

Regeringsuppdrag att främja utveckling av en nationell språkinfrastruktur för talbaserade tjänster



**Regeringsuppdrag att genomföra ett pilotprojekt i syfte att främja
utveckling av en nationell språkinfrastruktur för talbaserade tjänster**

Rapportnummer

1

Diarienummer

N2014/2840/ITP

ISSN

1650-9862

Författare

Post- och telestyrelsen

Post- och telestyrelsen

Box 5398

102 49 Stockholm

08-678 55 00

pts@pts.se

www.pts.se

Förord

För att Sverige ska följa den internationellt snabbvärande utvecklingen inom talteknologi och utveckla fler talbaserade tjänster på svenska krävs en satsning på nationell språkinfrastruktur för talbaserade tjänster. Personer med funktionsnedsättning är en målgrupp som denna teknik har potential att göra vardagen enklare för, bland annat med talsyntes och dikteringsprogram. Därför har PTS fått i uppdrag av regeringen att utreda förutsättningarna för att skapa en sådan infrastruktur. Regeringsuppdraget (N2014/2840/IIP) består av två huvudaktiviteter - skapandet av en talresursbank med befintligt tillgängliga talresurser och utvecklandet av en taligenkänningsprototyp för direkttextning av tv-program. Förutom dessa två huvudaktiviteter genomfördes sex olika utredningar som alla bidrog till att svara på regeringsuppdragets delaktiviteter.

PTS grupp med sakkunniga har arbetat med målsättningen att finna en lösning som är långsiktig, hållbar och ger maximal effekt för alla slutkonsumenter av talbaserade tjänster, inklusive personer med funktionsnedsättning. PTS fick regeringsuppdraget i juni 2014 och slutredovisar uppdraget den 31 mars 2017.

Resultatet av arbetet framgår av denna rapport. Uppdraget har utarbetats med insatser från handläggare Jörgen Ahlström, Anders Franzén och Adam Rosén på enheten för inkluderande IT och telefoni samt jurist Stina Hemberg från upphandlingsenheten. Styrgruppen för regeringsuppdraget, som består av Malin Wahlquist från PTS, Rickard Domeij från Institutet för språk och folkminnen samt Katarina Persson från Sveriges Television, har även varit involverade i arbetet med regeringsuppdraget. Slutrapporten överlämnas härmed.

Stockholm i mars 2017.

Catarina Wretman
Divisionschef
Post- och telestyrelsen

Innehållsförteckning

Förord	3
Sammanfattning	6
Ordlista	8
1 Bakgrund till regeringsuppdraget	10
1.1 Regeringsuppdraget och tidigare arbete	10
1.2 Metod och genomförande	10
1.3 Vägval i regeringsuppdraget	15
2 Skapandet av en talresursbank för långsiktigt främjande av svensk talteknologi	17
2.1 Nyttan med en talresursbank	17
2.2 Talresurser finns utspridda hos offentliga aktörer	18
2.3 Användningen av en talresursbank	20
2.4 Förvaltning av en framtida talresursbank	21
2.5 Finansiering och organisationsform av en framtida talresursbank	23
2.6 Strategi för en framtida talresursbank	26
3 Skapandet av en taligenkänningsprototyp för direkttextning	28
3.1 Taligenkänning klassas som standarteknologi mogen för massmarknad	28
3.2 Skuggtalarteknik nyttjar taligenkänning	30
3.3 Utvärdering av skuggtalarteknik som direkttextningsmetod	31
4 PTS förslag till nationell utveckling av infrastruktur för talteknologi	35
4.1 Förslag om en långsiktig finansiering av en talresursbank som förvaltas av SND med finansiering från statligt anslag	35
4.2 Förslag om användning av skuggtalarteknik som metod för direkttextning	36
5 Förteckning över bilagor	37

Figurer

Figur 1 - Tidslinje över regeringsuppdragets aktiviteter	11
Figur 2 - Ranking av förvaltningskandidater	22
Figur 3 - Kostnader för en talresursbank	24
Figur 4 - Organisationsform för en talresursbank	25
Figur 5 - Målsättning och strategi för en talresursbank	26
Figur 6 - Gartners hypekurva från 2014.....	29
Figur 7 - Gartners hypekurva från 2016.....	30
Figur 8 - Illustration över direkttextning med skuggtalarteknik.....	31

Sammanfattning

Det finns behov av infrastruktur för att främja talteknologi i Sverige

Utveckling av talteknologi och talbaserade tjänster går snabbt framåt, och taligenkänning anses idag vara en standardteknik. Däremot sker utvecklingen av talteknologiska applikationer huvudsakligen på kommersiellt mer attraktiva språk än svenskan, såsom engelska och spanska. Detta kan leda till att den tekniska innovationen inom talteknologi på svenska begränsas eller helt avstannar. För att det svenska språket ska prioriteras vid talteknikutveckling och att befintliga talbaserade tjänster ska bli användbara för bland annat personer med funktionsnedsättning, krävs en nationell satsning på talteknologi. Detta är nödvändigt om Sverige vill kunna ta en framträdande roll som en innovativ nation med en hög utvecklingstakt av talbaserade tjänster. Dessutom dominerar den svenska talteknologimarknaden idag av ett fåtal amerikanska bolag. En lösning på behovet att skapa en nationell språkinfrastruktur för talbaserade tjänster är att etablera en digital samlingsplats av talresurser, även kallad en talresursbank. Talbaserade tjänster som i framtiden kan nyttja en talresursbank är allt ifrån personliga assistenter såsom Apples Siri eller Google Assistant till skuggtalarsystem för direkttextning av tv-program.

KTH har skapat grunden till en talresursbank

För att skapa en talresursbank har Kungliga Tekniska Högskolan (KTH), på uppdrag av PTS, inventerat den svenska talteknologiska marknaden och skapat grunden till en öppen nationell talresursbank. KTH har även inrättat ett taligenkänningsystem av öppen källkod som är kopplat till talresursbanken, vilket möjliggör att även personer utan stor teknisk talteknologiexpertis kan nyttja talresurserna. PTS har även, via det oberoende konsultbolaget The Boston Consulting Group, utrett nio potentiella förvaltare av en framtida talresursbank och kommit fram till att Svensk Nationell Datatjänst (SND) är mest lämpade. SND föredras huvudsakligen på grund av sitt tydliga förvaltningsintresse, erfarenhet av metadatahantering, sitt aktiva deltagande i en rad internationella organisationer samt SND:s neutralitet på marknaden.

En prototyp har utvecklats för att utreda skuggtalarteknik på svenska

För att svara på hur väl taligenkänning fungerar som direkttextningsmetod har PTS, tillsammans med Sveriges Television (SVT), upphandlat en prototyp för

direkttextning av tv-program med hjälp av skuggtalarteknik. Uppdraget avslutades med en utvärderingsstudie som visar att taligenkänning är ett fungerande alternativ för direkttextning och att skuggtalarteknik fungerar väl för det svenska språket. Utvärderingen visar även att slutanvändare, inklusive personer med funktionsnedsättning, uppskattar den nya tekniken och att användandet av skuggtalarteknik kan innebära vissa positiva effekter för ett programbolag. Exempelvis är utbildningen av nya skuggtalare betydligt kortare än utbildningen för befintliga tekniker. Detta medför en ökad flexibilitet för tv-bolag som snabbare kan ta in nya resurser vid exempelvis resursbrist eller semestrar.

PTS två förslag för främjandet av talbaserade tjänster i Sverige

Det första förslaget som läggs fram i rapporten baseras på att det idag råder en brist på infrastruktur för talresurser på det svenska språket. PTS anser att detta behov kan tillfredsställas vid finansiering av en öppen nationell talresursbank. Rapportens förslag är att SND bör förvalta talresursbanken och initialt erhålla 5-8 miljoner kronor per år i en tvåårig implementeringsfas och sedan 4-6 miljoner kronor årligen. Finansieringen bör vara i form av ett statligt anslag för att garantera oberoende och öppet nyttjande. PTS förslag är även att införa ett forskningsråd och ett expertråd som tillsammans ansvarar för riktning och strategi för talresursbanksförvaltaren. KTH föreslås vara samordnare för forskningsrådet och Institutet för språk och folkminnen föreslås vara samordnare för expertrådet.

Det andra förslaget som läggs fram i rapporten baseras på att taligenkänning är ett fungerande alternativ för direkttextning. PTS förslag är att programbolag i Sverige bör överväga skuggtalarteknik som textningsmetod, både förinspelat och direktsänt. För att underlätta övergången från dagens tekniker till skuggtalarteknik kan det även vara rimligt för staten att stimulera fortsatt utveckling av skuggtalarteknik. Detta skulle i slutändan gynna slutanvändarnas möjligheter att få en större del av programutbudet textat. Ytterligare en fördel är att talresursbanken skulle få fler talresurser från sådana taligenkänningsprojekt. Förutsatt att programbolagen då överför skapade talresurser till talresursbanken. Slutligen är PTS positiva till SVT:s intentioner att fortsätta arbetet med att utveckla skuggtalarteknik för att dels testa fler genrer och dels genomföra reella textningsprojekt som visas inför publik.

Ordlista

Direkttextning – Textning som sker i realtid. Vid direktsända tv-program skrivs textningen ut direkt efter att orden är artikulerade i tv-sändningen.

Kaldi – Kaldi är ett öppet taligenkänningsystem som låter utvecklare skriva så kallade recept som tillåter enkel träning av modeller givet en väldefinierad resurs. Kaldi finns fritt tillgängligt under The Apache License.

Korpus – En korpus är en stor samling språklig data, vanligen en samling text eller transkriptioner av talat språk.

Multimodal data – En benämning som syftar på en kombination av olika datakällor, såsom tal, text och bild.

N-gram – Är en följd av n antal ord i en ordföljd. Om $n=3$ är det tre efterföljande ord som analyseras i förhållande till varandra.

