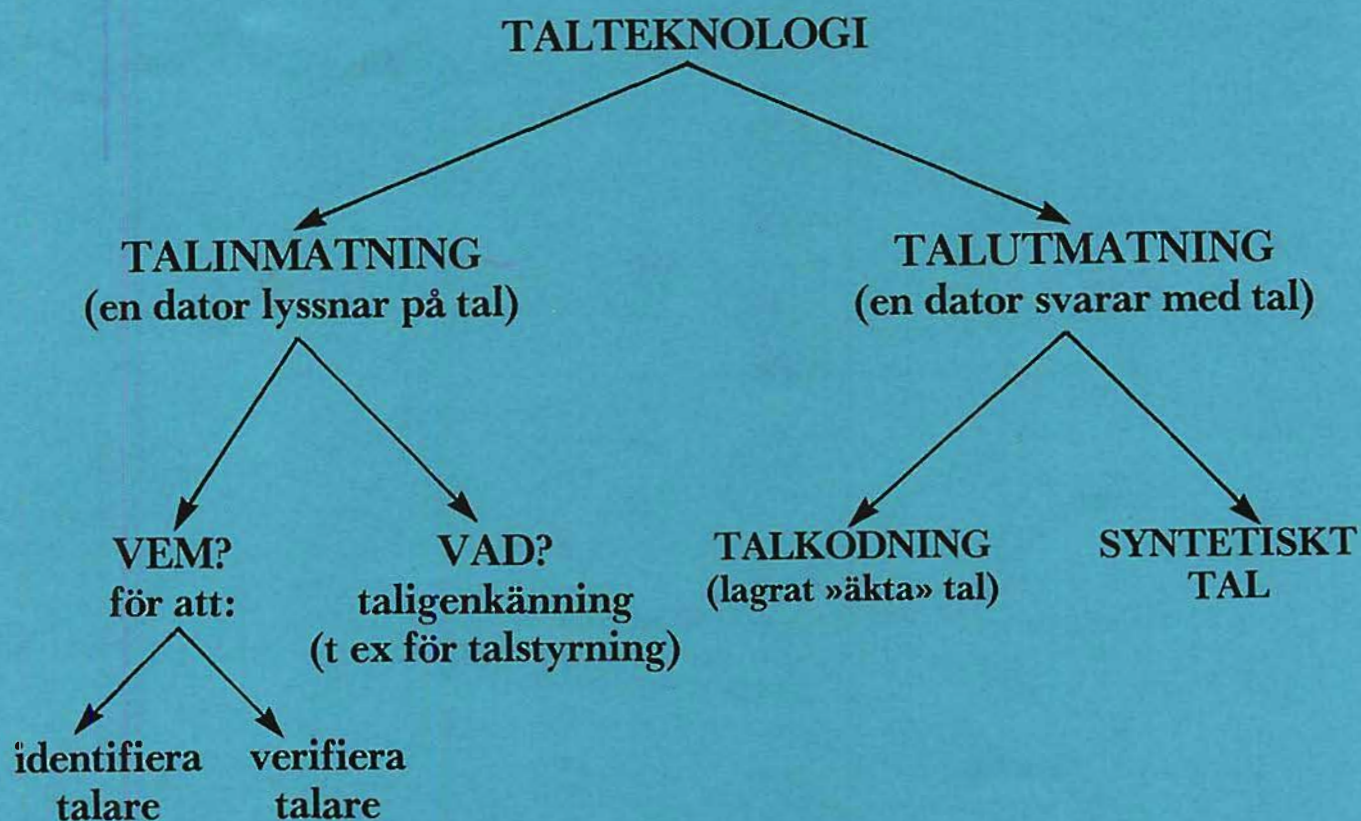


# TELDOK-info

Nyhetsblad som utges av TELDOK (se s 2)

Nr 1 • Årgång 1 • ISSN 0280-9567

*Det här numret handlar om:*



## UR INNEHÅLLET

Vad är TELDOK?	2
I det här numret behandlar vi tal	3
Ska man räkna med tal?	4
Telefonvakt, telefonsvarare – och sen?	5
Röstbrevlådor – i Sverige också!	6
Så här fungerar några röstbrevlådor	6

Tillverkare av röstbrevlådor	8
Röstupplysning	9
Så här fungerar talutmatning	9
Så här fungerar talinmatning	11
Några olika taligenkännare	13
Röstkontoret	14



\*\*\*Nr 1 \*\*\* Årgång 1 \*\*\* November 1982 \*\*\*

TELDOK-INFO utges av TELDOK Redaktionskommitté

Bertil Thorngren (ordförande), Televerkets huvudkontor (Gdp), 123 86 FARSTA

Materialet i det här numret har utarbetats av P G Holmlöv med hjälp av synpunkter från Hans Hise och Fred Lundin m fl vid televerkets huvudkontor

\*\*\* TELDOK-INFO beräknas utkomma 3 gånger per år \*\*\*

## VAD ÄR TELDOK?

TELDOK är telestyrelsens stöd till dokumentation och spridning av erfarenheter som rör teleanknutna informationssystem.

Televerkets styrelse (telestyrelsen) beslutade den 26 augusti 1980 att inrätta ett särskilt anslag om 5 miljoner kr för att medverka till att dokumentation av olika teleanknutna informationssystem snabbt och lättfattligt kan spridas till intresserade. Detta anslag till dokumentation och informations-spridning kallas TELDOK och är avsett att fördelas under 5–6 år. TELDOK gäller framför allt kontorsfunktioner men kan även avse andra teleanknutna informationssystem för tal, data, text och bild.

Genom TELDOK vill telestyrelsen ge möjlighet till:

- 1 dokumentation vid tidigast möjliga tidpunkt av praktiska tillämpningar av teleanknutna informationssystem, företrädesvis för kontorsfunktioner.
- 2 publicering och spridning, i förekommande fall översättning, av annars svåråtkomliga erfarenheter av teleanknutna informationssystem, företrädesvis för kontorsfunktioner, samt kompletteringar avsedda att öka användningsvärdet för svenska förhållanden och svenska läsare.
- 3 studieresor och konferenser i direkt anknytning till arbetet med att dokumentera och sprida information beträffande praktiska tillämpningar av teleanknutna informationssystem, företrädesvis för kontorsfunktioner.



För TELDOK finns en redaktionskommitté som sköter utgivningen av rapporterna liksom av TELDOK-INFO. I TELDOK Redaktionskommitté ingår:

**Datadelegationen,**

Agneta Owerin, tel 08–763 2372

**Data- & Elektronikkommittén,**

Jan Carlsson, tel 08–763 2908

**Dataeffektutredningen,**

Lars Loman, tel 08–21 98 01

**Forskningsrådsnämnden,**

Bengt-Arne Vedin, tel 08–23 25 20

**LO,**

Birgitta Frejhagen, tel 08–22 55 80

**Riksdataförbundet,**

Leif Jonas, tel 08–52 07 20

**TCO,**

Peter Magnusson, tel 08–14 24 00

**Televerket,**

Bertil Thorngren (ordf), tel 08–713 3077,

Göran Fredriksson, tel 713 2319,

P G Holmlöv, tel 713 4131

Ring gärna någon av dem om du har synpunkter på vad som har publicerats eller bör publiceras.



