

Utvärdering av några text-till-talomvandlare som läshjälpmedel

Malin Persson
malinp@stp.ling.uu.se

Examensarbete i datorlingvistik
Språkteknologiprogrammet
Uppsala universitet · Institutionen för Lingvistik och Filologi

17 juni 2004

Handledare:

Bertil Lyberg, Uppsala universitet
Lars Sönnebo och Jesper Klein, Talboks- och punktskriftsbiblioteket, Enskede

Sammandrag

Vid produktion av läshjälpmiddel talböcker används idag mänskliga inläsare. I en empirisk, kvalitativ utvärdering, med en liten grupp slutanvändare av talböcker, har jag undersökt huruvida syntetiskt tal kan användas som läshjälpmiddel. Text-till-talomvandlarna som utvärderades är Babils röster *AnnMarie* och *Ingmar*, BrightSpeechs röst *Emma* samt RealSpeaks röst *Ingrid*. En mänsklig röst användes till referens. Genom subjektiv skattning av längre texter och parvis jämförelse av meningar studerades lyssnarnas ansträngning, acceptans och preferens, samt text-till-talomvandlarnas taltempo och röstkvalitet. I utvärderingen har vissa av de syntetiska rösterna lika höga medelvärden som den mänskliga rösten, vilken redan används i befintliga läshjälpmiddel. Sålunda är det möjligt att använda syntetiska röster som läshjälpmiddel. I undersökning av ansträngning och acceptans är ingen röst markant bättre än de andra. *Ingmar* har jämnast talhastighet. För röstkvalitet samt preferens i den parvisa jämförelsen får *Ingrid* och den mänskliga rösten högst betyg. Ingen av rösterna i undersökningen, inte heller den mänskliga rösten, får höga betyg av testgruppen. För att det syntetiska talet ska få höga betyg av slutanvändarna behöver text-till-talomvandlarna förbättras.

Innehåll

Figurer	iv
Tabeller	v
Tack	vi
1 Introduktion	1
1.1 Inledning	1
1.2 Syfte	1
1.3 Disposition	1
2 Läs hjälpmedel	2
2.1 Producenter av läshjälpmedel	2
2.2 Talböcker	2
2.3 Talboksanvändare	3
2.3.1 Studenter med läshandikapp	3
3 Text-till-talomvandlare	5
3.1 Prosodiska företeelser som behöver beaktas av text-till-talomvandlare	5
3.2 Text-till-talomvandrares uppbyggnad	5
3.2.1 ”Text-till”: Processning av naturligt språk	6
3.2.2 ”Till-tal”: Syntes	8
3.3 Text-till-talomvandlare för svenska	9
3.3.1 Babil	9
3.3.2 BrightSpeech	10
3.3.3 RealSpeak	10
3.4 Utvärdering av text-till-talomvandlare	11
4 Metod	13
4.1 Utvärderingsmetod	13
4.2 Mål med utvärderingen	13
4.3 Utformning av utvärdering	13

4.3.1	Urval av respondenter	13
4.3.2	Enkät	14
4.3.3	Testmaterial	14
4.3.4	Röster	15
4.4	Genomförande av utvärdering	15
5	Resultat	16
5.1	Respondenterna	16
5.1.1	Personuppgifter	16
5.2	Erfarenheter av talböcker	18
5.2.1	Kontakt med talböcker	18
5.2.2	Användningsfrekvens talböcker	19
5.2.3	Talböcker som läshjälpmedel	19
5.3	Erfarenheter av syntetiskt tal	20
5.3.1	Användning av syntetiskt tal	20
5.3.2	Syntetiskt tal som läshjälpmedel	21
5.3.3	Inställning till syntetiskt tal	22
5.4	Subjektiv skattning	22
5.4.1	Testmaterial	23
5.4.2	Felaktigheter	23
5.4.3	Ansträngning	24
5.4.4	Röstkvalitet	25
5.4.5	Taltempo	25
5.4.6	Acceptans	27
5.5	Parvis jämförelse	30
5.6	Avslutande frågor	33
6	Analys och diskussion	35
6.1	Respondenterna	35
6.2	Ansträngning	35
6.3	Röstkvalitet	35
6.4	Taltempo	36
6.5	Acceptans	36
6.5.1	Tidigare erfarenheters påverkan av acceptans	36
6.6	Parvis jämförelse	37
6.7	Den mänskliga rösten	37
6.8	Kognitiv belastning	37
6.9	Text-till-talomvandlare som läshjälpmedel	38

6.9.1	Krav på text-till-talomblandaren	38
6.10	Möjliga felkällor	39
6.10.1	Testmaterial	39
6.10.2	Talsyntesröster	39
6.10.3	Respondenter	40
6.10.4	Undersökningsparametrar	40
6.10.5	Användarundersökning	40
6.10.6	Svårtolkade kommentarer	40
6.11	Vidareutveckling	40
7	Egna observationer	42
7.1	Befintliga talsyntesapplikationers möjligheter	42
7.2	Sådant som befintliga talsyntesapplikationer inte kan hantera	42
	Litteraturförteckning	45
A	Intervjufrågor	47
B	Svar och kommentarer från respondenterna	49
C	Meningar som användes vid den parvisa jämförelsen	57
D	Kommentarer till den parvisa jämförelsen	60

Figurer

3.1	Innehåll i en text-till-talomvandlare	6
5.1	Åldersfördelning	17
5.2	Har du använt talböcker? (Fråga 5)	18
5.3	Är talböcker till stor hjälp för dig när du läser? (Fråga 7)	19
5.4	Har du använt dig av syntetiskt tal? (Fråga 8a)	20
5.5	Hur ofta har du använt syntetiskt tal för läsning av kortare texter? (Fråga 8b)	20
5.6	Hur ofta har du använt syntetiskt tal för läsning av längre texter?(Fråga 8c)	21
5.7	Är syntetiskt tal till stor hjälp för dig när du läser? (Fråga 9)	21
5.8	Är du positivt inställd till syntetiskt tal? (Fråga 10)	22
5.9	Upplevde du felaktigheterna som störande? (Fråga 12c)	24
5.10	Behövde du anstränga dig för att förstå vad som sades? (Fråga 13)	24
5.11	Var rösten behaglig? (Fråga 14)	25
5.12	Var talhastigheten lagom snabb, snabbare eller långsammare än vad du föredrar? (Fråga 15 a)	26
5.13	Var talhastigheten lagom jämn? (Fråga 15 b)	26
5.14	Skulle du kunna lyssna på längre texter som spelats in med denna typ av röst? (Fråga 16 a) 27	
5.15	Skulle du kunna använda en DAISY-talbok som spelats in med denna röst? (Fråga 16 b) .	27
5.16	Skulle du kunna använda en DAISY-talbok som spelats in med denna röst, om DAISY-talboken, förutom det inspelade talet, innehöll hela den tryckta bokens text i form av e-text (elektronisk text) som är kopplad till talet? (Fråga 16 c)	28
5.17	Jämförelse mellan svar på frågorna 16 a, 16 b och 16 c för hela testgruppen.	29
5.18	Jämförelse mellan svar på frågorna 16a, 16b och 16 c för respondenter med och utan erfarenhet av regelbunden användning av syntetiskt tal.	30
5.19	Resultatet av den parvisa jämförelsen	31
5.20	Jämförelse av medelpoäng för röster med och utan användarlexikon.	31
5.21	Resultatet av den parvisa jämförelsen sorterat efter respondenternas kön	32
5.22	Resultatet av den parvisa jämförelsen sorterat efter respondenternas modersmål	32
5.23	Resultatet av den parvisa jämförelsen sorterat efter respondenternas tidigare erfarenheter av syntetiskt tal som läshjälpmedel.	33

Tabeller

2.1	Den procentuella fördelningen av producerade titlar på TPB Högskoleservice år 2002 . . .	2
2.2	Utlåning av läshjälpmedel till högskolestuderande år 2002	3
2.3	Studenter med läshandikapp 1 januari - 31 december 2003.	3
2.4	Fördelning av studenter med läshandikapp på utbildningsområden, år 2002.	4
3.1	Standardinställningar för rösterna AnnMarie och Ingmar i Infovox Desktop version 1.1 .	10
3.2	Standardinställningar för rösten Emma i BrightSpeech Studio version 1.1	11
3.3	Standardinställning för rösten Ingrid i Voxit Macaw version 2.0	11
5.1	Läshandikapp och kön	16
5.2	Läshandikapp och modersmål	17
5.3	Ämnesstudier	18
5.4	Omvandling av enkätens svarsalternativ till siffervärden, för frågorna 12c, 13, 15 och 16.	22
5.5	Omvandling av enkätens svarsalternativ till siffervärden, för fråga 14.	23

Tack

Tack till min handledare Bertil Lyberg som har hållt mig på rätt köl, hjälpt mig att avgränsa arbetet och delat med sig av sina erfarenheter. Tack till alla på TPB som har hjälpt till, mina handledare Lars Sönnebo och Jesper Klein, Catarina Hummelstrand som var min kontakt till slutanvändarna, Berit Winblad för statistikuppgifter och Tomas Johansson som fixade testmaterial i e-textformat, för att nämna några.

Slutligen stort tack till Sara Hallbergsson, Jenny Rahbek och Tina Ohlander som läst igenom uppsatsen och gett stöd, uppmuntran och många goda råd.

1 Introduktion

1.1 Inledning

I Sverige finns det många läshandikappade som är beroende av någon form av hjälpmedel för att kunna läsa. Lëshjälpmedel finns till exempel som elektronisk text, punktskrift eller talböcker. Talboks- och punktskriftsbiblioteket (TPB) har ansvar för att högskolestuderande med läshandikapp får sin kurslitteratur i ett för dem anpassat medium. Vid produktion av talböcker används idag mänskliga inläsare, men TPB vill veta om det går att använda syntetiskt tal som läshjälpmedel, istället för dessa mänskliga inläsare. Syntetiskt tal används redan idag som läshjälpmedel i olika applikationer, till exempel uppläsning av hemsidor. Syntetiskt tal kan produceras snabbare, och i längden billigare, än tal som produceras av en människa. Men är det syntetiska talet så pass bra att talboksanvändare kan tänka sig att använda det som läshjälpmedel?

1.2 Syfte

Syftet med examensarbetet är att undersöka om det går att använda syntetiskt tal från befintliga text-till-talomvandlingsprodukter för svenska som läshjälpmedel. Undersökningen görs genom att låta läshandikappade studenter på högskolenivå, som använder talböcker som läshjälpmedel, utvärdera några text-till-talomvandlerare. I utvärderingen undersöks de läshandikappades acceptans av syntetiskt tal, d.v.s om de är välvilligt inställda till att använda syntetiskt tal. Föredrar de en typ av talsyntes framför andra? Utifrån det textmaterial som används i utvärderingen med de läshandikappade studenterna görs även en inventering av vad talsynteserna i undersökningen kan och bör klara av språkligt för att fungera som läshjälpmedel för kurslitteratur. Utvärderingen fokuserar på användarfrågor, och inte på text-till-talomvandlerares tekniska förutsättningar. Uppdragsgivare för undersökningen är TPB.

1.3 Disposition

I bakgrundskapitlen ingår en beskrivning av talböcker och talboksanvändare (kapitel 2), hur talomvandlerare generellt sett är uppbyggda, samt vilka svenska talomvandlerare som ingår i undersökningen (kapitel 3). I kapitel 4 beskrivs valet av utvärderingsmetod, samt genomförandet av utvärderingen. Resultatet av utvärderingen med talboksanvändarna redovisas i kapitel 5, därefter analyseras och diskuteras resultaten i kapitel 6. I kapitel 7 redovisas inventeringen av talsyntesernas språkliga prestanda.

2 Lëshjälpmedel

I detta kapitel beskrivs vilka som är ansvariga för lëshjälpmedlet talböcker, vad talböcker är och vilka som använder dem.

2.1 Producenter av lëshjälpmedel

Talboks- och punktskriftsbiblioteket (TPB) är en statlig myndighet med ansvar att ge människor med lëshandikapp tillgång till litteratur i form av exempelvis talböcker, punktskrift eller elektronisk text (e-text). TPB har ansvar för att högskolestuderande som är lëshandikappade får sin kurslitteratur på ett för dem anpassat medium. Kurslitteratur för högskolestuderande görs idag endast som DAISY-talböcker på cd-rom (se avsnitt 2.2). Produktionen av kurslitteraturen finansieras av Utbildningsdepartementet. Lagen om lika behandling av studenter i högskolan innebär bland annat att lëshandikappade studenter skall ha samma service i litteraturfrågor på högskole- och universitetsbibliotek som andra studenter (*Talboks- och Punktskriftsbibliotekets hemsida 2003*).

I tabell 2.1 kan man se den procentuella fördelningen av de lëshjälpmedel som TPB Högskoleservice producerade år 2002, vilket visar att den största delen av DAISY-talböckerna som producerades var för svensk litteratur. Antalet utlån av lëshjälpmedel till högskolestuderande år 2002 finns dokumenterat i tabell 2.2.

Tabell 2.1: Den procentuella fördelningen av producerade titlar på TPB Högskoleservice år 2002

Lëshjälpmedel	Fördelning i procent	
	svensk litt.	utländsk litt.
DAISY-böcker	65%	35%
E-textböcker	59%	41%
Punktskrift	60%	40%
Förstoringar	73%	27%

Enligt dessa uppgifter utgör utlånen av DAISY-talböcker ca 72% av det totala utlånet till lëshandikappade studenter. Informationen är hämtad från TPB:s biblioteksdatasystem Elibri samt TPB:s Accessdatabas ”Studentregister 2002” och ”Studentregister 2003”.

2.2 Talböcker

En talbok är en bok som är inläst av en människa. Det finns både analoga och digitala talböcker. De analoga talböckerna är lagrade på kassetter. De digitala talböckerna har DAISY-format (DAISY står för

Tabell 2.2: Utlåning av läshjälpmedel till högskolestuderande år 2002

Läshjälpmedel	Antal utlån	Procent
Talböcker, analoga	3 636	24,5%
DAISY-talböcker	10 638	71,8%
E-text	383	2,6%
Punkt	13	0,1%
Förstoringer	141	1,0%
Totalt	14 811	100,0%

Digitalt Anpassat InformationsSystem), som är strukturerat så att det till exempel går att navigera mellan kapitel, hoppa direkt till rätt avsnitt, sätta ut egna bokmärken eller söka efter sidnummer i talboken. Det krävs en särskild hårdvara för att kunna spela en DAISY-talbok, till exempel VICTOR som är en fristående cd-romspelare. Spelarna har ett antal funktioner som underlättar talboksläsningen, så användaren kan utnyttja de fördelar som struktureringen av texten innebär och spela upp talet snabbare. Med ett särskilt läsprogram, t.ex. Playback 2000, kan man spela DAISY-talböcker i en cd-romspelare som är ansluten till en dator. Med ett sådant läsprogram har talboksanvändare tillgång till fler funktioner. Det är till exempel möjligt att söka efter ord i texten, skriva egna anteckningar, läsa in röstanteckningar, samt markera viktiga delar av talboken och bara lyssna till de delar användaren själv markerat.

2.3 Talboksanvändare

I avtal mellan staten och Sveriges Författarförbund (1977-02-21) finns uppräknat de grupper av personer som har rätt att låna talböcker: synskadade, rörelsehindrade, personer med läs- och skrivsvårigheter, personer med afasi, personer med utvecklingsstörning, hörselskadade (för hörselträning), långtidssjuka (även psykiskt sjuka) och konvalescenter. Däremot har invandrare, som inte har någon av de nämnda svårigheterna, inte rätt att låna talböcker (*Talboks- och Punktskriftsbibliotekets hemsida 2003*). De talböcker som finns på vanliga bibliotek är dock tillgängliga för alla.

2.3.1 Studenter med läshandikapp

Studenter på universitetsnivå med någon form av läshandikapp har rätt till talböcker. I tabell 2.3 visas antalet studenter med läshandikapp som är registrerade hos TPB Högskoleservice 1 januari - 31 december 2003. För att få låna läshjälpmedel av TPB Högskoleservice så krävs det att man är registrerad. Under år

Tabell 2.3: Studenter med läshandikapp 1 januari - 31 december 2003.

Läshandikapp	Antal	Procent
Synskadade	136	12,0%
Rörelsehindrade	57	5,0%
Dyslektiker	912	80,4%
Övriga (reumatiker m.fl.)	30	2,6%
Totalt	1 135	100%

2003 ökade studentantalet med 29%, jämfört med tidigare år. Tabell 2.4 visar fördelningen av studenter

med läshandikapp som anser sig behöva läshjälpmedel på utbildningsområdena för år 2002. Talböcker läses in av människor, men skulle kunna läsas in med syntetiskt tal. Hur det fungerar tar jag upp i nästa kapitel.

Tabell 2.4: Fördelning av studenter med läshandikapp på utbildningsområden, år 2002.

Utbildningsområde	Antal studerande	Procent
Humanistiskt	84	9.5%
Juridiskt	30	3.4%
Media	5	0.6%
Medicinskt	15	1.7%
Musik	1	0.1%
Naturvetenskapligt	31	3.5%
Samhällsvetenskapligt	301	34.2%
Tekniskt	93	10.6%
Teologiskt	13	1.5%
Undervisning	170	19.3%
Vård	91	10.3%
Övrigt	46	5.2%
Totalt	880	100%

3 Text-till-talomvandlare

Text-till-talomvandlare är program som omvandlar godtycklig maskinläsbar text till utdata i form av syntetiskt tal. Ibland används förkortningen TTS-system som står för det engelska "text-to-speech system". Syntetiskt tal används bland annat för att hantera information som ständigt varierar, såsom e-brev (Gibbon et al. 1997). I detta kapitel beskrivs prosodi, varför prosodi bör beaktas när text omvandlas till tal, hur talomvandlare är uppbyggda, vilka kommersiella svenska talsynteser som finns samt hur talomvandlare kan utvärderas.

3.1 Prosodiska företeelser som behöver beaktas av text-till-talomvandlare

Prosodi ger struktur åt det talade språket så att det blir mer än en samling språkljud. Prosodi anger grammatiska funktioner, samhörighet eller gränser mellan ord och ordled, känsloläggningar samt skiljer olikbetydande ord åt. Det fungerar på så vis underlättande för kommunikationen för både talare och lyssnare. För att få syntetiskt tal naturligt och begripligt måste man kunna styra prosodin (Bruce 1998).

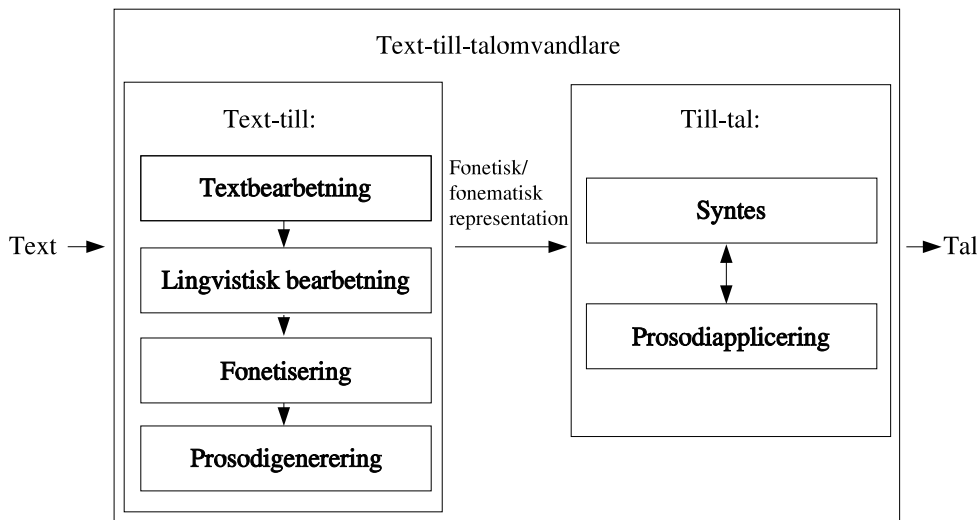
Prosodin ges av talets rytmiska, melodiska och dynamiska egenskaper. De rytmiska egenskaperna realiseras med temporala förhållanden, t.ex. kvantitet och final förlängning. I svenska är kvantitet begränsat till betonad stavelse (Bruce 1998). Exempel från svenskan är *här* respektive *herr*, samt *Polen* respektive *pollen*. Tryckstyrka kan läggas på ord och stavelser. Jämför till exempel *kör på kanten* med *kör på kanten*. De melodiska egenskaperna i talet visas genom tonhöjdsförhållanden, d.v.s. tonhöjdens förlopp över tiden. Det används för att visa ord- och satsaccent (Elert 2000). Exempelvis fokal accent, en typ av satsaccent, används när något nytt introduceras, och det markeras med både grundtonsförhållande och starkare tryck. Satsaccent av denna typ är svår att automatiskt identifiera i skriven text.

Svenska har de ordprosodiska egenskaperna betoning, kvantitet och ordaccent (Bruce 1998). Betoning av stavelser görs genom relativt hög tryckaccent och/eller tonhöjd, ofta med åtföljande förlängning. Skillnaden mellan olika ordaccenter realiseras i första hand genom tonhöjdsförloppet (*Nationalencyklopedin: ett uppslagsverk på vetenskaplig grund utarbetat på initiativ av Statens kulturråd 1989-*).

För att fånga de prosodiska företeelserna i en text så behöver texten analyseras lingvistiskt. Exempelvis så måste heterofona homografer, d.v.s. ord som stavas lika men uttalas olika, analyseras morfologiskt och ibland kontextuellt, syntaktiskt eller semantiskt. Genom sådan analys kan lämpligt uttal identifieras vilket gör det möjligt för lyssnaren att förstå vilket ord det är fråga om.

3.2 Text-till-talomvandrares uppbyggnad

Talet som kommer ut ur text-till-talomvandlare är maskinellt producerat. Det kan skapas på olika sätt, men text-till-talomvandlare består vanligen av två delar: "text-till" och "till-tal", se figur 3.1 (Dutoit 1997). I den första delen, "text-till", processas det naturliga språket. Där ingår omvandling av texten



Figur 3.1: Innehåll i en text-till-talomvandlare

till grafem (bokstäver), lingvistisk bearbetning för att fånga prosodiska företeelser i texten, en fonetisk modul där grafemen görs om till någon form av fonetisk representationen, samt en modul där information om prosodi läggs på den fonetiska representationen. I ”till-tal”-delen görs den fonetiska representationen om till tal genom syntetisering och applicering av prosodi. De ingående delarna är inte obligatoriska för en talsyntes, varje steg kan utföras mer eller mindre fullständigt och med varierande metoder. I detta avsnitt ges en generell beskrivning av hur text-till-talomvandlare är uppbyggda, baserat på (Dutoit 1997).

3.2.1 ”Text-till”: Processning av naturligt språk

I systemets första del översätts indata, i form av godtycklig mängd text, till lingvistisk kod. Vissa talomvandlare kan hantera indata i form av fonetisk eller fonematisk representation, som visar uttalet, exempelvis IPA (International Phonetic Association) eller SAMPA (Speech Assessment Methods Phonetic Alphabet), som är ett datorläsbart fonetiskt alfabet.

Textbearbetning

I textbearbetningen identifieras alla ord och meningsslut genom teckenläsning. För detta krävs språk-specifika regler som omskrivning från tecken till motsvarande bokstäver. ASCII-tecken segmenteras till ord separerade av blanktecken. Meningsslut identifieras med sådana tecken som används för att markera meningsslut: ”.”, ”;”, ”!”, ”?” och ”?”. Tecknen disambigueras, om så krävs, och meningsslut markeras. Förkortningar skrivs ut, för att inte förväxla dem med meningsskiljare. Specialtecken, fysiska enheter och tecken för bland annat valuta disambigueras och skrivs ut. Akronymen som innehåller punkt behöver också hanteras i textbearbetningen, för att inte förväxlas med meningsskiljare, och skrivs ut. Vanligtvis räcker sekvensiell läsning av varje tecken separat, exempelvis TPB, CSN och BMW, men ibland måste någon form av uttalsregler ta hand om de akronymen som inte läses sekvensiellt, exempelvis NATO och GATT. I exemplet ovan med akronymen som läses sekvensiellt har de två första finalbetoning, medan det sista exemplet har icke-final betoning. För att få rätt betoning krävs mer än teckenläsning. Siffror identifieras och tolkas som datum, klockslag, telefonnummer, ordningstal m.m. utifrån dess kontext, och skrivs sedan ut.

Lingvistisk bearbetning

Det behövs någon form av lingvistisk analys för att kunna identifiera en texts prosodiska ledtrådar, exempelvis fastställa grundton och duration m.m. Detta avsnitt tar upp några prosodiska företeelser, samt vilka analyser som bör göras för att fånga upp de prosodiska ledtrådarna. Denna lingvistiska bearbetning kan göras med hjälp av lexikongångar, lexikonregler, parsning, ordklasstaggning, dolda Markovmodeller (HMM:er), neurala nätverk, träna talomvandlaren med korpusmaterial m.m. I denna uppsats kommer dessa olika metoder inte att beskrivas närmare, för mer information om dem, se Dutoit (1997).