Skuggtalarteknik – En metod som används för att texta direktsända tv-program där en professionell inläsare talar in den text som sägs i tv-programmet. Skuggtalarens tal omvandlas sedan till text i tv-sändningen.

Språkresursbank – En databas som bland annat innehåller språklig data såsom text, tal och ljud i digital form. För mer information, se kapitel 1.3.2

Talarberoende taligenkänningsystem – Ett taligenkänningsystem som är anpassat efter en viss individ, och får därmed hög precision då systemet har anpassats efter talaren.

Talaroberoende taligenkänningsystem – Ett taligenkänningsystem som inte är anpassat efter en viss individ, utan kan användas av många olika människor.

Taligenkänning – Processen att elektroniskt eller datatekniskt tolka mänskligt, naturligt tal.

Talresursbank – En databas som bland annat innehåller språklig data såsom text, tal och ljud i digital form. En databas som kan användas för att utveckla intelligenta talbaserade tjänster.

Talresurser – Inspelat tal som används som grundsten och resurs till talteknologiska applikationer, såsom dialogsystem och dikteringssystem.

Talsyntes – Artificiellt tal som ska efterlikna människans röst skapat med datorer.

Talteknologi – Tal som används som grundsten och resurs till talteknologi, såsom skuggtalarteknik.

Textresurser – Text som används som grundsten och resurs till textteknologiska applikationer, såsom lexikon och stavningskontroll.

Velotype – Ett ackordbaserat tangentbord där stavelser trycks ned istället för enskilda bokstäver. Velotype används idag huvudsakligen som direkttextningsmetod av tv-program.

Word Error Rate – En mätmetod som används för att beräkna antalet felaktiga ord i undertexter. 96 % WER betyder exempelvis att 96 % av orden i undertexterna var korrekt tolkade.

1 Bakgrund till regeringsuppdraget

I detta kapitel beskrivs bakgrunden till nuvarande regeringsuppdrag, vilka metoder som har använts och vilka vägval som har gjorts.

1.1 Regeringsuppdraget och tidigare arbete

Under de senaste 15 åren har olika statliga utredningar genomförts med ambitionen att främja marknaden och infrastrukturen för talbaserade tjänster. ”Mål i mun”, som var ett förslag till handlingsprogram för det svenska språket togs fram av Kulturdepartementets Kommitté för svenska språket 2002 (SOU 2002:27)¹. Efter detta har olika former av utredningar gjorts, bland annat PTS utredning ”Nyttoaspekter av tillgången till en språkresursbank” från 2009. Den senaste större utredningen, var i form av ett regeringsuppdrag, och genomfördes av Institutet för Språk och Folkminnen (ISOF).

Regeringsuppdraget slutrapporterades år 2012 och resulterade i slutrapporten ”Infrastruktur för språken i Sverige”². PTS var samrådande i utredningen, och baserat på ISOF:s rapport låg grunden för PTS regeringsuppdrag. Regeringen valde i juni 2014 att ge PTS i uppdrag att genomföra ett pilotprojekt i syfte att främja utveckling av en nationell språkinfrastruktur för talbaserade tjänster (N2014/2840/ITP)³. Uppdraget slutredovisades i samband med denna rapports överlämning, den 31 mars 2017.

1.2 Metod och genomförande

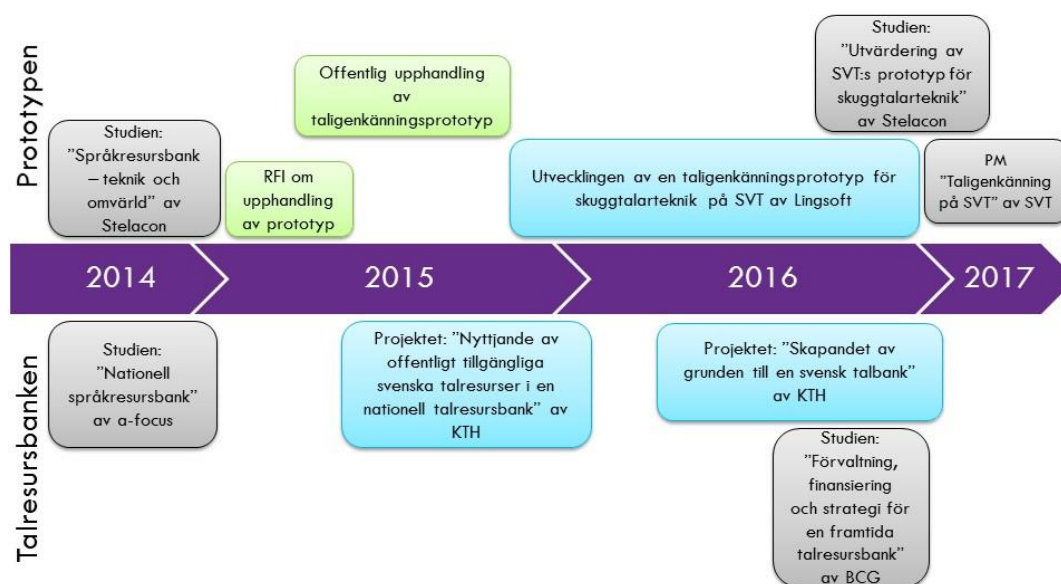
PTS har genomfört alla regeringsuppdragets aktiviteter med hjälp av utlagda konsultutredningar, genomförda upphandlingar och transfererade medel till forskare på KTH. Figur 1 nedan illustrerar en överblick av regeringsuppdragets olika aktiviteter. Aktiviteterna ovanför tidslinjen i figur 1 är primärt kopplade till prototypen, medan aktiviteter under tidslinjen är primärt kopplade till talresursbanken. Bredden på figurerna indikerar hur länge en viss aktivitet pågick. Likaså indikerar färgerna i figuren vilken typ av aktivitet som genomfördes. Svart indikerar en studie eller PM, blått indikerar utveckling eller projekt och grön indikerar PTS upphandlingsarbete.

¹ <http://www.regeringen.se/contentassets/36c88c95edf44b31be6cd524ec81bab1/del-1-mal-i-mun---forslag-till-handlingsprogram-for-svenska-spraket-t.o.m.-kap.-7>

² <http://www.sprakochfolkminnen.se/sprak/sprak-och-it/spraklig-infrastruktur/nationell-sprakdatabank/rapporten-infrastruktur-for-spraken-i-sverige.html>

³ <http://www.regeringen.se/regeringsuppdrag/2014/06/n20142840itp/>

Figur 1 - Tidslinje över regeringsuppdragets aktiviteter



1.2.1 Förankring av regeringsuppdraget

Genomgående i regeringsuppdraget har PTS arrangerat styrgruppsmöten för att strategiskt säkerställa uppdragets riktning och framfart. Alla stora beslut och vägval i projektet har fört förankrats i styrgruppen som bestod av representanter från PTS, ISOF och SVT.

PTS har stämt av progress i uppdragets aktiviteter hos uppdragsgivarna på Närings- och Kulturdepartementet. Exempel på frågor som har behandlats är förskjutningen av slutleverans för regeringsuppdraget från 31 mars 2016 till 31 mars 2017 samt stora vägval.

Vid tre tillfällen har PTS presenterat arbetet med regeringsuppdraget och lyssnat på funktionshinderorganisationernas synpunkter vid PTS samrådsmöten. Alla samrådsdeltagare blev dessutom inbjudna till en specifik workshop för att undersöka hur funktionshinderorganisationernas medlemmar upplever direkttextning. Slutligen har flertalet intresseorganisationer blivit tillfrågade och intervjuade i de olika konsultutredningar som genomförts inom ramarna för regeringsuppdraget.

PTS har också varit stödjande till andra myndigheters utredningar och arrangemang. När myndigheten för press, radio och TV exempelvis presenterade sin utredning om krav på tillgänglighet till tv-sändningar för personer med funktionsnedsättning var PTS behjälpliga och närvarade bland annat på ett remissmöte för att svara på frågor kopplat till PTS regeringsuppdrag⁴.

PTS har även presenterat regeringsuppdraget på olika nätverksträffar och diskuterat med olika aktörer och nyckelpersoner på talteknologimarknaden. Ett sådant exempel var den nätverksträff för myndighetssamverkan inom IT-tillgänglighet som Arbetsförmedlingen arrangerar⁵. Detta ansåg PTS var ett bra tillfälle att sprida information kring uppdraget och uppmärksamma övriga myndigheter om talresursbanken.

Under uppdragstiden har PTS löpande publicerat resultat på webbsidan www.pts.se/sprakresursbank. Detta för att förankra projektets progress och för att samla in eventuella åsikter och kommentarer från branschen.

1.2.2 Offentlig upphandling av en taligenkänningsprototyp

I november 2014 utlyste PTS en Request for information (RFI) med syfte att få marknadens åsikter och kommentarer kring möjlig öppenhet i ett taligenkänningsystem, genre för prototypen samt rimligt mått på träffsäkerhet och fördröjning vid nyttjande av skuggtalarteknik. PTS fick svar på ställda frågor från sex olika aktörer, varav fysiska möten bokades in med fyra utav respondenterna för att diskutera frågorna mer djupgående. Beslut som togs efter genomförd RFI var bland annat att utveckla en prototyp med vädersändningar som genre samt att alla talresurser som nyttjas i utvecklingen av prototypen ska tillgängliggöras i en framtida nationell talresursbank. Som ett resultat av att PTS genomförde en RFI med efterföljande diskussion medförde att den efterföljande offentliga upphandlingen var realistisk och förankrad hos marknaden.

