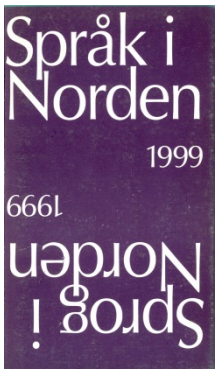
Sprog i Norden

Titel: Taleteknologi

Forfatter: Torbjørn Svendsen

Kilde: Sprog i Norden, 1999, s. 9-19

URL: http://ojs.statsbiblioteket.dk/index.php/sin/issue/archive



Søgbarhed

Artiklerne i de ældre numre af Sprog i Norden (1970-2004) er skannet og OCR-behandlet. OCR står for ’optical character recognition’ og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Taleteknologi

Torbjørn Svendsen

Innledning

Mennesker kommuniserer med hverandre på mange ulike sett, for eksempel med kroppsbevegelser, skrevet tekst, bilder og tegninger. Likevel er nok stemmen det viktigste kommunikasjonsmediet for de fleste. Sir Richard Paget har en interessant teori på opprinnelsen til talemålet: *What drove man to the invention of speech was, as I imagine, not so much the need of expressing his thoughts (for that might have been done quite satisfactory by bodily gesture) as the difficulty of "talking with his hands full"*. Tale er et enestående medium for å formidle tanker og idéer, og behovet for å kunne kommunisere med tale over avstander i rom og tid ga oss i sin tid de første taleteknologiske oppfinnelsene med stor utbredelse: Telefon og lydopptakere.

Bruken av tale var ikke begrenset til formidling av informasjon mennesker imellom. Tale ble (og blir) også benyttet til å kontrollere datidens "maskiner". En hest kunne for eksempel styres med et lite sett av talte kommandoer. Ikke fordi man ikke hadde noen annen metode, men fordi det var mer hensiktsmessig å kunne ha hender og føtter fri til å utføre andre oppgaver, og at man ikke var bundet til å være i fysisk nærhet til hesten. Naturligvis hadde man reserveløsninger som f.eks. tømmer dersom hestens talegjenkjenner skulle svikte!

Etter som menneskene har blitt mer og mer omgitt av maskiner, har også behovet for å kunne kommunisere med maskinene ved bruk av tale blitt større. Dette har bakgrunn i at tale har en rekke fordeler framfor andre kommunikasjonsformer. For eksempel:

- Tale er for de fleste formål den enkleste og mest naturlige måten å kommunisere på.

- Tale krever ikke at hender og/eller øyne brukes til kommunikasjonen.
- Talekommunikasjon med maskiner kan være svært nyttig for funksjonshemmede.
- Telefonen kan benyttes for å kommunisere med maskiner som er plassert på et annet sted.

Taleteknologien har fascinert oppfinnere og forskere gjennom lang tid. Telefonen ble oppfunnet for vel 100 år siden, og den første enkle talegjenkjenneren ble demonstrert på Verdensutstillingen i 1929 – en leketøyshund som hoppet ut av hundehuset sitt når navnet dens ble ropt. Talesyntese har enda lenger historie – den første mekaniske talemaskinen ble laget av von Kempelen i 1791, en sinnrik innretning der luftstrømmen fra lungene ble etterlignet av en belg, stemmebåndenes operasjon ble utført av ei flis som på en treblåser, og bevegelsene i munnhule og av leppene ble kontrollert ved å forme et resonansrør av lær med hånden. Talemaskinen krevde en dyktig operatør i tillegg til en hjelper som opererte belgen som drev lydproduksjonen.

Likevel er det først i løpet av de siste 30 år at utviklingen av taleteknologien har skutt fart. Mye av årsaken til dette er de framskritt som er gjort innenfor mikroelektronikken. Digitale datamaskiner og stadig kraftigere mikroprosessorer har gjort det mulig å eksperimentere med nye teknikker, og å utvikle produkter som tidligere var utopiske. Taleteknologien er basert på matematisk beskrivelse og behandling av tale, og tilgangen på regnekraft har gitt ny innsikt i hvordan man kan lage systemer som setter oss i stand til å ha verbal kommunikasjon med maskiner.