Den lingvistiska bearbetningen består av morfologisk, kontextuell och syntaktisk analys, och utifrån dessa analyser identifieras prosodiska ledtrådar.

Morfologisk analys I en morfologisk analys tilldelas alla individuella ord alla dess möjliga ordklasser, och dess ingående delar. Innehållsord kan p.g.a. det kontinuerliga nyskapandet inte täckas av ett lexikon. I stället krävs någon form av analys, för att identifiera vilka morfem som orden består av, och därigenom identifiera ordens uttal. En morfologisk analys kan innebära att man identifierar böjningsmorfem, avledningsmorfem och sammansättningar för att nå fram till ordets stam. Böjningsmorfem läggs som affix till ett ords stam och ger ordet värde för exempelvis genus, numerus, tempus, modus, kasus, species. Exempelvis kan substantivet *tomten* antingen vara bestämd form av lemmat *tomt* eller lemmat *tomte*. Beroende på vilket lemma det är fråga om så varierar ordaccenten, lemmat *tomt* har akut accent medan lemmat *tomte* har grav accent. Sådana flerformtydiga graford inom samma ordklass kallas interna homografer. Ett annat exempel på intern homografi är substantivet *planet* som kan vara bestämd form av lemmat *plan* eller bestämd form av lemmat *planet*. Båda har akut accent men *plan* har betoning på första stavelsen och *planet* har betoning på andra stavelsen .

Sammansättningar består av två eller flera lexem som sätts samman till ett nytt lexem. Sammansättningar är mycket vanliga i svenska. Det är viktigt att veta var junkturen, d.v.s. initialgränserna för rotmorfemen, sitter i sammansatta ord (McAllister 2000), eftersom de kan påverka uttal, ordaccent och betoning. Till exempel kan *bildrulle* vara en sammansättning av *bil+drulle* eller *bild+rulle*. För att disambiguera sammansättningen behövs morfologisk och semantisk analys. Listan med möjliga ordklassstillhörigheter kan sedan reduceras med hänsyn till omgivande ords ordklassstillhörighet, i den kontextuella analysen.

Kontextuell analys Lexikal flertydighet i indata kan hanteras genom analys av ett ords kontext. Givet de omgivande ordens ordklassstillhörighet kan listan med ett ords möjliga ordklassstillhörigheter, som gavs i den morfologiska analysen, reduceras.

Syntaktisk analys En syntaktisk analys strukturerar texten i fras- och satsliknande konstituenten. Detta behövs för att fånga vissa av de prosodiska företeelser som finns i texten. Homografer, ord som stavas likadant men som har olika betydelse, kräver disambiguering om uttalet varierar beroende på lemmat. Genom att identifiera omgivande ords ordklass- eller frastillhörighet kan lämpligt lemma hittas. Exempelvis så kan de interna homograferna *planet* och *tomten* disambigueras genom identifiering av eventuella framförställda determinerare inom samma fras.

Omvandling till fonetisk representation

Att automatiskt omvandla text till fonetisk representation är inte trivialt. Det finns två strategier för att omvandla bokstäver till ljud, lexikonbaserad fonetisering och regelbaserad fonetisering. Dessa kan användas var för sig eller i en kombination. Den lexikonbaserade fonetiseringen lagrar så mycket fonologisk kunskap som möjligt i lexikon, till exempel morfem samt fonologiska regler för hur morfemen får avledas, böjas och sammansätts till ord. De regelbaserade systemen använder grafem-till-fonemregler. Ord

med avvikande uttal kan hanteras med lexikon eller särskilda undantagsregler. Beroende på böjning eller avledning av ett ord varierar uttal och placering av betoning, ex. *arbeta* - *bearbeta*.

Beroende på vilken syntesmetod som tillämpas för talomvandlingen så används olika uttalsrepresentationer. Vid formant- och artikulationsbaserad syntes (se avsnitt 3.2.2) krävs fonetisk representation, som återger uttal så exakt som möjligt. I konkateneringsbaserad syntes (se avsnitt 3.2.2), som använder naturligt tal, räcker det med fonematisk representation, som inte beaktar allofonskillnader, eftersom det naturliga talet som används i syntesen innehåller dessa varianter naturligt.

Prosodisk representation

Efter omvandling till fonetisk eller fonematisk representation läggs information om prosodiska parametrar, som den lingvistiska analysen genererat, på den fonetiska transkriptionen. Information om bland annat rytm och intonation läggs in i textens fonetiska representation.

3.2.2 ”Till-tal”: Syntes

I den andra delen av text-till-talomvandlingen omvandlas den fonetiska representationen till tal genom digital signalbehandling och eventuellt med tidigare kodad taldata. Omvandlingen från fonetisk representation till tal görs genom någon form av syntes.

Syntesmetoder

Nedan beskrivs de tre syntesmetoderna formantsyntes, artikulationssyntes och konkateneringssyntes. Både forskning och applikationer inom talsyntes idag är fokuserat på främst syntesmetoder som är baserade på konkatenering av talenheter. Forskningen fokuserar på hur man på effektivt sätt väljer ut talenheter och konkatenerar dem, samt hur man skapar optimala taldatabaser (Carlson, Sigvardson och Sjölander 2002).

Formantbaserad talsyntes I formantbaserad talsyntes försöker man efterlikna spektrum hos signalen som kommer från munnen vid tal. Det syntetiska talet skapas genom manipulering av akustiska parametrar, t.ex. formantfrekvens, grundtonskurva m.m. Formantbaserad syntes förenklar den naturliga talapparatusens akustiska möjligheter. Parametrarna i modellen ger inte samma flexibilitet som hos människan för att producera akustiska signaler. Val av modellparametrar och särskilt optimering av dessa är ett problem som måste lösas i formantbaserad talsyntes. Det är svårt att hitta styrparametrar för optimering som inte är ad hoc-lösningar. Ingen formantbaserad talsyntes är med i den här undersökningen.

Artikulationsbaserad talsyntes Artikulationsbaserad talsyntes innebär att man genom simulering med elektriska nät bygger upp en modell av den mänskliga talapparaten. Man använder en mängd virtuella rörliga artikulatorer (tungan, munnen, läpparna, gommen och struphuvudet) som kan manipuleras oberoende från varandra. Artikulationsbaserad talsyntes liknar formantbaserad syntes på så sätt att modellen skapar artificiellt tal genom att manipulera parametrar med hjälp av regler, och liksom för formantbaserad syntes är det problem med val av parametrar och särskilt med optimering av dessa. De artikulationsbaserade modellerna är bara förenklingar av naturens original och är inte så framgångsrika pga att de är okompleta och samtidigt dyra (Carlson 1993). I den här undersökningen finns ingen artikulationsbaserad talsyntes med.

Konkateneringsbaserad talsyntes I konkateneringsbaserad talsyntes sätts delar av mänskligt tal ihop till större, godtyckliga, talsekvenser genom att återanvända inspelade talenheter. Huvuduppgiften i konkateneringssystemet är att hantera det inspelade talet, och att välja rätt talenhetssekvens för en given sekvens av fonembeskrivningar. Datainsamling görs genom inspelning av textmaterial. För att få homogent tal krävs att inläsningen görs av en och samma person. Det är viktigt att välja rätt person som talar in materialet till taldatabasen. Det blir generellt sett bättre resultat med en professionell uppläsare som talar jämnt, en mansröst blir oftast bättre än en kvinnoröst (Black och Campbell 1995).

Talet kan antingen lagras i en databas som talenheter eller som en ljudfil med löpande naturligt tal. Storleken på talenheterna är viktig. Ju större talenheterna är i taldatabasen, desto mer lagringsutrymme behövs. Det är vanligt att använda difoner som talenheter. En difon är uppbyggd av två halva foner som gränsar till varandra. Klippet ligger mitt i fonerna, där påverkan från intilliggande foner är som minst. Exempelvis så ger "tut" fyra difoner: *tystnad-tt-ulu-tt-tystnad*. (- anger gräns mellan foner; | anger klipp mitt i fonen). *Unit selection* är en form av konkatenering där taldatabasen består av en ljudfil med uppmärkt naturligt tal. Systemet väljer ut så långa talenheter som möjligt från ljudfilen. Talenheterna i databasen kan vara märkta med etiketter som har en vektor med särdragsvärden för exempelvis särdragen duration, position i fras och riktning på tonhöjd.

Ett av de största problemen med konkateneringen är att sätta ihop talenheterna på ett naturlig sätt. Regler styr hur talenheterna väljas ut ur ljudmaterial och hur enheterna sedan konkateneras. Efter konkateneringen modifieras talsekvensen så att gränserna mellan talenheterna görs så mjuka som möjligt, och spektra för de intilliggande talsekvenserna matchar varandra. Ju mer lika de intilliggande talsekvenserna är desto mer liknar utdata naturligt tal. Ju mindre talenheterna är desto fler enhetsgränser måste hanteras. Signalbehandling kan göras med exempelvis PSOLA (Pitch-Synchronous Overlap and Add), som är en metod för att manipulera tonhöjd och duration hos en akustisk talsignal (Black och Campbell 1995).

Prosodiapplicering

Efter syntes appliceras prosodi på talet utifrån den information om prosodi som genererats av den språkliga bearbetningen. Prosodiska regler appliceras på talsekvensen för att ge rätt intonation och ge talet flyt. I vissa fall är prosodiappliceringen integrerad i syntesen.

3.3 Text-till-talomvandlare för svenska

TPB var intresserade av att veta hur bra befintliga, icke webbaserade text-till-talomvandlare. De talsynteser som kommit ifråga för den här undersökningen är *Babil* och *BrightSpeech*, som tillhandahålls av företaget *BaBel Infovox AB* (fr.o.m. våren 2004 *Acapela Group*), *RealSpeak* som distribueras av företaget *ScanSoft*, samt företaget *Loquendo* som har en talsyntesprodukt för svenska. *Loquendo* utgick ur undersökningen då de inte levererade någon demo i rimlig tid. Övriga produkter som användes i undersökningen beskrivs nedan i generella drag.

3.3.1 Babil

Babil har en kvinnlig och en manlig röst för svenska, *AnnMarie* och *Ingmar*. Syntesmetoden är konkatenerande av difoner, baserad på MBROLA (Multi Band Resynthesis Overlap Add), som är en algoritm för konkatenering. Endast en instans av varje difon är lagrad i databasen och signalbehandling görs med MBROLA för att modifiera duration och tonhöjdskurvorna (*Babel Infovox AB hemsida 2004*). Lexikonet är regelbaserat.

Mjukvaran för *Babil*, som användes i undersökningen, var Infovox Desktop version 1.1 av BaBel Infovox AB 2002. Indata kan vara ren text eller innehålla Microsoft SAPI-taggar (Speech Application Programming Interface). Med Infovox Desktop version 1.1 kan följande manipuleras: läsinställning (bokstav för bokstav alternativt mening för mening), talhastighet, volym, pauslängd för interpunktion (kan göras kort/lång), tonhöjd, dynamik, frekvenskorrigering (equalizer) samt samplingfrekvens. Standardinställningarna för *AnnMarie* och *Ingmar* visas i tabell 3.1. Talhastighet 100% motsvarar 180-190 ord/minut. Tonhöjd 100% motsvarar för män 100-110 Hz, och för kvinnor 180-190 Hz. I Infovox Desktop ver-

Tabell 3.1: Standardinställningar för rösterna AnnMarie och Ingmar i Infovox Desktop version 1.1

Infovox Desktop version 1.1		
Röst	AnnMarie	Ingmar
Talhastighet	67%	67%
Volym	100%	100%
Tonhöjd	72	50
Dynamik	47	32
Pauslängd	lång	lång
Samplingfrekvens	22 kHz	22 kHz
Frekvenskorrigering (equalizer)	-	-
Läsinställning	mening	mening
Användarlexikon	-	-

sion 1.1 finns ett användarlexikon i vilket man kan lägga in en-ordningar i fullformer. Notationen för det fonetiska alfabetet i användarlexikonet heter RULESYS (*BabTTS User Guide* 2003).

3.3.2 BrightSpeech

BrightSpeech har en kvinnlig röst för svenska, *Emma*. Syntesen är konkateneringsbaserad, med unit selection av talenheter som är minst en difon långa (polyfoner).

BrightSpeechs huvudsakliga lexikon är uppbyggt genom träning på en transkriberad korpus. Dessutom finns det ett litet, specifikt lexikon (okänt innehåll) och ett namnlexikon. Notationen för det fonetiska alfabetet i BrightSpeech är en icke-standardiserad variant av SAMPA, som är ett datorläsbart fonetiskt alfabet.

Applikationen för BrightSpeech heter BrightSpeech Studio. I BrightSpeech Studio version 1.1, som används i undersökningen är följande manipulerbart: röst, talhastighet, pauslängd för interpunktion(.,:;!?), frekvenskorrigering (equaliser) samt samplingsfrekvens. Standardinställningar för Emma visas i tabell 3.2. I BrightSpeech Studio version 1.1 finns ett användarlexikon i vilket man kan lägga in en-ordningar i fullformer (*BabTTS User Guide* 2003).

3.3.3 RealSpeak

Talsyntesen RealSpeak har en kvinnlig röst för svenska som heter *Ingrid*. Syntesmetoden är konkateneringsbaserad och taldatabasen består av difoner. Difondatabasen innehåller hundratals instanser för varje difon, d.v.s. taldatabasen är mycket stor. Den språkliga processningen och syntesen är korpusbaserade (Lindh 2001). Det har varit svårt att få information om RealSpeak från ScanSoft, vilket gjort att jag fått använda mig av en sekundär källa.

Tabell 3.2: Standardinställningar för rösten Emma i BrightSpeech Studio version 1.1

BrightSpeech Studio version 1.1	
Röst	Emma
Talshastighet	100%
Pauslängd	kortast möjliga
Samplingfrekvens	22 kHz
Frekvenskorrigering (equaliser)	-
Användarlexikon	-

I undersökningen användes applikationen Voxit Macaw version 2.0 som är baserat på ScanSofts RealSpeak version 2.1. Följande inställningar kan manipuleras med Voxit Macaw version 2.0: röst, läsinställning (stava texten bokstav för bokstav under uppläsning, eller läsa hela meningar), volym, talshastighet samt användarlexikon. I användarlexikonet läggs ord in i fullform. Standardinställningar för *Ingrid* visas i tabell 3.3.

Tabell 3.3: Standardinställning för rösten Ingrid i Voxit Macaw version 2.0

Voxit Macaw version 2.0	
Röst	Ingrid
Talshastighet	5 (av 9)
Volym	8 (av 9)
Läsinställning	mening
Användarlexikon	MLSwedish1053.bdc

3.4 Utvärdering av text-till-talombvandlare

Det finns en rad olika sätt att utvärdera talsyntessystem. Man använder sig av olika metoder beroende på om utvärderaren känner vill vad systemet innehåller eller ej. Om man utvärderar ett systems kapacitet eller prestanda, utan att se till de ingående modulernas respektive kapacitet kallas det för en black box-utvärdering. Denna form av utvärdering används vanligen när man inte vet vilka de ingående modulerna är eller hur dessa ser ut och interagerar med varandra. En glass box-utvärdering kan man göra när man vet vilka de ingående delarna är i ett system, och kan mäta respektive moduls kapacitet (Gibbon, Moore och Winski 1997). Utvärdering med mänskliga testare kan exempelvis göras med slutanvändarna av ett system. Nackdelen med tester med mänskliga lyssnare är att de inte är konsekventa vad gäller bedömning eller prestation. Det är även tidskrävande och kostsamt att utföra tester med mänskliga testare. Automatiserade tester kan användas för att mäta avvikelser mellan talsystems utdata och naturligt tal, mätt i akustiska termer. Nackdelen är att man inte vet vad som gör att en lyssnare urskiljer skillnader mellan två olika realisationer av samma meddelande, vilket gör det svårt att avgöra vilka parametrar i den automatiserade testningen som man ska observera. Vid bedömande utvärdering får en grupp lyssnare ge sitt omdöme och bedöma utdata från talsyntes i ett antal skalor som rankas.

Det finns ingen vedertagen metod för att utvärdera användarnas acceptans av syntetiskt tal. I (Gibbon et al. 1997) beskrivs många olika metoder för utvärdering av system som är svarta lådor, för utvärdering av längre talenheter beskrivs endast två testmetoder: *SAM Overall Quality Test* och *ITU-T Overall Quality Test*. De kan användas för att göra en jämförande utvärdering av totala kvaliteten för längre talenheter.

Fokus i dessa tester är acceptabilitet, förståelighet och naturlighet. Man kan även studera vilken kognitiv belastning syntetiskt tal innebär, jämfört med mänskligt tal. Mer om detta står i avsnitt 6.8.

4 Metod

I detta kapitel beskrivs val av utvärderingsmetod och respondenter samt det praktiska vetenskapliga arbetet, d.v.s. insamling och bearbetning av data med mera.

4.1 Utvärderingsmetod

Jag har valt att göra en empirisk, kvalitativ utvärdering för att ta reda på om syntetiska tal kan användas som läshjälpmedel. Utvärderingen grundas sålunda på erfarenheter och iakttagelser av ett fåtal användare och resultaten ger ingen generell bild utan visar snarare exempel på hur subjekten uppfattar det som utvärderas (Allwood 1998). Det är viktigt att få fram vilka aspekter hos systemet som användarna tycker om respektive inte tycker om. Genom att låta läshandikappade studenter lyssna på syntetiskt tal och därefter svara på frågor försöker jag få en bild av de tänkta användarnas inställning till syntetiskt tal, deras acceptans av syntetiskt tal för läsning, och eventuella röstpreferenser. Genom insamling av användarnas subjektiva åsikter kan man få reda på om talsynteserna är bra nog för att fungera som läshjälpmedel. Användarna kan även jämföra utdata från de olika talsyntessystemen samt genom åsikter förmedla sina krav på det som undersöks (Preece et al. 1994). Förutom användarundersökningen har jag dokumenterat mina egna iakttagelser av det syntetiska talet.

4.2 Mål med utvärderingen

Målet med utvärderingen är att ta reda på läshandikappade studenters attityder till och acceptans av syntetiskt tal. Hur uppfattar testpersonerna syntetiskt tal? Är det skillnad på testpersonernas acceptans av talet från de olika systemen, och av de två rösterna som kommer från samma system? Vilken attityd har de till syntetiskt tal före respektive efter undersökningen? Förståelse och naturlighet är inte i fokus i denna utvärdering och det fanns inte utrymme att undersöka kognitiv belastning.

4.3 Utformning av utvärdering

4.3.1 Urval av respondenter

Respondenterna i undersökningen valdes utifrån några givna kriterier. De var registrerade som studenter med läshandikapp på TPB Högskoleservice och bodde eller studerade i Uppsala januari 2004. Urvalet baserades på råd som nämns i (Trost 2001) och (Gibbon et al. 1997) och var icke-slumpmässigt, strategiskt, klustrat och heterogent. Målet var att få en jämn könsfördelning och att alla läshandikapp skulle representeras. Ålder och modersmål var inte med i urvalskriterierna.

4.3.2 Enkät

Det finns ingen vedertagen metod för utvärdering av talomvandlerare som läshjälpmiddel. När jag utformade enkätfrågorna utgick jag från International Telecommunication Unions (ITU) rekommendationer för subjektiv utvärdering av kvalitet på utdata från talsynteser, i *ITU-T Recommendation P.85: A method for subjective performance assessment of the quality of speech voice output devices* (1995). Jag utarbetade egna frågor som fokuserar mer på acceptans än ITU-T P.85, som är inriktat på att utvärdera förståelighet och naturlighet. Enkäten är strukturerad med både öppna och bundna frågor, d.v.s. några frågor besvarades med egen kommentar från respondenten och andra frågor besvarades genom val av svarsalternativ som var fastställda i förväg. Före intervjuerna med respondenterna genomfördes en provintervju för att kontrollera att intervjun innehöll relevanta frågor, som föll sig naturliga, för att säkerställa validiteten. I utformningen av enkäten och i intervjusituationen följde jag de råd som ges i (Preece et al. 1994). Jag har även följt tips på hur enkäter och intervjuer bör utformas, som ges i (Andersson 1985), (Lantz 1993), (Trost 2001) och (Langlet och Wärneryd 1980).

Enkäten består av fyra delar:

1. Frågor om respondenten
2. Subjektiv skattning av längre texter
3. Parvis jämförelse
4. Avslutande frågor

Första delen utgörs av frågor om respondenten själv och dennes erfarenheter av talböcker och syntetiskt tal. I den andra delen får respondenterna utvärdera rösterna genom subjektiv skattning. Efter genomlysning av ett textstycke fick respondenterna svara på frågor om rösten de just hade lyssnat på. De parametrar som där undersöks är talsyntesernas röstkvalitet, felaktigheter i talet och taltempo, samt användarnas acceptans och ansträngning. Röstkvalitet innefattar ljudbilden, språkljuden på fonemnivå, hur behaglig rösten låter m.m. Felaktigheterna i talet handlar om uttal på ordnivå, d.v.s. accent, betoning och borttappade segment, samt prosodi på enheter större än ett ord, d.v.s. melodi, satsmelodi och huruvida rösten är monoton eller inte. Taltempo beaktar hur jämn talhastigheten är och om talhastigheten är lagom snabb. Acceptansen handlar om huruvida respondenterna godtar att använda rösterna för läsning av längre texter och för läsning med DAISY-talbok. Den tredje delen i enkäten är en parvis jämförelse. Respondenterna får höra en mening två gånger med olika röster och därefter svara om föredrar att lyssna på den ena eller andra rösten, eller ingen av dem. Enkäten avslutas med frågor om syntetiskt tal och undersökningen. Samtliga enkätfrågor och svarsalternativ ligger i appendix A.

4.3.3 Testmaterial

Testmaterialet i undersökningen är hämtat från första kapitlet i "Elevens värld" (Imsen 2000) som var litteratur på kursen Pedagogik A vid Uppsala universitet hösten 2003. Majoriteten av de läshandikappade studenterna som vill ha lässtöd läser ämnen inom utbildningsområdena undervisning och samhällsvetenskap, se figur 2.4.

För utvärdering av respektive röst användes texter som är ett stycke långt. Varje textstycke innehöll 100 till 200 ord. Fem olika testmaterial användes för den subjektiva skattningen av längre texter, med samma testmaterial men med olika röster som läste upp texterna. Varje respondent fick lyssna på ett av testmaterialen. För den parvisa jämförelsen användes textenheter som är en till två meningar långa. I den parvisa jämförelsen användes två olika sorters testmaterial, Test 1 och Test 2. Halva testgruppen lyssnade på Test 1 i den parvisa jämförelsen och den andra halvan lyssnade på Test 2. Båda innehöll

samma testmeningar, men ordningen på var olika och den inbördes ordningen mellan de två rösterna som läste upp varje testmening var omvänd. Test 1 innehöll sålunda testmeningar med rösterna A-B och Test 2 innehöll samma meningar i annan ordning, med omvänd ordning på rösterna, B-A. I appendix C ligger samtliga meningar som användes i den parvisa jämförelsen.

4.3.4 Röster

Tre talsynteser användes i undersökningen. Totalt användes fem olika röster: *Ingrid* (RealSpeak), *Ingmar* (Babil), *AnnMarie* (Babil), *Emma* (BrightSpeech) och en mänsklig röst. Den mänskliga rösten hämtades ur en befintlig talbok, och användes som referens för de andra rösterna. *Ingmar* och *AnnMarie* användes även tillsammans med ett användarlexikon. I användarlexikonet lades sådant in som var fanns med i textmaterialet som användes i undersökningen, som uttalades felaktigt men som inte var unikt för den texten, d.v.s. ad hoc-lösningar undveks. Förändringar gjordes bland annat av accent, placering av accent, placering av betoning och fonologisk längd. Sammanlagt användes tre versioner av samma talmaterial: syntetiskt tal direkt från talsyntessystemet, syntetiskt tal med användarlexikon samt mänskligt tal. Rösterna med användarlexikon användes endast i den parvisa undersökningen. Samtliga röster spelades upp med sina standardinställningar. Den mänskliga rösten som var hämtad från en befintlig talbok hade något lägre inspelningsvolym än övriga röster.

4.4 Genomförande av utvärdering

Respondenterna fick i förväg information om att de garanterades anonymitet. De fick även i förväg veta syftet med undersökningen och hur undersökningen skulle gå till.

Testpersonerna intervjuades en och en. Intervjuerna tog 1,5–7 timmar och spelades in. För att efterlikna en användarsituation så mycket som möjligt så gavs de seende respondenterna möjlighet att ha texten framför sig. Fem av respondenterna (en rörelsehindrad och fyra dyslektiker) valde att ha texten framför sig när de läste, så som de oftast brukar använda talböcker. De tre övriga (två synskadade och en av dyslektikerna) brukar använda talböcker utan att ha texten framför sig, och gjorde så även i utvärderingssituationen. I den parvisa jämförelsen hade respondenterna inte möjlighet att ha texten framför sig.