I augusti 2015 utlyste PTS, tillsammans med SVT, en offentlig upphandling avseende utvecklandet av en prototyp för direkttextning av tv-program. Efter resultatet från RFI:n valde PTS och SVT att kravställa en fördröjning på minst

⁴ <http://www.radioochtv.se/sv/projekt/tillgangliga-tv-sandningar/>

⁵ <http://www.arbetsformedlingen.se/Om-oss/Om-Arbetsformedlingen/Samverkan/Digital-tillganglighet.html>

2,8 sekunder och en träffsäkerhet från 95 procent WER⁶. Leverantörerna betygsattes efter 15 skriftliga bör-krav och en presentation där utvalda områden presenterades. De två viktigaste kraven gällde vad leverantörerna kunde garantera för träffsäkerhet och fördröjning, vilket också gav höst poäng i utvärderingen av anbuden. I båda dessa krav fick anbudsgivarna fler poäng desto kortare fördröjning och högre träffsäkerhet som kunde garanteras. Två aktörer svarade på upphandlingen, båda dessa aktörer presenterade även sitt anbud hos PTS. Det finska språkteknologiföretaget Lingsoft vann upphandlingen och utvecklade prototypen från november 2015 till och med december 2016. Lingsoft kunde garantera en fördröjning på maximalt 2,5 sekunder och en träffsäkerhet på minst 96 procent WER, vilket var lägsta fördröjningen och högsta träffsäkerheten av de två anbuden som inkom. Förutom att acceptanstesta prototypen genomfördes även en utvärderingsrapport, se bilaga 6: ”Utvärdering av SVT:s prototyp för skuggtalarteknik”.

Hela processen med prototypupphandlingen fick i slutet av 2016 ytterligare uppmärksamhet när Upphandlingsmyndigheten lyfte fram upphandlingen med tillhörande RFI som ett gott exempel på en väl genomförd funktionsupphandling⁷.

1.2.3 Externa utredningar och kostnader för regeringsuppdraget

Konsulter från Stelacon anlätades för att utreda den kommersiella utvecklingen avseende talteknologisk infrastruktur på svenska samt en internationell omvärldsanalys. Detta resulterade i en konsultrapport (bilaga 1) och sammanfattas bland annat i kapitel 3.2.

Konsulter från a-focus anlätades för att utreda kommersiella och offentliga aktörers intresse av att använda en nationell talresursbank som bas för utveckling av produkter och tjänster för personer med funktionsnedsättning. Vidare görs en beskrivning av samhällsnyttan av en nationell talresursbank ur ett tillgänglighetsperspektiv. Detta resulterade i en konsultrapport (bilaga 2) och sammanfattas bland annat i kapitel 2.4.

⁶ WER = ”Word Error Rate” är metoden som används för att mäta antalet felaktiga ord i undertexter. 96 % WER betyder att 96 % av orden i undertexterna var korrekt tolkade.

⁷http://www.upphandlingsmyndigheten.se/globalassets/publikationer/exempelblad/exempelblad_post_telestyrelsen_funktionsupphandling_webb.pdf

Forskare från KTH anlätades för att utreda och inventera vilka talresurser som finns tillgängliga och är av tillräckligt hög kvalitet för att ingå i en framtida talresursbank. Detta resulterade i en rapport (bilaga 3) och sammanfattas bland annat i kapitel 2.2.

Forskare från KTH anlätades även för att samla in rekommenderade talresurser med syftet att skapa grunden för en svensk talresursbank. Förutom att samla in befintliga talresurser arbetade även KTH för att långsiktigt säkerställa olika resursintagsprocesser kopplat till talresursbanken samt skapa ett recept för taligenkänning av öppen källkod. Detta resulterade i en rapport (bilaga 4) och sammanfattas bland annat i kapitel 2.3.

Konsulter från The Boston Consulting Group anlätades för att utreda vem som bör förvalta en framtida talresursbank samt ge förslag på hur organisationen bör vara konstruerad, hur finansieringen bör se ut samt hur strategin bör vara utformad för en framtida nationell talresursbank. Detta resulterade i en rapport (bilaga 5) vars innehåll ligger till grund för PTS förslag om framtida talresursbank och sammanfattas bland annat i kapitel 2.5 – 2.7.

Konsulter från Stelacon anlätades för att utvärdera vad personer med funktionsnedsättning anser om taligenkänningsteknik som undertextmetod. Utvärderingen genomfördes med användartester och en workshop. Detta resulterade i en rapport (bilaga 6) vars innehåll ligger till grund för det förslag som PTS har gjort i samband med taligenkänning som direkttextningsmetod och sammanfattas bland annat i kapitel 3.4.

SVT har kommenterat skuggtalartekniken utifrån programbolagets perspektiv, mot bakgrund av erfarenheterna av pilotprojektet. Detta resulterade i ett PM (bilaga 7) och sammanfattas i kapitel 3.4.

PTS redovisade kostnader för regeringsuppdragets uppgick till 8 910 920 kronor, vilket var under det maximala belopp på 10 miljoner kronor som uppgavs i regeringsbeslutet. För utförligare information om kostnaderna, se bilaga 8.

1.3 Vägval i regeringsuppdraget

PTS har gjort tre huvudsakliga vägval som har påverkat uppdragets genomförande och de förslag som PTS ger i kapitel 4. Dessa vägval följer nedan.

1.3.1 Uppdelning i två huvudaktiviteter

Enligt regeringsuppdraget ska PTS utveckla en prototyp av en språkresursbank med talresurser (även kallad talresursbank) för att utreda möjligheten att använda taligenkänning som direkttextningsmetod av tv-program. Efter att ha genomfört konsultrapporter och RFI valde PTS att dela upp aktiviteten i två olika huvudaktiviteter. Den ena är skapandet av en talresursbank, dvs. en digital samlingsplats av talresurser, och den andra är utvecklingen av en taligenkänningsprototyp för direkttextning av tv-program. Uppdelning har gjort det lättare att fokusera på respektive aktivitetens skilda behov av bland annat talresurser, vilket har lett till bättre slutprodukter. Trots uppdelningen har ändå synergieffekter uppstått mellan huvudaktiviteterna. Exempelvis har alla talresurser som har nyttjats för att förfinas och träna upp taligenkänningsprototypen levererats till talresursbanken.

1.3.2 Definition och benämning av talteknologi och talresursbank

Enligt regeringsuppdraget ska PTS utreda förutsättningarna för utveckling och finansiering av en heltäckande nationell språkresursbank i syfte att främja utveckling av nationell språkinfrastruktur för talbaserade tjänster. I PTS uppdrag behandlas talresurser och andra typer av data (även text) ur ett talperspektiv, dvs. talteknologi, och inte språkteknologi. Den forskning och utveckling som mestadels avses när man talar om språkteknologi handlar om skrivet språk, och resurserna är huvudsakligen text. Talteknologin, å andra sidan, handlar om det talade språket. De resurser som är mest värdefulla i talteknologin har med ljud och med registreringar av mänskliga beteenden att göra, och behandlas sällan på samma sätt som det skrivna språket. Därmed används benämningarna talresursbank och talteknologi.

Avgränsningen från språkteknologi och en språkresursbank till talteknologi och en talresursbank genomfördes halvvägs in i projektet, vilket bland annat betyder att de tidigare konsultutredningarna i regeringsuppdraget gör benämningen språkteknologi och språkresursbank. Fokus i uppdraget har dock alltid varit mot talbaserade tjänster, så de analyser och rekommendationer som

gjort i tidigare utredningar anses fortfarande vara relevanta. Beslutet att använda benämningen talteknologi och talresursbank har tagits i samråd med styrgruppen för regeringsuppdraget.

1.3.3 Förlängning av tiden för genomförande av regeringsuppdrag

PTS hade som initial ambition att slutrapportera regeringsuppdraget den 31 mars 2016, men detta blev förlängt till den 31 mars 2017. Skälen till förlängning var att uppdragets omfattning var större än väntat och att ytterligare tid behövdes för att säkerställa kvalitet och förankring hos berörda aktiviteter. Förlängningen var nödvändig, både för upphandling och utvecklingen av taligenkänningsprototypen som tog ca 2 år och för skapandet av grunden till talresursbanken som tog ca 1,5 år.

2 Skapandet av en talresursbank för långsiktigt främjande av svensk talteknologi

I detta kapitel beskrivs huvudsakligen den talresursbank som KTH, på uppdrag av PTS, har lagt grunden till. En talresursbank som är öppen och fri för alla att använda, med syftet att främja utveckling av talbaserade tjänster i Sverige.

2.1 Nyttan med en talresursbank

En talresursbank är en digital samlingsplats av talresurser som består av språklig data, innehållande bland annat text-, tal- och multimodala korpusar. Talresursbanken kan även bestå av analysverktyg såsom lexikon, språkbeskrivningar och olika basverktyg för att märka upp språkligt material med metadata samt statistiska modeller av akustiska signaler. En talresursbank kan även inkludera deriverade talteknologiska resurser, vilket åsyftar anonymiserade versioner av ursprungsdata som gör att den personliga integriteten skyddas och eventuell upphovsrätt upprätthålls.

Talresursbankens huvudsyfte är att tillgängliggöra och sprida fria talresurser till alla, inklusive företag, enskilda applikationsutvecklare, privata aktörer och forskningsinstitut. Dessa användare kan bland annat använda talresurserna till att träna upp sina applikationer, vilket i sin tur ökar medborgarnas tillgång till högkvalitativa talbaserade tjänster eller produkter. Talresursbankens ska ha ett brett nationellt intresse och eftersträva en bredd som inkluderar talgrupper och talstilar som annars riskerar att uteslutas från kommersiella satsningar, exempelvis minoritetsspråk och uttalsvariationer på grund av funktionsnedsättning. När talresursbankens data tillgängliggörs kan de bli sökbara, återanvändbara samt långtidsbevarande och därmed främja vidareutveckling av talteknologi. Ju mer talresurser av hög kvalitet som finns kopplat till talresursbanken, desto bättre blir de talteknologiska tjänster som tränas upp av talresursbanken. För mer information om infrastruktur för talbaserade tjänster och dess innehåll, se bilaga 3: ”Nyttjande av offentligt tillgängliga svenska talresurser i en nationell talresursbank” och bilaga 5: ”Förvaltning, finansiering och strategi för en framtida talresursbank”.

