999). *NLP Meets the Jabberwocky: Natural Language Processing in Information Retrieval*, ONLINE
<http://www.infoday.com/online/OL1999/feldman5.html>
4. *Guide to Speech Recognition*.(1998). ZDNet UK, Reviews
<http://www.zdnet.co.uk/pcmag/supp/1998/speech/>

Bilaga: Intervjufrågor

1. Vilka produkter arbetar ni med?
2. Inom vilka miljöer kan de används?
3. Hur ser den svenska marknaden ut idag?
4. Finns det något intresse för taligenkänningsprodukter?
5. Var finns det största efterfrågan?
6. Vilka är mest intresserade av produkterna?
7. Har det funnits stort intresse hos investerare?
8. Stämmer försäljningen med förväntningarna om den?
9. Vad får ni för respons från kunder?
10. Uppfattas produkter som användbara?
11. Vilka är stora för- och nackdelar med taligenkänning?
12. Vilket företag utvecklade tjänsten?
13. Används den så mycket som Ni har hoppats på?
14. Anser kunder att tjänsten är lätt att använda? /Vill de hellre prata med en levande människa?
15. Har införande av tjänsten inneburit stora omstruktureringar i företaget?
16. Fick Ni friställa personalen p.g.a. detta eller infördes tjänsten enbart för att avlasta personalen?
17. Kommer Ni på något sätt att utöka tjänsten?
18. Tror Ni att flera företag kommer att införa liknande tjänster?
19. Framtidsvision? / Hur kan taligenkänning tänkas utvecklas i framtiden?
20. Kommer taligenkänning att få ett genombrott?
21. Vem tror Ni att framtidens kund blir?