För den subjektiva skattningen fick respondenterna lyssna på fem längre textstycken, fyra av texterna lästes med syntetiskt tal och en av texterna lästes med mänsklig uppläsare. I intervjun kunde respondenterna kommentera sina svar, om de så ville.

Respondenterna fick inte veta vilka talsynteser som användes i undersökningen, inte heller att naturligt tal fanns med som referensmaterial, förrän efter de lyssnat igenom allt ljudmaterial. För sju av de åtta respondenterna spelades ljudfilerna upp på Windows Media Player. I intervjun med den åttonde respondenten spelades ljudfilerna upp med en DAISY-spelare. Samma respondent svarade endast på 39 av de 43 frågorna i den parvisa jämförelsen, p.g.a. tekniska problem. I den parvisa jämförelsen ville en av respondenterna vid några tillfällen svara att båda rösterna var bra, och fick så göra om det motiverades. I två av intervjuerna påverkades inspelningen av yttre störningar, vilket gjorde att delar av intervjuerna inte gick att transkribera. Jag antecknade emellertid svar och kommentarer i samband med intervjuerna och respondenternas svar var på så sätt registrerade.

5 Resultat

I detta kapitel presenteras en sammanfattning av resultatet av utvärderingen med talboksanvändarna. Först beskrivs användargruppen genom att deras svar på de personliga frågorna i enkäten redovisas. Därefter redovisas respondenternas svar på frågorna om de olika rösterna. Slutligen visas resultatet av den parvisa jämförelsen. Användargruppens resultat redovisas dels allas svar sammanlagt, dels grupperat efter läshandikapp, för att se om resultaten skiljer sig åt för de olika läshandikappen. I redovisningen av de bundna enkätfrågorna är resultaten för de rörelsehindrade inbakade i resultatet för den sammanlagda gruppen eftersom de rörelsehindrade består av en person. Respondenternas svar på de öppna frågorna och deras kommentarer till de bundna frågorna ligger i appendix B.

5.1 Respondenterna

5.1.1 Personuppgifter

Läshandikapp och modersmål

Fråga 1: Vilket läshandikapp har du?

Totalt deltog åtta personer i utvärderingen. Fem är dyslektiker. Tre är män och fem är kvinnor, vilket framgår av tabell 5.1 nedan.

Tabell 5.1: *Läshandikapp och kön*

Läshandikapp	Antal respondenter	varav kvinnor	varav män
Synskadade	2	1	1
Rörelsehindrade	1	1	0
Dyslektiker	5	3	2
Totalt	8	5	3

Fråga 2: Har du svenska som modersmål?

Tre av åtta har inte svenska som modersmål, se tabell 5.2 på nästa sida.

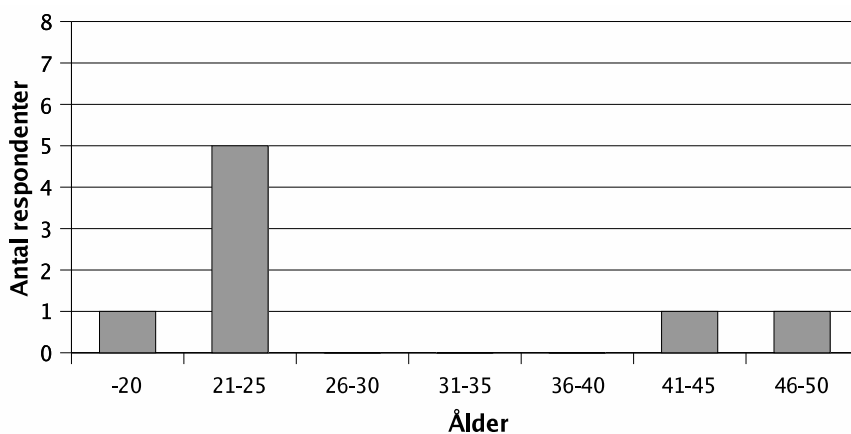
Tabell 5.2: *Läshandikapp och modersmål*

Läshandikapp	Antal resp. utan svenska som modersmål
Synskadade	1
Rörelsehindrade	1
Dyslektiker	1
Totalt	3

Ålder

Fråga 3: Hur gammal är du?

Majoriteten av respondenterna är mellan 21 och 25 år, se figur 5.1 nedan.



Figur 5.1: *Åldersfördelning*

Utbildning

Fråga 4: Inom vilket/vilka ämnesområden studerar du/har du studerat?

I tabell 5.3 på nästa sida ser man vilka ämnen som respondenterna studerar eller har studerat. De två synskadade har båda studerat språk och två dyslektiker läser tekniskt ämnen.

Tabell 5.3: Ämnesstudier

Ämnesområde	Antal respondenter
Data	1
Humanistisk	2
Juridik	1
Naturvetenskap	1
Samhällsvetenskap	1
Språk	2
Teknisk	2
Teologi	1
Vård	1

Övrigt

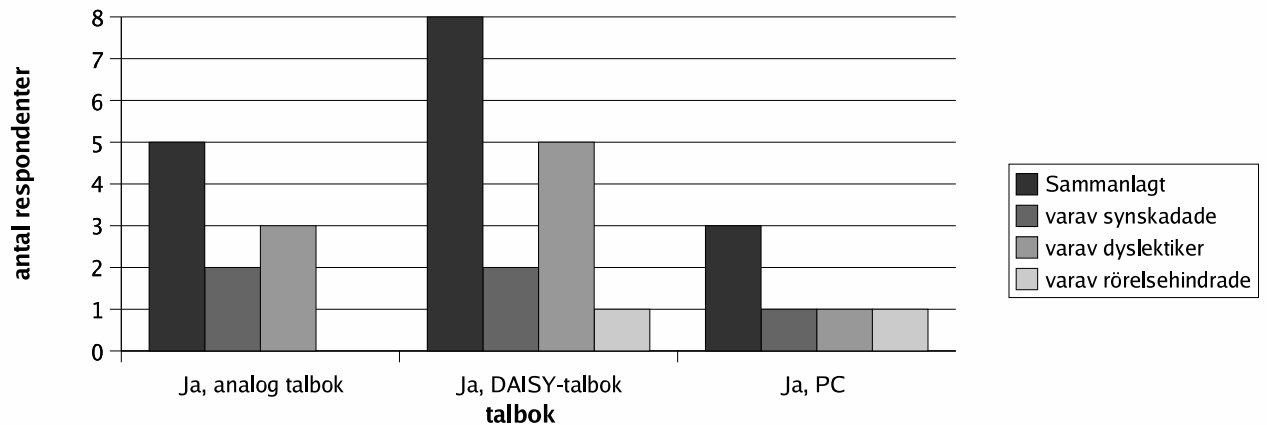
I intervjun framkom även att en av dyslektikerna har en lätt hörselskada på ena örat.

5.2 Erfarenheter av talböcker

5.2.1 Kontakt med talböcker

Fråga 5: Har du använt talböcker?

Samtliga respondenter har använt digitala talböcker. Fem använder analoga talböcker, som spelas upp med kassetbandspelare. Tre använder PC med programvara för läsning när de läser talböcker. Svaren kan ses i figur 5.2 nedan.



Figur 5.2: Har du använt talböcker? (Fråga 5)

5.2.2 Användningsfrekvens talböcker

Fråga 6 a: Hur ofta har du använt talböcker i dina studier?

Tre av respondenterna använder alltid talböcker när de ska läsa kurslitteratur. En av åtta använder talböcker för läsning av ca 80% av kurslitteraturen, en av åtta ca 75% och två av åtta ca 50%. En av personerna i testgruppen har i sina studier endast använt talböcker för en kurs på 5 poäng under de senaste två årens heltidsstudier.

Fråga 6 b: Vad är orsaken till att du inte använt talböcker på alla kurser?

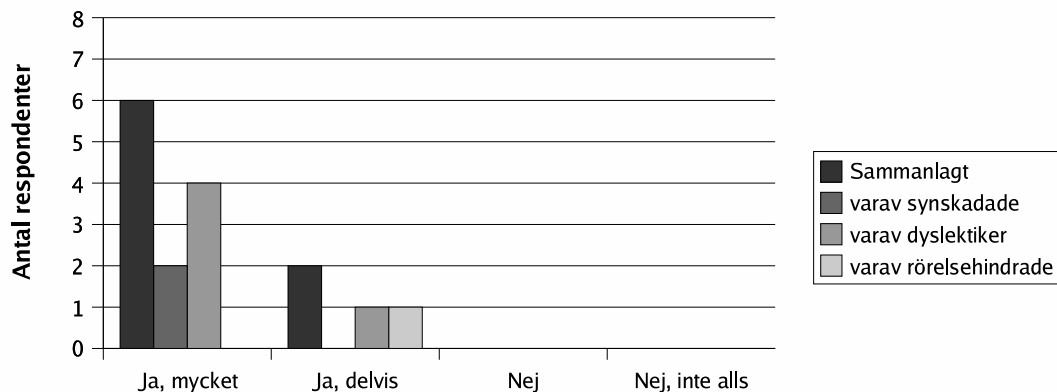
Det finns flera orsaker till att respondenterna inte har använt talböcker till all litteratur i studierna beror på flera saker.

- De hinner inte få litteraturen inläst på talbok i tid, vilket beror på att de med kort varsel får veta vilken kurslitteratur som ska användas på kursen de läser eller för att produktions- eller leveranstid av talboken är för lång. (Fyra av åtta respondenter kommenterade detta.)
- De inte är nöjda med hur talboken är inläst. (Två av åtta respondenter kommenterade detta.)
- Valet av läshjälpmedel styrs även av vilken kurslitteratur de använder. En av de synskadade respondenterna föredrar att läsa viss språkvetenskaplig kurslitteratur, så som lexikon och grammatiker, med läshjälpmedel där det är möjligt att ta reda på stavning, d.v.s. med punktskrift eller elektronisk text. En dyslektisk respondent läser hellre kurslitteratur själv om texten är strukturerad och skriven på ett sådant sätt att respondenten klarar av att läsa det. En tredje respondent säger sig inte behöva läsa alls för att klara sina studier, utan får all information genom att närvara på sina föreläsningar. En fjärde respondent vill öva sig i att läsa utan läshjälpmedel.

5.2.3 Talböcker som läshjälpmedel

Fråga 7: Är talböcker till stor hjälp för dig när du läser?

Som framgår i tabell 5.3 nedan har alla respondenterna angett att talböcker är till stor hjälp för dem när de läser. Sex av åtta säger sig ha mycket stor hjälp av talböcker när de läser. De övriga två, som angett att talböcker delvis är till stor hjälp för dem när de läser, använder talböcker till 75% och 50%. En av dyslektikerna använder talböcker endast för att läsa skönlitteratur. En av de synskadade använder talbok som avkoppling från det syntetiska talet och punktskrift, för att läsa mer litterära texter.



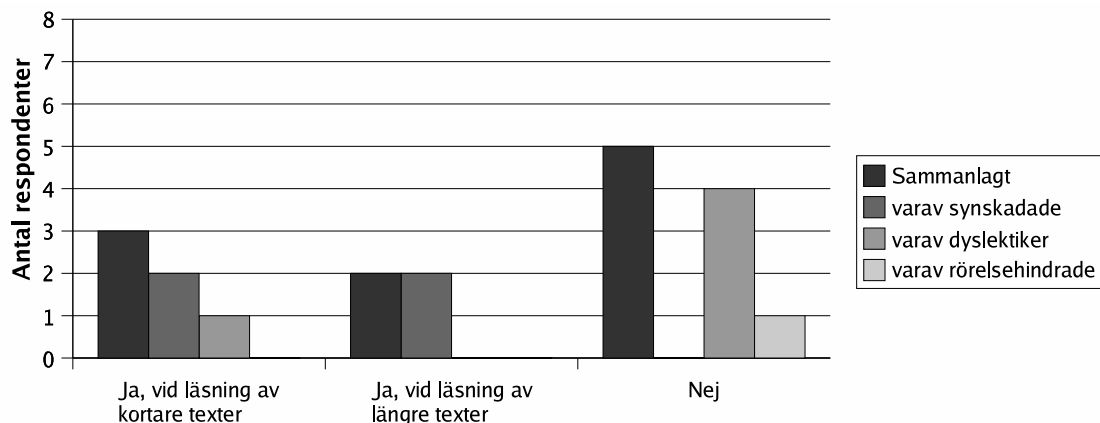
Figur 5.3: Är talböcker till stor hjälp för dig när du läser? (Fråga 7)

5.3 Erfarenheter av syntetiskt tal

5.3.1 Användning av syntetiskt tal

Fråga 8 a: Har du använt dig av syntetiskt tal?

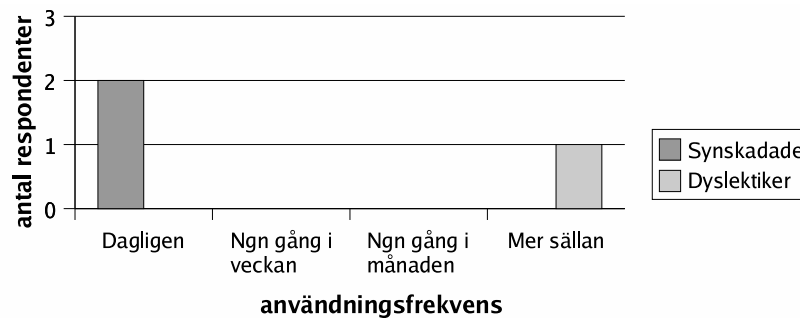
Som figur 5.4 nedan visar så har fem av respondenterna aldrig använt syntetiskt tal. Totalt tre respondenter svarade att de hade erfarenhet av syntetiskt tal. De båda synskadade har använt det både för läsning av kortare och längre texter. En annan respondent har använt syntetiskt tal för att läsa kortare texter. De båda synskadade har använt syntetiskt tal som läshjälpsmedel. De uppgav att de använder sig av *Ingmar* och *Ingrid*. De tre respondenter som svarade att de hade tidigare erfarenhet av att använda syntetiskt tal som



Figur 5.4: Har du använt dig av syntetiskt tal? (Fråga 8a)

läshjälpsmedel fick svara på följdfrågor om detta.

Fråga 8 b: Hur ofta har du använt syntetiskt tal för läsning av kortare texter?

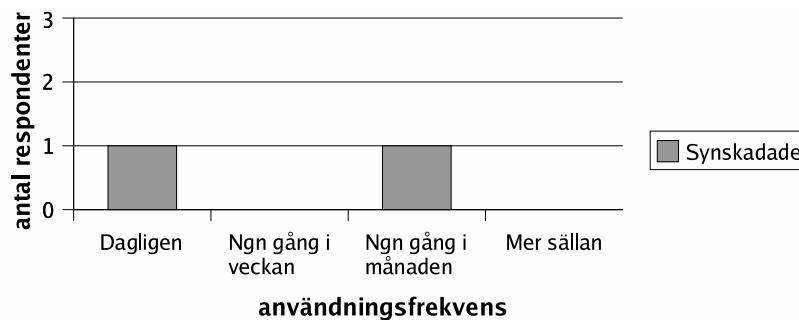


Figur 5.5: Hur ofta har du använt syntetiskt tal för läsning av kortare texter? (Fråga 8b)

Svaren går att se i figur 5.5 ovan. Av de tre personer som använt syntetiskt tal för läsning av kortare texter använder de två synskadade det dagligen för läsning av exempelvis e-meddelanden och webbsidor. Den

tredje personen med erfarenhet av syntetiskt tal är dyslektiker. Han har lyssnat på syntetiskt tal vid endast ett tillfälle, men ”la av direkt” för han tyckte att det var så dåligt.

Fråga 8 c: Hur ofta har du använt syntetiskt tal för läsning av längre texter? De båda synskadade använder



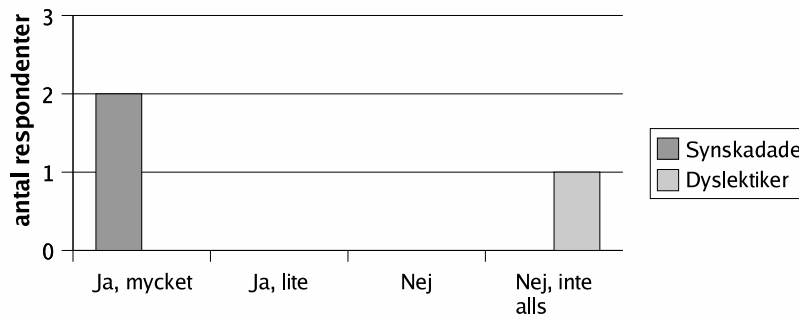
Figur 5.6: Hur ofta har du använt syntetiskt tal för läsning av längre texter? (Fråga 8c)

syntetiskt tal för läsning av längre texter enligt figur 5.4 på föregående sida. Den ena av de två respondenterna använder det syntetiska talet på detta vis någon gång i månaden, medan den andre synskadade respondenten dagligen använder syntetiskt tal för läsning av längre texter, se figur 5.6 ovan.

5.3.2 Syntetiskt tal som läshjälpmiddel

Fråga 9: Är syntetiskt tal till stor hjälp för dig när du läser?

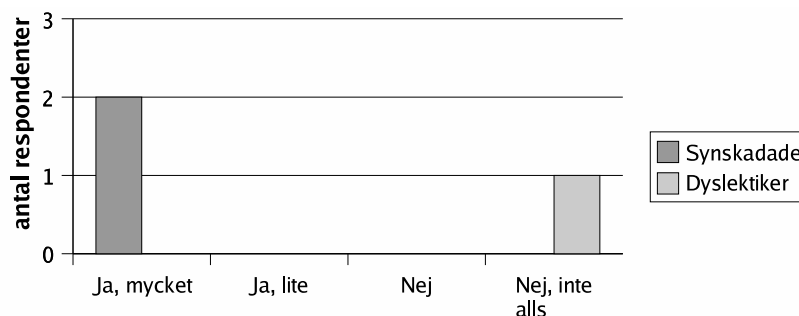
De båda synskadade anger att syntetiskt tal är till stor hjälp för dem när de läser, se figur 5.7 nedan. Den tredje respondenten tycker sig inte alls ha stor hjälp av syntetiskt tal vid läsning.



Figur 5.7: Är syntetiskt tal till stor hjälp för dig när du läser? (Fråga 9)

5.3.3 Inställning till syntetiskt tal

Fråga 10: Är du positivt inställd till syntetiskt tal?



Figur 5.8: Är du positivt inställd till syntetiskt tal? (Fråga 10)

De två som använder syntetiskt tal dagligen, och som har stor hjälp av det, är mycket positivt inställda till syntetiskt tal, se figur 5.8 ovan. De enda i undersökningen som använder syntetiskt tal regelbundet som läshjälpmiddel är de två synskadade.

5.4 Subjektiv skattning

I detta avsnitt redovisas resultaten av testgruppens subjektiva skattning av rösterna i undersökningen. De parametrar som undersökningen fokuserat på är röstkvalitet, ansträngning, taltempo och acceptans, vilket tidigare nämnts i metodkapitlet.

Svarsalternativen på de slutna frågorna i enkäten har omvandlats till siffror, för att få fram medelvärden av testpersonernas skattning. Svaren till frågorna 12c, 13, 15 och 16 har omvandlats till siffervärden enligt tabell 5.4 nedan. Fråga 14 har en femgradig skala som omvandlats till siffervärden enligt tabell 5.5 på nästa sida. Mittpunkt för den skalan är 3.

Tabell 5.4: Omvandling av enkätens svarsalternativ till siffervärden, för frågorna 12c, 13, 15 och 16.

Enkät svar	Siffervärde
Nej, inte alls	1
Nej	2
Ja, delvis	3
Ja, mycket	4

I stapeldiagrammen, som visar resultatet av de slutna frågorna, har origo placerats på skalans mittpunkt, d.v.s. 2,5 respektive 3. Medelvärde i diagrammen visar genomsnittligt svar från respondenterna. Den enda öppna frågan är 12b, men respondenterna har fått möjlighet att kommentera sina slutna svar. Svaren och kommentarerna ligger i sin helhet i appendix B.

Tabell 5.5: Omvandling av enkätens svarsalternativ till siffervärden, för fråga 14.

Enkätsvar	Siffervärde
Mycket långsammare	1
Långsammare	2
Lagom snabbt	3
Snabbare	4
Mycket snabbare	5

5.4.1 Testmaterial

Fråga 11: Har du läst litteraturen tidigare?

Ingen av respondenterna hade tidigare läst litteraturen som användes i undersökningen.

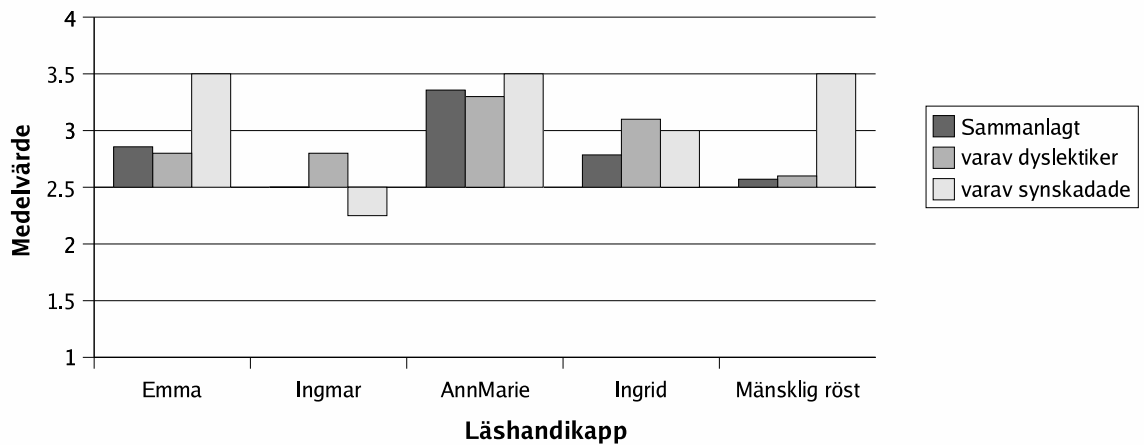
5.4.2 Felaktigheter

Fråga 12 a och b: Upptäckte du några felaktigheter? Vilka?

- **Emma:**
Uttal på ordnivå: Betoning och uttal var felaktigt enligt fyra av åtta respondenter. Prosodi: Två av åtta respondenter tyckte att det lät som om hon talade med dialekt.
- **Ingmar:**
Prosodi: Fyra av åtta repondenter tyckte att uppläsningen var monoton.
- **AnnMarie:**
Prosodi och uttal: Fyra av åtta respondenter tyckte att rösten var monoton och betoningen tråkig.
- **Ingrid:**
Prosodi: Rösten är hoppig och satsmelodin konstig enligt fem av åtta respondenter.
- **Mänsklig röst:**
Uttal: Tre av åtta respondenter angav att orden var avkapade.

Fråga 12c: Upplevde du felaktigheterna som störande?

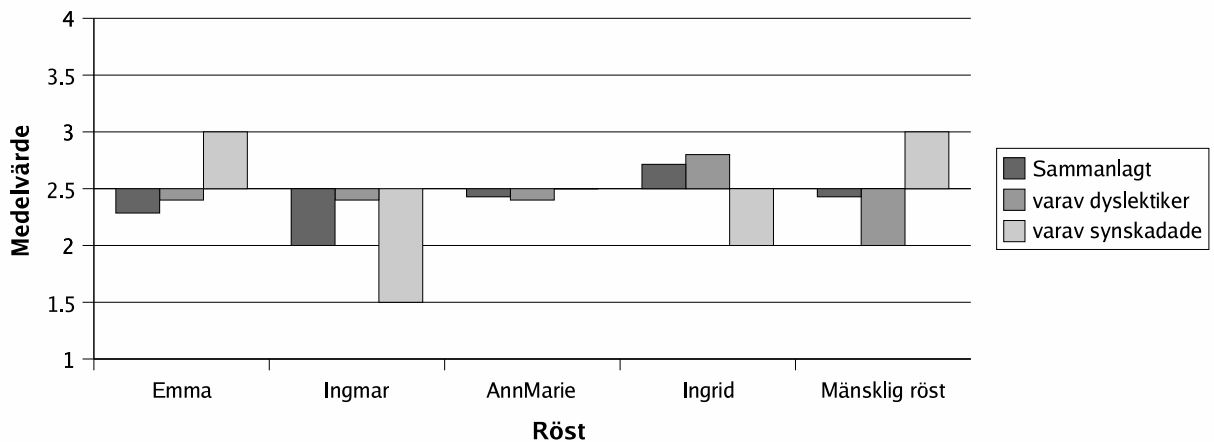
Ju högre medelvärde desto mer störande upplevde respondenterna att felaktigheterna var. Enligt testgruppen innehåller alla rösterna störande felaktigheter, se figur 5.9 på nästa sida. *AnnMarie* har mest störande felaktigheter. Minst störande är *Ingmar* enligt de synskadade. Dyslektikerna anser att den mänskliga rösten har minst störande felaktigheter.