2.2 Talresurser finns utspridda hos offentliga aktörer

För att skapa en talresursbank genomförde KTH, på uppdrag av PTS, en inventering av svenska offentligt tillgängliga talresurser med syfte utreda vilka resurser som fanns tillgängliga och höll tillräcklighet hög kvalitet. Efter att inventeringen var genomförd samlades även dessa in av KTH i ett efterföljande uppdrag. Den framtagna talresursbanken är uppbyggd på befintlig teknik och med befintliga talresurser eftersom detta är mer kostnadseffektivt än att spela in helt nya talresurser.

2.2.1 KTH:s inventering av talresurser

KTH:s inventering fann ett flertal offentliga talresurser av hög kvalitet som lämpar sig väl i en offentlig talresursbank. Dessa återfinns huvudsakligen hos universitets- och forskningsinstitutioner, myndigheter, bibliotek och museer. De talresurser som KTH ansåg var av hög kvalitet och sedermera samlades in var bland annat talresurser från prototypprojektet som genomfördes på SVT, data från Nasjonalbiblioteket i Oslo samt Riksdagens talresurser. Alla insamlade talresurser ligger för närvarande på en dedikerad hårdvara på KTH och kan därmed överföras till en framtida förvaltare i sin helhet. KTH avser även att förvalta och underhålla talresurserna tills ett beslut är taget om en ny förvaltare av talresursbanken.

KTH har även arbetat aktivt med att samverka med andra talteknologiprojekt, med syfte att finna synergieffekter och potentiellt tillföra mer data till talresursbanken. Dessa samverkansprojekt var både projekt som avsågs starta inom kort, såsom projektförslag om att translitterera texter tillsammans med Uppsala universitetsbibliotek samt projekt som pågick parallellt med KTH:s uppdrag. Ett exempel på sådant projekt är Wikispeech. Wikispeech är ett samarbete med Wikimedia Sverige om att skapa en talsyntes globalt på Wikipedia, även detta är finansierat av PTS, men utanför regeringsuppdraget⁸. Wikispeech-projektet är ett tydligt exempel på ett projekt som vid en existerande talresursbank blir billigare och mer kostnadseffektivt att genomföra.

Slutligen skapade KTH också ett recept på akustiska modeller för taligenkänning av systemet Kaldi, så att likväl experter som utvecklare utan talteknologiexpertis kan nyttja talresurserna. Kaldi är ett öppet

⁸ <http://www.pts.se/pts/templates/newspage.aspx?id=56330&epslanguage=sv>

taligenkänningsystem som låter utvecklare skriva recept som tillåter enkel träning av modeller givet en väldefinierad resurser. Det öppna taligenkänningsystemet är av godkänd nivå, men erhåller inte samma kvalitet som de kommersiella lösningarna som finns på marknaden. Däremot kan systemet bidra till mindre avancerade taligenkänningsprojekt och bistå vid olika typer av prototypprojekt. För mer information om KTH:s inventering och insamling av talresurser, se bilaga 3: ”Nyttjande av offentligt tillgängliga svenska talresurser i en nationell talresursbank” och bilaga 4: ”Skapandet av grunden för en svensk talbank”.

2.2.2 Rättigheter och integritetshänsyn i en öppen talresursbank

Ett problem som uppstod i KTH:s uppdrag var att rättigheter och integritetshänsyn i många fall visade sig olösliga utan ny lagstiftning. Många resursförvaltare undviker problematiken genom att inte tillgängliggöra rådata, utan tillgängliggör istället olika former av derivat - till exempel analyser, modeller, och statistik. Genom att processa talresurserna kan man säkerställa att identiteten hos deltagarna skyddas, och genom att tillgängliggöra statistik och modeller baserade på talresurserna kan man dra nytta av det utan att bryta mot upphovsrätten. På myndigheter som hanterar resurser som är relevanta ur ett talteknologiskt perspektiv kan man lägga till steg i de befintliga processerna som gör det möjligt att dynamiskt och fortlöpande dela information till en liten kostnad. Och genom att tillhandahålla verktyg som hjälper de som arbetar med talresurser kan man dra nytta av att verktygen genererar data som en sidoeffekt.

En viktig utgångspunkt för skapandet av grunden till talresursbanken är att den ska vara helt öppen och fri för alla att använda, oavsett om nyttjaren är en offentlig aktör, kommersiellt företag, forskningsinstitut eller privatperson. Öppenheten kan dock i vissa hänseenden ställa till problem. Många nuvarande förvaltare av talresurser har rådande avtal med inspelningsobjekten om att endast använda talresurserna till specifika ändamål, exempelvis forskning. Eftersom att det inte är möjligt att i efterhand individuellt återkoppla med alla inspelningsobjekt för att utöka ändamålet till en talresursbank är det många talresurser som inte kan användas. Det är därför viktigt för producenter av talresurser, framförallt projekt som är offentligt finansierade, utökar rättigheterna till att möjliggöra nyttjandet i en talresursbank. För mer information om rättighetsproblematiken vid insamling av talresurser, se bilaga

3: ”Nyttjande av offentligt tillgängliga svenska talresurser i en nationell talresursbank”.

2.2.3 En talresursbank främjar den svenska talteknologimarknaden

Enligt KTH finns det en obalans på talteknologimarknaden där ett fåtal amerikanska bolag styr marknaden. Det råder även en osäkerhet i hur befintlig lag ska tolkas gällande satsningar på fritt tillgängliga data. De försök som gjorts med kommersiella lösningar för datadelning (t.ex. ELRA i Europa och LDC i USA) har inte lyckats lösa situationen. I stället säljs stats- och forskningsfinansierade data ofta till höga priser. Kostnaden för att samla in den här typen av data är mycket hög, vilket begränsar möjligheten att de kommer till maximal nytta. En offentlig svensk talresursbank som tillgängliggör data, arbetar för att driva på utvecklingen av nya data samt har ett juridiskt fokus skulle således vara ett steg i riktningen för att minimera den obalans som KTH har identifierat. Likaså skulle införandet av en talresursbank främja den svenska talteknologimarknaden. För mer information om vikten av en hållbar och långsiktig talresursbank, se bilaga 4: ”Skapandet av grunden för en svensk talbank”.

2.3 Användningen av en talresursbank

Det svenska språket är relativt sett litet, och därmed inte kommersiellt prioriterat av internationella språkteknologiföretag. Detta leder till att allt fler nya innovativa talbaserade tjänster utvecklas globalt, men inte lanseras på det svenska språket. Ett exempel på detta är Skype Translate som möjliggör direktkommunikation på olika språk via videokommunikation⁹. Ett annat exempel är YouTube's automatiska undertextningsfunktion, som automatiskt transkriberar ljudfilerna i ett videoklipp och möjliggör automatiska undertexter på YouTube-videos¹⁰. Båda dessa tjänster finns tillgängligt på tio olika språk, men ingen av tjänsterna finns på det svenska språket. Införandet av en nationell talresursbank med svenska talresurser garanterar i och för sig inte att liknande tjänster i framtiden tillgängliggörs på svenska, men det ökar sannolikheten att internationella företag väljer att implementera sina talteknologiska applikationer på den svenska marknaden. Även svenska småföretag och offentliga aktörer främjas av en talresursbank, om det finns fritt tillgängliga talresurser ökar sannolikheten att dessa företags talteknologiska

⁹ <https://support.skype.com/en/faq/FA34543/what-languages-are-available-in-skype-translator>

¹⁰ <https://support.google.com/youtube/answer/6373554?hl=en>

idéer kan genomföras. Således är framtagandet av en talresursbank främjande för både privat och offentlig sektor samt för stora och små aktörer.

Noterbart är att en talresursbank inte är en konkurrent till någon befintlig aktör på den svenska talteknologimarknaden. Talresursbanken kommer inte att leverera någon produkt eller tjänst och avser därmed inte konkurrera om marknadsandelar. Talresursbankens syfte är att bistå så många som möjligt med så mycket talresurser som möjligt, oavsett vem mottagaren eller ägaren till en viss produkt är. Pondera att ett stort kommersiellt företag som Microsoft exempelvis skulle vilja implementera Skype Translator på den svenska marknaden och addera det svenska språket. De ska då ha möjlighet att använda talresursbanken och dess resurser för att träna upp sin applikation, oavsett hur mycket interna talresurser de har sedan innan.

För slutanvändare kan en etablerad talresursbank bland annat innebära att fler talsyntes- och taligenkänningslösningar skapas för den svenska marknaden. Dessa talbaserade tekniker möjliggör ett nytt gränssnitt där slutanvändare kan nyttja rösten för att interagera med maskiner för att bland annat hantera kundtjänstärenden, fylla i blanketter, beställa biljetter och boka fysiska möten. Sådana tjänster kan innebära att många målgrupper i samhället får större frihet och bättre möjligheter att klara av vardagssituationer, bland annat personer med funktionsnedsättning. Utvecklingen av talbaserade tjänster kan även innebära att slutanvändaren inte hänvisas till utpekade speciallösningar, utan kan använda samma standardtekniker som alla andra, så fler kan bli fullt delaktiga i samhällslivet. Ökad tillgänglighet i aktuella standardtjänster kan skapa förutsättningarna för självständighet och självbestämmande som kan resultera i ökad upplevd delaktighet och större känsla av samhörighet med samhället i stort. En förutsättning för detta är att det sker en samverkan mellan akademi, kommersiella företag och offentlig sektor, och att talresursbanken även har ett fokus på att bistå tjänster som avses utvecklas för grupper som inte kommersiellt är prioriterade, såsom språkliga minoriteter och personer med funktionsnedsättning. För utförligare information om samhällsnyttan av en nationell språkinfrastruktur se bilaga 2: ”Nationell språkresursbank”.