## I DET HÄR NUMRET BEHANDLAR VI TAL

Vad betyder det, att behandla tal? Det är inte fråga om talteknik som hos en logoped där man får säga bo-bö-ba-bu och bolla med rösten. Det gäller istället en parallell till ord- och textbehandling. Med ord- och textbehandling kan man ju lagra, ta fram, ändra, stryka och kopiera ord och text; talbehandling och talteknologi betyder att ungefär samma operationer kan utföras med tal som utgångspunkt. Det här numret handlar om det. Snart kommer nämligen pratorerna . . . talande och lyssnande datorer.

**Digitaliseringen** — mycket enkelt uttryckt, den kodning av talet i »nollor» och »ettor» som förekommer redan i vissa av dagens telefonsystem — gör det möjligt att snabbare än förut leta i, nästan bläddra i, lagrade röstmeddelanden. I förlängningen t o m att kombinera allt innehåll som översänds i telefonledningar så att man t ex kan lagra både ett brevutkast (i text) och egenhändigt intalade kommentarer (i tal) tätt intill varandra på något s k minnesmedium, som ett skivminne eller en kiselbricka (chip). Den som läser brevutkastet får också höra kommentarerna. En annan sak som bidrar till utvecklingen på området är att **kostnaderna för minnesenheter** — särskilt de som kallas chips — stadigt minskar samtidigt som det går att lagra fler detaljer i en minnesenhet.

Man kan grovt urskilja två huvudvarianter av talbehandling (men det finns fler). Den **ENA** är: att **känna igen tal**. Det kallas **talinmatning**. Antingen vill man känna igen **vem** som pratar, eller också vill man att datorn så att säga ska förstå **vad** som sägs. När det gäller **vem** som säger något så är det för att kunna identifiera eller verifiera, alltså bekräfta, vem som t ex ska få tillträde till en datorhall eller till en bankavdelning. När det gäller **vad** som har sagts så är det förstås för att instruera och styra datorn, t ex när man med ett fåtal korta och enkla kommandon kan få en eldriven rullstol att svänga först åt höger och sedan åt vänster eller ett bagageband att skicka en resväska till platsen där en lastvagn står som ska packas med väskor för ett visst flygplan.

Den **ANDRA** varianten är: att automatiskt **åstadkomma** tal. Det kan man göra

- 1) genom att **lagra** och sedan spela upp (kanske i andra kombinationer av de inlästa orden) mänskligt tal, **eller**
- 2) genom att utgå från regler för bl a uttalet av mänskligt tal och sedan få datorn att **åstadkomma helt syntetiskt** tal. (Det händer att båda varianterna kallas talsyntes.) Om det förra, att känna igen vad som sägs eller vem som talar, kan kallas taligenkänning eller talinmatning, så brukar man oftast kalla det här **talutmatning** eller ibland **talsvar**. Man kan tänka sig att sådana talsvar — lagrade eller syntetiska — kan beställas på olika sätt, antingen genom att folk ringer ett telefonnummer och får ett slags röstbrev eller får höra ett allmänt tillgängligt budskap (som Fröken Ur eller TT-Nytts telefonsvarare) eller genom att man trycker på en knapp. Det finns ju talande kalkylatorer, motsvarande rättstavningsapparater (Speak-and-Spell) och kassaapparater som kan uttala priset på varorna samtidigt som apparaten läser av den streckade s k EAN-koden som numera finns tryckt på en hel del förpackningar.

Forskningen har kommit längre beträffande talutmatning än ifråga om talinmatning. Vi presenterar i det här numret exempel på båda typerna av system, hur de fungerar praktiskt och tekniskt, och — närmast — en översikt av varför talbehandling behövs. Vad är det bra för?



## SKA MAN RÄKNA MED TAL?

Det finns tre faktorer som talar till röstens fördel om man jämför med andra sätt att överföra information och kommentarer, som t ex med text eller bilder.

- Talet är ibland *naturligare*. Skriftspråket är bara ett förkonstlat sätt att lagra tal på.
- Talet är ibland *effektivare*. Dvs, det finns många situationer när man kan tala men inte skriva, liksom det finns meddelanden som på ett med effektivt sätt kan överföras med rösten än med text.
- För en del är skriftspråk eller olika arbetsmoment som måste utföras med händer och kropp en *barriär*. Tex: genom att styra maskiner med tal istället för med armarna kan rörelsehindrade hitta fler arbetstillfällen.

**Det naturliga talet.** Flera undersökningar av hur människor tillbringar sin tid på arbetet — främst gäller det kontor — visar att den allra största andelen går åt till att *prata*. Hämta uppgifter på telefon, svara på frågor från folk som ringer upp, sitta i sammanträde, besöka andra. Det gäller särskilt »managers» — chefer, handläggare, direktörer — som ofta vill få underlag för beslut presenterade för sig i tal hellre än i text. Tabellen visar hur stor del av tiden — enligt fyra stora studier som refereras av Ray Panko i *Telecommunications Policy* 1981 (vol 5, 265 — 272) — som handläggare ägnar åt att tala respektive göra annat. Över halva tiden går åt till tal, det mesta personligt.

### % av arbetstiden som man:

Samtalar personligt	} 46	54	} 52	54
Talar i telefon		6		9
Läser/skriver	21	28	24	24
Gör annat	33	12	24	20
(Antal personer)	299	160	76	66

Telefoner finns där det finns människor, allrahelst på kontor. Textterminaler, ordbehandlare och telefaxapparater finns däremot mer sparsamt utplacerade i kontorsmiljö (och nästan aldrig i hemmen).

**Det effektiva talet.** Kjell Elenius och Mats Blomberg sysslar med system för taligenkänning vid KTH (Tekniska Högskolan i Stockholm), närmare bestämt institutionen för talöverföring och musikakustik. I en artikel ur *Elteknik* 1980 (nr 16) anger de vilka tidsbesparingar som man kan uppnå även med mycket enkla talinmatningssystem. Mellan 30 och 95 % av tiden kan sparas. Tal är ibland snabbare än text eller knapptryckning. Den som talar har händer och ögon fria.

Talinmatning vid...	Sparar:
Kvalitetskontroll i tillverkningsindustri	30 %
Godssortering	50 %
Lagerkontroll, inventering	90 %
Omprogrammering av tillverkningsrobotar	95 %

**Talet som övervinner barriärer.** För många handikappade är tal lämpligare än andra sätt att informera eller få information. För handikappade skulle talstyrning (talinmatning) av lampor, spis, radio och tv, rullstolar m m vara bekvämare än att manövrera med händer och fötter. För blinda är det istället talutmatning som är mest intressant — att kunna beställa upplysningar av alla slag per telefon, att höra siffrorna man just slår in på en kalkylator, att tidningstext automatiskt omvandlas till inspelat tal.

**Några användningsområden.** Ett par tillämpningsexempel har redan getts. Några andra: vissa minidatorer eller persondatorer (PET och den dator som säljs av affärskedjan Radio Shack i USA) kan mata ut talmeddelanden, och till persondatorn Apple II kan man ge talkommandon.

● Det finns exempel på maskiner som lyssnar på inmatat tal och som kan användas för att godkänna — »auktorisera» — t ex kontokort (dvs som hjälper till vid kontroll att inte ett visst kort är stulet eller kontot överskridet). En amerikansk detaljistkedja har installerat ett system där en talutmatningsenhet ropar upp priset på en vara samtidigt som personen i kassan slår in priset eller läser av EAN-koden på förpackningen. Utmataren säger alltså, t ex, Heinz vita bönor, femtiofyra cent.