Figur 5.9: *Upplövde du felaktigheterna som störande? (Fråga 12c)*

5.4.3 Ansträngning

Fråga 13: Behövde du anstränga dig för att förstå vad som sades?



Figur 5.10: *Behövde du anstränga dig för att förstå vad som sades? (Fråga 13)*

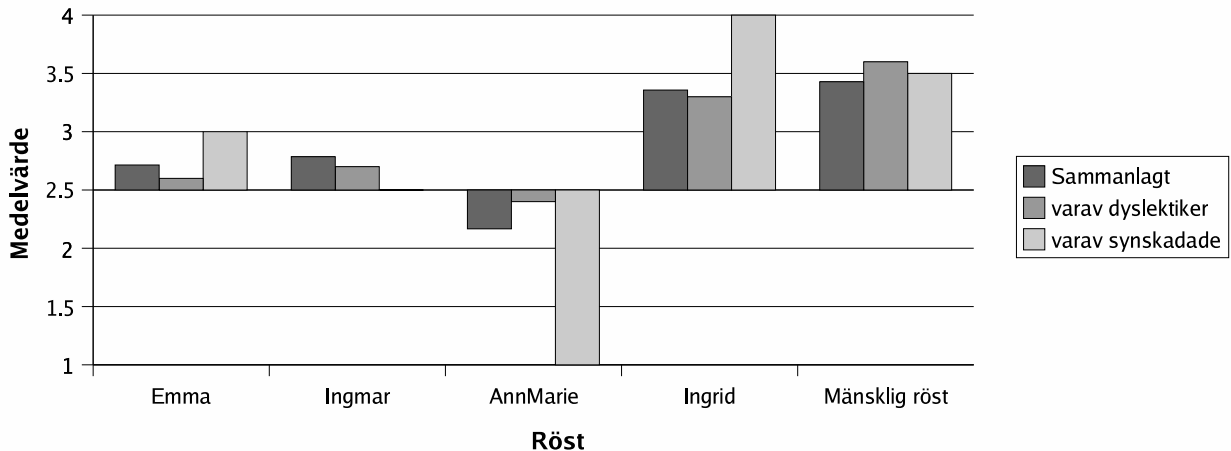
Ju lägre medelvärde desto mindre ansträngande ansåg respondenterna att det var förstå vad som sades. Den röst som fick bäst medelvärde av de synskadade var *Ingmar*. Dyslektikerna tycker att den mänskliga rösten är minst ansträngande att förstå. Sammantaget fick *Ingmar* bäst medelvärde, se figur 5.10 ovan. Störst skillnad i resultat mellan dyslektiker och synskadade har den mänskliga rösten.

I kommentarerna sade två av respondenterna att *Ingrid* hade en konstig rytm i talet vilket var mycket störande, eftersom de hade svårt att koncentrera sig på och höra vad som sades. Två respondenter tyckte

att den mänskliga röstens läsning var alltför ojämn. Två respondenter hade svårt att höra vad *Emma* sade.

5.4.4 Röstkvalitet

Fråga 14: Var rösten behaglig?



Figur 5.11: Var rösten behaglig? (Fråga 14)

Ju högre medelvärde desto mer behaglig röst. De röster som har högst medelvärden för hela testgruppen är *Ingrid* och den mänskliga rösten, se figur 5.11 ovan. *AnnMarie* har lägst medelvärde för hela gruppen.

I respondenternas kommentarer går följande åsikter att urskilja om röstkvaliteterna. Två av respondenterna klagade på språkljuden hos *Emma*, de tyckte att s-ljuden var för skarpa. Två ansåg även att *Ingmars* s-ljuden var skarpa. Tre respondenter tyckte att ljudet hos *AnnMarie* låter syntetiskt och två att rösten var nasal.

5.4.5 Taltempo

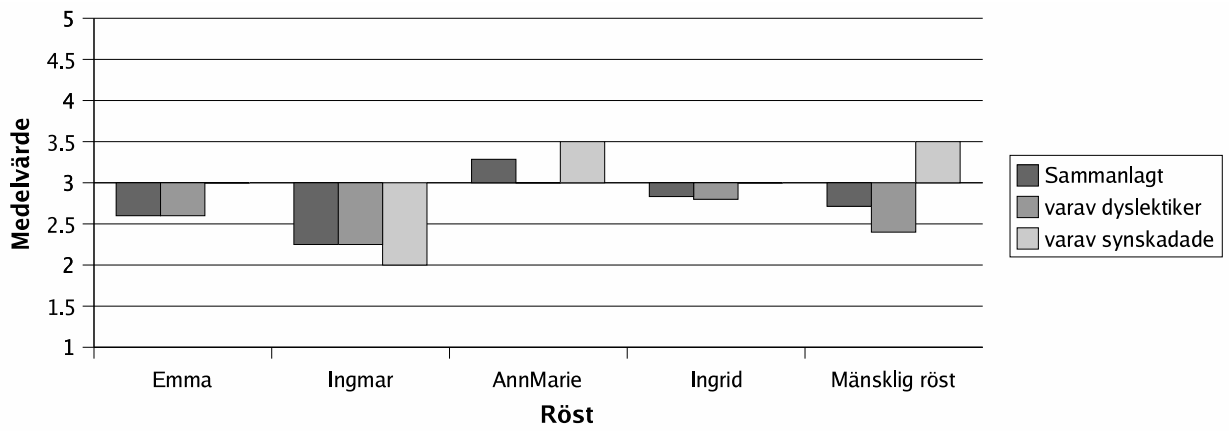
Fråga 15 a: Var talhastigheten lagom snabb, snabbare eller långsammare än vad du föredrar?

Högre värde än mittpunkten anger att talhastigheten var för snabb, lägre att talhastigheten var för långsam. *Ingmar* var något för långsam enligt alla, se figur 5.12 på nästa sida.

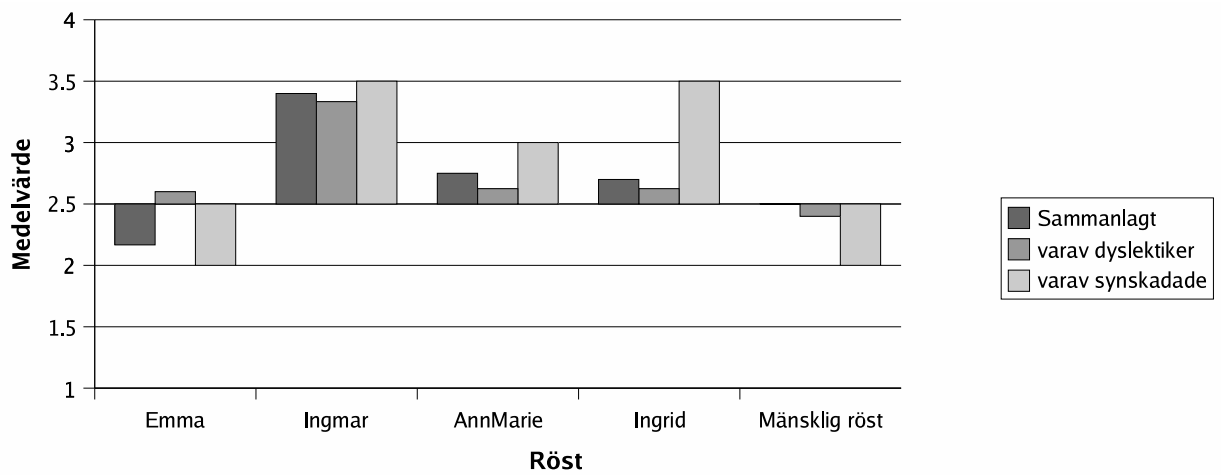
Fråga 15 b: Var talhastigheten lagom jämn?

Ju högre medelvärde desto jämnare uppfattades talhastigheten, se figur 5.13 på nästa sida. Den röst som alla anser ha jämnast talhastighet är *Ingmar*. Minst jämn talhastighet har *Emma* och den mänskliga rösten.

Tre respondenter av åtta sade i kommentarerna att talhastigheten hos den mänskliga rösten var för ojämn. De tyckte att det är för många pauser mitt i meningarna.



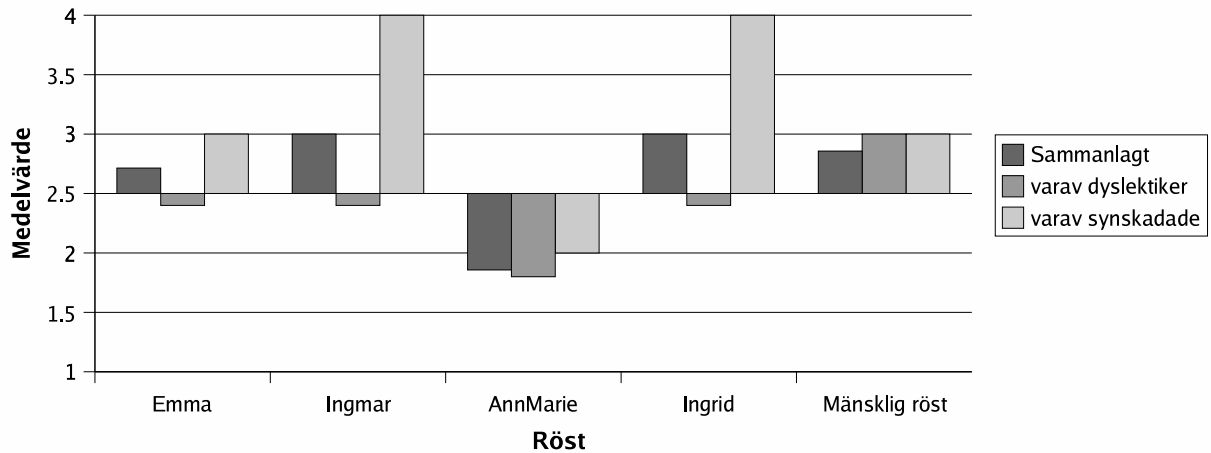
Figur 5.12: *Var talhastigheten lagom snabb, snabbare eller långsammare än vad du föredrar? (Fråga 15 a)*



Figur 5.13: *Var talhastigheten lagom jämn? (Fråga 15 b)*

5.4.6 Acceptans

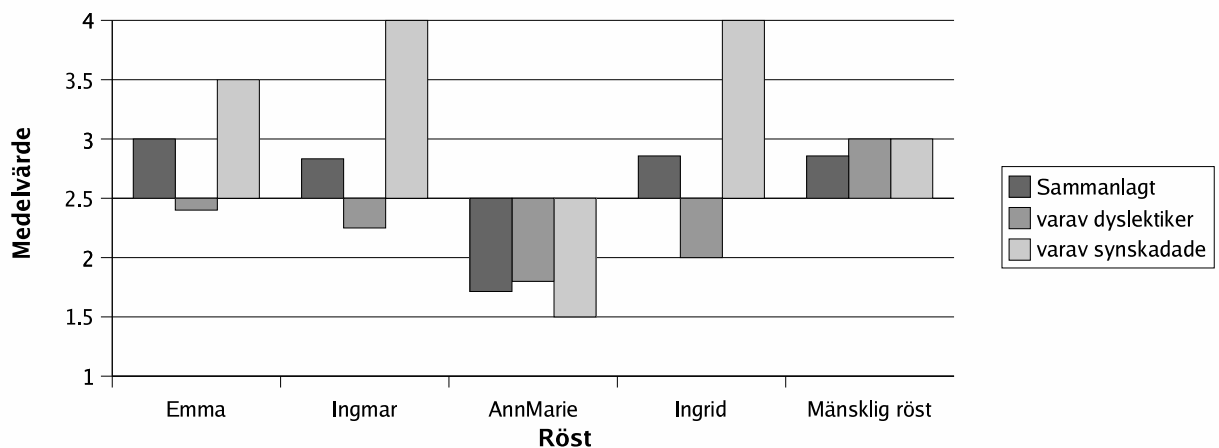
Fråga 16 a: Skulle du kunna lyssna på längre texter som spelats in med denna typ av röst?



Figur 5.14: Skulle du kunna lyssna på längre texter som spelats in med denna typ av röst? (Fråga 16 a)

Ju högre medelvärde desto mer tror sig respondenterna kunna lyssna på längre texter som spelats in med den typen av röst, se figur 5.14 ovan. Den röst som sticker ut ur mängden är *AnnMarie*, som har lägst medelvärde. Dyslektikerna föredrar den mänskliga rösten, och de synskadade visar tydligt att de föredrar rösterna *Ingmar* och *Ingrid*.

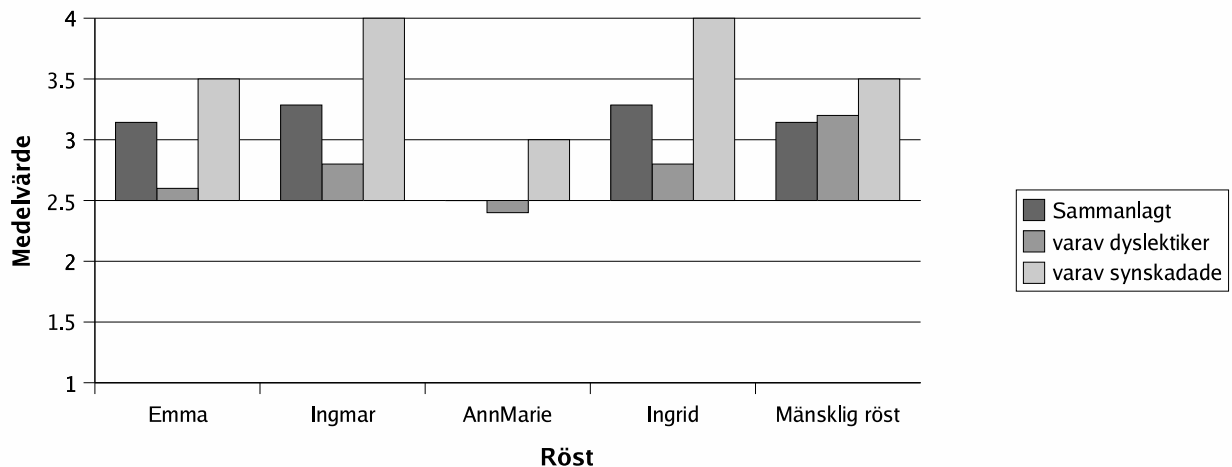
Fråga 16 b: Skulle du kunna använda en DAISY-talbok som spelats in med denna röst?



Figur 5.15: Skulle du kunna använda en DAISY-talbok som spelats in med denna röst? (Fråga 16 b)

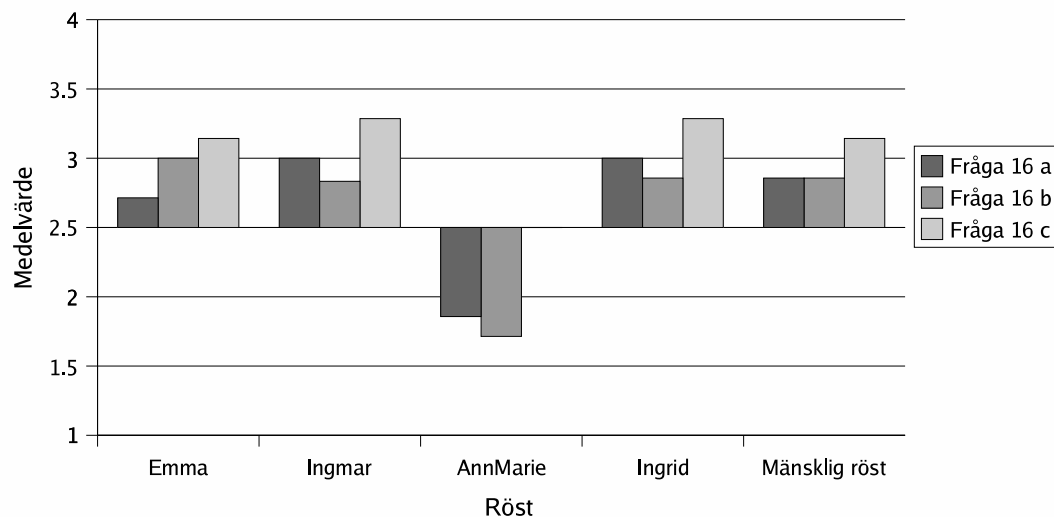
Ju högre medelvärde desto mer kan användarna tänka sig att använda en DAISY-talbok som spelats in med den rösten. Det man kan säga generellt är att rösten *AnnMarie* får betydligt sämre medelvärde än resten av rösterna. De synskadade föredrar, och ger höga betyg till, rösterna *Ingmar* och *Ingrid*. I övrigt ligger resultaten i mitten, varken bra eller dåligt. Dyslektikerna föredrar den mänskliga rösten. För alla sammanslaget får *Emma* högst medelvärde, se figur 5.15 på föregående sida. Skillnaden mellan att läsa en längre text med eller utan DAISY-format är att DAISY-talboken är strukturerad så att den är navigerbar, och att den vanligtvis används av användarna i en studiesituation.

Fråga 16 c: Skulle du kunna använda en DAISY-talbok som spelats in med denna röst, om DAISY-talboken, förutom det inspelade talet, innehöll hela den tryckta bokens text i form av e-text (elektronisk text) som är kopplad till talet?



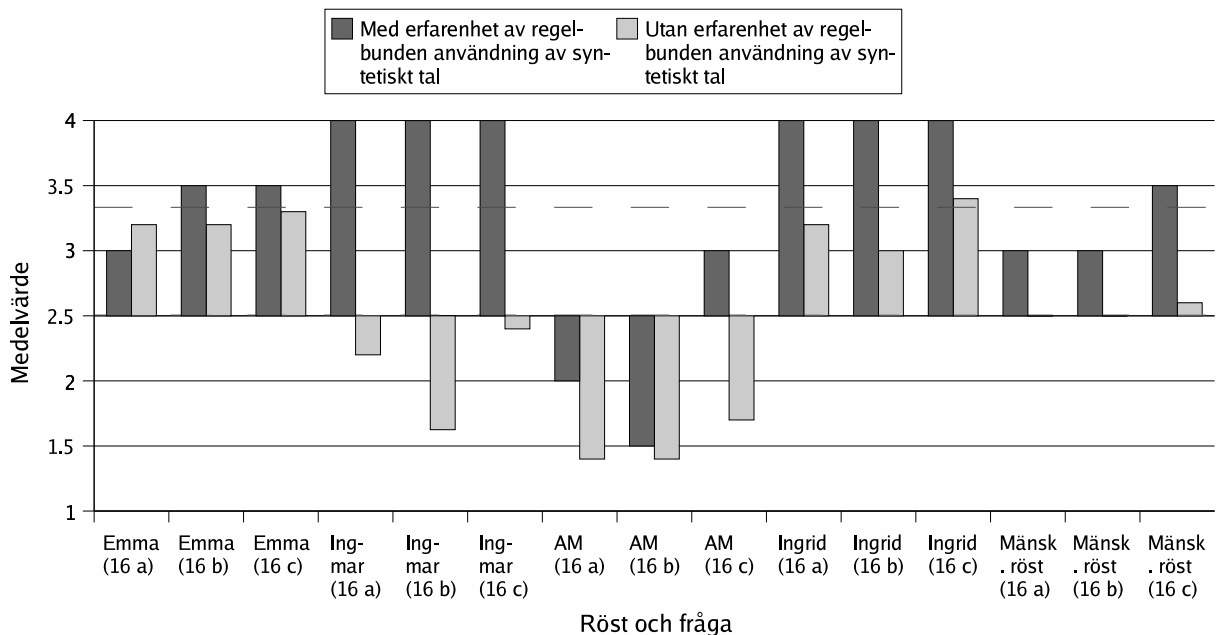
Figur 5.16: Skulle du kunna använda en DAISY-talbok som spelats in med denna röst, om DAISY-talboken, förutom det inspelade talet, innehöll hela den tryckta bokens text i form av e-text (elektronisk text) som är kopplad till talet? (Fråga 16 c)

AnnMarie får sämst medelvärde (2,5) av alla respondentgrupper. De synskadade föredrar och ger höga betyg till rösterna *Ingmar* och *Ingrid*. Dyslektikerna föredrar den mänskliga rösten. Resultaten har sammanställts i figur 5.16 ovan.



Figur 5.17: Jämförelse mellan svar på frågorna 16 a, 16 b och 16 c för hela testgruppen.

I en jämförelse av resultaten på fråga 16 a, 16 b och 16 c kan man, i figur 5.17 ovan, se att respondenterna generellt sett var mer positiva till att använda talböcker med länkat text och tal än vanliga talböcker. De motiverade detta med att de med text och ljud länkat till varandra skulle kunna identifiera de ord som de hade problem med att förstå (två av åtta) och det skulle vara lättare att följa med eftersom det skulle gå att se var i texten de är (tre av åtta). Den av respondenterna som i vanliga fall inte använder talböcker för läsning av kurslitteratur uttryckte sig mycket positiv till möjligheten att i en vetenskaplig bok lätt kunna hoppa över formler och figurer, och endast läsa texten emellan. För rösterna *Ingmar*, *AnnMarie* och *Ingrid* verkar testgruppen mer positiv till att använda de syntetiska rösterna till läsning av längre texter, än till att använda rösterna i talbok utan länkat text och tal.



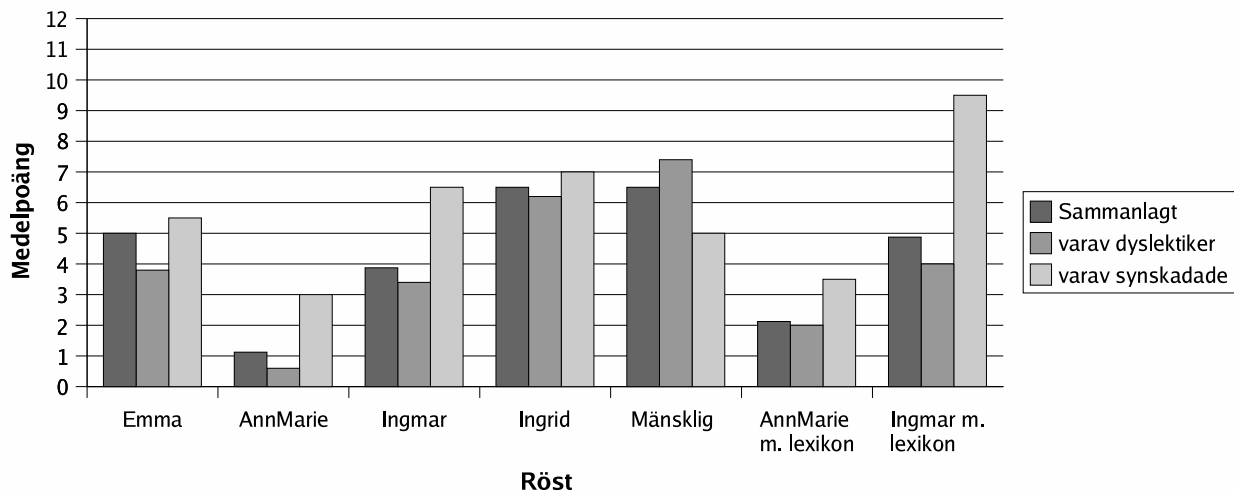
Figur 5.18: Jämförelse mellan svar på frågorna 16a, 16b och 16c för respondenter med och utan erfarenhet av regelbunden användning av syntetiskt tal.

I figur 5.18 ovan kan man se resultaten på frågorna 16 a, 16 b och 16 c för respondenter med respektive utan erfarenhet av regelbunden användning av syntetiskt tal som läshjälpmiddel. Den mörkgrå streckade linjen (ca 3,3) visar medelvärdet för sammanlagt utdelade poäng från respondenter med erfarenhet av regelbunden användning av syntetiskt tal. Deras acceptans av syntetiskt tal är högre än acceptansen hos respondenter utan erfarenhet av regelbunden användning av syntetiskt tal, medelvärdet av skattningen från den senare gruppen visas med den ljusgrå streckade linjen (2,5) på baslinjen.