2.4 Förvaltning av en framtida talresursbank

För att utreda hur en talresursbank bäst bör vara organiserad och vem som anses vara mest lämpad att förvalta en framtida talresursbank anlätade PTS

konsultbolaget The Boston Consulting Group (BCG) för att utreda frågan. BCG har i sitt uppdrag utvärderat nio olika potentiella förvaltningskandidater och betygsatt dem efter 22 olika kriterier med olika prioritet. Den förvaltare som BCG ansåg vara mest lämpad för att förvalta en nationell talresursbank är Svensk Nationell Datatjänst (SND). Se figur 2 nedan.

Figur 2 - Ranking av förvaltningskandidater

Kriterier	Förvaltarekandidat							
	Svensk Nationell Datatjänst	Kungliga Tekniska Högskolan	Institutet för språk och folkminnen	Språkbanken	Kungliga Biblioteket	Nasjonalbiblioteket	Digisam	Terminologicentrum
Grundläggande kriterier								
Verksam i Sverige	✓	✓	✓	✓	✓	✗	✓	✓
Statligt ägt	✓	✓	✓	✓	✓	✓	✓	✗
Förvaltningsintresse	●	●	●	○	●	●	●	●
Oberoende kommers/forsk	●	●	●	●	●	●	●	●
Etablerad aktör/ Förtroende	●	●	●	●	●	●	●	●
Långsiktighet/ Stabilitet	●	●	●	●	●	●	●	●
Forskningsnära	●	●	●	●	●	●	●	●
Nyckelkompetens								
Talteknologisk kompetens	●	●	●	●	●	●	○	○
Teknisk kompetens	●	●	●	●	●	●	●	●
Juridisk kompetens	●	●	●	●	●	●	●	○
Samordning talinfrastruktur	●	●	●	●	●	●	●	●
Admin/ budgetansvar	●	●	●	●	●	●	●	●
Distribution /kommunikatör	●	●	●	●	●	●	●	●
IT säkerhet	●	●	●	●	●	●	●	●
Operativa förmågor								
Insamling	●	●	●	●	●	●	●	○
Datahantering	●	●	●	●	●	●	●	●
Uppdatering	●	●	●	●	●	●	●	○
Lagring	●	●	●	●	●	●	●	●
Tillgängliggörande	●	●	●	●	●	●	●	●
Infrastruktur								
Maskinkapacitet	●	●	●	●	●	●	○	○
Skalskydd/ hostinglösning	●	●	●	●	●	●	○	●
Fysisk lokal	●	●	●	●	●	●	●	●
Viktat betyg (%)	95%	92%	80%	76%	73%	71%	65%	43%
Ranking	1	2	3	4	5	6	7	8

BCG rankar SND högst på grund av sitt tydliga förvaltningsintresse, erfarenhet av metadatahantering, sitt aktiva deltagande i en rad internationella organisationer samt SND:s neutralitet. SND har sedan 2008 anförtrots ett nationellt uppdrag från Vetenskapsrådet att samordna data i forskningssyfte. I den nuvarande verksamheten hos SND finns tydliga visioner om långsiktighet, öppenhet och tillgänglighet, vilket går i linje med talresursbankens syfte. SND är certifierade enligt Data Seal och Approval, vilket omfattar processer för hur data lämnas ut, tas emot och struktureras till hur de återanvänds och tillgängliggörs. Slutligen har SND ingen egen användning av talresurser, vilket säkerställer en oberoende inställning till förvaltningen. Detta gör att SND

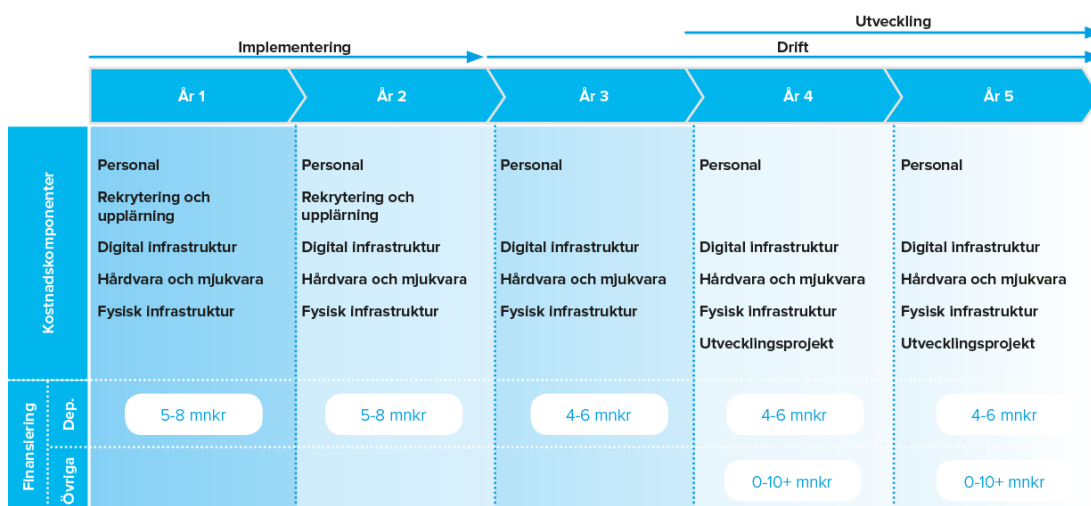
prioriterar alla forskningssätten och kommersiella aktörer likvärdigt och säkerställer en neutralitet på marknaden. SND arbetar även aktivt med att etablera kontaktnät mot flertalet lärosäten i Sverige, vilket ligger i linje med den samordnarroll som efterfrågas på talresursbanken.

2.5 Finansiering och organisationsform av en framtida talresursbank

BCG:s konsultutredning utredde även vilken organisationsform och vilken finansiering en framtida talresursbank bör ha. I syfte att eliminera kostnader som inte är värdeskapande för slutanvändarna har talresursbanken organiseras efter de mest väsentliga operativa funktionerna. En sådan modell innebär kontinuerlig uppdatering av de talresurser som motsvarar industrins minimikrav på relevans och kvalitet, men kan även öppna upp för storskaliga satsningar på projektbasis när finansiering och samarbetsparter matchas.

Vid kartläggning av finansieringsbehovet har kostnader associerade med implementering, drift och utveckling åtskiljts, vilket tydliggörs i figur 3 nedan. Implementeringsfasen är bedömd till 2 år med en årlig kostnad på 5-8 miljoner kronor beroende på förvaltarens initiala förutsättningar och existerande infrastruktur. Denna årliga finansiering minskar sedan till 4-6 miljoner kronor per år efter den initiala implementeringsfasen. BCG gör bedömningen att dessa årliga implementerings- och driftkostnader bör finansieras av statligt anslag för att säkerställa en oberoende och långsiktigt position. För att en investering i en framtida talresursbank ska vara meningsfull bedöms att finansieringen bör säkerställas för minst tio till tjugo år framåt i tiden.

Figur 3 - Kostnader för en talresursbank



Av möjliga finansieringsformer för drift anses statligt anslag som mest relevant. Detta för att säkerställa oberoende och långsiktighet. BCG har även identifierat tre relevanta departement som skulle kunna agera finansjär:

- Utbildningsdepartementet, som är beredande departement till den föreslagna förvaltaren SND och har koppling till forskningsdimensionen.
- Näringsdepartementet, som är beredande departement till regeringsuppdraget och har koppling till den marknadsfrämjande dimensionen.
- Kulturdepartementet, som är involverade i regeringsuppdraget och har koppling till den språkpolitiska dimensionen.

Talresursbanken bör lämpligen placeras på en egen enhet inom en större etablerad organisation, för att på bästa sätt nyttja existerande infrastruktur och möjliggöra bemanningen efter behov. Baserat på utvärderingen av användarens behov och övergripande målsättning för talresursbanken har väsentliga funktioner kartlagts. Av de kostnadskomponenter som finns vid etablerandet av en talresursbank är det personalkostnader som förväntas vara den största kostnadsposten. Som visas i figur 4 nedan skulle den operativa verksamheten inledningsvis kräva cirka tre till fem personer, som vid behov kompletteras av

insatser från externa råd. Den organisationsform som BCG har rekommenderat för en fungerande talresursbank är följande:

- En till två personer bör till en början ansvara för all samordning och extern kommunikation på talresursbanken. Samordnaren har i huvuduppgift att samverka med utvalda aktörer, utföra omvärldsbevakning samt utvärdera talresursbankens möjligheter att tillgodogöra sig utfallet från befintliga talteknologiprojekt i Sverige.
- En till två personer bör vara talteknologer och arbeta med alla uppgifter relaterade till datahantering, såsom insamling, komplettering, metadatahantering och uppdatering av talresurser.
- En person bör vara drift- och webbansvarig med huvuduppgiften att drifva talresursbanken och säkerställa en användarvänlig webbkatalog med slagkraftig sökfunktion.