I løpet av de siste par årene er det kommet en rekke taleteknologiske produkter beregnet på konsumentmarkedet, for eksempel talegjenkjenningssystemer for diktering. Disse produktene, understøttet av uttalelser som denne fra Microsoft-sjefen Bill Gates: *Speech is not just the future of Windows, it is the future of computing itself*, har gitt stor mediaoppmerksomhet. Business Week slo for eksempel bombastisk fast at *Speech is the next big thing in computing* i en større artikkel i fjor vinter. Ikke

desto mindre har taleteknologien knapt passert krabbestadiet, og det er fortsatt langt fram før menneske/maskin-kommunikasjonen kan foregå slik som det er framstilt i enkelte science fiction filmer (f.eks. HAL i "2001 – en romodyssé" eller R2D2 i "Star Wars").

Hva er taleteknologi?

Etter en liten tur på det lokale kjøpesenteret, finner du ut at du skal stikke innom tante Olga. Du tar opp mobiltelefonen og ringer til den automatiske informasjonssystemet for busstrafikk. Her møtes du av en vennlig stemme som ønsker velkommen og ber om å få vite hva slags informasjon du er ute etter. Du forteller at du skal fra Risvollan til Byåsen, og ber om å få vite hvilke busser du skal ta, når neste buss går og når du er framme. Stemmen gir deg informasjonen, og du setter kursen mot nærmeste bussholdeplass for å rekke bussen, som går om fire minutter.

Dette scenariet viser et eksempel hvordan et system som utvikles i et forskningsprosjekt ved NTNU og SINTEF i Trondheim skal kunne brukes. Dette er et automatisk informasjonssystem der all kommunikasjon mellom bruker og maskin foregår med tale. Et slikt dialogsystem inneholder flere taleteknologiske komponenter, som automatisk talegjenkjenning og talesyntese. I tillegg til de taleteknologiske komponentene i dialogsystemet, benytter også mobiltelefonen seg av taleteknologi for å overføre talen på en effektiv måte.

Taleteknologi er i prinsippet all teknisk og automatisk behandling av tale. Dette omfatter en lang rekke anvendelser og teknologier.

Talekoding

Talekoding går ut på å komprimere informasjonsinnholdet i tale, slik at talesignalet kan overføres eller lagres på billigst mulig måte, uten at kvaliteten blir særlig skadelidende. Talekoding

benyttes spesielt innen telefoni. En GSM-telefon benytter f.eks. vel 13 000 bit for å overføre ett sekund tale, mens en vanlig telefonforbindelse krever 64 000 bit. Talekoding er også viktig for internett-telefoni, som griper mer og mer om seg. Prinsipper fra talekoding anvendes også i mer generell lydkomprimering, som f.eks. anvendes i digital kringkasting, MiniDisc og lydkompresjon for multimedia.

Talesyntese

Talesyntese er tale generert av maskiner. Syntetisk tale kan være laget ved sammenskjøting av ord og/eller delsetninger, og være begrenset til et relativt lite antall meldinger, noe vi kjenner fra f.eks. bankenes kontofoner. Sammenskjøtet tale kan være av relativt god kvalitet, men har begrenset anvendelse. Det som vanligvis menes med betegnelsen syntetisk tale er tale som er generert fra tekst, såkalt tekst-til-tale syntese (TTS). TTS innebærer at datamaskinen må foreta en rekke operasjoner, omforming fra skrevet form til uttale, setningsanalyse for å definere prosodi, dvs. trykklegging, timing og toneleie og selve lydgenereringen. I prinsippet kan TTS omforme enhver tekst til tale, og har derfor stor fleksibilitet, og mange anvendelsesområder. Imidlertid er kan kvaliteten fortsatt forbedres en god del. Det er spesielt prosodi og lydgenerering som er de svakeste leddene i dagens talesyntese, og disse komponentene har vesentlig betydning for hvor naturlig den syntetiserte talen blir. Det er fortsatt behov for en stor forskningsinnsats før syntetisk tale vil ha en naturlighet (og forståelighet) som er sammenlignbar med menneskelig tale.

Dagens talesyntetisatorer har typisk én eller et lite utvalg av stemmer, og genererer tale med standardisert uttale. Hvis Telenor og Netcom, som er konkurrenter på mobiltelefonimarkedet i Norge, skulle lage informasjonstjenester basert på syntetisk tale, ville de trolig ønske å ha stemmer som markerer at det er forskjellige bedrifter. Effektive metoder for å lage nye stemmer er derfor viktig for mange anvendelser. For funksjonshemmede med

talevansker, er bruk av talesyntese som en taleprotese interessant. For slike anvendelser er det også viktig at den syntetiserte stemmen kan bli tilpasset brukerens kjønn og dialektområde.