5.5 Parvis jämförelse

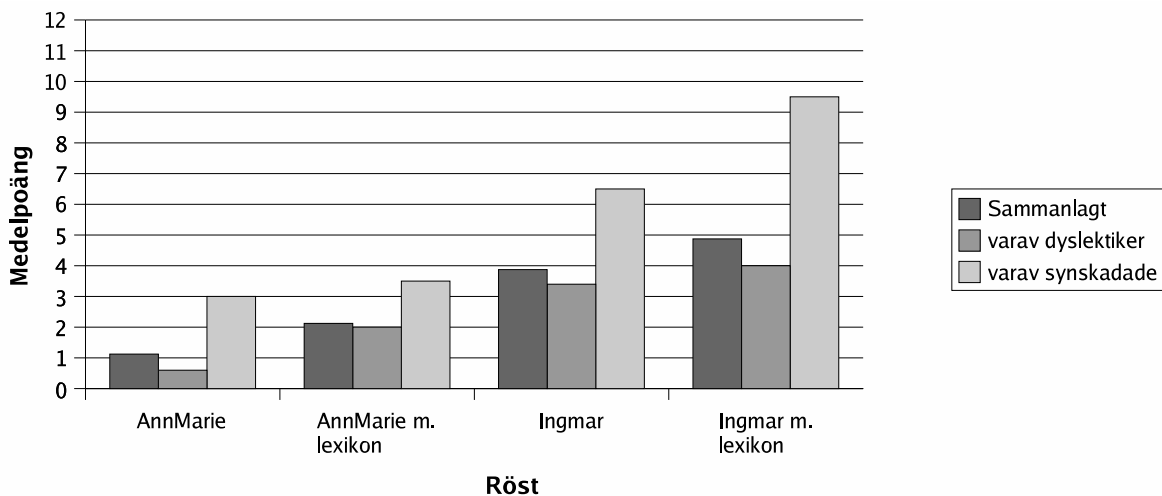
I den parvisa jämförelsen fick respondenterna svara på vilken av två röster som de helst lyssnade på. Fråga 17 var formulerad ” Vilken av A eller B lyssnar du helst på?” Den röst de valde fick ett poäng. De kunde även avstå från att välja någon av rösterna, eller välja båda två. Sammanlagt var 42 meningsspar med i den parvisa undersökningen. Sammanlagt antal poäng som en respondent kunde dela ut var 42 gånger 2 poäng, d.v.s. 84 poäng, om de gav båda rösterna poäng för alla meningar. En röst kan maximalt få 12 poäng från en och samma respondent, eftersom den är med i 12 jämförelser. *AnnMarie med lexikon* och *Ingmar med lexikon* är samma röster som *AnnMarie* och *Ingmar*, men med tillägget att de har ett användarlexikon som förbättrar uttal. Användarlexikonet är samma för både *AnnMarie med lexikon* och *Ingmar med lexikon*. Resultatet av den parvisa jämförelsen visas med stapeldiagram. Respondenternas kommentarer ligger i appendix D. Medelpoängen visar hur många poäng som respondenterna i genomsnitt delat ut till rösterna.

Den mänskliga rösten och *Ingrid* får i den parvisa jämförelsen högst medelpoäng av alla respondenter sammanslaget. Dyslektikerna ger flest poäng till den mänskliga rösten och de synskadade ger markerat



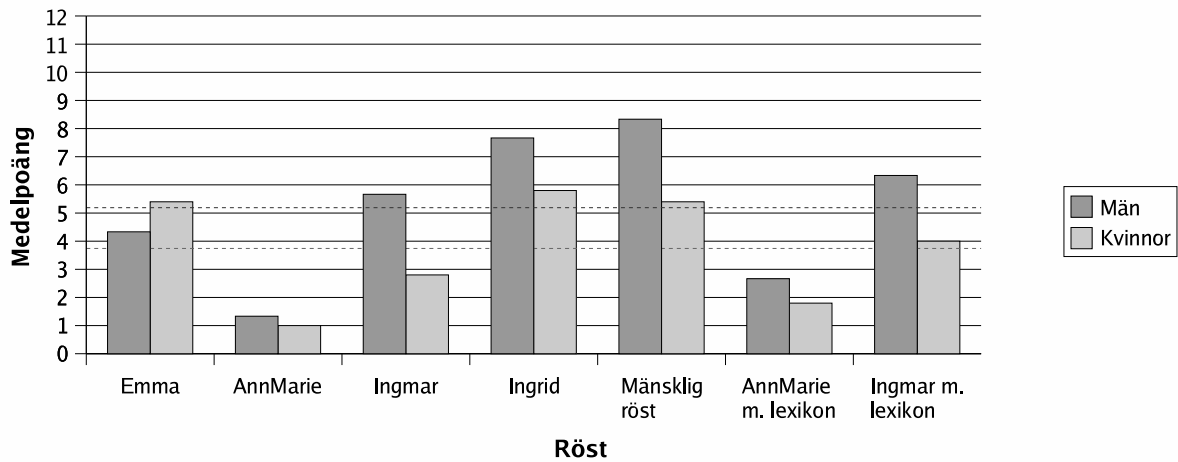
Figur 5.19: Resultatet av den parvisa jämförelsen

högst medelpoäng till *Ingmar med lexikon*. Lägst medelpoäng får rösten *AnnMarie*, vilket gäller för både hela testgruppen och testgruppen uppdelad i de två läshandikappnen, se figur 5.19 ovan.



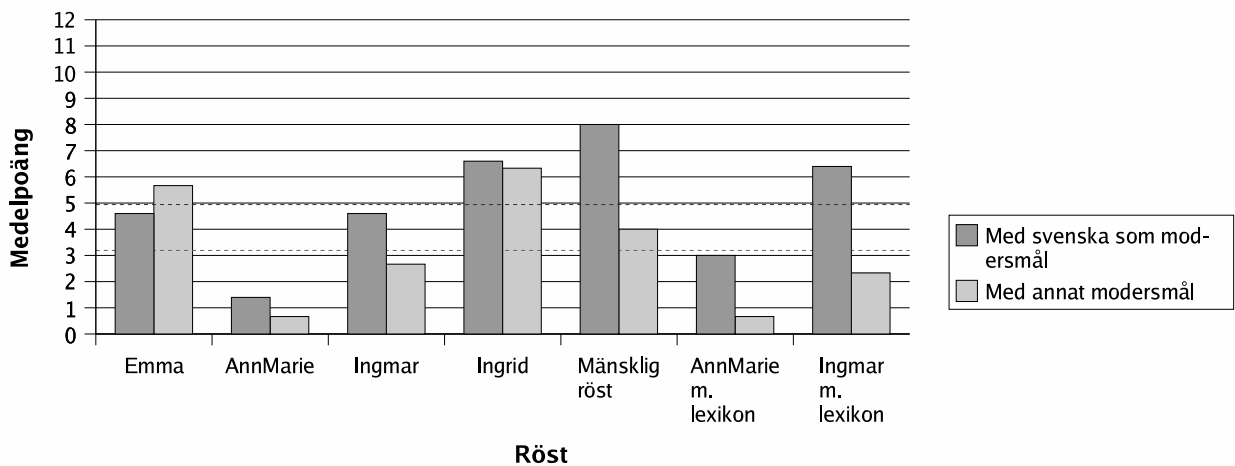
Figur 5.20: Jämförelse av medelpoäng för röster med och utan användarlexikon.

I figur 5.20 ovan visas en jämförelse i hur hög medelpoäng röster fått med och utan användarlexikon. Rösterna med användarlexikon har fått högre medelpoäng än sina motsvarande röster utan användarlexikon. *Ingmar med lexikon* har högst medelpoäng för både dyslektiker, synskadade och alla sammanslaget. *AnnMarie* har lägst medelpoäng för samtliga. *AnnMarie* och *Ingmar* kommer från samma produkt, men med olika röster. *AnnMarie* är en kvinnlig röst och *Ingmar* är en manlig. Testgruppen föredrar tydligt den manliga rösten framför den kvinnliga.



Figur 5.21: Resultatet av den parvisa jämförelsen sorterat efter respondenternas kön

I figur 5.21 ovan jämförs medelpoäng utdelade av män respektive kvinnor, oavsett läshandikapp. Både männen och kvinnorna ger högst medelpoäng till den mänskliga rösten och *Ingrid*, för kvinnorna ligger även *Emma* på högsta medelpoäng. Medelvärde för poängutdelning skiljer sig åt mellan män och kvinnor. De två streckade horisontella linjerna visar medelvärde för poängutdelning av män (mörkgrått, 5,2) och kvinnor (ljusgrått, 3,7).

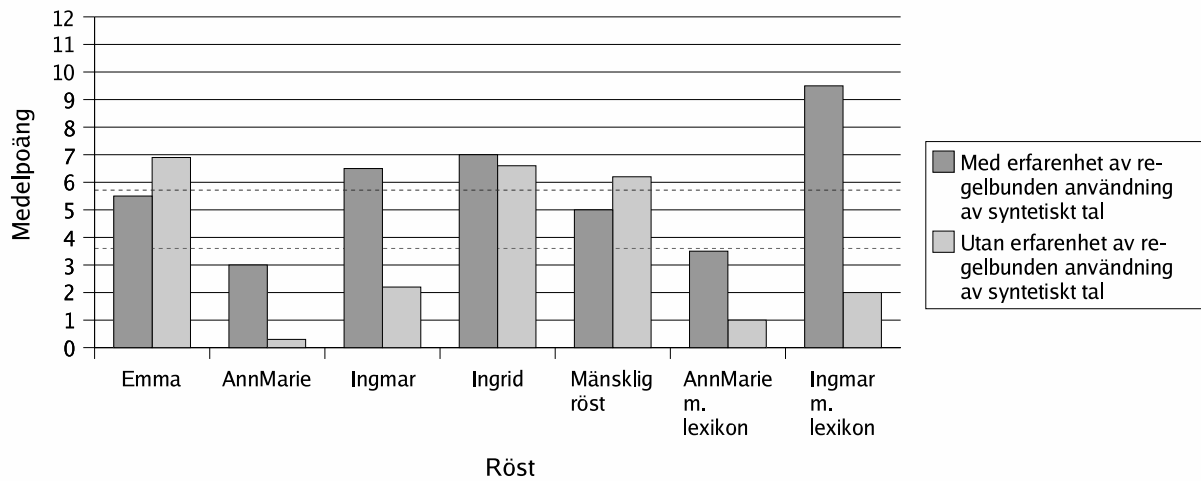


Figur 5.22: Resultatet av den parvisa jämförelsen sorterat efter respondenternas modersmål

I en jämförelse mellan de respondenter som har svenska som modersmål och de respondenter som har annat språk som modersmål kan man se att gruppen med annat modersmål än svenska ger högst medelpoäng till rösterna *Emma* och *Ingrid*, se figur 5.22 ovan. Den del av testgruppen som har svenska som modersmål föredrar den mänskliga rösten. Respondenter med svenska som modersmål delar ut fler poäng än övriga (4,9 respektive 3,2).

I figur 5.23 på nästa sida visas en jämförelse mellan medelpoäng utdelat av respondenter med respektive

utan tidigare erfarenhet av regelbunden användning av syntetiskt tal som läshjälpmiddel. De respondenter som har erfarenhet av att regelbundet använda syntetiskt tal delar ut fler poäng än de andra (5,7 respektive 3,6). Störst skillnad i utdelade medelpoäng är det för *Ingmar med lexikon* (9,5 respektive 2).



Figur 5.23: Resultatet av den parvisa jämförelsen sorterat efter respondenternas tidigare erfarenheter av syntetiskt tal som läshjälpmiddel.

5.6 Avslutande frågor

För att respondenterna skulle kunna svara på de avslutande frågorna fick de veta att en av rösterna som använts i undersökningen tillhörde en mänsklig inläsare. Därefter fick de svara på frågor om deras inställningen till syntetiskt tal förändrats, samt vilka egenskaper de anser är viktiga hos syntetiskt tal för att de ska lyssna på det.

Fråga 18: Har din inställning till syntetiskt tal förändrats?/ Vad tycker du om det syntetiska tal som du just fått lyssna på?

Av de tre respondenter som hade tidigare erfarenhet av syntetiskt tal sade en sig vara lite mer positiv. De andra två hade samma inställning som tidigare. De övriga fem respondenterna, som saknade tidigare erfarenheter av syntetiskt tal, fick svara på vad de tyckte om de röster som de just hade hört. En var positivt överraskad av hur det syntetiska talet hade låtit. Två respondenter utan tidigare erfarenheter av syntetiskt tal uttrycker att de säkert kommer att kunna använda syntetiskt tal i framtiden, men att rösterna måste förbättras för att de ska använda dem. En hade svårt att förstå rösterna eftersom de pauser och det tonfall som underlättar läsning och förståelse saknades, enligt respondenten.

Fråga 19: Vilka egenskaper anser du är viktiga hos syntetiskt tal, för att du ska lyssna på det?

Sammanfattningsvis tycker respondenterna att följande egenskaper är viktiga hos syntetiskt tal:

- **Naturlighet:**

Talet ska vara som mänskligt tal och låta naturligt. (Totalt fem av åtta kommenterade naturlighet)

- **Uttal:**

Det syntetiska talet ska ha rätt uttal (fem av åtta) och vara lätt att förstå (två av åtta). (Totalt fem av åtta kommenterade uttal)

- **Taltempo:**

Talet ska vara flytande och ha jämn hastighet (tre av åtta). Det ska inte finnas pauser på fel ställen (en av åtta) men däremot andningspauser som visar när det kommer något nytt (en av åtta). (Totalt fyra av åtta kommenterade taltempo)

- **Prosodi:**

Talet ska inte vara monotont och ha ”bra intonation”. (Totalt fyra av åtta kommenterade prosodi)

- **Röstkvalitet:**

Rösten ska inte ha metalliska ljud (två av åtta) eller låta nasal (en av åtta), utan vara varm (två av åtta) och harmonisk (en av åtta). (Totalt tre av åtta kommenterade röstkvalitet)

Enligt fyra respondenter lämpar sig rösterna bättre eller sämre som läshjälpmiddel beroende på vilken sorts litteratur rösten ska användas till. Kraven på rösterna är olika om det är faktaböcker eller skönlitteratur. Två andra i testgruppen kommenterar att kurslitteratur används vid inläring och att detta ställer extra höga krav på läshjälpmidlet. Röster som de inte vill använda vid inläring skulle de kunna tänka sig att använda vid andra tillfällen, som inte kräver lika mycket.

6 Analys och diskussion

I detta kapitel analyseras resultaten från utvärderingen från olika vinklar. Resultaten för undersökningsparametrarna sammanfattas och analyseras. Vidare analyseras läshjälpmedel i kombination med syntetiskt tal, och vad man bör tänka på. Därefter diskuteras möjliga felkällor, och slutligen tas några förslag på vidareutveckling upp.

6.1 Respondenterna

Undersökningen visar att läshandikappade studenter är en heterogen grupp vad gäller typ av läshandikapp och erfarenheter av syntetiskt tal. Gemensamt för hela testgruppen är att alla har vana att använda talböcker i DAISY-format och anser sig ha stor eller mycket stor hjälp av talböcker när de ska läsa.

6.2 Ansträngning

För hela testgruppen sammanslaget ligger medelvärdena för samtliga röster mellan 2 och 3 (på skalan 1–4 där 1 motsvarar ”inte alls ansträngande” och 4 motsvarar ”ja, mycket ansträngande”).

De synskadade verkar tycka att den röst de är vana att lyssna på dagligen, *Ingmar*, är minst ansträngande att lyssna på. *Ingmar* fick av de synskadade medelvärde 1,5.

Tidigare erfarenheter verkar påverka ansträngning som krävs för att lyssna på syntetiskt tal på ett positivt sätt. De respondenter med erfarenhet av att regelbundet använda syntetiskt tal som läshjälpmedel anser att det är minst ansträngande att lyssna på *Ingmar*, som de använder dagligen. De som inte har erfarenhet av dessa syntetiska röster ger inte lika höga betyg till dem, vilket kan ses i figur 5.10. Medelvärdet för de synskadade, som har erfarenhet av att regelbundet använda syntetiskt tal som läshjälpmedel, ligger på 1,5 och på 2,4 för dyslektikerna.

6.3 Röstkvalitet

De röster som fick högst medelvärde från den sammanlagda testgruppen för behaglighet var den mänskliga rösten och *Ingrid*, som fick ca 3,5 på skalan 1–4 där 4 är bästa medelvärde och 1 är sämst.

De synskadade gav lägsta möjliga betyg till *AnnMarie*, medelvärdet 1.

Jag fick en känsla av att endel respondenter fastnade på hur behaglig rösten var, och att de inte brydde sig om huruvida de förstod vad som sades. Detta har jag dock inte några belegg för, eftersom undersökningen inte omfattade parametern förståelighet. I en annan studie, av Sluijter et al.(1998), fick de system som var mest förståeliga sämst resultat i acceptabilitetstesten, vilket stöder min teori.

6.4 Taltempo

Hela testgruppen tyckte att rösten *Ingmar* hade något långsammare talhastighet än vad de föredrog. Medelvärde ligger mellan 2 och 3,5 (på skalan 1–5, där 1 motsvarar ”mycket långsammare” än vad respondenten föredrar och 5 motsvara ”mycket snabbare” än vad respondenten föredrar), vilket indikerar att hastigheten inte var alltför avvikande hos någon av rösterna.

Den röst som testgruppen ansåg ha jämnast talhastighet var *Ingmar*.

Taltempot för rösterna *AnnMarie* och *Ingmar* uppfattas som olika av respondenterna. Talhastighet och paus (se tabell 3.1) är emellertid samma för de båda rösterna, däremot har *AnnMarie* högre tonhöjd och mer dynamik än *Ingmar*. Det verkar som om tonhöjden och/eller dynamiken skulle kunna påverka hur talhastigheten uppfattas.

6.5 Acceptans

Testgruppen sammanslaget ger ungefär samma medelvärde för alla röster, förutom *AnnMarie* som får lägsta medelvärde för både läsning av längre texter, DAISY-talbok och DAISY-talbok med länkat text och tal. Lägst medelvärde, 1,7 (på skalan 1–4 där 4 är bästa medelvärde och 1 är sämst), får *AnnMarie* för läsning med DAISY-talbok.

De synskadade ger högsta möjliga betyg (4) till *Ingmar* och *Ingrid* för både läsning av längre texter, DAISY-talbok och DAISY-talbok med länkat text och tal.

Dyslektikernas högsta medelvärde är 3,2 och det ges till den mänskliga rösten för läsning med DAISY-talbok med länkat text och tal.

Som figur 5.17 visar så varierar acceptansen beroende på vilken applikation som den syntetiska rösten kan tänkas användas i. För några röster är medelvärdet lägst för läsning med DAISY-talbok. En av orsakerna till detta kan vara att respondenterna ställer högre kognitiva krav på läshjälpmiddel för kurslitteratur, eftersom de ska lära sig något av den litteraturen.

6.5.1 Tidigare erfarenheters påverkan av acceptans

Tidigare erfarenheter av regelbunden användning av syntetiskt tal verkar positivt påverka respondenternas acceptans av syntetiskt tal som läshjälpmiddel. De skattar de syntetiska rösterna högre än de som saknar sådan erfarenhet, 3,3 respektive 2,5, se figur 5.18. De med tidigare erfarenhet föredrar genomgående de röster de är vana vid, och ger dessa röster högsta möjliga betyg. De respondenter som inte har erfarenhet av syntetiskt tal ger inte alls lika höga medelvärden till dessa röster. Skillnaden är mest påtaglig för *Ingmar* som i fråga 16 b har medelvärde 4 respektive 1,6.

Dyslektiker ger sina högsta medelvärden till den mänskliga rösten, inte bara för acceptans utan för de flesta parametrarna i den subjektiva skattningen. Det kan bero på att de har mest erfarenhet av att lyssna på röster från talböcker. Vissa kände till och med igen rösten, vilket framkom i deras kommentarer. De synskadade använder talböcker procentuellt sett lika ofta för studier som övriga läshandikappade, men är inte lika favoriserande av den mänskliga rösten. Det kan bero på att de använder syntetiskt tal dagligen och att det finns ett begränsat antal syntetiska röster, medan det finns en mängd olika mänskliga inläsare av talböcker. På så sätt lär de känna de syntetiska rösterna mer ingående än de mänskliga rösterna i talböckerna.

De med tidigare erfarenhet av syntetiskt tal är även mycket positivt inställda till syntetiskt tal, se figur 5.8. Frågan är vad som är orsak och verkan. Använder de syntetiskt tal regelbundet för att de från början är positivt inställda, eller är de positivt inställda för att de använder syntetiskt tal regelbundet? En av de

synskadade ser stora framtidsmöjligheter för syntetiskt tal: ”Det är även viktigt att det finns många olika röster att välja mellan, för olika texter, olika situationer, man bör kunna kombinera flera olika röster till en talbok, så att t. ex. en pjäs kan läsas upp med olika röster för de olika rollerna.”, vilket måste vara tecken på mycket positiv inställning till syntetiskt tal.

6.6 Parvis jämförelse

Den sammanlagda testgruppen gav högsta medelpoäng (ca 6 av 12) till den mänskliga rösten och *Ingrid*, medan *AnnMarie* fick lägst medelpoäng (ca 1 av 12).

När man jämför resultaten för tal med och utan användarlexikon, se figur 5.20, så får tal med användarlexikon högre medelvärde än tal utan förbättrat uttal. Skillnaden i medelpoäng är störst för *Ingmar* som av de synskadade får ca 3 medelpoäng högre med användarlexikon än utan.

I både den subjektiva skattningen och den parvisa jämförelsen fick *AnnMarie* genomgående sämre betyg än *Ingmar*, se figur 5.20, trots att de kommer från samma produkt, Babil. Standardinställningarna för talhastighet, volym och paus är samma för de båda rösterna, däremot har *AnnMarie* högre tonhöjd och mer dynamik än *Ingmar*. För konkatenerande talsynteser verkar det således vara viktigt vem som är inläsare av taldata basen och hur röstmaterialen hanteras av systemet. Standardinställningarna i en syntesapplikation kan också tänkas påverka hur en röst tas emot av användarna.

Ett alternativt sätt att analysera resultatet av den parvisa jämförelsen skulle kunna vara att se vilken röst som prefereras kopplat till den andra rösten som är med. T.ex. skulle det kunna vara så att en röst A får poäng endast i jämförelse med en röst B, men att den inte får poäng i jämförelse med andra röster.

6.7 Den mänskliga rösten

Den mänskliga rösten var tänkt att fungera som referens i utvärderingen. Dyslektikerna föredrog den mänskliga rösten i de flesta fall, men för hela testgruppen sammanslaget fick den inte högre betyg än de syntetiska rösterna. Respondenterna kommenterade att de inte tyckte om lästekniken som den mänskliga inläsaren använde och några i testgruppen påpekade att rösten påminde om talböcker de använt men som de inte längre lyssnade på eftersom de inte tyckte om lästekniken. Hela testgruppen sammanslaget ger den mänskliga rösten ungefär samma medelbetyg som de bästa talsynteserna i både den parvisa jämförelsen och den subjektiva skattningen.

De synskadade, som regelbundet använder talböcker, föredrar syntetiska röster framför den mänskliga rösten. För dem är nästan alla syntetiska röster mindre ansträngande att förstå, och de syntetiska röster som de synskadade använder regelbundet får högre medelvärde för acceptans. I den parvisa jämförelsen ger de den mänskliga rösten medelpoäng som endast är högre än *AnnMarie* och *AnnMarie med lexikon*.

6.8 Kognitiv belastning

Med kognitiv belastning menas ”en mental tankeansträngning, orsakad av miljöfaktorer, som går ut på att tolka och bearbeta information för att kunna bestämma handlingsalternativ inom en given tidsrymd” (Gulliksen och Göransson 2002). Två testpersoner uttryckte att de inte vill ha för hög kognitiv belastning när de läser kurslitteratur, och en tredje trodde sig behöva pauser för att lyssna på länge texter med *Ingmar*. De synskadade uttryckte att de ville ha punktstöd till vissa av rösterna, för att kunna kontrollera vad som sades, vilket kan indikera att de inte uppfattar allt som sägs.

Mätmetoder och resultat

Om en talsyntes kräver mycket koncentration och ansträngning för att förstås minskar uppmärksamheten som lyssnaren kan rikta mot andra pågående aktiviteter. Människan har begränsade kognitiva resurser och den kognitiva belastning som syntetiskt tal innebär kan leda till längre reaktionstid när lyssnare ska ge svar på enkla frågor, sämre resultat och prestanda p.g.a. trötthet, samt att de blir sämre på att utföra enkla sysslor i samband med lyssnandet, enligt (Gibbon et al. 1997). Utan att beakta den kognitiva belastningen är det svårt att tolka förståelighetstester på korrekt sätt (Silverman et al. 1990). Klatt (1987) nämner flera olika undersökningar, som andra personer har gjort, vilka intresserat sig för den kognitiva belastningen vid lyssning av syntetiskt tal. Enligt en av undersökningarna är reaktionstiden något längre för syntetiskt tal än vid naturligt tal. En annan undersökning har mätt försämrade exakthet och hastighet vid omedelbart svar på frågor ställda i syntetiskt tal istället för naturligt tal. Andra undersökningar som Klatt tar upp visar att korttidsminnet för tidigare element i en lista uppläst med syntetiskt tal är sämre jämfört med om listan läses upp av en mänsklig inläsare. För att mäta kognitiv belastning krävs någon form av automatiserat test, som noggrant kan mäta exempelvis fördröjning i reaktionstid.

6.9 Text-till-talomvandlerare som läshjälpmedel

Kan man sålunda använda text-till-talomvandlerare som läshjälpmedel? I utvärderingen har vissa av de syntetiska rösterna lika höga medelvärden som den mänskliga rösten, vilken används i befintliga läshjälpmedel. Därav kan man dra slutsatsen att det är möjligt att använda syntetiska röster som läshjälpmedel.