Figur 4 - Organisationsform för en talresursbank

Bemanning	Ansvarsområden	Kompetenskrav
Samordnare / extern kommunikatör 1-2 pers	<ul style="list-style-type: none"> • Samordna talprojekt med utvalda aktörer, forskning • Samverka nätverk • Övervaka hela värdekedjan 	<ul style="list-style-type: none"> • Skriva juridiska avtal, upphovsrättspecialisering • Samordning nationella och internationella nätverk <p>Rollen kräver projektledning och juridisk kompetens för skrivande av avtal och licenser med databaser</p>
Talteknolog 1-2 pers	<ol style="list-style-type: none"> 1. Insamling av talresurser 2. Datahantering 3. Uppdatering 4. Lagring/ hosting 5. Tillgängliggörande 	<ul style="list-style-type: none"> • Teknisk kompetens • Juridisk kompetens • Datahantering <p>Rollen kräver talteknologisk systemutvecklare med juridisk kompetens, för effektiv datahantering.</p>
Drift och webbansvarig 1 pers	<ul style="list-style-type: none"> • Admin/ budgetansvar • Distribution /kommunikatör • IT säkerhet • Tillgängliggöra resurser • Support function 	<ul style="list-style-type: none"> • Ledning/ controller • Distribution /kommunikatör • Marknadsföring <p>Rollen kräver erfarenhet av drift av stora databaser och webb, samt kompetenser inom admin, ledning och marknadsföring.</p>
Talråd	<ul style="list-style-type: none"> • Säkerställa huvudsyfte: bevaka konsumentnytta • Tillgodose strategisk riktning; i linje språkpolitiska prioriteringar 	<ul style="list-style-type: none"> • Representanter från ISOF, funktionsgrupper, kommersiella företag och internationella taldatabaser t.ex. Norges Språkbank
Forskningsråd	<ul style="list-style-type: none"> • Säkerställa industrirelevans, kvalitet och utvecklingstempo • Behovsanlys, skapa akustiska modeller, utvärdera metoder 	<ul style="list-style-type: none"> • Talteknologer från KTH, språkteknologer, Språkbanken GU etc <p>■ Eventuellt outsourca</p>

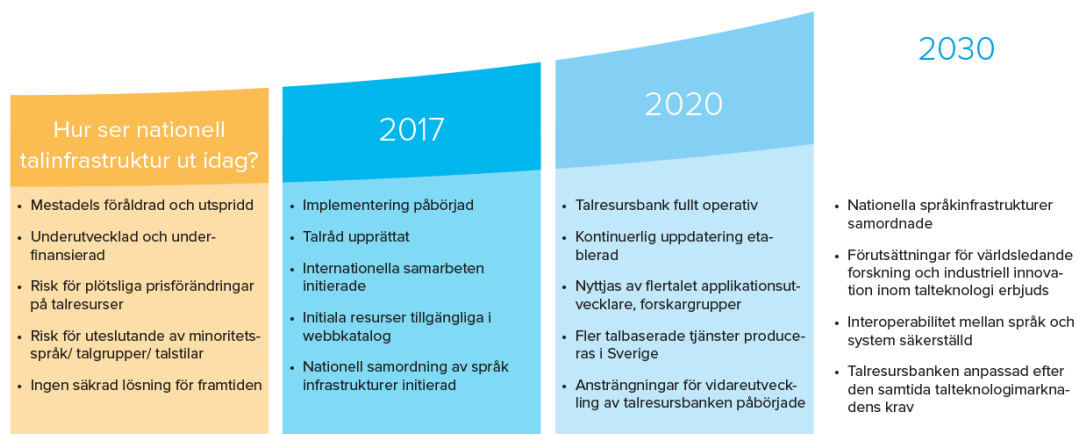
Kopplat till talresursbanken ska det även finnas ett talråd och ett forskningsråd. Talrådets har som uppgift att bevaka det kommersiella huvudsyftet samt besluta om talresursbankens strategiska inriktning. Talrådet ska bestå av representanter från universitet, talteknologiföretag, myndigheter samt

intresseorganisationer. ISOF föreslås vara samordnande och ansvariga för att talresursbankens övergripande strategiska riktning är i samklang med de politiska prioriteringar som föreligger svensk språkteknologi. Forskningsrådet har som huvuduppgift att bevaka forskningsrelevansen, avgöra vilken metod för utveckling som bör prioriteras samt vilken talinfrastruktur som är mest relevant att investera i. Avdelningen för tal, musik och hörsel på KTH föreslås vara samordnare och ansvara för att säkerställa industrirelevans, kvalitet, forskningsnytta och att ett tillräckligt högt utvecklingstempo hålls. För mer information om föreslagen organisationsform, se bilaga 5: ”Förvaltning, finansiering och strategi för en framtida talresursbank”.

2.6 Strategi för en framtida talresursbank

Målsättningen för talresursbanken är att etablera en hållbar och långsiktig infrastruktur för talresurser i Sverige. Detta genom att samordna, nyttja och tillgängliggöra talteknologiska satsningar och insamlingar. I figur 5 nedan presenteras BCG:s föreslagna målsättning och strategi för en talresursbank från idag till och med år 2030.

Figur 5 - Målsättning och strategi för en talresursbank



Enligt föreslagen strategi ska implementeringen påbörjas i slutet av år 2017. De talresurser som KTH har tagit fram i uppdraget att skapa grunden till en talresursbank ska lämnas över till vald förvaltare. Både ett tal- och forskningsråd ska ha etablerats och långsiktiga strategier för talresursbanken

ska initierats. Slutligen ska webbkatalogen vara färdigställd och ska vara tillgänglig för spridning.

År 2020 ska talresursbanken vara fullt operativ och komplett personalstyrka för uppdraget ska vara på plats. Talresursbanken uppdateras kontinuerligt och nya resurser tillförs, vilket håller banken aktuell. Det finns en bred kännedom om talresursbanken, både i och utanför Sverige. Fler talbaserade tjänster utvecklas i Sverige, eftersom att infrastruktur finns på plats som möjliggör en snabb och effektiv träning av talbaserade tjänster.

År 2030 ska nationella språkinfrastrukturer inom terminologi (Rikstermbanken), text (Språkbanken) och tal (Talresursbanken) vara samordnade och möjliggör skapandet av en gemensam vision som säkerställer utveckling i samma riktning. Talresursbanken skapar förutsättningar för världsledande forskning och industriell innovation vilket stärker svensk talteknologi internationellt. Slutligen ska talresursbanken vara anpassad efter den samtida talteknologimarknaden och främjar utvecklingen av talbaserade tjänster i Sverige. För mer information om framtida strategi för en talresursbank, bilaga 5: ”Förvaltning, finansiering och strategi för en framtida talresursbank”.

3 Skapandet av en taligenkänningsprototyp för direkttextning

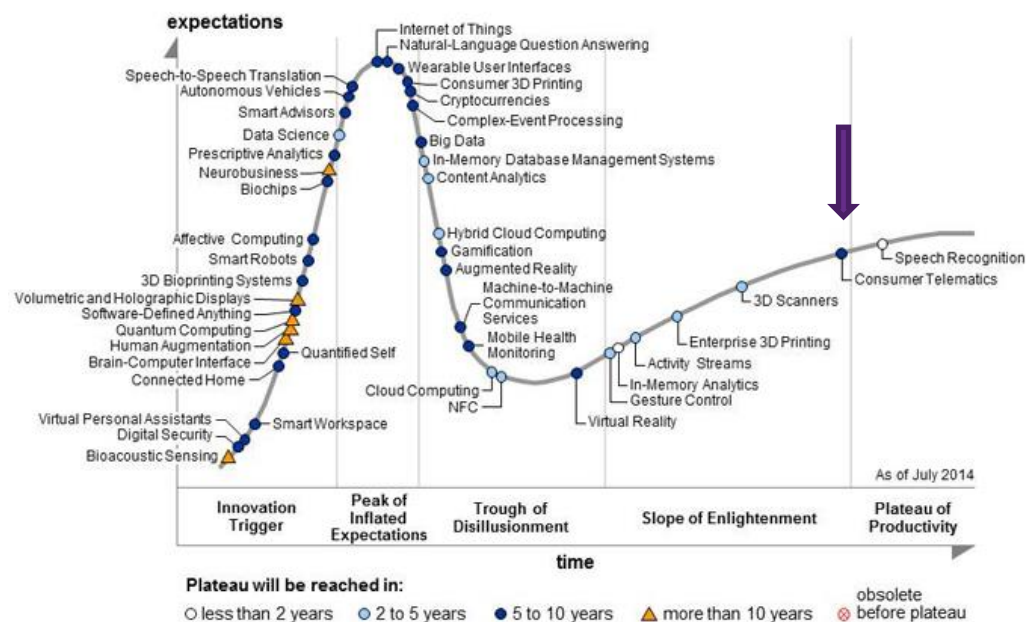
I detta kapitel beskrivs huvudsakligen marknaden för talteknologi och den taligenkänningsprototyp som Lingsoft har tagit fram och testat i SVT befintliga direkttextningssystem.

3.1 Taligenkänning klassas som standarteknologi mogen för massmarknad

Talteknologi och dess närliggande områden genomsyras av en positiv utveckling med hög innovationstakt. Vilket styrks av att taligenkänning är en mogen teknologi som används kommersiellt med framgång. Även talsyntes, som nyttjar text-till-tal istället för tal-till-text som sker vid taligenkänning, används i stor utsträckning och tillgängliggör texter via textuppläsning. IT-konsultbolaget Gartner publicerar årligen en kurva över nya framväxande teknologier. I Gartners kurva från 2014 var taligenkänning längst ut på kurvan, vilket tolkas som att det är en mogen teknologi som är färdig för massmarknad¹¹. Detta bekräftas även från KTH, som framhåller att tekniken bland annat har skapat en stor mängd applikationer och tjänster inom många olika industrier, inklusive tillgänglighet. Se figur 6 för Gartners hypekurva från 2014.