● Det rapporteras att biltillverkaren Chrysler beställt utrustning från elektronikföretaget Texas Instruments för att ha talande instrumentpanel på fyra av sina bilmodeller av årgång -83, bland an-



nat på två Dodge-varianter. Genom bilradion ska en liten talutmatningsapparat kunna lämna elva olika varningsmeddelanden när systemet känner av att något är fel på eller fattas bilen. Från konkurrenten Ford kommer ett par märken att kunna köpas med talutmatningsenheter som kan ge tre olika meddelanden för att varna föraren om han t ex lämnar bilen med startnyckeln i. På Ford funderar man också på talinmatning, dvs röststyrning. Å andra sidan har det ryktats länge om att talutmatning skulle komma på några bilmodeller, men inga har setts — eller hörts — ännu. I 1983 års Audi Quattro kommer en syntetisk röst att på tyska ge ett drygt dussin varningsmeddelanden.

● Den mest långtgående tillämpningen är förstås en blandning av talutmatning och talinmatning. Stephen Levinson och Mark Liberman vid Bell Labs i USA berättar i *Scientific American* om en prototyp som finns på Bell Labs. Systemet påstås kunna »förstå» 127 ord (till och med i sammanhängande meningar och sagda per telefon!) och svarar — korrekt men efter en viss paus — på frågor så länge de handlar om flygtider mellan städer i USA. Det talas ibland om femte generationens datorer som skulle kunna översätta direkt från ett talmeddelande på ett språk till dess talade motsvarighet på ett annat. Det pågår arbete på så-

dana system i Japan, men det är naturligtvis oerhört svårt att korrekt översätta mellan olika språk.

*När får vi höra mer?* *Speak & Spell* — en färgglad dosa stor som en bok, från Texas Instruments, som uttalar bokstäver och ord när man trycker in dem med tangenter på dosan — var den första talutmatningsenhet att få stor spridning. Några likartade produkter från andra tillverkare har emellertid inte sålts lika bra.

Talutmatning har nog en ljusare framtid än talinmatning, men i både fallen kommer utvecklingen att ta tid. De produkter som kan introduceras och användas under 1980-talet kommer sannolikt att finnas inom arbetslivet — industrier och i arkivarbeten, kan man tänka — och ha mycket begränsad vokabulär. Varningssystem eller mätinstrument som talar förefaller inte otroliga ens inom en handfull år. Men de flesta talutmatningskomponenterna tillverkas i USA och Japan, av de båda ländernas företag och alltså för engelska språket eller för japanska. Det tar tid innan vi får höra lagrad svenska i stor skala. Däremot finns den syntetiska text-till-talmaskinen Ove IV — som också kan taligenkänning — vid KTH, institutionen för talöverföring. Ganska snart kommer erfarenheterna av Ove IV att användas i snabbtelefoner: numren ska kunna sägas i stället för att knappas in.

## TELEFONVAKT, TELEFONSVARARE - OCH SEN?

*Telefonvakt* kallas en service som kan beställas från televerket (90 110). Det innebär att telefonsamtal när den uppringde är borta och inte svarar automatiskt kopplas till en speciell person (på ett annat ställe i stan) som vet vems nummer som är uppringt (och kan lämna vissa meddelanden). I Sverige har ganska få företag och privatpersoner abonnemang på Telefonvakt, medan sådan prenumeration är betydligt vanligare i USA.

*Telefonsvare* är en liten apparat som träder i funktion när den uppringde inte vill eller inte kan svara i telefon. Ett kort bandinspelat meddelande läses upp. Eventuellt kan den uppringande lämna ett eget meddelande på mellan 30 och 90 sekunder. I USA beräknas att bara var fjärde som ringer till en telefonsvare verkligen pratar in ett meddelande. När den uppringde ska konstatera vilka som har ringt blir han tvungen att lyssna ige-

nom alla konversationer på hela det sparade bandet. 1980 fanns det 100 000 telefonsvare i Sverige som hyrdes via televerket, plus ytterligare 15–20 000 privata.

Telefonvakt och telefonsvare används helt enkelt för att *inte missa viktiga telefonsamtal*. Amerikanska beräkningar tyder på att bara två av tre lokalsamtal kommer fram (inte får upptagetton eller spärrton), och att bara en dryg tredjedel av dessa når den som skulle ringas upp — alltså att man i bara 25 % av alla uppringningar når den person man vill. Motsvarande undersökningar i Storbritannien talar om att bara 40–50 % av alla samtal går fram till rätt person. Siffror från Sverige tyder på att man når ungefär tre av fyra personer som man söker (men då ringer man förmodligen många fler gånger).



Ett alternativ till telefonvakt och telefonsvarare är – en *kombination*. Den moderna telefonvakten kan kallas hänvisningsdator. En sorts hänvisningsdator som kan lagra tal tillverkas av flera företag i USA. Många abonnentnummer tillsluts till ett slags telefonistcentral där flera telefonister turas om att svara på de nummer som blir uppringda. Telefonisterna vet vem som blir uppringd genom att datorn ger ett meddelande på en textskärm där det t ex kan stå vad abonnenten heter, sysslar med, när han väntas in igen osv. Telefonisterna kan då ta emot beställningar eller skriva ner andra meddelanden. Det här är ett *manuellt* system. Den som ringer upp blir mottagen med mänsklig värme och tror rentav att han talar med någon på företaget som han söker.

Den moderna telefonsvararen är en sorts röstbrevlåda. Det här är ett *automatiskt* system. Användaren måste kunna eller begripa sig på det: man bör alltså vara van för att klara av det. En person som ringer upp någon som inte finns på plats hänvisas – automatiskt eller av telefonist eller genom telefonsvarare – till ett annat telefonnummer. Där kan man tala in ett kort meddelande som man kan ändra i genom att slå vissa koder på en knapptelefon (eller genom att använda en tillsats, en sk tonvalssändare, om man har telefon med nummerskiva). Detta meddelande förvaras för den uppringde tills han ringer upp systemet för att hämta sin röstpost eller skicka egna talbrev.

## RÖSTBREV LÅDOR – I SVERIGE OCKSÅ

Det finns en automatisk telefontjänst som kan kallas »Röstbrevlåda». En mer teknisk beskrivning skulle vara ett system för mellanlagring av tal, dvs att intalade meddelanden kan lagras ett tag, innan de skickas vidare när abonnenten är beredd att ta emot dem. Man kan säga att det är ett sätt att skicka just röstbrev, intalade meddelanden som lämnas utan att den som ska höra dem behöver vara på plats. Systemet kan också beskrivas som en telefonsvarare som inte finns hos den enskilde abonnenten utan i

- a) en abonnentväxel, typ företagsväxel, eller
- b) helt enkelt i det allmänna telefonnätet.

Under ett provår, med början i september 1982, kommer en sådan automatisk teletjänst att finnas i Göteborg – Röstbrevlådan. I brevlådan lagras intalade meddelanden tills adressaten kan lyssna på dem. Den som får meddelandet kan »bläddra» i det, spara det, skicka det vidare så andra får lyssna på det och enkelt sända tillbaka en intalad kommentar. Det enda som behövs, förutom att abonnenten talar in ett meddelande, är att systemets funktioner anropas med knapparna på en Diavox med 13 knappar eller att abonnenten har en särskild tillsats, en sk tonvalssändare, som passar alla telefonapparater. De som prenumererar på Röstbrevlådan betalar för det och tilldelas en hemlig ID-kod för att kunna lyssna på och tala in meddelanden. Sverige blir först i Europa med en Röstbrevlåda!