Om man ser till resultaten från hela den sammanlagda testgruppen så går det inte att entydigt urskilja en röst som bättre än de övriga. I undersökning av ansträngning och acceptans är det ingen röst som är markant bättre än de andra. *Ingmar* har jämnast talhastighet. För röstkvalitet samt preferens i den parvisa jämförelsen får *Ingrid* och den mänskliga rösten bäst betyg. Således får *Ingrid* något högre betyg än övriga syntetiska röster av den sammanlagda testgruppen.

Kvaliteten hos den mänskliga rösten kan dock ifrågasättas utifrån respondenternas negativa kommentarer om den, se avsnitt 6.7. En mänsklig inläsare garanterar inte hög kvalitet på talet. Man kan fråga sig vilka krav som är uppställda för läshjälpmedel och vilka krav som måste vara uppfyllda för att ett läshjälpmedel ska kunna sägas vara bra.

För att de syntetiska rösterna ska få bra resultat i undersökningen, d.v.s. få högsta skattnings- och preferensvärden av hela testgruppen, krävs det mer av dem. I kapitel 7 redovisas resultatet av min inventering av vad talsynteserna i undersökningen har för språkliga möjligheter och brister. Det visar vad som kan behöva åtgärdas för att rösterna ska bli bättre läshjälpmedel. Vidare är det lämpligt att definiera vilka krav som finns på läshjälpmedel i allmänhet. Det borde även studeras hur stor hjälp läshjälpmedlen är för de läshandikappade.

6.9.1 Krav på text-till-talomvandleraren

För att kunna avgöra om en talomvandlerare är lämplig som läshjälpmedel bör kraven på läshjälpmedlet vara definierade. Därefter kan man se om det syntetiska talet uppfyller kraven eller vad som behöver åtgärdas före det kan användas som läshjälpmedel. Nedan följer några förslag på vad som behöver tas ställning till i en kravspecifikation för syntetiskt tal som läshjälpmedel.

- Användningsområde
 - För vilken form av litteratur är det syntetiska talet tänkt att användas? Kurslitteratur omfattar mycket varierande typ av litteratur, allt från formelsamlingar till skönlitteratur kan användas

som litteratur på högskolenivå. Det innebär att stora krav måste ställas på talomvandlerare om den ska kunna omvandla all sorts kurslitteratur till tal.

- Slutanvändarna
 - Ta ställning till huruvida litteraturkvaliteten för slutanvändarna får försämrats.
 - Hur stor kognitiv belastning är rimlig att lägga på slutanvändarna?
 - Hur lång tid är rimlig att låta slutanvändarna vänta på läshjälpmiddel?
- Text-till-talomvandleraren
 - Vilka språkliga krav ska vara uppfyllda före det syntetiska talet används som läshjälpmiddel? Var går gränsen för vilka språkliga felaktigheter i talet som är acceptabla och vilka som inte är acceptabla?
 - Ord som inte finns med i systemets lexikon bör identifieras, för att ha möjlighet att ge alla ord lämpligt uttal.
 - Vilka former på indata ska talomvandleraren klara av? DAISY-formatet? Utvecklarna av talomvandlerarna får åtgärda sådant som inte kan hanteras. Tabeller och figurer? För att läsa ut tabeller med syntetiskt tal krävs att det finns algoritmer för hur proceduren ska gå till. De algoritmer och riktlinjer för hur tabeller och grafer läses upp av mänskliga inläsare skulle kunna användas för att automatisera omvandling från tabell m.m. till tal.
 - Ska slutanvändarna ha möjlighet att anpassa talet efter sin egna preferenser, så som det är möjligt med de applikationer som använts i undersökningen? T.ex. välja längden på pauserna vid interpunktion, grad av dynamik, tonhöjd m.m., eller t.o.m. välja bland flera olika röster.
- Utdata
 - Ska störande ljud i talet, som begränsar slutanvändarnas lyssningsmöjligheter, vara accepterat eller ej? Exempelvis skarpa s-ljud hindrar användaren att lyssna på ljudfilerna med hörlurar.
 - Ska det gå att reglera talhastigheten med det syntetiska talet på ett tillfredställande sätt?

6.10 Möjliga felkällor

6.10.1 Testmaterial

Testmaterialet som användes i undersökningen är hämtat från utbildningsområden inom vilket majoriteten av de studerande talboksanvändarna läser, Undervisning och Samhällsvetenskap. Kanske hade det varit mer intressant att välja litteratur från ett utbildningsområde inom vilket få talboksanvändare studerar, eftersom det är billigare per talboksanvändare att producera läshjälpmiddel som många kommer att använda än sådant som endast en enda person kommer att använda. Om man vill få ner kostnaderna för talboksproduktionen skulle det löna sig att använda syntetiskt tal för litteratur som få personer använder än sådant som många använder.

6.10.2 Talsyntesröster

I undersökningen hade jag helst använt röster av ett och samma kön, för att undvika alltför många parametrar som kan påverka testpersonerna. Eftersom två av de talsynteser som utvärderades, BrightSpeech och RealSpeak, endast finns med kvinnliga röster, och talboken som användes som testmaterial hade manlig inläsare, så beslutades att använda röster av båda könen. Ett av talsyntessystemen, Babil, hade

både en kvinnlig och en manlig röst. Resultaten för dessa röster jämfördes för att undersöka om respondenterna eventuellt föredrar röst av specifikt kön. De båda rösterna fick mycket skilda resultat, till fördel för den manliga rösten.

6.10.3 Respondenter

Respondenterna valdes ut från ett och samma område, Uppsala, vilket inte kan sägas vara representativt för den användargrupp som de representerar eftersom det finns läshandikappade studenter i hela landet, inte bara i Uppsala.

Tre av de åtta respondenterna hade annat modersmål än svenska. Detta verkar vara en ovanligt stor andel, och är antagligen inte representativt för användargruppen. Detta går dock inte att bekräfta eftersom TPB Högskoleservice inte har uppgift om de läshandikappades modersmål.

6.10.4 Undersökningsparametrar

Den mänskliga rösten är hämtad från en befintlig talbok och hade lägre inspelningsvolym än övriga röster. Volymen kan ha påverkat resultatet och det hade sålunda varit bra att ha med volym som en parameter i undersökningen.

I den parvisa jämförelsen vet man inte om respondenterna tyckt att båda rösterna var lika bra eller lika dåliga när respondenten inte valt ut en framför den andra. Det hade varit bra att ge respondenterna fyra möjliga val till frågan ”Vilken av A och B lyssnar du helst på?”: ingen av rösterna / A / B / både A och B.

6.10.5 Användarundersökning

Ett problem med attitydundersökningar är att det inte är säkert att man genom frågor om attityd får en rättvis bild av personens beteende. Verbalt uttryck är en helt annan sak än hur man i verkligheten beter sig.

Användartesterna kan endast visa exempel på användarnas acceptans, det behövs fler användare i undersökningen för att få statistiskt underlag.

6.10.6 Svårtolkade kommentarer

Flera av respondenterna uttryckte att rösterna i undersökningen var hackiga. Vad som avses med detta är oklart eftersom både taltempo, prosodi, och systemet som gör syntesen kan vara hackig. Hade varit bra att ha följdfrågor när respondenterna uttryckte sig om vilken sorts hackighet de kommenterade.

6.11 Vidareutveckling

I framtida studier vore det lämpligt att undersöka acceptansnivån hos användarna efter längre, regelbunden användning av syntetiskt tal som läshjälpmedel.

Det skulle även vara intressant att studera skillnaden i acceptans av talomvandlare med olika syntesmetoder. Enligt Fahlstedt (2004) har det visat sig att talsynteser med unit selection får höga acceptansvärden till en början och lägre efter längre exponering, medan talsynteser med difonsyntes till en början har låga acceptansvärden, men att de gradvis blir högre.

I användarundersökningen fanns inte resurser till att studera den kognitiva belastningen, eftersom sådan undersökning kräver tillgång till automatiska tester med exempelvis pekskärm. Det vore dock av intresse att ta reda på hur stor kognitiv belastning som syntetiskt tal innebär för användarna, i jämförelse med mänskligt tal.

Förväntningar verkar påverka respondenterna, och det hade därför varit intressant att undersökningen studera testgruppens förväntningar. Exempelvis är en respondent mycket positiv till *Emma* trots att respondenten uppenbarligen inte förstår vad som sägs och även anser felaktigheterna hos rösten som störande, då respondenten tror att det är en talsyntes som han redan känner till och tror på. En annan respondent pratade mycket om att dennes förväntningar på talet som inte uppfylldes. Det hade därför varit intressant att studera testgruppens förväntningar.

7 Egna observationer

De talsynteser som används i undersökningen har en rad brister i den textuella och lingvistiska bearbetningen. I detta kapitel tar jag upp mina egna observationer av dessa tillkortakommanden, som jag gjorde när jag lyssnade igenom textmaterialet som användes till användarundersökningen. Först tar jag upp sådant som kan hanteras med de befintliga talsyntesapplikationer som användes i undersökningen. Därefter tar jag upp vad som skulle kunna hanteras med olika former av språkliga analyser. Vissa problem kan åtgärdas relativt enkelt, några går att hantera men är svårare, andra ligger på forskningsnivå, t.ex. identifiering av egennamn och utländska ord, och är på intet sätt lösta språkliga problem. Eftersom den inre strukturen på talsynteserna inte är öppen för insyn är det inte möjligt att säga hur och var varje språkligt problem ska hanteras i respektive talomvandlare. Ingen av talsynteserna uppfattades som bättre än någon annan och jag redovisar här alla de olika talomvandlarnas tillkortakommanden gemensamt.

7.1 Befintliga talsyntesapplikationers möjligheter

Applikationerna, i vilka talsynteserna förekommer, tillhandahåller användarlexikon. Dessa är fullformslexikon, d.v.s. de har ingen automatisk generering av ett ords böjningsformer utifrån dess lemma, och varje böjningsform måste läggas in i lexikonet var för sig. I de undersökta text-till-talomvandlarna förekom följande som kan hanteras med användarlexikonet:

- Den icke-ambiguösa **förkortningen** t.ex. läses som om punkterna är meningsskiljare. Detta kan hanteras genom att läggas in i de befintliga användarlexikonet.
- **Specialtecken**. Parenteser läses inte ut av någon av talomvandlarna, vilket de bör i ett läshjälpmedel.
- **Egennamn** fick felaktiga uttal, t.ex. *Imsen*, *Kelly* och *Haider*, vilket kan åtgärdas genom att de läggs in i lexikon tillsammans med korrekt uttal.
- **Uttalsfel**. *Planet* uttalas med felplacerad betoning. Andra ord har fått fel accent, d.v.s. grav accent istället för akut eller tvärtom, exempelvis *steget* uttalades med grav accent istället för akut av flera talsynteser. Andra ord får segment med felaktig fonologisk längd i talsyntesen, exempelvis *datalära* som uttalas *dattalära* har för kort fonologisk längd på det första *a*:et.

7.2 Sådant som befintliga talsyntesapplikationer inte kan hantera

I de undersökta text-till-talomvandlarna förekom följande som inte kan hanteras med befintliga talsyntesapplikationer:

- **Siffror** utläses av vissa talomvandlare som årtal, andra som tal. Det kan hanteras genom regler för i vilken kontext siffror läses ut som årtal alternativt ett tal.
- **Specialtecknet ”-”** tolkas olika av talsynteserna. I 17-22 läses ”-” ut som *minus* av en talsyntes, som bindestreck, d.v.s. tystnad, av en annan och som *till* av en tredje. Citattecken läses inte ut av någon av talomvandlarna, inte heller när läsinställningen är satt till läsning bokstav för bokstav. Citattecken bör läsas ut så att den som läser vet vad som är citat utan att behöva ha texten framför sig.
- **Extern homografi**, ordklassflertydighet, förekommer i talet som produceras av talsynteserna. Exempelvis *drog* i sammansättningen *drogproblematiken* uttalas som verbet *dra* i preteritum när det i själva verket handlar om substantiv i utrum singular obestämd form. Alla möjliga tolkningar måste finnas med i lexikon och genom exempelvis syntaktisk analys och sammansättningsanalys kan rätt lexem och uttal väljas.
- **Intern homografi**, formflertydighet, förekommer också. *Planet* uttalas som substantiv neutrum singular bestämd form när det är ett substantiv i utrum singular obestämd form, d.v.s. den andra stavelsen har betoning istället för den första stavelsen. I viss kontext kan rätt lexem identifieras och lämpligt uttal fås fram genom exempelvis syntaktisk analys. En förkortning, *övers*, som i texten är förkortning till *översättarens*, är homograf till ett ord i fullform, det oböjliga substantivet *övers* som i *till övers*. Genom exempelvis syntaktisk och eventuellt även semantisk analys kan disambiguering ske. Med standardiserad användning av förkortningar är det lättare för ett system att automatiskt hantera dessa.
- I texten förekommer **lexikaliserade uttryck**, flerordsuttryck, som inte läses ut som del av samma fras, utan som delar av olika fraser eller satser. Exempelvis *till och med* och *saker och ting* läses inte ut som nominalfraser. Skulle kunna identifieras med syntaktisk parsning.
- Talsynteserna kan inte hantera **verbpartiklar** som prepositioner som hör till ett framförstått verb. Betoning av verbpartikel skiljer sig från andra prepositioner genom att de bildar en fonologisk fras med verbet och uppbär frasens huvudbetoning, men talsynteserna gör inte denna betoning. I texten förekommer bland annat följande fraser med verbpartiklar: *tränga igenom*, *hjälpa till*, *orka med*, *klara av*, *ställa upp*, *dyker upp* och *ta upp*.
- Vissa interna homografer av samma ordklass, som förekommer i texten, skulle inte gå att disambiguera genom syntaktisk analys. Till exempel *kön* får av systemen uttalet för substantiv neutrum singular obestämd form när det i själva verket är substantiv utrum singular bestämd form. I kontexten *de unga hamnar sist i kön* krävs att man får fram textens betydelse för att disambiguera tolkningen av ordet, genom någon form av **semantisk analys**. Detta är komplicerat och idag finns inget automatiserat sätt att semantisk analysera godtyckliga texter.
- Vissa språkliga fenomen kan hanteras genom att utifrån en korpus statistiskt undersöka vad som är vanligast förekommande i ett språk. **Statistiska beräkningar** på en ordklasstaggad korpus kan exempelvis ligga till grund för disambiguering av interna homografer. Man skulle kunna få statistik för hur ofta t.ex. *kön* förekommer som nominalfras i species obestämd form respektive species bestämd form, i attribut till prepositionsfras, och utifrån resultaten välja den analys som är mest sannolik.
- Idiom hanteras som ett helt lexikalt uttryck, snarare än att varje konstituent hanteras var för sig, exempel från texten: *på sätt och vis*. **Lexikaliserade uttryck** måste ligga i lexikon för att de ska läsas ut som sammanhängande enheter. Ett alternativ är att texten analyseras semantiskt.

- Talsynteserna varierar inte betoning på verb som står verbfinalt t.ex.*har*. **Prosodin** i talet är viktig för förståelsen av det som sägs. Om prosodin är felaktig kan det vara mycket förvirrande för den som lyssnar, och svårt att förstå vad som sägs.
- **Utländska ord**, såsom *black box* uttalas som svenska. Utländska ord förekommer ofta i kurslitteratur och måste på något sätt hanteras av talsynteserna. Utländska ord skulle kunna märkas upp manuellt med språktillhörighet, och låta talomvandlingssystemen ge orden lämpligt uttal utifrån uppmärkningen.
- Vissa **sammansatta ord** uttalas felaktigt av talsynteserna. Exempelvis *massundervisning* uttalas *massunder-visning*. Detta kan hanteras genom att de felaktiga orden läggs in i lexikon. Ett annat sätt att hantera sammansatta ord skulle kunna vara att göra en sammansättningsanalys.
- Några av talsynteserna lägger inte in paus efter rubriker som saknar avslutande interpunktion. **Texters struktur** måste kunna hanteras av system som ska omvandla litteratur till tal.

Litteraturförteckning

- Allwood, M. (1998). *Människa-datorinteraktion: Ett psykologiskt perspektiv*, Studentlitteratur, Lund.
- Andersson, B.-E. (1985). *Som man frågar får man svar - en introduktion i intervju- och enkätteknik*, Rabén & Sjögren, Kristianstad.
- Babel Infovox AB hemsida* (2004). <http://www.babeltech.be/>. (15 maj 2004).
- BabTTS User Guide* (2003). BaBel Technologies S.A. version 1.2.
- Black, A. W. och Campbell, N. (1995). Optimising selection of units from speech databases for concatenative synthesis, *Proc. Eurospeech '95*, Madrid, Spain, pp. 581–584. <http://citeseer.ist.psu.edu/black95optimising.html>.
- Bruce, G. (1998). *Allmän och svensk fonologi*, Reprocentralen, Lund.
- Carlson, R. (1993). Models of speech synthesis, STL-QPSR 1/1993, pp 1–14. <http://www.speech.kth.se/rolf/papers/wnnasqpsr.pdf>.
- Carlson, R., Sigvardson, T. och Sjölander, A. (2002). Data-driven formant synthesis, *Proc of Fonetik 2002*, Vol. 44, TMH-QPSR, pp. 121–124.
- Dutoit, T. (1997). *An introduction to text-to-speech synthesis*, Kluwer Academic Publishers, Nederländerna.
- Elert, C.-C. (2000). *Allmän och svensk fonetik*, Norstedts, Stockholm.
- Fahlstedt, E. (2004). *SVT Talande textremsa. En användarutvärdering av två talsynteser i en ny tillämpning.*, Examensarbete, Institutionen för tal, musik och hörsel, Kungliga Tekniska Högskolan. http://www.speech.kth.se/ctt/publications/exjobb/exjobb_efahlstedt.pdf.
- Gibbon, D., Moore, R. och Winski, R. (eds) (1997). *Handbook of Standards and Resources for Spoken Language Systems*, Mouton de Gruyter, Berlin.
- Gulliksen, J. och Göransson, G. (2002). *Användarcentrerad systemdesign*, Studentlitteratur.
- Imsen, G. (2000). *Elevens värld: introduktion till pedagogisk psykologi*, Studentlitteratur, Lund.
- ITU-T Recommendation P.85: A method for subjective performance assessment of the quality of speech voice output devices* (1995). International Telecommunication Union. www.itu.int/ITU-T/publications/recs.html.
- Klatt, D. (1987). Review of text-to-speech conversion for english, *Journal of the Acoustic Society of America* **82**(3): 737–793.

- Langlet, P. och Wärneryd, B. (1980). *Att fråga*, Statistiska centralbyrån/LiberFörlag, Stockholm.
- Lantz, A. (1993). *Intervjumethodik*, Studentlitteratur, Lund.
- Lindh, J. (2001). Introductory Evaluation of the Swedish RealSpeak System, Term Paper for Speech Technology 1, GSLT. http://www.speech.kth.se/rolf/gslt/_papers/JonasLindh.pdf.
- McAllister (2000). *Talkommunikation*, Studentlitteratur, Lund.
- Nationalencyklopedin: ett uppslagsverk på vetenskaplig grund utarbetat på initiativ av Statens kulturråd* (1989-). Bra Böcker.
- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. och Carey, T. (1994). *Human-computer interaction*, Addison-Wesley.
- Silverman, K., Basson, S. och Levas, S. (1990). Evaluating synthesiser performance: is segmental intelligibility enough?, *Proceedings of the International Conference in Spoken Language Processing, ICSLP90*, Kobe, Japan.
- Sluijter, A., Bosgoed, E., Kerkhoff, J., Meier, E., Reitveld, T., Sanderman, A., Swerts, M. och Terken, J. (1998). Evaluation of speech synthesis systems for dutch in telecommunication applications, *Proceedings of the Third International Workshop on Speech Synthesis, ESCA/COCOSDA*, Blue Mountains, Australien.
- Talboks- och Punktskriftsbibliotekets hemsida* (2003). <http://www.tpb.se>. (20 oktober 2003).
- Trost, J. (2001). *Enkätboken*, Studentlitteratur, Lund.

A Intervjufrågor

Respondenterna

Fråga 1: Vilket läshandikapp har du? Svartaltnativ: dyslexi/synskada/röreslehinder/övrigt

Fråga 2: Vilket modersmål har du? Svartaltnativ: svenska/annat

Fråga 3: Hur gammal är du?

Fråga 4: Inom vilket/vilka ämnesområden studerar du / har du studerat? (flervalsfråga)

Erfarenheter av talböcker

Fråga 5: Har du använt talböcker? Svartaltnativ: ja, analogaja,DAISY-spelare/ja, pc med programvara för läsning/nej (flervalsfråga)

Fråga 6 a: Hur ofta har du använt talböcker i dina studier? Svar anges i procent.

Fråga 6 b: Vad är orsaken till att du inte använt talböcker på alla kurser? (öppen fråga)

Fråga 7: Är talböcker till stor hjälp för dig när du läser? Svartaltnativ: ja, mycket/ja, delvis/nej/nej, inte alls/vet ej, obestämd

Erfarenheter av syntetiskt tal

Fråga 8 a: Har du använt dig av syntetiskt tal? Svartaltnativ: ja, för läsning av kortare texter/ja, för läsning av längre texter/nej

Fråga 8 b: Hur ofta har du använt syntetiskt tal för läsning av kortare texter? Svartaltnativ: dagligen/någon gång i veckan/någon gång i månaden/mer sällan

Fråga 8 c: Hur ofta har du använt syntetiskt tal för läsning av längre texter? Svartaltnativ: dagligen/någon gång i veckan/någon gång i månaden/mer sällan

Fråga 9: Är syntetiskt tal till stor hjälp för dig när du läser? Svartaltnativ: ja, mycket/ja, delvis/nej/nej, inte alls/vet ej, obestämd

Fråga 10: Är du positivt inställd till syntetiskt tal? Svartaltnativ: ja, mycket/ja, delvis/nej/nej, inte alls/vet ej, obestämd

Subjektiv skattning

Fråga 11: Har du tidigare läst litteraturen som används i undersökningen? Svartaltnativ: ja/nej

Fråga 12 a: Upptäckte du några felaktigheter? Svartaltnativ: ja/nej

Fråga 12 b: Vad upplevde du som felaktigt? (öppen fråga)

Fråga 12 c: Upplevde du felaktigheterna störande? Svartaltnativ: ja, mycket/ja, delvis/nej/nej, inte alls/vet ej, obestämd

Fråga 13: Behövde du anstränga dig för att förstå det som sades? Svartalernativ: ja, mycket/ja, delvis/nej/nej, inte alls/vet ej, obestämd

Fråga 14: Var rösten behaglig? Svartalernativ: ja, mycket/ja, delvis/nej/nej, inte alls/vet ej, obestämd

Fråga 15 a: Var hastigheten på talet lagom snabbt, snabbare, långsammare än vad du föredrar? Svartalernativ: mycket snabbare/snabbare/lagom snabbt/långsammare/mycket långsammare

Fråga 15 b: Var talhastigheten lagom jämn? Svartalernativ: ja, mycket/ja, delvis/nej/nej, inte alls/vet ej, obestämd

Fråga 16 a: Skulle du kunna lyssna på längre texter som spelats in med denna typ av röst? Svartalernativ: ja, mycket/ja, delvis/nej/nej, inte alls/vet ej, obestämd

Fråga 16 b: Skulle du kunna använda en DAISY-talbok som spelats in med denna röst? Svartalernativ: ja, mycket/ja, delvis/nej/nej, inte alls/vet ej, obestämd

Fråga 16 c: Skulle du kunna använda en DAISY-talbok som spelats in med denna röst, om DAISY-talboken, förutom det inspelade talet, innehöll hela den tryckta bokens text i form av e-text (elektronisk text) som är kopplad till talet? Svartalernativ: ja, mycket/ja, delvis/nej/nej, inte alls/vet ej, obestämd

Parvis jämförelse

Fråga 17: Vilken av A eller B lyssnar du helst på?

Avslutande frågor

Fråga 18: Har din inställning till syntetiskt tal förändrats under utvärderingen? / Vad tycker du om det syntetiska tal som du just fått lyssna på? (öppen fråga)

Fråga 19: Vilka egenskaper anser du är viktiga hos syntetiskt tal, för att du ska lyssna på det? (öppen fråga)

Fråga 20: Har du ytterligare synpunkter på denna undersökning och dess frågor? (öppen fråga)

B Svar och kommentarer från respondenterna

Fråga 4: Inom vilket/vilka ämnesområden studerar du/har du studerat?

Naturvetenskap: 1 dyslektiker(R3); Vård: 1 dyslektiker (R6); Teologi: 1 synskadad (R5); Teknisk: 2 dyslektiker (R4, R6); Data: 1 dyslektiker (R4); Humanistisk: 2 synskadade (R5, R7); Språk: 2 synskadade, 1 dyslektiker (R5, R7, R8); Samhällsvetenskap: 1 dyslektiker, 1 rörelsehindrad (R1, R2); Juridik: 1 dyslektiker (R8).