¹¹ <http://www.gartner.com/newsroom/id/2819918>

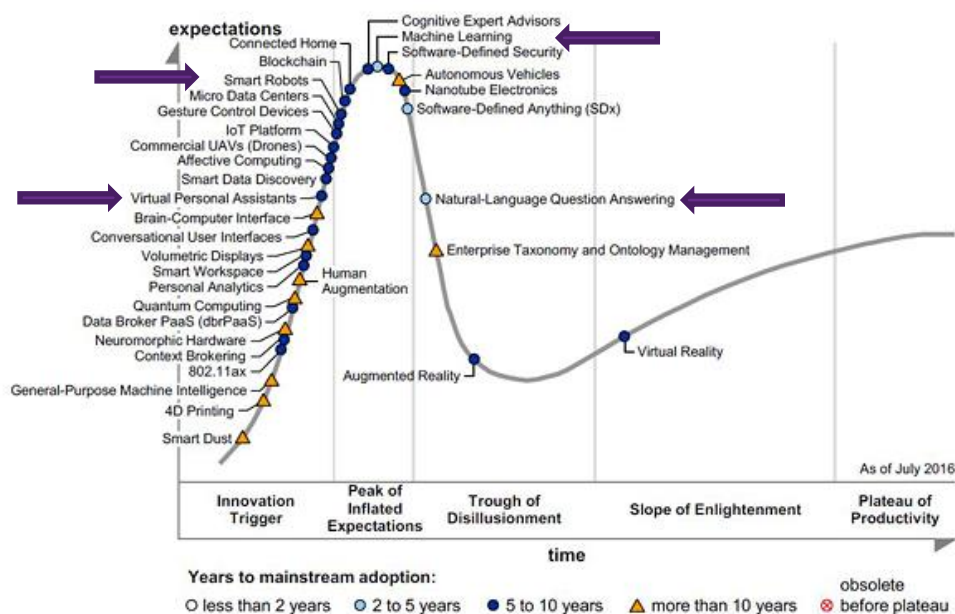
Figur 6 - Gartners hypekurva från 2014



År 2014 var sista året som taligenkänning placerades på Gartners hypekurva, vilket betyder på att taligenkänning sedan dess har klassats som en standardteknik, även detta bekräftas av KTH. Att taligenkänning anses vara en standardteknik är ett tydligt tecken att taligenkänning kommer bli allt vanligare som gränssnitt mellan teknik och människan, vilket också PTS har märkt av i sitt främjandearbete kring digital tillgänglighet. I Gartners senaste hypekurva från 2016 finns dessutom flertalet teknologier som innehåller starka inslag av talteknologi, såsom ”Natural-Language Question Answering”, ”Virtual Personal Assistants”, ”Machine Learning” samt ”Smart Robots”¹², se figur 7 nedan. Flertalet av dessa områden når sin produktiva plåtå om fem år eller mer, vilket ytterligare belyser den framfart och utveckling som har skett och förmodans ske inom talteknologimarknaden i framtiden. Taligenkänning och dess tillämpningar kan idag ge stor nytta för alla, inklusive personer med funktionsnedsättning. Enligt KTH finns det taligenkänningsapplikationer inom en mängd industrier, bland annat inom tillgänglighet, och Sverige har en framskjuten position inom flera av dem. För mer information om talteknologin och dess utveckling, se bilaga 3: ”Nyttjande av offentligt tillgängliga svenska talresurser i en nationell talresursbank”.

¹² <http://www.gartner.com/newsroom/id/3412017>

Figur 7 - Gartners hypekurva från 2016



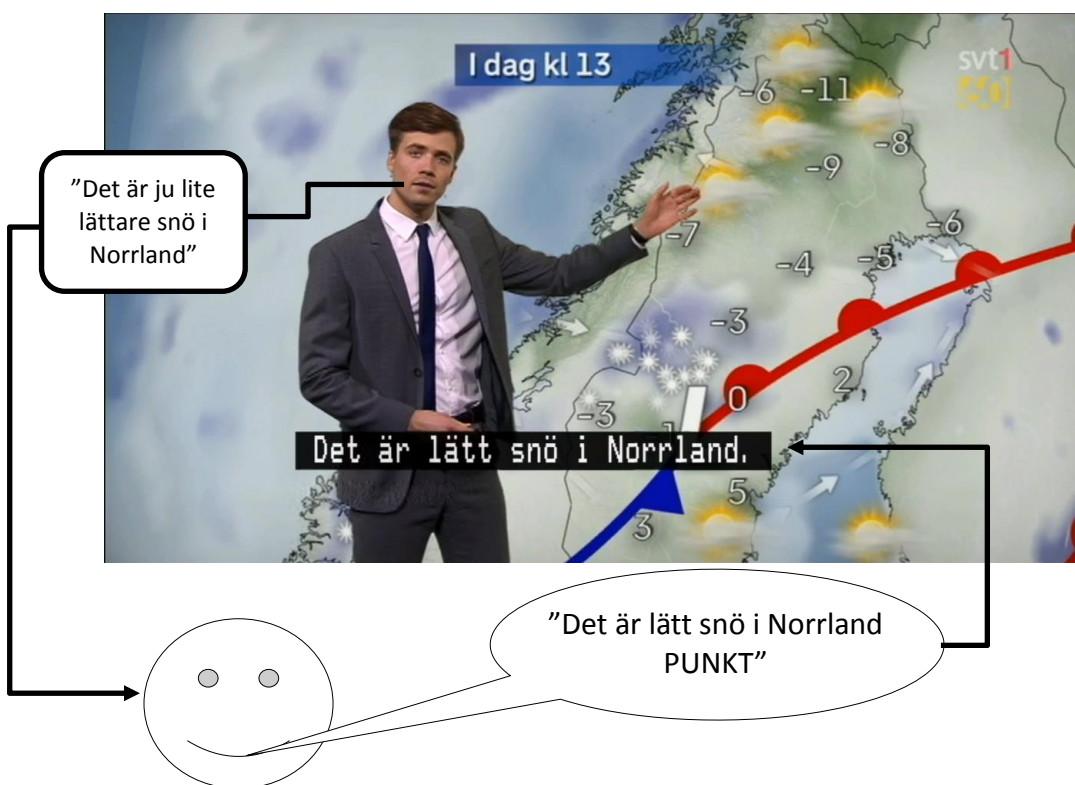
3.2 Skuggtalarteknik nyttjar taligenkänning

I ett taligenkänningssystem identifieras talat ljud och omvandlas till den mest sannolika ordföljden med hjälp av statistiska modeller. Det finns två huvudsakliga typer av taligenkänningssystem, talarberoende och talaroberoende. Är precision av stor betydelse i en tillämpning, som direkttextning av tv-program, är det mer fördelaktigt att använda ett talarberoende system som kan tränas upp efter specifika talare.

Skuggtalarteknik är ett exempel på ett sådant talarberoende taligenkänningssystem. Skuggtalarteknik används bland annat för direkttextning av tv-program. En skuggtalare sammanfattar det som sägs i tv-programmet och talar in den text hen önskar sända ut som undertexter till tv-tittarna. För illustration av skuggtalarteknikens flödesschema, se figur 8 nedan. Eftersom normalt tal sker snabbare än hastigheten för skrift, har denna teknik potential att vara mer effektiv än befintliga tekniker som används för att producera direkttexter idag. Då systemet dels är baserat på den talades röst och dels att skuggtalaren anpassar sitt röstläge till taligenkänningssystemet, kan låg fördröjning och hög träffsäkerhet erhållas. Ett taligenkänningssystem som nyttjar skuggtalarteknik kan dessutom använda den data som produceras i

systemet för att anpassas och vidareutvecklas, vilket betyder att ju mer ett taligenkänningsystem används, desto bättre kvalitet kan erhållas. Skuggtalarteknik har bland annat använts av BBC i Storbritannien sedan år 2001. Ett urval av övriga länder som i någon utsträckning använder skuggtalarteknik idag är Frankrike, Italien, Tyskland, Spanien, Österrike, Australien och Tjeckien. För mer information om skuggtalarteknik som direkttextningsmetod, se bilaga 1: ”Språkresursbank – teknik och omvärld” och bilaga 6: ”Utvärdering av SVT:s prototyp för skuggtalarteknik”.

Figur 8 – Illustration över direkttextning med skuggtalarteknik



3.3 Utvärdering av skuggtalarteknik som direkttextningsmetod

Som en av regeringsuppdragets aktiviteter har PTS upphandlat en prototyp av ett taligenkänningsystem som nyttjar skuggtalarteknik, se illustration av skuggtalarteknik i figur 8 ovan. Syftet med upphandlingen var att utvärdera prototypen för att svara på om taligenkänning, via skuggtalarteknik, är en

fungerande metod för att direkttexta tv-sändningar. Det finska språkteknologiföretaget Lingsoft vann upphandlingen och garanterade en fördröjning på maximalt 2,5 sekunder och en träffsäkerhet på minst 96 procent WER. Snittfördröjning i de uppmätta acceptanstesterna låg på 1,06 sekunder och den genomsnittliga träffsäkerheten låg på 98,23 procent, vilket var betydligt bättre resultat än vad som uppgavs i upphandlingen. Notera dock att detta inte är den totala fördröjningen för ett komplett taligenkänningsystem, utan det tillkommer tid för bland annat bearbetning. Likaså är det upphandlade taligenkänningsystemet endast en prototyp som har testats inom en begränsad genre. Vidare har SVT och dess personal aldrig använt skuggtalarteknik som direkttextningsmetod tidigare, utan SVT har sedan 1980-talet använt Velotype¹³. Velotype har dock begränsningar, bland annat tillgång till personer som behärskar tekniken. Trots att skuggtalartekniken endast har testats utifrån en begränsad prototyp ger projektet ändå en god inblick och en tydlig indikation för hur skuggtalarteknik fungerar på det svenska språket.

3.3.1 Slut användarna uppskattar den nya tekniken och anser att kort fördröjning är viktigast att prioritera

För att genomföra en utvärderingsstudie anlät PTS konsultbolaget Stelacon. Utvärderingens resultat tyder på att tittarens upplevelse av direkttextning inte kommer att förändras avsevärt om SVT väljer att implementera skuggtalartekniken. Den skillnaden som tittaren kan märka vid byte till skuggtalarteknik är att texten presenteras i ett annat flöde och tempo än tidigare samt att felen i texten blir annorlunda. Felen som uppstår med skuggtalartekniken upplevs däremot mindre problematiska än de fel som uppstår vid textning med den befintliga direkttextningsmetoden Velotype. Vid användning av skuggtalarteknik blir misstagen snarare felaktiga ord än felstavade ord, vilket är fallet med Velotype. Detta eftersom att det inte finns några felstavade ord inlagda i systemet. Med dagens teknik kan felen bli totalt oförståeliga, där det kan stå ”sleon” istället för ”solen”. Med skuggtalartekniken skulle felet istället kunna vara ”kjolen”. Användarna uppger i utvärderingsstudien att det är enklare att förstå skuggtalarteknikens fel. Dessa slutsatser är inte heller unikt för Sverige, utan liknande observationer har gjorts i andra länder.