## SÅ HÄR FUNGERAR NÅGRA RÖSTBREV LÅDOR

En tillverkare av utrustning som kan mellanlagra tal – »Röstbrevlådor» – är ECS Telecommunications i Texas. ECS har flera typer av utrustningar som kallas VMX och finns installerade hos minst 10 stora kunder. VMX används av ca 26 000 personer, varav 2 600 använder ECS egen Voice Mailbox, alltså ett system där man inte köper en komplett dator utan prenumererar på ett antal (minst 25) röstbrevlådor. Det ska ECS sluta med för att koncentrera sig på att sälja den utrustning som behövs.

Den som lämnar ett röstbrev kanske slår telefonnumret till en person som begärt att alla hans telefonsamtal ska kopplas vidare till ett Voice Mailbox-telefonnummer. Eller så kopplar den uppringdes sekreterare eller kolleger vidare samtalet. Eller också vet den som ringer upp att han helst ska ringa en röstbrevlåda – direkt.



Om den uppringande har en knapptelefon så använder han den för att komponera och lagra meddelanden till Voice Mailbox (och den uppringde kan senare plocka fram och lyssna på sina »röstbrev» på samma sätt). Varje knapp är ett kommando, ibland flera kommandon. När den uppringande tryckt på knappen för att starta inspelning (knapp 1 på bild 1) hör han en svarssignal och kan tala in sitt brev (och trycka \* om han ångrar sig eller 1 en gång till för att stanna och tänka). Den som inte har knapptelefon får tala in ett brev utan att göra paus, som till en vanlig telefonsvarare.

Bild 1 visar några av funktionerna hos Voice Mailbox (om man har knapptelefon med 13 knappar). Störst möjlighet har den som lyssnar på »brev». Han kan »bläddra» (hoppa fram i brevet eller hoppa till nästa), arkivera, svara direkt osv. Avsändaren kan kontrollera att brevet kommit fram och plockats upp. Dessutom kan han skicka brev till flera olika personer på en gång.

Bild 2 visar funktionen hos Voicemail, som är likt Voice Mailbox. Voicemail lanseras av TeleVoice International i Santa Clara, Californien. Här finns inga uppgifter om prenumerationskostnader. TeleVoice säljer inga datorer som ECS utan bara »röstbrevlådor». Man använder DEC-datorer med skivminnen (där finns röstbrevlådorna) och dit kan 64 personer ringa samtidigt (dvs det finns 64 ingående linjer). Man kan lagra 8 000 meddelanden på en halv minut var och har plats för 32 000 »brevlådor».

Flera likartade system finns. I det svenska försöket med Röstbrevlåda används ett system från IBM. Det amerikanska telefonbolaget AT & T har utvecklat ett system kallat Advanced Calling. Det skiljer sig från alla de övriga genom att det vänder sig till allmänheten och inte till företagskunder. Meningen är att man inte behöver prenumerera på Advanced Calling. Än så länge är det bara fråga om planer. I mer avancerade system, som t ex IBM använder själva, kan man bli snabba upp talhastigheten — utan att det låter som Kalle Anka eller smurfarna! — när man lyssnar igenom brevsörden.

STARTA/ STANNA INSPELNING 1	STARTA/ STANNA UPPSPELNING 2	BACKA 10 SEK 33 = BACKA TILL BÖRJAN 3
HOPPA FRAM 10 SEK 44 = HOPPA TILL SLUTET 4	SÄND/ BEKRÄFTA MOTTAGANDE 4	6
SPARA/ HOPPA ÖVER 7	8	9
ÅNGERKNAPP/ ADJÖ *	SLUT PÅ ADRESSER FÖR RÖSTBREV 0	□

011 = Lyssna på posten

021 = Skicka vidare

022 = Besvara ett brev

Bild 1

SKICKA 1	KOLLA MOTTAGANDE 2	KOLLA AVSÄNDARE & TID 3
TALA IN BREV PÅ FÖRHAND 4	SKICKA VIDARE 5	6
HOPPA ÖVER/ SPARA 7	BACKA TILL BÖRJAN 8	TA BORT 9
STANNA/ SPARA BREV *	HJÄLP AV TELEFONIST 0	ÅNGERKNAPP □□ = ADJÖ □

Bild 2



## TILLVERKARE AV RÖSTBREVLÅDOR

Röstbrevlådor — »Voice Mail» — kan vara av flera slag. Det kan vara fråga om en servicebyrå där vem som vill får prenumerera på ett minimi-antal röstbrevlådor. Men det kan också vara fråga om särskild datorutrustning som köps för att installeras på ett större företag. Denna utrustning kan vara mer eller mindre integrerad med det företags egen telefonväxel — och dessutom med det egna telefonnätet som företaget kanske hyr.

Utrustningar som inte är integrerade kallas »stand-alone», dvs fristående. Flest sådana har sålts av ECS Telecommunications, näst flest av IBM. Därnäst kommer Wang.

I tabellen visas några egenskaper hos de fristående utrustningar för mellanlagring av tal som kan köpas från de tillverkare som förekommer i USA. Uppgifterna har hämtats bland annat från ett nyhetsblad — Voice Processing — som utges av Probe Research i Morristown, New Jersey.

Det finns ytterligare specialdetaljer till flera av systemen i tabellen. IBM-utrustningen uppges passa till många andra IBM-system och har dessutom alfabetisk adressering. Wangs DVX-system ska kunna kopplas ihop med en kombinerad ordbehandlare och diktafonutrustning, Wangs s k Audio Work Station. På Out Voice finns en lampa som talar om att röstbrev finns att hämta.

Den totala marknaden i USA för »Voice Mail» under 1981 uppskattas till totalt 5,5 miljarder dollar (nära 35 mkr) i försäljning och hyra av utrustning samt prenumeration på »röstbrevlådor». Det motsvarar 140 000 linjer, dvs 140 000 samtal med röstbrevlådor kan pågå samtidigt. Prognoser för 1985 talar om en marknad på minst 300 miljarder dollar bara i USA.

Tillverkare	System	Pris i 1000 \$	Antal användare	Antal linjer	Minne i megabyte
ECS	VMX	195	1000	16	800
		525	3000	64	1200
IBM	Audio	115	50	2	500
	Response	235	1000	10	?
Wang	DVX	117	250	4	275
		246	2000	16	1100
Voice & Data	Out	97	100	8	?
	Voice	235	500	32	?
		360	3100	64	?

Några kommentarer till begreppen i tabellen. Megabyte anger hur stort minnesutrymme som finns; som framgår på sidan 10 kan tal kräva 64 kilobitar per sekund, beroende på hur det kodas, och med skiftande talkvalitet som följd. Som tumregel gäller: ju fler bitar per sekund, desto bättre talkvalitet. Antalet linjer anger hur många som samtidigt kan tala in eller lyssna på röstbrev.