Fråga 6 b: Vad är orsaken till att du inte använt talböcker på alla kurser?

- Orsaken till att jag inte använt talbok hela tiden är för att de varit för dåligt inlästa. Det har även varit svårt att få fram talboken i tid.(R1)
- Ibland vill jag inte använda talböcker, utan istället läsa in själv. Ibland har talböckerna varit inlästa med sövande röster, det låter som en vaggvisa, så jag somnar när jag lyssnar om jag är trött. Andra röster låter som om de druckit, de vill jag inte heller lyssna på. (R2)
- Det beror på hur böckerna är upplagda. Om de förklarar på ett lätt sätt så är det lättare att läsa, inom biologiämnen har jag läst själv, men inom kemiämnen har jag använt Victor. Rösten har inte spelat någon roll. (R3)
- Kurslitteraturen finns inte som talbok. Vi har ofta nyskrivna böcker. Har vid ett tillfälle beställt en bok i data, då tog det 8 månader att få boken inläst (augusti-mars). Jag har använt talböcker för studielitteratur till en enda kurs, då fanns all litteratur redan inläst. Jag använder talböcker till skönlitteratur, när man inte har så bråttom. När man läser tekniska ämnen behöver man inte läsa så mycket, det gäller bara att gå på föreläsningarna, då föreläsaren går igenom allt muntligt. Att sedan förstå en formel på fem bokstäver, det kan vem som helst göra utan att ha speciellt lätt att läsa. Man måste bara gå på alla föreläsningar. När jag studerar tekniska ämnen är talböcker inte till stor hjälp, men om man läser t. ex. samhällsvetenskap när det går ut på att plöja igenom en bok för att kunna sammanfatta den efteråt, då är det klart att talböcker är till stor hjälp, när man måste ta sig igenom stora textmassor. Ett annat problem är att det blir väldigt konstigt när man ska läsa en massa tekniska formler, integraltecken och x upphöjt i två och sånt där. Det gör sig inte i talboksform, man kan inte se sambanden lika lätt när man får dem upplästa för sig bara. Jag har väldigt svårt att hänga med i "Read Along", där man får boken och kassetten och ska läsa med. Då kan jag inte hänga med för fem öre, klarar inte av att hålla det koordinerat. (R4)
- Till endel kurser finns det inte inläst material, för jag läser ett nytt program, och då får jag talböckerna i slutet av kursen för det tar så lång tid för dem att läsa in, det är lite tufft att hänga med. (R6)

- Det är lättare att läsa grammatiktexter eller latin där jag kan kolla av mot punktskriften om det är långa meningar, då använder jag inte talsyntesen. I längre texter som skönlitterär roman för engelskan läser jag hellre talbok, det är jobbigt att läsa stora mängder punktskrift. Det är lättare att göra grammatiska övningar, se grammatiska strukturer, ta ut satsdelar. Det brukar vara lättare att få tag i en talbok än i elektroniskt format. (R7)
- Jag läser nästan all litteratur med talböcker, men om det inte finns inläst (rättsfall som inte är inspelade, material som läraren kopierar på kursen) så får jag läsa själv, men det är svårt och tar lång tid. Samtidigt vill jag inte förlora förmågan att läsa utan apparat, för jag har den här möjligheten på skolan, men efteråt måste jag kunna läsa utan apparat. En gång fick jag kurslitteraturen efter det att tentan varit.(R8)

Fråga 12 b: Upptäckte du några felaktigheter? Vilka?

- **Mänsklig röst:**

Helst klart bäst, men störande att han stannade så många gånger. Han pausade där det varken fanns komma eller punkt.(R6)

Jag tyckte inte att det var jämn hastighet, han läste lite fort ibland. Annars var det en ganska behaglig röst. (R7)

Det kändes litet avhugget i slutet av meningar och i slutet av de längre orden ibland. Men det låter väldigt bra tycker jag. Även om det är monotont, för det är samma tempo i varje mening. Jag tycker att det liknar endel inläsare som endel böcker jag lyssnat på, som jag slutat lyssna på för att inläsaren varit så tråkig. Och det är så när man läser en bok monotont, men det var en skön röst annars. Behaglig röst. Det sluddrade lite lätt. Felaktigheterna var lite störande, men inte särskilt mycket. Han sluddrade, han talade lite i skägget.(R4)

Långa pauser mitt i meningar, ex. ”skoldokument...är knapp...”. Inget i uttalet var störande. Bästa rösten! (R1)

Många pauser, man blir trött av att vänta. Den här rösten skulle kunna fungera på skönlitteratur men inte på kurslitteratur. Det är skillnad om man lyssnar på en roman eller om man läser kurslitteratur. Om jag ska lära mig något vill jag inte ha så mycket pauser, men de kan fungera i romaner. (R2)

Ojämma pauser, skyndade på i slutet av meningar. Det är lättast att lyssna på den här rösten. (R3)

Rösten var inte så aktiv, så om man lyssnar på honom så kommer man somna. Ibland uttalar han inte sista ordet i meningen. ”eleverna” uttalar han ”elev”. Han tar pauser som är bra. Man hör inte vad han uttalar ibland, för han säger bara några bokstäver i ordet. Han läser fort. (R8)

- **Ingrid:**

De första två meningarna var mycket bra, sedan blev det hackigt, det gick upp och det gick ner och istället för att säga ”pla:net” så sa det ”plane:t”. Hade jag inte haft texten hade jag nog blivit väldigt snurrig. Det låter som om datorn inte har tillräckligt med minne för att spela upp ljudet som det ska. (R6)

Jättebra, behaglig att lyssna på, lät realistiskt. Det enda felaktiga var att den hoppade ibland, det lät som om den stakade sig. Jag störde mig lite grann på det. (R7)

Oktavfel i bland annat ”skolporten”. Det låter som en tonåring som pratar fast det är med en gammal dams röst, för rösten spricker hela tiden. Hon känns lite halvfull, det är småsluddrigt och lite konstigt sätt att tala på, konstig rytm. Bättre på slutet av orden, att det inte gick upp och ner så mycket i tonläge. Det var helt klart störande, jag hörde knappt texten och hade svårt att koncentrera mig. Det är som när en föreläsare säger ett och samma ord många gånger, det är jobbigt, det är saker man kan haka upp sig på och det var samma med den här rösten tycker jag. Jag hakade upp mig på att det hoppade upp och ner. (R4)

Mindre hackig än (Emma och Ingmar), monoton, bättre än Ingmar och Emma. Mindre störande s-ljud. (R1)

Vet ej, upplevde inget som störande. Rösten hade bra harmoni.(R2)

Ojämn, det kom så snabbt på varandra så det lät konstigt. Det lät kul. Man skulle störa sig på det i början, sedan skulle jag säkert vänja mig. Några ord här och där behövde jag anstränga mig för att förstå. (R3)

Hon läser utan paus, eller pausen kommer på fler ställe. När det är punkt behöver man lite längre paus, för att veta att det är slut på meningen, och att det börjar en ny. Hon betonar olika på olika ord, det hänger liksom inte ihop, d.v.s. talmelodin är konstig, går emot det hon säger. Melodin ser konstig ut. Det låter som ett enda långt ord när det i själva verket är många ord. Hon tar inte paus vid komma och för kort paus vid punkt. Jag förväntar mig en annan melodi, längre pauser. (R8)

- **AnnMarie:**

Det låter mycket syntetiskt och artificiellt. Mycket obehaglig att lyssna på, låter mycket skarpt och oklart, låter som om den läspar. Ojämn läsning, nasal, chunky, läser inte jämnt. Mycket ointressant att lyssna på. Låter robotaktigt. Måste gissa mig till vad som sägs i vissa partier, är osäker på om jag gör rätt gissning. mycket obehaglig röst. Nasal, ojämt, godkänt men inte bra representation av svenska. Det kräver att en person som är svensk, med svenska som modersmål för att kunna använda en sådan röst. (R5)

Början var inte alls bra, sedan blev det bättre. Det lät som om hon pratade i näsan. Det var hackigt på några ord, låter som om datorn inte har tillräckligt med minne för att spela upp ljudet. Meningen "Tanken...läroplaner" började bra, blev värre och värre, lät ansträngt.(R6)

Det flöt på väldigt bra, och det var lätt att förstå, men det var ganska entonigt. Betoningen, det skulle vara tråkigt att lyssna på det jättelänge, det lät konstgjort, det var samma tryck på allt.(R7)

Rösten hade svårt att uttala vissa ord, ex "planet". Accentproblem. Den har mindre oktavproblem [än 5.2], men går upp och ner, speciellt på "e". Men rösten är bekvämare. Det är skillnad om man får höra en liten bit text eller om man ska lyssna på en hel roman. Om det är en roman vill man ha så att det är stabilt. Den här rösten är bättre för kortare texter, för längre behövs mer inlevelse. Visserligen behöver inte den här texten någon inlevelse, för den är så faktaspäckad. (R4)

Syntetiskt ljud, hackig, inget fel på s-ljuden. Monotont. Tänkte inte på talhastigheten för jag var för irriterad på rösten.(R1)

Tonen är mycket tråkig och lite obehaglig. Jag undrar om det går att spela den rösten långsammare utan att hon låter som om hon druckit? Ljuden var mycket störande. Rösten låter omogen.(R2)

Konstigt uttal, vissa ord blev fel t.ex. "kön". Det lät som en robot, mycket syntetiskt och ojämnt. Ibland gick det snabbare i slutet av ord. Pauserna var ojämna. Om jag inte haft texten skulle det nog varit svårare att förstå vad som sades.(R3)

Hon sjunger monotont när hon läser, och lägger melodin felaktigt. Rösten har inget fel, det är bara _hur_ hon läser som är fel. Jag kan följa talet. Hon betonar inte. Hennes sätt att läsa är som om hon sjunger. Kan det bero på vilken del av landet hon kommer ifrån? Hon tar bra pauser på punkt och komma, men säger inte när det är citat. Klarar inte av förkortningen "t. ex." utan ser det som punkter. Hon klustrar orden tillsammans, så tre ord läses som ett enda ord. Det läggs på rad i en linje och har samma nivå i hennes röst. Hon följer inte satsmelodin utan lägger felaktig satsmelodi. Rösten är inte jämn.(R8)

- **Emma:**

Det där var inget vidare. Hon hackade sig, och det gick upp och ner i själva textläsningen. Det gick lite långsamt, jag hann läsa i förväg. Rösten var inte behaglig att lyssna på. Det var skrattretande. (R6)

Mer hackig än Ingrid. Ojämn i rytmen. Den lät väldigt realistisk, det var en behaglig röst att lyssna på.(R7)

Hon kunde inte läsa texten, det var ju vissa ord som hon missade, som hon uttalade felaktigt

tycker jag. Missade ett "o" i början. "Omständigheter" lästes "ständigheter" eller nåt sånt, som inte passade in i texten.. Mycket slätare tonhöjd på det här, det var bra, men det var lite ryckigt istället, på ett konstigt sätt, även om det var pauserat med mer fart på vissa delar så var det paus på konstiga ställen, kändes onaturligt. En sak jag misstänker är att man vänjer sig vid det om man lyssnar tillräckligt mycket, det är ju anledningen till att vi känner det tal som vi använder nu som naturligt, att man blir indoktrinerad oavsett vad det är, hör man tillräckligt mycket genom öronen så rinner det ut genom munnen, det är ju så man lär sig språk.(R4)

S-ljuden mer markanta än resten, hack mellan orden. Monoton.(R1)

Det enda felet med rösten var att den tryckte på vissa ställen, hade felaktiga betoningar: ak_tiv_, fri_tid_. Annars bra tempo och bra ljud.(R2)

Det lät som om hon hade dialekt. Hon hade konstiga uttal på vissa ställen.(R3)

Paus för punkter och komma är ok. Men melodin måste man vänja sig vid för att kunna hänga med. Kanske beror det på vilken del av landet kan komma ifrån, om man är van vid det så är det kanske ok. En sak, som jag även har problem med i mina talböcker, är att hon har skarpa s-ljud. Om man har hörlurar när man lyssnar med Victor så är det mycket störande. Det låter som om hon är ovan läsare, som om hon läser för första gången. Det låter inte som jag förväntar mig.(R8)

- **Ingmar:**

Lite bättre än den förra [Emma], men fortfarande hackigt. Säger "ettuseniohundra"istället för "nitiohundra"vid årtal, om jag inte haft texten framför mig hade jag nog blivit väldigt fundersam på vad som sades. Lite långa pauser. (R6)

Rösten var ganska entonig, läste lite för långsamt, men den var ok att lyssna på.(R7)

Den där var riktigt bra, men det låter som om han pratar med en klädnyppa på näsan, det var väl det enda, det var det man tänkte på, och så var det hans s-ljud som var lite roliga, annars var det inget, själva uttalet var jättebra, även om det var vissa ställen som var lite tveksamma. Jag tycker att den sista rösten Bm/Ingmar är väldigt bra. Problemet är att det känns som om rösterna är inspelade bokstav för bokstav. De verkar inte känna till hur ett ord ska uttalas.(R4)

Mer hackigt, orden flyter inte jämnt, lite problem med s-ljuden. Jag tror det är den här rösten jag har hört en gång förut. Hackigheten var enormt störande. Monoton.(R1)

Rösten är inte bra och inte dålig. Den har samma rytm hela tiden och det är väldigt tråkigt. Jag hittade inget passande tempo trots att hastigheten ändrades. Rösten var värre än den andra mansrösten.(R2)[Anmärkning: Victor-spelare användes vid lyssningen]

Jag upplevde inget felaktigt, men det lät väldigt gammeldags. Det lät väldigt roligt, men det var inte störande. Det lät som om man gått tillbaka i tiden. Det var lite svårt att följa med pga detta.(R3)

Lagom hastighet. Melodin är inte den jag föredrar. Hans sätt att läsa, det tar lång tid innan han tar paus. Alla ord uttalar han på samma sätt. Långa ord, korta ord, entonigt, monotont. Jag förväntar mig betoning för att ge ord värde. Jag vill att han ska betona, visa hur viktigt ett ord är i texten. Han läser på samma nivå. Så jag skulle inte välja den här rösten pga melodin är så dålig. För att hänga med beror av hur man uttalar ordet och hur man betonar, och hur ordet hänger i meningen. Han uttalar rätt men betonar inte. När man betonar så får man smak på språket.(R8)

Fråga 13: Behövde du anstränga dig för att förstå det som sades?

- **Mänsklig röst:**

Behövde anstränga mig lite för volymen var låg och läsningen var ojämn.(R7).

Ja, mycket, eftersom jag blev vilse när det inte lästes som jag förväntade mig.(R8)

- **AnnMarie:**

Ja, delvis, eftersom hon krymper en mening till ett enda långt ord. (R8)

- **Emma:**

Ja, delvis, för att den uttalade så konstigt så var jag tvungen att tänka efter.(R4)

Ja, delvis. Jag är inte van vid melodin när hon pratar.(R8)

Nej, men om jag bara hade lyssnat utan att ha texten framför mig så hade det nog varit mkt svårare.(R6)

- Ingmar:

Ja, delvis. För att få helheten av det han läser så måste jag anstränga mig eftersom det saknas smak i språket, brist på andningspauser. Jag hinner inte hänga med när det är så monotont och entonigt.(R8)

Fråga 16 a: Skulle du kunna lyssna på längre texter som spelats in med denna typ av röst?

- Mänsklig röst:

Nej, hastigheten är för ojämn.(R7)

Jag skulle kunna lyssna på längre texter som man korrigerar de långa pauserna i talet.(R1)

Om det inte fanns någon annan möjlighet skulle jag lyssna på text med den här rösten.(R2)

- AnnMarie:

Jag skulle inte lyssna på den här rösten om det var enda alternativet. Rösten är som ett slag mot kroppen, mkt obehaglig. En röst ska vara mjuk, då kan man ta den till sig lättare.(R2)

Om det fanns ett annat alternativ med en bekvämare röst där jag inte behöver anstränga mig för att följa med så skulle jag välja det andra. Om jag fick en cd som är inspelad skulle jag vara så illa tvungen.(R8)

Nej, det var för mycket näsprat.(R6)

Ja, om det var faktaböcker, men inte för skönlitteratur. Ok för när man ska lära sig saker, där man ska memorera fakta, men inte för drama där det händer spännande saker.(R7)

- Emma:

Nej, det är för störande med hackigheten, det är svajigt. Intonationen var lite konstig, det var för mycket betong på många ställen, det lät överdrivet.(R7)

Vill inte använda den här rösten för att läsa kurslitteratur, men kanske för andra texter då jag inte är lika beroende av att ta in informationen som finns i texten.(R1)

Ja, delvis. Men jag måste då begränsa mig till att inte lyssna med hörlurar. Om jag hade annat, bättre alternativ skulle jag välja en annan röst. t.ex. med mer riksuttal som jag är van vid. Jag är inte bekant med skånska dialekten, det skulle vara svårt att förstå det också.(R8)

Nej, inte alls. Skulle ha svårt att hålla mig för skratt.(R6)

- Ingmar:

Om jag behövde så skulle jag kunna göra det. Jag har hört sämre, jag har sämre hemma, med mäsklig inläsare, han läser jättetråkigt och jättekonstigt, så man somnar efter fem minuter. Det här är också smått sövande, det läser stelt det som står utan att känna någonting för texten. Det ger ett annat intryck om den som läser läser för att den tycker att det är kul.(R4)

Ja, det kändes bekvämt att lyssna på.(R7)

skulle endast kunna lyssna på kortare texter med den här rösten.(R8)

Ja, delvis. Så länge jag kan ta pauser med jämna mellanrum. Skulle vilja snabba upp talet.(R6)

- Ingrid:

Jag vill inte använda en sådan röst för kurslitteratur, eftersom det är krävande i sig att läsa kurslitteratur. Monoton röst och sånt som kan göra mig irriterad vill jag undvika. Däremot kan jag tänka mig att läsa annan typ av texter med denna röst. (R1)

Jag skulle föredra att en människa läste upp texten.(R3)

Ja, absolut. den är behaglig, det låter realistiskt. Skulle kunna använda rösten till långa texter, talbok och talbok med länkat tal-text. (R7)

Nej, om jag hade möjlighet att välja skulle jag inte använda det. (R8)

Fråga 16 b: Skulle du kunna använda en DAISY-talbok som spelats in med denna röst?

- **Mänsklig röst:**
Om jag kunde välja en annan röst som är bättre så skulle jag inte välja den här, men om jag inte hade något alternativ skulle jag använda det som fanns.(R8)
- **AnnMarie:**
Nej, jag sitter hellre och tragglar själv. Om det inte varit näsprat så hade det gått.(R6)
Nej, det skulle bli lite för platt och tråkigt. DAISY är avkoppling mot det syntetiska talet på datorn, DAISY använder jag för mer skönlitterära texter.(R7)
- **Emma:**
Nej, inte med den rösten, inte på det sättet som det lästes, det var väldigt hackigt. Om det var enda alternativet hade det väl gått bra. (R6)
- **Ingmar:**
Vet ej, just i inläringssyfte skulle det vara irriterade, men på barnbok eller skönlitteratur så skulle det nog gå, den var lite väl störande för kurslitteratur. Det skulle kräva för mycket av mig.(R6)
- **Ingrid:**
Nej, inte som kurslitteratur. Det skulle kanske gå i nödfall.(R6)

Fråga 16 c: Skulle du kunna använda en DAISY-talbok som spelats in med denna röst, om DAISY-talboken, förutom det inspelade talet, innehöll hela den tryckta bokens text i form av e-text (elektronisk text) som är kopplad till talet?

- **Mänsklig röst:**
Nej, hastigheten är för ojämn. Med länkat tal-text skulle jag dock delvis kunna tänka mig att använda den här rösten, för då skulle jag kunna kolla upp texten om det var något jag inte hängde med på.(R7)
Länkat tal-text i en vetenskaplig bok om man hade möjlighet att slippa formlerna, genom hänvisning till figur och talboken strukturerad så att man lätt kan hoppa över dem, och läsa texten emellan. Det skulle vara en stor fördel för ibland vill man bara läsa ett par sidor, beror dock på vilken typ av litteratur det är.(R4)
Ja, delvis. Det skulle nog gå snabbare att läsa då för då skulle jag se var han var i texten. Men jag skulle bara använda det om det inte fanns något bättre alternativ, för det skulle det bli en krock mellan vad jag ser och vad jag hör.(R8)
- **AnnMarie:**
Nej, näspratet är för störande.(R6)
Jag skulle endast kunna använda denna röst för att få information.Man kan inte njuta av att lyssna på denna röst. Kräver att lyssnaren kan svenska mycket bra. Vill ha punktstöd till rösten. (R5)
- **Emma:**
Bättre med länkat än utan. (R3)
Jag skulle inte välja den här rösten, skulle endast lyssna på den som jag inte hade några alternativ.(R8)
Ja, delvis. Mer intresserad av det än utan länkar, om jag hakar upp mig eller vänder på ord eller inte hör så kan jag se samtidigt som jag hör. (R6)

- Ingmar:
 Det vore bra att ha tal och text länkat eftersom det var svårt att följa med, eftersom det lät så gammeldags. Men sånt kanske man lär sig.(R3)
 Jag kunna absolut kunna använda den här rösten för länkat text-tal, för kortare texter. För längre texter skulle jag aldrig använda länkat text-tal, men jag skulle förmodligen kunna läsa en kortare bit så ur en längre, och bara läsa det som jag behöver. Som ett uppslagsverk. Man behöver oftast inte läsa hela kursboken [i de ämnen jag läser], utan tar bara de två sidor man behöver, väntar en vecka och läser två sidor igen.(R4)
 Skulle inte välja det, men om jag fick en cd inläst med denna röst skulle jag inte ha något val.(R8)
 Ja, delvis. Det skulle nog gå.(R6)
- Ingrid:
 Jag skulle välja den första rösten istället, om jag fick välja [d.v.s. den mänskliga rösten].(R4)
 Ja, det skulle vara lättare att följa med om det är länkat.(R3)
 Ja, delvis. Jag tror det.(R6)
 Ja, delvis, för då kan jag inte tappa raden, då skulle det gå att använda. Ibland hoppar jag in på fel rad när jag läser, när jag byter rad. Med länkar skulle jag veta var jag är. Det hjälper mig att se att det finns kommatecken och punkt, även om jag inte kan höra det. Om jag hade möjlighet så skulle jag välja en röst med vilken jag känner mig trygg, vilket jag inte gör med den här rösten. [svarade nej på a och b] (R8)

Fråga 18: Har din inställning till syntetiskt tal förändrats/ Vad tycker du om det syntetiska tal som du just fått lyssna på?

Ändrade sig från att inte alls vara positivt inställd till syntetiskt tal till att vara lite mer positiv.(R1)
 Jag tyckte om en av mansrösterna, antagligen den mänskliga, de andra lät lite som på rutin, robotaktiga. de hade alltid samma uttryck.(R2)
 Rösterna varierade väldigt mycket, någon röst lät väldigt syntetisk, någon gammaldags, någon som vanligt.(R3)
 De syntetiska talet är mkt bättre än vad jag trodde innan. De syntetiskt rösterna kommer säkert kunna bli bra i framtiden, men det är ingen av dessa som jag vill ha.(R4)
 Jag är lika positiv till syntetiskt tal nu som före. Syntetiska röster ger större möjligheter att förbättra situationen, styra vad jag får för röst, man kan använda olika röster för olika situationer, på så vis kan vi talboksanvändare få bättre situation.(R5)
 Jag kände igen en av rösterna. Det skulle säkert funka, bara man får det att flyta ordentligt. (R6)
 Nej, samma positiva inställning.(R7)
 Det är svårt att hänga med en röst som har dialekt. Om det bara är dialekt så är det inte störande, men när det är utan pauser och monotont så är det svårt att hänga med. Jag förväntar mig att höra något annat än det som jag får höra.(R8)

Fråga 19: Vilka egenskaper anser du är viktiga hos syntetiskt tal, för att du ska lyssna på det?
 Öppen fråga.