Talegjenkjenning

Talegjenkjenning innebærer at datamaskinen skal identifisere innholdet i det som blir sagt, og kople resultatet av identifikasjonen til en aksjon. Formen på aksjonen vil variere avhengig av formålet med talegjenkjenningen. I et dikteringssystem skal talen omformes til tekst, dvs. at det identifiserte ordet blir skrevet på dataskjermen. I et system for talestyring av et Windows-grensesnitt skal den gjenkjente talen knyttes til en systemkommando.

Måten de ulike talelydene blir uttalt på er blant annet avhengig av den enkelte taleren, av den fonetiske og semantiske kontekst lyden opptrer i, og av omgivelser og sinnsstemning. Selv om den samme setningen skulle uttales av samme taler og under identiske forhold, vil den akustiske realiseringen av språklydene variere. For å kunne håndtere variasjonene er den underliggende motoren i dagens talegjenkjennere basert på en statistisk modell av den akustiske realiseringen av de ulike talelydene.

Yteevne og kompleksitet for en talegjenkjenner er avhengig av bruksområde og de krav som kan settes til brukeren. I mange anvendelser vil talegjenkjenneren brukes av én enkelt person, f.eks. i mange PC-baserte anvendelser. Gjenkjenneren kan da tilpasses brukerens stemme og uttale gjennom at brukeren gir eksempler på sin talemåte i en opptreningssesjon. Siden variabiliteten da blir mindre, vil gjenkjenneren ha færre feil. I andre anvendelser, f.eks. i telenettet, må systemene kunne gjenkjenne alle potensielle brukere. Det har også mye å si for kompleksitet og gjenkjenningsrate om en setning kan uttales naturlig, uten pause mellom ordene, eller om ordene må sies isolert. La oss se på en gjenkjenner som skal kunne gjenkjenne åttesifrede telefonnummer. Utgangspunktet er at sifrene uttales enkeltvis (dvs. at et telefonnummer som 82 11 22 32 uttales som *åtte to en en to*

to tre to). Dersom vi krever at brukeren tar en pause mellom hver ord, vil talegjenkjenneren mellom hver pause ha ti mulige valg, siden pauser er relativt lette å detektere. Hvis vi derimot lar brukeren uttale tallstrengen uten pause mellom hvert siffer, vil talegjenkjenneren i prinsippet bli nødt til å velge mellom 100 millioner forskjellige setningsalternativer. Heldigvis finnes det metoder som håndterer kompleksitetsproblemet på en effektiv måte, ellers hadde det vært tilnærmet umulig å lage praktiske talegjenkjennere for f.eks. diktering.

I løpet av de siste par årene har det kommet flere dikteringssystemer på markedet. Dikteringssystemene gjør det mulig å legge vekk tastatur og datamus, og i stedet bruke mikrofon når en skal forfatte brev, artikler og andre dokumenter. Systemene er alle i stand til å gjenkjenne kontinuerlig tale, og har vanligvis ordforråd på 20 000–50 000 ord. Også andre anvendelser av talegjenkjenning er i praktisk bruk. I USA har AT&T erstattet noen av sine operatørtjenester i telenettet med en enkel talegjenkjenner og sparer årlig millioner av dollar på denne teknologien.

En oversikt over ytelsen til dagens talegjenkjennere for noen anvendelser er vist i tabellen nedenfor. Tabellen illustrerer tydelig at feilraten er avhengig av størrelsen på ordforrådet, men at en enda viktigere faktor er hvorvidt talen som skal gjenkjennes er spontan eller lest. Den nederste raden i tabellen viser hvor vanskelig det er å få til automatisk gjenkjenning av en vanlig telefonsamtale mellom mennesker. Siden det meste av menneskelig tale er spontant generert, er det tydelig at det fortsatt er vesentlige oppgaver som må løses før tale er et fullgodt alternativ til tradisjonelle brukergrensesnitt som tastatur.