## RÖSTUPPLYSNING

Nästan alla länder har sin Fröken Ur, sin Talande Klocka. Det är ett exempel på röstupplysning. Och det är det vanligaste, mest efterfrågade exemplet. I alla länder som det finns statistik över är Fröken Ur (eller motsvarande) den mest populära röstupplysningstjänsten. I Västeuropa ringer varje hushåll – i genomsnitt – åtminstone 10 gånger per år till sitt lands Fröken Ur. Samtalen för att få reda på vad klockan är slagen utgör alltid minst 2/3 av alla samtal till röstupplysningstjänster.

Röstupplysning, det är inspelad information – *talutmatning* – som man oftast kommer in på någonstans i mitten och får vänta på tills bandet startar om från början. Det är alltså inte bara Fröken Ur utan också väder och temperatur, nyhets-sammmandrag, börsnoteringar, sagor, hitskivor, utflykts- och nöjestips (Frida), vinguides. De flesta sådana system är svåra att uppdatera (allt måste spelas in på nytt) och trädiga att lyssna på (man

måste vänta tills det börjar om igen). Alla system är lika dyra och utrymmeskrävande – mekanismen, plus att en särskild telefonlinje behövs – trots att Fröken Ur får överlägset flest samtal.

Med syntetiskt tal (se sid 9–11) går det att ganska automatiskt uppdatera innehållet i vissa telefonupplysningstjänster. Om man använder apparatur som kan omvandla skriven text till tal, så behöver ingen människa läsa in texten. Till ytterligare extra kostnad kan systemet konstrueras så att den som ringer upp får höra meddelandet garanterat från början. Om systemet dessutom förses med utrustning för talinmatning (se sid 11) så kan den som ringer upp begära fram inspelad information av olika slag, t ex genom att svara genom att säga siffror när en bandspelare läser upp en meny med flera (numrerade) innehållstyper. Olika innehållstyper kan också beställas som när man manövrerar en »röstbrevlåda»: genom knapptryckningar på en knapptelefon.

## SÅ HÄR FUNGERAR TALUTMATNING

Så kallat syntetiskt tal kan man åstadkomma på i huvudsak två olika sätt:

1) Antingen genom att en människa, vem som helst, talar in meddelanden som kodas och lagras enligt någon av flera speciella principer. Detta brukar kallas *talkodning*.

2) Eller genom vad som är *egentlig talsyntes*. Med hjälp av uttalsregler som lagras i ett elektroniskt minne samt en elektronisk apparat som kan frambringa vad som låter som mänskliga talljud. Man utgår t ex från en text som finns lagrad i en dator, beskriver den fonetiskt och går vidare till olika inställningar av en elektronisk röst. Se bild 3.

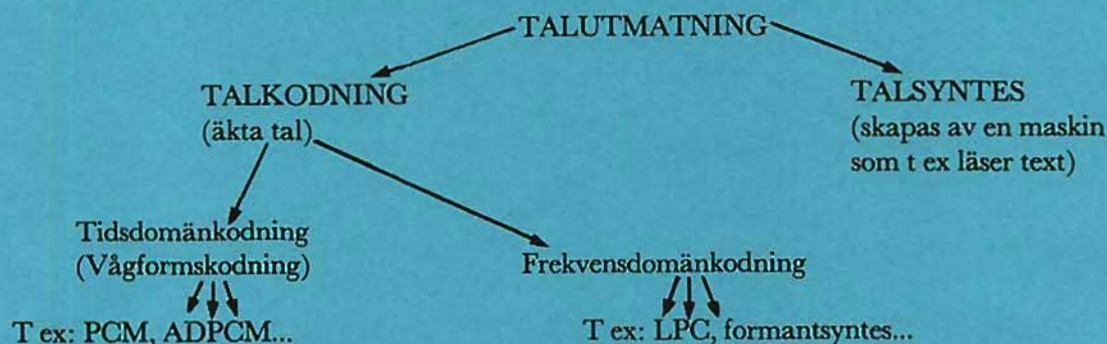


Bild 3



Talkodningen först. Som regel bygger den på att talvågen förändras ganska lite — och på ett sätt som går att förutsäga rätt så bra! — från en bråkdel sekund till nästa. Talvågen, det är de svängningar (kring ett jämviktsläge) som luftmolekylerna i kroppens talorgan gör när stämbandena vibrerar och »knuffar till» luft som stiger upp och ut ur munnen. Det är alltså fråga om röstenergi, eller enkelt uttryckt, med vilken styrka en person pratar.

Med tidsdomänkodning tar man »prover» av talvågen, när denna omvandlas med hjälp av en mikrofon — från mekaniska svängningar hos luftmolekyler, till elektrisk ström som varierar på samma sätt som luften. »Proverna» tas regelbundet, flera tusen gånger i sekunden. Sedan sparas de och kan användas för att (nästan) återskapa den ursprungliga talvågens form så som den förändras över tiden. Därav namnet vågformskodning. Det fungerar som i bild 4.

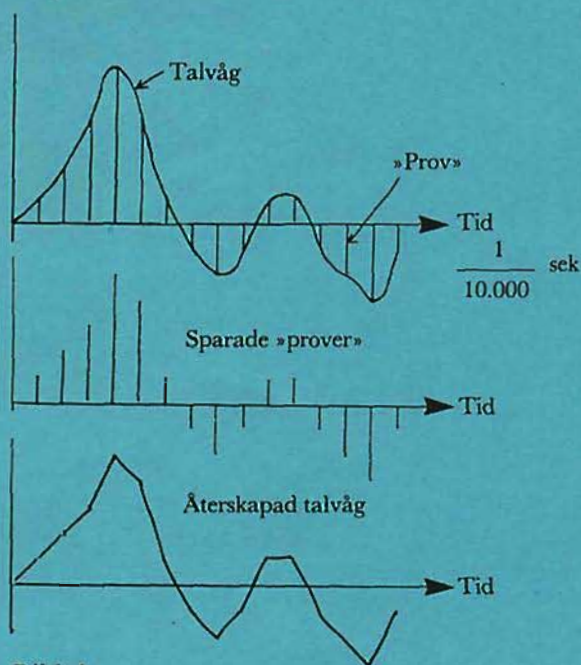


Bild 4

Proverna omvandlas — digitaliseras — för att kunna lagras i minnen. Olika stort minnesutrymme behövs, beroende på vilken speciell teknik som används. Med PCM — som är vanligast i digitala telefonnät — krävs upp till 64 kilobitar per sekund för att spara och återskapa tal. (Kilobitar är ett mått på hur mycket som behöver sparas.) Ljudet blir bra med PCM men talet kräver alltså ett stort minnesutrymme. Med en annan teknik (ADPCM) sparar man bara skillnaderna mellan två röstprover och kan på så sätt komma ner till kanske 16 kilobitar per sekund.

I fallet med frekvensdomänkodning bryr man sig inte om hur talvågen i sig själv ser ut i tiden. Istället omvandlar man den för att beskriva hur många och vilka delvibrationer av olika hastighet och styrka som den består av. Den övre vågen i bild 5 är sammansatt av de båda nedre delvågorna. En talvåg består naturligtvis av många fler delkomponenter, men i princip kan den beskrivas som i bild 5.