¹³ Velotype är en teknik som bygger på snabbinskrivning av text och använder ett tangentbord där stavelser trycks ned istället för enskilda bokstäver.

Användartesterna i utvärderingsstudien indikerar också att textens fördröjning är den aspekt som har störst påverkan på användbarheten av undertexterna, detta resultat är oberoende av vilken målgrupp testpersonen tillhör. För många målgrupper är fördröjningen helt avgörande för om testpersonen använder undertexter. Givet behovet av minskad fördröjning bland alla tittargrupper så kan det eventuellt finnas ett behov av att utvärdera till vilken målgrupp undertexterna vänder sig till. Om målgruppen anses vara bredare än personer som har grav hörselnedsättning eller dövhet så bör ett övervägande göras om det finns möjlighet att göra någonting för att minska fördröjningen, exempelvis att senarelägga livesändningarna eller att förbereda mer material innan livesändningen. För personer som använder texten som stöd till ljudet så är det vitalt att ljud och text kommer i nära anslutning till varandra, annars har de ingen användning av texten. För mer information om hur användarna uppfattar skuggtalarteknik, se bilaga 6: ”Utvärdering av SVT:s prototyp för skuggtalarteknik”.

3.3.2 Skuggtalarteknik leder till ökad flexibilitet och bättre effektivitet vid direkttextning

Enligt Stelacon är den tydligaste positiva effekten vid införande av skuggtalarteknik att utbildningen av direkttextare är betydligt kortare än utbildning av personal i dagens teknik. Istället för en flerårig utbildning krävs nu endast ett par månaders upplärning innan en person kan arbeta som direkttextare. Eftersom det inte kräver lika tydliga specialkunskaper och lång utbildning kan antalet direkttextare potentiellt öka. Detta kan i sin tur leda till minskade lönekostnaderna och minskade rekryteringskostnader för programbolag. Detta medför framförallt en ökad flexibilitet för tv-bolag som snabbare kan ta in nya resurser i organisationen vid exempelvis resursbrist eller semester.

Enligt Stelacon finns det även belägg för att produktiviteten av undertexter skulle öka om ett tv-bolag bestämmer sig för att texta även förinspelade program med skuggtalartekniken. Erfarenheter från andra länder visar en ökning i produktivitet på cirka 20-30 procent. Detta framkommer från intervjuer med Lingsoft och Ericson. Mycket tyder på att detta även skulle kunna vara möjligt på den svenska marknaden. Detta innebär att en minuts tal, som med den tidigare tekniken tagit cirka 10 minuter att texta, kan textas på sju till åtta minuter med skuggtalarteknik. Det vore också möjligt att fördela

arbetet så att en mer senior kompetens ansvarar för slutbearbetning och juniorer arbetar med skuggtalningen. För mer information om effekten av skuggtalartekniken, se bilaga 6: ”Utvärdering av SVT:s prototyp för skuggtalarteknik”.

Förutom Stelacons utvärdering av prototypen har SVT själva kommenterat sina erfarenheter från taligenkänningsprojektet. SVT:s kommentarer liknar de slutsatser som dras av Stelacon, att taligenkänningstekniken inte per automatik bidrar till ökad mängd producerad text för publiken, men att utbildningstiden för taligenkänningspersonal är kortare än för andra tekniker. Enligt SVT är de största skillnaderna mot nuvarande teknik är att felen troligen blir färre, men de fel som uppkommer kan vara svåra att förstå. Vidare kommenterar SVT att den sammantagna fördröjningen troligtvis blir oförändrad, då texten visserligen genereras snabbare men att ytterligare ett steg tillkommer i sändningskedjan.

SVT anser det lämpligt att fortsätta utveckla tekniken och vill även påbörja reella textningsprojekt som visas inför publik under 2017. Tester behöver genomföras för att se vilka övriga genrer som är bäst lämpliga för taligenkänning (sport, politik, nyheter etc.). SVT anser även att övergången från Velotype till taligenkänning bör ske gradvis, likt hur flera andra länder har genomfört teknikskiftet. För att vidareutveckla skuggtalarteknik till att bli mer heltäckande och täcka fler genrer behövs talresurser och språklig infrastruktur, vilket är ett argument för att satsa på en nationell talresursbank. För mer information om SVT:s erfarenheter av skuggtalartekniken, se bilaga 7: ”Taligenkänning på SVT”.

4 PTS förslag till nationell utveckling av infrastruktur för talteknologi

I detta kapitel beskrivs de förslag som PTS har tagit fram för hur främjandet av talbaserade tjänster på den svenska marknaden bör se ut. Både med hänsyn till infrastrukturen för talresurser och med hänsyn till skuggtalarteknik som direkttextmetod för tv-sändningar.

4.1 Förslag om en långsiktig finansiering av en talresursbank som förvaltas av SND med finansiering från statligt anslag

Talteknologimarknaden utvecklas i hög takt och allt fler talbaserade tjänster tas fram, både nationellt och internationellt. Däremot sker utvecklingen av talteknologiska applikationer idag främst på kommersiellt mer attraktiva språk än svenskan idag, såsom engelska och spanska. Detta kan leda till att den tekniska innovationen inom talteknologi på svenska begränsas eller helt avstannar. För att det svenska språket ska prioriteras vid talteknologiutveckling, och att befintliga talbaserade tjänster ska bli användbara för bland annat personer med funktionsnedsättning, krävs en nationell satsning på talteknologi. Den svenska talteknologimarknaden domineras dessutom idag av en handfull kommersiella aktörer, vilket betyder att det idag råder en oligopolsituation från amerikanska bolag på den svenska marknaden. Detta kan leda till en ohälsosam konkurrenssituation med risk för prishöjningar som resulterar i att svenska nyttjare tvingas betala höga ersättningar för talbaserade tjänster som är inlåsta i befintliga system. För att undvika ett sådant scenario och om Sverige vill ta en framträdande roll som en innovativ nation med en hög utvecklingstakt av talbaserade tjänster krävs en nationell språkinfrastruktursatsning.

Det största hindret vid utveckling av talbaserade tjänster för aktörer idag är bristen på talresurser, detta på grund av att det idag råder en brist på infrastruktur för talresurser på det svenska språket. PTS anser att detta behov kan tillfredsställas vid finansiering av en öppen nationell talresursbank. KTH har, på uppdrag av PTS, tagit fram grunden till en sådan talresursbank, och PTS följer den oberoende konsultutredningens rekommendation som pekar ut SND som den framtida förvaltaren av talresursbanken. PTS instämmer även med konsultutredningens estimering av kostnad för förvaltning och drift.

Förslaget är 5-8 miljoner kronor per år i en tvåårig implementeringsfas och sedan 4-6 miljoner kronor årligen. Finansieringen bör vara i form av ett statligt anslag för att säkerställa oberoende och långsiktighet. Slutligen föreslår PTS att det införs ett forskningsråd lett av KTH och ett expertråd lett av ISOF, som tillsammans ansvarar för riktning och strategi för talresursbanken.

4.2 Förslag om användning av skuggtalarteknik som metod för direkttextning

Taligenkänning är ett fungerande alternativ för direkttextning och de erfarenheter som PTS och SVT har av regeringsuppdraget har varit mycket goda. Resultatet från prototyptesterna, där fördröjning i taligenkänningssystemet uppgick till ca 1 sekund och träffsäkerheten uppgick till ca 98 procent, tyder på att skuggtalarteknik fungerar väl på det svenska språket. Utvärderingsstudien, som genomfördes efter att prototypen var framtagen, tyder på att tittarens upplevelser av direkttextning inte påverkas markant av vilken direkttextningsteknik som har använts. Däremot är den tydligaste effekten av att implementera skuggtalarteknik att utbildningen av personal för direkttextning minskar betydligt med tekniken. Detta medför en ökad flexibilitet för tv-bolag som snabbare kan ta in nya resurser i organisationen vid resursbrist eller semesterar samt att eventuellt minska lönekostnaderna.

PTS förslag är att programbolag i Sverige bör överväga skuggtalarteknik som textningsmetod, både förinspelade och direktsända program. För att underlätta övergången från dagens tekniker till skuggtalarteknik kan det även vara rimligt för staten att stimulera fortsatt utveckling av skuggtalarteknik. Detta skulle i slutändan gynna slutanvändarnas möjligheter att få en större del av programutbudet textat. Ytterligare en fördel är att talresursbanken skulle få fler talresurser från sådana taligenkänningsprojekt. Förutsatt att programbolagen då överför skapade talresurser till talresursbanken. Vidare är PTS positiva till SVT:s intentioner att fortsätta arbetet med att utveckla skuggtalarteknik för att dels testa fler genrer och dels genomföra reella textningsprojekt som visas inför publik. Vid reella tv-sändningar bör även slutanvändare involveras, främst personer med funktionsnedsättning, för att utvärdera hur användarna upplever teknikskiftet. Slutligen instämmer PTS med SVT att språklig infrastruktur krävs för ökad användning av taligenkänningsteknik, vid exempelvis utökning till fler genrer. Detta anser PTS bör finnas i form av en nationell talresursbank.

5 Förteckning över bilagor

Bilaga 1: ”Språkresursbank – teknik och omvärld” en rapport framtagen av Stelacon

Bilaga 2: ”Nationell språkresursbank” en rapport framtagen av a-focus

Bilaga 3: ”Nyttjande av offentligt tillgängliga svenska talresurser i en nationell talresursbank” en rapport framtagen av KTH

Bilaga 4: ”Skapandet av grunden för en svensk talbank” en rapport framtagen av KTH

Bilaga 5: ”Förvaltning, finansiering och strategi för en framtida talresursbank” en rapport framtagen av BCG

Bilaga 6: ”Utvärdering av SVT:s prototyp för skuggtalarteknik” en rapport framtagen av Stelacon

Bilaga 7: ”Taligenkänning på SVT” ett PM framtaget av SVT

Bilaga 8: ”Redovisade kostnader för regeringsuppdraget” ett PM framtaget av PTS