Oppgave	Type tale	Ordforråd	Ordfeil-rate %
Tallstrenger	Lest	10	< 0,3
Flyreiseinformasjon	Spontan	2 500	2
Wall Street Journal	Lest	64 000	7
Radionyheter	Blandet	64 000	30
"Ring hjem"	Konversasjon	10 000	50

Tabell 1. Ordfeilrate for "state-of-the-art" talegjenkjenning for ulike oppgaver. (Kilde: IEEE Spectrum, desember 1997.)

Talergjenkjenning

Talergjenkjenning har som formål å bestemme identiteten til taleren. En skiller gjerne mellom to former for talergjenkjenning: Taleridentifikasjon og talerverifikasjon. I talerverifikasjon er systemets oppgave er å avgjøre om en taler er den personen han eller hun gir seg ut for eller ikke. Talerverifikasjon benyttes typisk for adgangskontroll, enten fysisk (få adgang til et rom e.l) eller logisk (gi adgang til et datasystem, f.eks. en bankkonto), gjerne i tillegg til en nøkkel eller et passord. I taleridentifikasjon har ikke taleren noen påstått identitet, og systemets oppgave er å identifisere en ukjent taler, f.eks. for overvåkning eller i kriminalsaker.

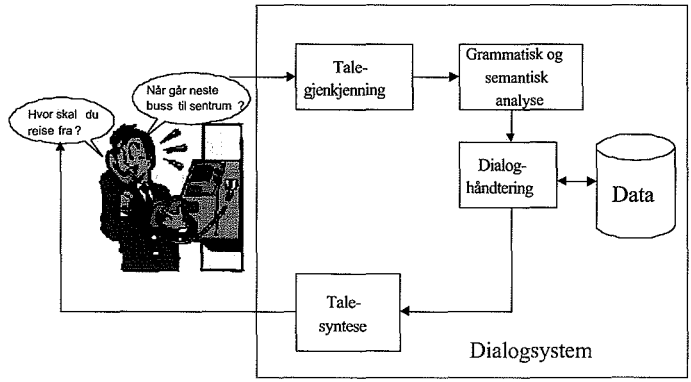
Språkidentifikasjon

Språkidentifikasjon går ut på å bestemme hvilket språk som tales. Anvendelsen er f.eks. som en delkomponent i et talegjenkjenningssystem som er designet for flerspråklig bruk, noe som aktuelt i mange land. Dialektidentifikasjon er et spesialtilfelle, som er vanskeligere, men som kan være viktig i land som Norge der dialektvariasjonen er stor.

Dialogsystemer

Dialogsystemer er systemer der all kommunikasjon mellom menneske og maskin er talebasert. Dette betyr naturligvis at ta-

lesyntese og talegjenkjenning er komponenter i et dialogsystem. I tillegg kan også talerverifikasjon og språkidentifikasjon benyttes. Taleteknologi er imidlertid ikke nok for å lage et godt dialogsystem. Dersom dialogen mellom menneske og maskin skal være naturlig, må brukeren ha anledning til å formulere seg naturlig. Dette innebærer at innholdet i den gjenkjente talen analyseres for å trekke ut den informasjonen som er relevant for å løse den oppgaven som dialogsystemet er laget for. I et system som er laget for å gi informasjon om busstrafikk er hensikten med dialogen at systemet skal kunne finne ut hva brukeren ønsker, og få den nødvendige informasjon fra brukeren for å kunne besvare spørsmålet.



Hvor står Norden i dag?

Forskningsmiljøene innen taleteknologi i Norden holder en internasjonalt høy standard. KTH i Sverige har i mange år vært betraktet som et av de ledende forskningsmiljøer innen talesyntese, og også universitets- og instituttforskningen i Danmark, Finland og Norge har levert viktige bidrag til forskningen innen taleteknologi. Taleteknologien (med unntak av

talekoding) krever språkspesifikke løsninger; teknologien må være språkkompetent. De nordiske forskningsmiljøene besitter den nødvendige fagkompetanse for å utvikle taleteknologiske produkter som "kan" våre nasjonalspråk.

En essensiell faktor for utvikling av taleteknologiske løsninger er tilgang til store mengder av språkdata. Spesielt gjelder dette talegjenkjenning, som baserer seg på statistisk modellering av talesignalet, og som i tillegg benytter statistisk basert språkmodellering for å kunne gi god ytelse for komplekse oppgaver. Uten en infrastruktur som inkluderer store tale- og tekstdatabaser i tillegg til leksikalsk informasjon i form av blant annet uttaleleksika, er det umulig å utvikle avansert taleteknologi.