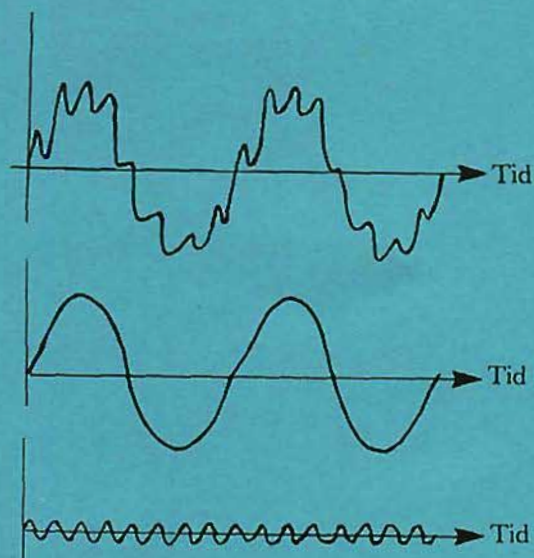


Bild 5

Med tekniker av det här slaget använder man så kallade filter för att dela upp talvågen i 10–20 (eller fler) frekvensområden. Frekvens är ett mått på vibrationshastighet, och mänskligt tal rör sig i frekvenser mellan 0 och ca 5 kHz. Man tar sedan »prover» (som när det gällde vågformskodning) inom varje frekvensområde för att se hur talvågen förändras. Med hjälp av dessa prover kan man återskapa något som låter ungefär som talvågen förändrar sig.

Proverna som man tar med någon frekvensdomänkodningsteknik är färre än med tidsdomänkodning. Med t ex Linear Predictive Coding (LPC) kan man spara mycket minnesutrymme; man kan komma ner till mellan bara 2 och 5 kilobitar som måste lagras per sekund. I gengäld blir den lagrade röstens ljudkvalitet dålig, klanglös och ibland förorenad av de mest häpnadsväckande ljud effekter. Riktigt vad människor tycker om att höra sådant tal vet man inte, men det kan antas att »maskinmässigt» tal duger bra i en hel del



sammanhang, åtminstone för tränade och vana lyssnare som väntar sig sådana ljud. Det finns inget tekniskt sätt att mäta ljudkvalitet: man måste helt enkelt lyssna .... Ännu används inte LPC i någon större utsträckning, men priset på LPC-kretsar kan komma att halveras under 80-talet och talkvaliteten förbättras.

Formantsyntes, som en annan metod kallas, är ännu mer minnesintensiv och kräver så få som 0.6 kilobitar per sekund. Med metoden lagras bara ett litet antal ljudkällor och frekvenser. Begreppet formant betyder svängningar eller resonanser som förekommer i vissa delar av talorganen — stämband, strupe, munhåla. I halsen finns öppningar och förträngningar som framkallar akustiska förändringar, och bakom formantsyntesen ligger antaganden om vilka av dessa skillnader i halsen som är viktigast, dvs vilka svängningar som »proverna» ska gälla.

Metoderna är inte alltid kompatibla, dvs de går inte alltid att förena med varandra. Inte ens två system som båda använder vågformsmodulering kan alltid användas med samma resultat.

Med så kallad *regelsyntes* — vad som kan kallas *egentlig talsyntes* — behöver man endast lagra 50–100 bitar per sekund (vilket är ungefär lika lite som behövs för att lagra text digitalt). Man utgår nämligen från en skriven text (eller, för vissa enklare maskiner, från en skriven fonetisk text). En apparat för *regelsyntes* — text-till-tal — innehåller

- 1) stavnings- och uttalsregler,
- 2) kanske ett »lexikon» med vissa vanliga ord som inte stavas eller uttalas helt enligt reglerna, samt
- 3) en elektrisk avbildning av det mänskiga talorganet.

Regelsyntesen påminner om vissa system för talinmatning (se sid 12) därigenom att den oftast utgår från s k fonem: talljud, talets minsta enheter. Fonem är alla de sätt som olika bokstäver eller delar av bokstäver kan uttalas på. Fonemen låter emellertid lite olika beroende på var i ordet eller meningen de uttalas; allofoner kallar man det ännu större antal småljud som kan beskrivas som fonem varierade med hänsyn till i vilka kombinationer med andra fonem som de uppträder i talat språk.

Ett system för *regelsyntes* avläser signaler: t ex text från en terminal eller en dator men lika gärna temperatur eller blodvärdet från särskilda instrument. Dessa signaler översätts sedan till tal. Systemet måste omprogrammeras med nya uttalsregler om det ska läsa text på något annat språk, som har annan intonation och annat uttal. Antalet fonem och allofoner, och hur de låter, är olika i olika språk. Talkvaliteten som uppstår med *regelsyntes* är fullt förståelig men långtifrån lysande.

## SÅ HÄR FUNGERAR TALINMATNING

Tal är tryckförändringar i luften. När stämbanden rör sig »knuffar» de till luftmolekyler som trängs upp genom de övriga talorganen och ut genom munnen. Talljudet är en mängd toner som fördelar sig i olika frekvensområden mellan 0 och ca 5 kHz (ett mått på frekvens). Frekvens — det är svängningar som ungefär bestämmer klangfärgen och tonen på det talade ljudet. Talljudet ändras dessutom hela tiden ifråga om hur mycket energi som används. Energi — det är hur kraftigt olika frekvensområden i det talade ljudet betonas.

Man brukar avbilda hur talet låter i något som kallas spektrogram. Bild 6 visar hur ett sådant kan se ut för ordet »teknik». Den vågräta axeln beskriver tiden: från vänster till höger ser man alltså hur ordet förändras när varje ny del av ordet uttalas. Den lodräta axeln anger frekvenser, svängningar, som alltså mäts i kHz. Skillnader i talenergi, betoning, visas som områden med olika svärta.



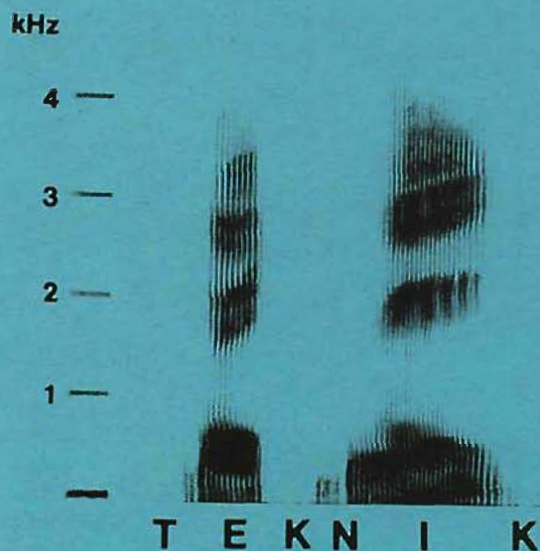


Bild 6

För apparater som ska reagera på talinmatning (taligenkännare) finns flera principer. Man talar ibland om a) **mönsteranalys**, som ska jämföras med så kallad b) **fonemanalys**. Gemensamt för både är att frekvensområdet där talet befinner sig delas upp i ett antal delområden (antalet varierar och påverkar metodernas precision). När någon talar till en taligenkännare tar denna sedan »prover» av talet inom varje frekvensdelområde. »Proverna» är så korta som mellan 10 och 50 millisekunder vardera.

den i sitt minne efter det fullständiga ord som mest påminner om *hela* ordet »noll» som det har uttalats den här gången.