Dessverre ser vi i dag at tilgangen på taleteknologiske produkter som er tilpasset de nordiske språkene er liten. Kostnadene ved å utvikle taleteknologi for ulike språk er noenlunde den samme for de fleste språk. Markedsøkonomiske mekanismer fører derfor til at nye og avanserte hjelpemidler først utvikles for de språkene som har det største markedet. Dette betyr i praksis at nye produkter utvikles først for engelsk. Deretter følger tyske, spanske og franske versjoner. De små markedene kommer sist, om de overhodet er interessante.

De nordiske land er små markeder. Selv om utbredelsen og bruken av informasjons- og kommunikasjonsteknologi relativt sett er svært høy, er likevel markedene for små til at det er lønnsomt å tilpasse taleteknologiske hjelpemidler til våre språk, og enda mindre lønnsomt å utvikle egne løsninger. Dette er bekymringsfullt, hovedsakelig av to årsaker:

- Språket er en viktig del av den nasjonale identitet. Dersom verktøy til støtte for tekstproduksjon og bruk av datamaskiner ikke er tilgjengelig på nasjonalspråket vil det være lettere å velge "internasjonale" løsninger, f.eks. bruk av engelsk i stedet for norsk, når hjelpemidlene er tilgjengelig for dette språket. Hvis framtidige versjoner av Windows har et velfungerende talebasert grensesnitt i den engelske versjonen, og den norske versjonen ikke har slike finesser, er det ikke urimelig å tro at

mange vil velge en engelsk Windows-versjon framfor den norske.

- Det eksisterer etterhvert en del hjelpemidler for funksjonshemmede som er basert på taleteknologi. Bortsett fra de enkleste systemene, som talertrente kommandosystemer for omgivelseskontroll, eksisterer disse hjelpemidlene bare for de store språkene. Hvis det ikke gjøres en nasjonal innsats, vil funksjonshemmede som ikke behersker fremmedspråk være avskåret fra å benytte av eksisterende teknologi som vil kunne forbedre livskvaliteten, og sette mange i stand til å føre et yrkesaktivt liv.

Veien framover

Det er derfor nødvendig med en nasjonal innsats for å gjøre utvikling av taleteknologi på våre nordiske nasjonalspråk økonomisk interessant. Dette kan gjennomføres ved at det på nasjonalt plan gjennomføres en planmessig innsamling av språkdata, med en stor grad av offentlig finansiering. Denne språkdatabasen, et nasjonalt språkteknologisk korpus, må inneholde tale, tekst og leksikalske data, og være konstruert slik at den danner en basis for såvel forskning som produktutvikling. Databasen må være tilgjengelig for alle som driver med forskning og utvikling. Bruk av databasen kan eventuelt være mot betaling, og med begrensninger på anvendelsen som forhindrer opphavsrettslige konflikter. For utvikling av nye produkter, eller tilpasning av eksisterende, fremmedspråklige produkter, vil det som regel bli nødvendig med ytterligere datainnsamling. Er det nasjonale korpuset godt designet, vil likevel kostnadene ved denne datainnsamlingen kunne være relativt beskjedne.

I tillegg til språkdata, er det også viktig å være klar over at det er nødvendig med kompetanse på såvel språk som på språk- og taleteknologi for våre egne språk. En satsing på språk- og taleteknologisk kompetanseoppbygging i form av forskningsprogrammer ved de nasjonale forskningsmiljøene er derfor en forutsetning. Spesielt gjelder dette universitetene, som utdanner

spesialister og forskere som vil kunne danne en grunnstamme for utvikling av nasjonal språkteknologi. Forskningsprogrammene kan gjerne være anvendelsesorientert, med et samarbeid mellom industri og forskning, men må inneholde langsiktige forskningsmål.

Hverken innsamling av språkdata eller forskningsprogrammer for språk- og taleteknologisk kunnskapsbygging kan forventes å bli finansiert av en pr idag så godt som ikke-eksisterende industri. Rammebetingelsene bestemmes av det offentlige, og det er av avgjørende betydning for framtida til tale- og språkteknologien i Norden at departementer og forskningsråd tar ansvaret for å hjelpe til å legge grunnlaget for utvikling av en nasjonal språkteknologi.