Mönsteranalysen är ganska språkoberoende: det ska gå att tala in ett urval ord på vilket språk som helst. Däremot är den något mindre effektiv än fonemanalys, och oftare beroende av att det är samma talare som använder apparaturen hela tiden. Med fonemanalys lagras inte hela ord utan delar av ord: fonem, de kortare talljud som är beståndsdelar i de talade orden, t ex alla sätt som en vokal kan uttalas på och de olika deljud som ingår i språkets konsonanter. Vid igenkänningen jämför därför apparaten *delar* av varje uttalat ord med sitt lexikon av lagrade fonem.

Talinmatning kan ha olika syften. De viktigaste framgår av bild 7 intill. Det kan vara fråga om att konstatera VEM som talar till apparaten. Vid *identifiering* gäller det att jämföra rösten hos en person med röstprover som har tagits från många olika personer, för att veta precis vem som t ex senast hade tillträde till ett laboratorium. Vid *verifiering* jämförs personens röst med prover på rösten hos den eller de personer som t ex är de enda som är berättigade att ta ut pengar på ett visst konto.

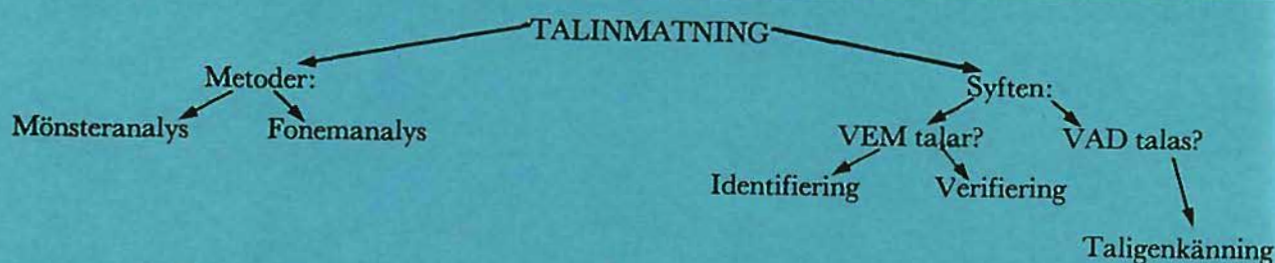


Bild 7

Mönsteranalysen utgår från att man lagrat hela ord i ett minne. Det kan ha gått till så att en enda talare läst in ett litet urval ord — minnesstorleken är begränsad — som upprepas några gånger så att taligenkännaren kan lägga märke till ungefär hur uttalet av samma ord kan variera. Samtidigt lagrar talaren en beskrivning av vad varje ord betyder. Säger han »noll» så skriver han samtidigt in »0». Varje ord kan beskrivas som ett ordmönster, på engelska »template». När samma person — ibland kanske till och med andra personer — sedan ger t ex kommandot »0!» till apparaten letar

När det gäller VAD som sägs — taligenkänning — brukar man tala om röst- eller talstyrning. *Styrning* innebär att vissa instruktioner har lagrats, och att maskinen programmeras att utföra vissa uppgifter enligt dessa instruktioner så fort de talas in på ett sätt så att apparaturen känner igen dem. Ännu så länge är repertoaren av kommandon ytterst begränsad. Det går inte heller att använda maskiner programmerade för t ex identifiering också till röststyrning, eller tvärtom.



## NÅGRA OLIKA TALIGENKÄNNARE

Bara under våren 1982 har nyhetsbladet Voice Processing rapporterat om att ett par företag i USA som tillverkat taligenkännare nu upphört med detta. Även om en hel del företag tillverkar utrustning för taligenkänning så är marknaden ännu ganska ung och försäljningen osäker.

Centigram Corporation startade t ex redan 1977 med att sälja sitt system, som kallades Mike. I januari 1982, när tillverkningen lades ner, rapporterades att totalt 300 Mikeapparater hade sålt. Heuristics startade samtidigt som Centigram och slutade tillverka sina 7000-enheter i februari 1982,

Man ser av tabellen att det är stora skillnader både vad gäller priser och prestanda. Kvaliteten försöker man bl a mäta genom att beräkna den så kallade *rättprocenten*: hur stor andel av alla gånger som en apparat korrekt kopplar samman («känner igen») ett intalat ord med en tidigare lagring av samma ord. Man ska observera att rättprocenten bara beräknas med utgångspunkt från de ord som maskinen verkligen uppfattar: vanligen måste ett kommando eller en instruktion upprepas många, många gånger innan maskinen ens börjar jämföra ordet med lagrade representationer av olika ord.

Tillverkare	System	Pris i \$	Antal ord	Rättprocent
Auricle	AUR 1	2 480	32-80	97 %
Interstate Electronics	VRM 102	1 982	100	99 % (96 %)
	VRT 103	6 595	100	99 %
NEC	DP 200	30 000	150	99 %
	SR 1301	35 000	128	?
Scott Instruments	VET 2	795	40	98 %
	Shadow	995	40	?
Threshold Technology	1500	8 700	350	99 % (98 %)
	1600	10 500	250	?
Verbex	1800	62 500	?	98 % (98 %)
Votan	V5000	5 000	256	?

då mindre än 1000 sådana hade sålts sedan starten. Som jämförelse kan nämnas att Scott Instruments — som fortfarande existerar och som tillverkar utrustning för Apple II-terminaler — hittills har sålt 400 taligenkännare.

Tabellen visar vilka företag i USA som för närvarande (våren 1982) tillverkar apparatur för talinmatning. Bara ett urval modeller (som fortfarande tillverkas och säljs) redovisas; Interstate tillverkar minst 5 olika modeller, övriga tillverkare en eller ett par vardera.

De procentsiffror som anges i tabellen är framför allt hämtade från tillverkarna själva och mycket imponerande (men kom alltså ihåg att de bara avser ord som varje system verkligen har hört . . .). Inom parentes anges för vissa system vilken rättprocent som konstaterats i testningar som Texas Instruments har genomfört.



Av tabellen framgår inte några ytterligare egenskaper som är viktiga att känna till för att bedöma systemens prestanda. Man brukar skilja mellan bl a *talberoende* och *taloberoende* system, där talberoende (som är vanligast) är sådana som bara kan användas av den person som talar in de ord som maskinen sedan ska känna igen. Alla system i tabellen är talberoende – dvs, de kan bara manövreras av en enda person – utom Verbex 1800 och NEC SR 1301, de två dyraste utrustningarna. Båda dessa har dessutom den fördelen att de reagerar på och känner igen ord som sägs på telefon, dvs där talkvaliteten är något lägre än i ett ostört rum där en ensam person lugnt pratar i en headset-mikrofon till en maskin.

Dessutom skiljer man mellan utrustning som bara kan känna igen enstaka ord – *diskret tal*, som det kallas – och sådan som känner igen *sammanhängande tal*. Det är naturligtvis mycket mer komplicerat att känna igen sammanhängande tal, där ordens uttal kan förändras beroende på när i meningen de uttalas osv. Mellan diskret tal och sammanhängande tal finns »sammanhängande ord», alltså grupper av ord eller ljud, som t ex postnummer. De flesta apparater, även de allra flesta i tabellen, kan bara känna igen enstaka ord, ett åt gången. Av dem i tabellen är det Verbex 1800 och en annan NEC-utrustning, den också ganska dyra DP 200, som sägs känna igen sammanhängande ord.

## RÖSTKONTORET

Det är en vårsval onsdag 2011, den dag i veckan då Rösterman jobbar inne i stan. De andra av veckans tre arbetsdagar sitter han ju på grannskapscentralen i byn där han bor, och fö disponerar hans familj bara ett par dagar i veckan den bil som bysamfälligheten äger gemensamt.

På infartsvägen till kontoret tänker Rösterman på andra saker så att bilens varningssystem upprepade gånger måste säga (med en räddhågad röst som Rösterman hatar): Sätt på bilbältet! Bensintanken är nästan tom – det finns bara en liter kvar! och Det finns ingen spolarvätska kvar! Väl framme vid kontorets parkeringshus legitimerar sig Rösterman genom att säga sitt namn i en anordning vid spärrbommen – där sitter det en enhet som lyssnar av att det verkligen är Rösterman som kommer – och när bommen öppnas för hans ekipage berättar en talutmatare om för honom var närmaste lediga parkeringsficka finns.

Givetvis lämnar Rösterman bilen olåst: för att starta den behövs ju att man säger en kod som bilen känner av och med en röst som bilen kan identifiera. Det är svårt de dagar Rösterman är snuvig och en vinternatt fick han lifta hem när bilen vägrade lyda honom. När han går från bilen ropar den räddhågade rösten efter honom: Vänster framdörr är öppen!

Hemma har Rösterman dikterat in några olika arbetspapper på sin freestyle-bandspelare och nu lämnar han bandet för automatisk utskrift, från tal till text. Utskriften blir naturligtvis inte perfekt – datorn kan inte stava så bra som jag, brukar han säga; men hans omgivning tycker inte att någon kan stava sämre än Rösterman – men Rösterman kan efter utskrift stava och stuva om orden för hand eller med röstinmatning, vilket han vill.

Sedan ringer Rösterman sin röstbrevlåda, ett särskilt interntelefonnummer som han når genom att säga Post! till sin telefon, och får ett tiotal inspelade meddelanden utpratade. Eftersom samma person har ringt tre gånger får Rösterman göra bruk av sina överhoppningsmöjligheter, han säger Skippa! till telefonen och vips hör han nästa telefonbesked. Ett röstbrev är så intressant att han skickar det vidare till sin kompis som jobbar i staten: Skicka till Svante! säger Rösterman, eftersom Skicka till! är ett av de 25 kommandon som systemet klarar av, och då letar telefonen reda på numret till Svantes röstbrevlåda och skickar sedan ljudet dit.



Nu ljuder Röstermans almanacka som finns lagrad i den totalterminal som han har stående på sitt sammanträdesbord. (Ingen människa säger skrivbord nu för tiden; och sitt, förresten — Rösterman delar ju rum med tre andra handläggare och med en söndagsskolegrupp.) Det är tio minuter kvar tills dagens första besökare ska komma, så namnet på besökaren både pratas ut och blinkar fram på totalterminalens bildskärm.

Besökaren kommer, man pratar en stund och Rösterman gör sin vanliga dragning om företagets produktfilosofi. På sin bildskärm beställer Rösterman fram bilder som han med rösten kallar från företagets teledata- och mikrobildsarkiv, och bilderna därifrån visas samtidigt som talutmatnings-systemet i terminalen läser upp texten som finns på varje bild. Med enkla tillsägelser eller knapptryckningar kan Rösterman »bläddra» och »hoppa» även i den talande texten.

När Rösterman hämtat doftande kaffe och wienerbröd med gult på kommer man fram till att man måste kolla saken med Ville. Sagt och gjort, en påringning till Ville: Rösterman säger Prata med Ville! men en krets i telefonnätet svarar Upptaget. Rösterman säger Upprepa! och efter ett par minuter svarar Ville; han har förstås massor att stå i men tar samtalet eftersom han på telefonens lilla display ser att ett samtal kommer från Röstermans telefon. Påringningen flyttas automatiskt till Byske för dit har Ville åkt just idag, och Rösterman slår på högtalartelefonen för att båda ska kunna prata med Ville.

Det tar en stund men när Röstermans besökare har gått hinner Rösterman ändå skriva en offert — dvs han skriver inte direkt men en offert blir det. På totalterminalen begär han fram en standardversion av ett offertformulär, säger det tilltänkta kundföretagets namn (och begär fram adress som hämtas fram ur en särskild telefonnummerdatabas), gör några röststyrda beräkningar och talar in de värden som offerten gäller. Därefter säger han till om att texten ska telefaxas i flera exemplar, dels förstås till det möjliga kundföretaget men också till andra enheter och personer inom företaget där Rösterman jobbar.

Rösterman frågar sin almanacka Besök? och får glädjande nog höra att nästa besökare kommer först veckan därpå. Han lutar sig tillbaka i rummets slitna men eleganta och helt sittriktiga arbetsstol, säger Skval! och radion väljer ett program där det finns lättsam musik som sedan flödar ut i rummet. Displayen på radion talar om att musiken kommer med satellit från Atlanta, Georgia. Medan rösterna från barnen som strömmar ut ur skolan på andra sidan gatan hörs som ett muntert sorl när skvalmusiken går i moll börjar Rösterman leta efter sin gamla tuschpenna och en bit papper.



**TELDOK**

**Televerkets huvudkontor, H32 :20**

**123 86 Farsta**



## VILL DU FÅ FLER SKRIFTER FRÅN TELDOK?

TELDOK planerar att ge ut fler nummer av det här nyhetsbladet — alla temanummer. TELDOK-INFO beräknas så i fortsättningen komma att innehålla temanummer om

- bredbandskommunikation
- presentationsteknik och bildhantering
- textkommunikation.

Tala om för oss ifall — och varför! — du ogillar det här numret, så kanske nästa utgåva av TELDOK-INFO blir bättre. Och om du hittar sakfel, oklarheter, påståenden som behöver kommenteras — berätta också det för oss, så att sakerna kan påtalas i nästa nummer.

Inget är bestämt om när nästa TELDOK-INFO kommer — men du står på utsändningslistan! Meddela oss om du vill bli struken, liksom om det finns andra som bör få TELDOK-INFO också eller istället. Kupongen här nedan kan användas för detta, och dessutom för att beställa exemplar av TELDOK Information nr 1.

Hittills har bara utkommit TELDOK Information nr 1: Om kontorsautomation i USA, men fler rapporter planeras. TELDOK Information nr 1 förtecknar 27 särskilt intressanta försök med kontorsautomation i USA: många exempel på textkommunikation men också bildkonferenser och några ljudsystem. Rapporten är utarbetad av Richard Dalton, då vid konsultföretaget Open Systems.

Dessutom har manus inkommit till vad som ska bli TELDOK Information nr 2 och 3. Dels en rapport från det pågående försöket TeleBild, ett svenskt försök med teledata för näringslivsinformation. Dels en inventering, liknande den i nr 1, av olika försök med kontorsautomation i England, Tyskland, Frankrike och Luxemburg. Projekt *pågår* som gäller att dokumentera: meddelandesystem i några svenska företag; kontorsautomation i Japan; bredbandskommunikation i USA; kommunikationssystem på advokatkontoret.



### KUPONG FÖR MEDDELANDE TILL TELDOK

Adress: Televerkets hk, H32:20, 123 86 Farsta

Kryssa för det som gäller!

- Skicka mig kommande TELDOK Information (nr 2, 3 osv).
- Skicka gärna TELDOK-INFO till följande personer:

---

---

- Jag vill helst inte få fler TELDOK-INFO.

Namn och adress:

---

---