- Jag vill inte ha några irritationsmoment, d.v.s. talet ska vara flytande, helst ej monotont, det ska ha rätt uttal och inget metalliskt ljud. Vill ha det mer varmt, mer som mänskligt tal.(R1)
- Jag vill ha harmoni, rösten ska visa känsla, inte maskinljud och vara monoton. Det är skrämmande att höra maskinljud (som en av kvinnorrösterna och en av mansrösterna), de låter som man brukar visa att röster låter i rymden, man känner ingen gemenskap med dem. Det är viktigt att välja en röst som ligger bra i hjärtat, som fungerar som en pusselbit och passar till kroppen och sitter bra på plats. För olika ämnen behöver man olika röster, t.ex. läroböcker och romaner kan ha olika röster.(R2)

- Jag vill inte ha pauser mitt i meningarna, jag vill ha jämna mellanrum mellan varje ord, helst ska samma takt hållas genom hela meningen. Det syntetiska talet ska låta så mänskligt som möjligt.(R3)
- Det ska låta som vanligt tal, det ska kunna uttala längre, svårare ord, och även korta ord, utan att bokstavera. Det ska klara av termer, tonläge. Huruvida en röst är behaglig eller ej har med personliga preferenser att göra. Till skönlitterära texter vill jag att rösten ska vara dynamisk, att den ska kunna tala tystare på vissa ord, betona vissa ord, kunna leva sig in i berättelsen. En faktabok däremot kan vara mer monoton, men då är problemet att det låter sövande. Det gäller även vanliga, mänskliga inläsare. det är trevligast med mänskliga inläsare.(R4)
- Vill att talet ska vara naturligt, lätt att förstå, ha rätt uttal eftersom talböcker är till för alla. Det ska inte spela någon roll om man har svensak som modersmål eller ej. Jämfört med mänskliga inläsare som kan tala med dialekt [vilket kan göra det svårt för vissa att förstå vad som sägs]. Man vill inte bli tvingad att lyssna på en viss dialekt. Det är inte viktigt att ha mänskligt inläst tal i alla situationer i en DAISY-talbok, det är viktigast att man förstår. Samtidigt är det viktigt att man kan koncentrera sig under en längre period, om man läser längre texter. (R5) Han tror att man i framtiden kommer använda talböcker även för icke läshandikappade, t.ex. i en bok har man att man har mjukvaran installerad med cd-spelaren i bilen. Att man beroende på vad man vill lyssna på, kan reglera själv vad man vill. Man behöver de billigare, sämre talsynteserna för att nå en större lyssnarpublik.(R5) Jag föredrar de mer mänskliga rösterna, tror mer på dem. Det är även viktigt att det finns många olika röster att välja mellan, för olika texter, olika situationer, man bör kunna kombinera flera olika röster till en talbok, så att t.ex. en pjäs kan läsas upp med olika röster för de olika rollerna.(R5)
- Det ska flyta bra, och det ska inte låta som om datorn inte har tillräckligt med minne för att köra uppläsningen, inget hack. Det ska vara en behaglig röst och inget nästal.(R6)
- Det ska låta mänskligt, uttala korrekt och ha bra intonation, så att det låter verkligt, så nära en människa som möjligt, så blir det behagligt att lyssna på, särskilt om man ska lyssna på längre texter.(R7)
- Det är inte bra att ingen av rösterna säger till när det är parentes eller citationstecken. Om jag lyssnar utan att ha boken framför mig vet jag inte när saker står i parentes eller inom citationstecken. Jag vill att sånt läses ut. Jag vill att den som läser läser tydligt, klart, betonar, har varierande pauser (punkt och kommatecken) andningspauser som visar när något nytt kommer. Om det kommer något störande så hakar jag upp mig på det och kan inte koncentrera. Då är det ungefär som när något yttre stör, då kan man flytta sig till en annan miljö, men om det är talet man stör sig på så kan man inte byta bort det man stör sig på.(R8)

C Meningar som användes vid den parvisa jämförelsen

1. Vi har alla varit elever en gång.
2. Därför kan det verka litet överflödigt att på nytt behöva lära sig, med hjälp av en teoretisk lärobok, på vilket sätt det upplevs.
3. En del år skiljer oss och eleverna, år med utbildning och personlig mognad för vårt eget vidkommande, samtidigt som samhällsutvecklingen har haft sin gång.
4. "Glömskan" skyller vi också på att vi är upptagna av andra saker och ting när vi kommer in i lärarrollen.
5. Hur skall vi åter upptäcka eleven från vår position bakom katedern?
6. Men vägen är längre. Man måste tränga igenom en urskog av sociala regler och socialt samspel.
7. Man måste se eleven i ett större sammanhang, som en människa med förväntningar, glädjeämnen och besvikelser, som en människa med förankring i ett hem och i en miljö utanför skolan.
8. Som lärare måste man känna sin egen situation så väl att man för ett ögonblick kan skaka av sig sina egna, nära problem för att möta elevens.
9. Det som var problem i början blir rutin och nya problem dyker upp.
10. Lärares och elevens värld är inte nödvändigtvis två åtskilda världar.
11. I detta kapitel skall vi i korta drag beskriva vad dessa problem kan bestå av.
12. Går vi till de officiella motiven för skolans verksamhet finner vi att den har en varierande uppgift.
13. Skolan skall förmedla allmänkunskap, men kanske väl så viktigt är det att den skall hjälpa till att uppfostra eleverna.
14. Skolan skall förmedla respekt för dem som tänker annorlunda och tolerans gentemot andra kulturer och traditioner än den egna.
15. Men individualiseringen i skolan har också ett mer politiskt-ideologiskt motiv: Massundervisning av den typ man finner i skolan innebär alltid en fara för likriktning, ja, kanske till och med indoktrinering av eleverna, som det kan vara nödvändigt att finna en motvikt till.
16. Att skolan skall stimulera och odla individuella egenskaper och individuell frihet skall också ses i sammanhang med den rotfasta plats som individualistiskt tänkesätt har i vårt västerländska samhälle.
17. Det man kan läsa sig till om skolans uppgift i officiella skoldokument är knappt halva sanningen om de uppgifter skolan faktiskt har.
18. Nya ämnesområden kommer ständigt in som en följd av nya samhällsproblem t. ex. konsumentfrågor, miljö- och naturskydd, alkohol- och drogproblematiken, data- och informationsteknologi, uppfostran till

”aktiv fritid”.

19. Den är samhällets stumma tjänare.

20. En av de tyngsta uppgifterna skolan har i dag och som inte är så synlig för utomstående är förvaringsfunktionen.

21. Men individualiseringen i skolan har också ett mer politiskt-ideologiskt motiv: Massundervisning av den typ man finner i skolan innebär alltid en fara för likriktning, ja, kanske till och med indoktrinering av eleverna, som det kan vara nödvändigt att finna en motvikt till.

22. Vidare skall skolan ha nyttoperspektivet för ögonen: eleverna skall få sådan undervisning och uppfostran att de skall bli till nytta för både hem och samhälle.

23. Att skolan skall stimulera och odla individuella egenskaper och individuell frihet skall också ses i sammanhang med den rotfasta plats som individualistiskt tänkesätt har i vårt västerländska samhälle.

24. Den förlängda skoltiden tjänar inte bara ett nationalekonomiskt syfte med att lyfta kunskapsnivån.

25. Motiven för detta individinriktade perspektiv kan vara flera: man vill försäkra sig om att ingen elev blir ställd inför krav den inte kan orka med. Därför är det viktigt att den enskilda elevens förutsättningar bildar utgångspunkt för skolans undervisning.

26. Bråk och oro på lektionerna, leda och brist på koncentration bland eleverna, oacceptabelt elevuppförande och ständiga konflikter är problem som ofta rapporteras.

27. Även om de flesta lärare rent ideologiskt helhjärtat sluter upp kring individualiseringsprincipen, upplevs det som svårt att genomföra anpassad inläring i praktiken.

28. Även om detta är allvarliga problem för dem som står mitt uppe i dem, är det dock till stor del problem som kunde ha varit lösta med bättre ekonomiska, materiella och fysiska resurser.

29. I några åldersperioder ser man vissa särdrag, som t ex trotsåldern i 3-årsåldern, konflikterna med vuxengenerationen i puberteten osv.

30. Några lämnar stor plats till undervisning och evaluering, medan andra anser att elevernas uppförande, karaktärsformning, miljö och trivsel är det viktigaste som skall behandlas.

31. Utgångspunkterna kan variera men vissa avsnitt är vanliga; inläring och minne, motivation, litet om hur barn och unga tänker och känner i de olika åldersstadierna, litet personlighetspsykologi och inte minst betydelsen av det sociala sammanhang som hela undervisnings- och fostransprocessen försiggår i.

32. Det psykologiska vetenskapsfältet är stort och överskådligt.

33. Det är få teorier som förfäktar extrema, renodlade synpunkter i dessa frågor men de flesta har en tyngdpunkt i den ena eller andra riktningen.

34. Även om framställningen är starkt förenklad, kan den likväl vara en nyttig orientering för läsare som är obekanta med psykologiämnet.

35. Överfört till psykologin leder detta till en objektivisering av människan.

36. Fenomenologisk psykologi har, med rätt eller orätt, avvisats som spekulativ för att den saknar etablerad objektivistisk vetenskapsstatus.

37. Individerna blir på detta sätt en slags ”black box”, en maskin, som reagerar automatiskt på yttre stimuli. Behavioristerna menar att det finns en lagbundenhet mellan stimulus och respons och forskningens mål är att finna dessa lagar.

38. Utifrån kraftlagen lyckas vi få den fartökning vi behöver för att variera kraften.

39. En viktig princip som präglar behavioristisk motivationsteori är nyttoprincipen (den utilitaristiska principen) : individen gör det som lönar sig i eget intresse.

40. Under det att den rena objektivismen begränsar sig till att studera yttre, observerbara förhållanden av mänskligt beteende, är subjektivismen öppen för den mänskliga, värderande medvetenheten både hos den som observerar och den som observeras.

41. Av psykologiska teoretiker som tar utgångspunkt i ett fenomenologiskt perspektiv kan nämnas Kelly och Heider.

42. Som utanför stående observatörer kan vi bara tolka en annan person utifrån personens handlingar, av det vi vet om personens förhistoria och bakgrund och inte minst det vi uppfattar som personens upplevelse i ögonblicket.

D Kommentarer till den parvisa jämförelsen

Kommentarer inom hakparentes klargör vilken röst som respondenten kommenterat.

Emma

- Det som är mellan det som låter fel låter mkt bra, och talhastigheten är jämn. Det svängde när hon talade, det är svårt att hålla sig för skratt när man lyssnar på henne. (R3)
- Behagligare [än AnnMarie]. Har behagligare röst, men [AnnMarie] klarar av att läsa texten bättre. Har helt kass pausering. Rösten är jättebra med uttalet är fruktansvärt. Bättre än [Ingmar] trots konstigt uttal på ”läroplaner”. Härligare röst än [Ingmar m. lexikon], hörde problem med uttal. (R4)
- Mkt bättre, mer användbar [än naturligt tal]. Trevligare röst än [AnnMarie], jämnare distribution av fraser Dialekt Uppsala/Stockholm, ojämn, låter drucken, fraserna sitter ihop vilket är dåligt, lättare att förstå [än den mänskliga rösten]. Mer ojämn än Ingrid, tonhöjden slår upp ibland, bruten pitch, förvirrande. Mer spännande, dynamisk röst men ojämn, ingen bra frasuppdelning eller tolkning, gav ingen helhet. Mer dynamisk, lättare att förstå än Ingrid, bra hastighet, röst och stil Bättre uttal än Ingmar, mer dynamisk vidd, renare. Texten var dock väldigt krävande för båda rösterna, för komplicerad. Interpunktionen gjorde det svårt att förstå vad som sades. Mer mänsklig [än Ingmar m. lexikon], lågt register vilket gjorde det svårt att höra vad som sades. Dynamisk, kanske lite för mycket, ej bra för akademiska texter. Trevligast röst [jämfört med Ingmar]. Ojämn men trevligare volym, mer mänsklig, jag tolererar mer [än av AnnMarie m. lexikon]. Modern, lämplig röst för talbok (R5)
- Lät bäst om än något hackigt [än Ingrid]. Rösten var bra, stundvis även bättre uppläsning [än den mänskliga rösten]. Sätt denna [Emmas] röst på den andres [AnnMaries] läsning, de är nog lika jobbiga att lyssna på i längden. Problem med uttal på ”drogproblematiken” Något snabbare [än Ingmar], problem med bokstavsuttal Lät som om hon skulle kvävas. Rösten var inte bra, den var hackig, sätt denna [Emmas] röst på den andres [AnnMarie m. lexikons] text. Rösten behaglig men svår att förstå.(R6)
- Tar mer hänsyn till interpunktion [än Ingmar], gjorde paus på rätt ställe Bättre intonation [än Ingmar], med tryck på rätt ställen. Bättre [än AnnMarie], mer varierad intonation, roligare att lyssna på, mer realistisk Bättre volym [än talboken], gjorde inte så många pauser, mer flyt.(R7)
- Verkar sakna punkt och komma, men rösten är ok. Läser bra men har pauser på fel ställe. För skarpa ”s”. Melodin går inte ihop med ordet. Låter som om hon snabbar sig när hon läser, låter inte

lugn. Låter bra men med konstig betoning exempelvis på problem". Man uppfattar inte riktigt vad som sägs. Felbetoning (R8)

Ingrid

- Betoningsfel(R3)
- Hackig, hade svårt med ändelserna. Har sämre uttal [än AnnMarie m. lexikon], men bättre röst. Ojämn i kvaliteten på uttalet av orden. Har mer tonvariation [än Ingmar], vilket inte är bra. Spricker, hoppar när den talar, låter som en tonåring i målbrottet. Har trevligare röst [än AnnMarie], det är enda skillnaden. Lika dålig [som AnnMarie m. lexikon] men trevligare tonläge. (R4)
- Mer personlig, reflekterande, lättare att förstå, jämnare tonal räckvidd, jag känner mer för rösten [än för AnnMarie]. Mest likt mänskligt tal, kräver inget ytterligare av lyssnaren, klar, ren, bra tekniskt. För skönlitteratur mer intressant röst, dock inte lika bra analys av boken, kräver ändringar av punkt och komma, låter som dialekt från västsverige, ev lämplig för text skriven av sydsvensk. Klarare [än Ingmar], mer som mänskligt tal, låter dock klustrat Mer naturlig [än AnnMarie], lite syntetiskt, ojämnt, reflekterar ej texten. Inläst på ett standardiserat sätt, närmre mänskligt tal. Mycket bättre [än Ingmar], mer mänsklig uppläsare, inflection och tal är mer jämnt, bra för en bok med mkt information i. Ok med nöjesläsning men jag är inte nöjd med den. låter dämpat, svårare att förstå, Svår att förstå, mkt oklart, lät förvirrat. Bättre, klarare, mer intressant och personlig, denna röst kan man använda för både längre och kortare texter. Trevligare, problem med inflection, mer spännande, långsammare. Låter mänsklig, mer jämnt mellanum mellan meningarna.(R5)
- Obehaglig röst, men läste bäst. Låter som om datorn inte har tillräckligt med minne. Början var ok, sedan lät det som om läsaren inte kunde hålla sig för skratt. Flöt bättre. Lät full i skratt, sätt Ingrids röst på Ingmars läsning. Sätt denna [Ingrids] röst på den andres [Ingmars] text. Rösten är bättre. Gillar den rösten bättre, men det hackar för mycket. Läste sämre [än talboken], men har bättre röst. (R6)
- Lät mer naturlig. Hackig, flöt ojämnare, Hackar. Betonade fel, ojämn talhastighet. Ojämnare läsning [än AnnMarie]. Jämnare intonation, gick inte upp och ner så mycket [som Emma]. Pausade på rätt ställe. (R7)
- Har fel betoning. Läser tydligt, betonar på rätt ställen. Uttalet är otydligt. Betonar på fel ställen, det känns inte rätt. Det låter som en sång snarare än uppläsning. Vissa ord har felaktigt uttal och betoning. kan det vara dialekt? Men rösten är behaglig. Monoton, rikssvenska. Läser som om det kommer vågor som blockerar det hon ska säga. Ungefär som stammning. det låter hackigt. Betonar för mycket. (R8)

AnnMarie

- Lät syntetiskt (R3)
- Den andra rösten [Emma] har behagligare röst, men denna [AnnMarie] klarar av att läsa texten bättre. Hackig och vass. Låter vassare [än Ingmar]. Kan inte uttala "kön". Har kortare pausering och läser snabbare [än AnnMarie m. lexikon]. [AnnMarie lästes först och AnnMarie m. lexikon efteråt](R4)
- Nasal, syntetiskt, ojämnt, svårare att förstå [än Ingmar m. lexikon]. Låter klarare, renare, bättre [än AnnMarie m. lexikon]. Båda rösterna är bra [AnnMarie lästes upp efter AnnMarie m. lexikon]. Låter gammal och omodern. Ej bra för långa texter, låter metallisk och artificiell. Bättre volym och

hastighet, mekanisk. Ger info, torrare, mer ointressant, låter objektivt. Mer personlig [än Emma], ev. mer jämnt tal. Svårt att förstå, opersonlig, eventuellt ok för akademiska texter. Alltför syntetiskt. Uppfyller inte mina förväntningar på vad jag skulle höra. Låter syntetiskt, svårt att förstå vad som sades. Lättare att förstå [än Emma], men dålig röst, men jag hade lättare att förstå vad som sades. Inte så bra frasstruktur, kräver mer av den som lyssnar, passar bättre för litteratur. Vill att rösten utvecklas.(R5)

- Jag vill inte lysna på den i längre sammanhang. Svårt att hänga med i, det lät som om den tappade delar av orden. Lättare att förstå [än Ingmar m. lexikon]. Sätt den andres [Emmas] röst på dennes [AnnMaries] läsning, de är nog lika jobbiga att lyssna på i längden.(R6)
- Bättre [än Ingrid] även om långsammare, entonig, lättare att lyssna på, behaglig. Sämre uttal [än Ingmar m. lexikon]. Jämnare läsning [än Ingrid]. Jämnare hastighet. (R7)
- För långa pauser, saknar betoning, låter som om hon sjunger entonigt när hon läser. Har konstig tonhöjdsvariation. Den höjer istället för att sänka tonhöjden i slutet av meningar. Det låter inte bekvämt. Verkar ha förstärkta r"[jämfört med AnnMarie m. lexikon]. Saknar betoning, monotont läst. Liten tonhöjning i slutet av meningen. (R8)

Ingmar

- Betoningsfel(R3)
- Svajar på bland annat "skola". Har skönare röst [än Emma]. [Ingmar] har skönare röst [än AnnMarie m. lexikon], pausering och uttal lika. Den andra rösten [AnnMarie m. lexikon] bättre än denna [Ingmar], för jag hörde inte vad sista ordet var [med Ingmar]. Har behagligare röst [än AnnMarie]. Kan inte uttala "kön" (R4)
- Mer pauser, ger mer mänsklig tolkning av texten, är lättare att förstå [än AnnMarie m. lexikon]. Renare/klarare [än AnnMarie], mer mänsklig, har för långa pauser. [Ingmar] låter mer personlig än [Ingmar m. lexikon], dess inflection är mer. standardiserat, låter mer likt mänskligt tal. Lättare att förstå och tolka [än Emma]. Endel ord är feluttalade. Klarare, lättare att se de olika delarna, uppfyller mina förväntningar på vad jag skulle höra, [till skillnad från AnnMarie]. Läses upp så som jag vill, bra frasuppdelning, lättare att komma ihåg vad som sades. [Ingmar lästes upp före AnnMarie] För ojämn inflection, upphackad, svårt att se helheten av det som sades. Läser mer proficient, bättre teknisk presentation [än Ingrid]. Bättre uttal, lite långsam, klarare än det mänskliga talet, uttalet mer standardiserat.(R5)
- Behagligare röst [än Emma]. Lät mkt bra. [Ingmar] hade behagligare röst [än AnnMarie m. lexikon], men mkt hackigt. Både röst och läsning var mest behaglig [jämfört med AnnMarie]. Dock mkt svårförståeligt med "kön".(R6)
- Lät mer naturlig [än AnnMarie]. Jämnare hastighet och intonation [än det mänskliga talet]. Jämnare rytm [än det mänskliga talet]. Behagligare [än Ingrid], hackar inte. Mer realistisk, mindre entonig. (R7)
- Tonhöjdshöjning på slutet, då jag vill att den ska gå er. Läser lite långsamt, sjunger monotont utan tryck. Läser utan paus, tonhöjden går upp på slutet, jag vill att den ska gå ner. Avsaknad av tryck. Betonar inte ord. (R8)

AnnMarie med lexikon

- Kvinnorösten har högre volym, märker den tydligare [än Ingmar m. lexikon]. (R1)
- Lät konstgjort (R2)
- Mer burkig [än talboken]. Bra pausering, dålig röst. Bättre uttal, sämre röst [än Ingrid]. [AnnMarie m. lexikon] har längre pausering och är släpigare [än AnnMarie]. [Ingmar] har skönare röst [än AnnMarie m. lexikon], pausering och uttal lika. Bättre [än Ingmar], för jag hörde inte vad sista ordet var med den andra rösten [Ingmar]. Bättre [än Ingmar m. lexikon], även om det låter burkigt. Vass. (R4)
- Syntetiskt Syntetiskt, mekaniskt, onaturligt. Långsammare, syntetiskt, tolkar mer standardiserat, låter renare. Lättare att förstå, gillar det inte, ok för korta texter. Ok för kort text, men inte bra. Bra. Lät lite klarare, förbättrat [jämfört med AnnMarie]. Syntetiskt artificiell. Mer syntetisk . Gammeldags. Mer syntetisk [än Ingmar]. Har begränsad inflection . Låter gammeldags, låter inte mänsklig. Skulle aldrig använda denna röst.(R4)
- De lät lika [AnnMarie och AnnMarie m. lexikon]. De [AnnMarie och Ingmar] är båda obehaglig röster. Skulle inte stå ut med rösten, men hade bättre flyt [än den mänskliga rösten]. Sätt den [mänskliga] rösten på dennes [AnnMarie m. lexikons] text. Fick texten att flyta, skulle ej klara av att lyssna på den rösten. Lät bäst [jämfört med AnnMarie]. Ingen var bra för längre texter. Ingen av rösterna var bra. Den andre [Emma] var hackig och denna [AnnMarie m. lexikon] nasal, sätt den andra [Emmas] rösten på dennes [AnnMarie m. lexikon] text. Ingen var bra. [AnnMarie m. lexikon] läste bäst, medan den andre [Ingmar] hade behagligare röst men mkt hackigt. Jag gillade inte rösten, men den andre [Ingmar m. lexikon] var värre.(R4)
- Bättre intonation, bättre pauser. Uttalade mer korrekt [än Emma], bättre intonation. Mer realistisk intonation och pauser mer som jag själv skulle ha läst det, det lät som om den förstod vad den sa utan att bara babbla. Mindre hackig [än Ingrid], flöt jämnare, lät mer naturlig. Långsam, entonig. Läste jämnare, tyckte bättre om den [än den mänskliga rösten]. Läste med ett mer statiskt tempo [än den mänskliga rösten]. Tyckte inte att det var någon skillnad [mellan AnnMarie och AnnMarie m. lexikon]. (R7)
- Hörs bra men tonhöjden höjs på sista ordet. Läser med entonig röst och monotont. Tonhöjning i slutet av meningen, där jag vill ha en tonsänkning. Tonhöjden är hackig, jag vill ha mjukare övergångar mellan högre och lägre tonhöjd. Läser med dialekt. Läser utan betoning och tryck. Tonhöjdshöjning på slutet, då jag vill att den ska gå er. Entonigt till slutet då tonhöjden går upp istället för att gå ner, som jag vill.(R8)

Ingmar med lexikon

- Bäst tekniskt [jämfört med AnnMarie m. lexikon].(R1)
- Jag föredrar denna [Ingmar m. lexikon] framför den andra rösten [Ingmar], men ingen av dem var bra. Mest naturlig [jämfört med AnnMarie m. lexikon]. (R3)
- Bra. [Ingmar m. lexikon] uttalar sista ordet bättre [än Ingrid]. Jämmt dålig. Behagligare [än AnnMarie], ingen annan skillnad. Föredrar denna [Ingmar m. lexikon], men båda [Emma och Ingmar m. lexikon] har tonhöjdsproblem. De [AnnMarie m. lexikon och Ingmar m. lexikon] har lika uttal men olika röster.(R4)

- Lättare att förstå än talboken, jag vill ha den uppläst så, lämplig för lyssnare som inte har svenska som modersmål. Bättre uttal, mer dynamisk, bättre tolkning [än Ingmar]. För akademiska texter, interpunktionen är bättre, bättre analys av texten. Mer syntetisk Lämplig för akademisk text, långsammare, jämnare Bra frasstruktur, andra delen av texten är mkt bra. Tolkade texten mer standardiserat. Tolkar texten, större inflection vilket är bra, låter inte mänsklig. Bättre för tolkning av kort text. Bra för långa texter. Mer klar och ren röst [jämfört med AnnMarie m. lexikon], bättre för akademiska röster. (R5)
- Jag vill inte lyssna på någon i längre sammanhang. Svårt att hänga med den andra rösten [AnnMarie], det lät som om den tappade delar av orden. [Ingmar m. lexikon] lät alldeles för ansträngd. Uppläsning är bäst, flyter bättre än talboken. Ingen bra läsning, rösten var dock ganska behaglig. Det var lättare att förstå [än AnnMarie m. lexikon]. Ingen [av Ingmar m. lexikon och AnnMarie m. lexikon] var bra, men denna [Ingmar m. lexikon] var mest behaglig. Bäst röst [jämfört med AnnMarie]. Föredrar denna [Ingmar m. lexikon], jag gillade inte den andre [Emma]. Lät mkt bra, klarade något ord bättre [än Ingmar]. Jag gillade inte rösten [på AnnMarie m. lexikon], men den andra [Ingmar m. lexikon] var värre.(R6)
- Mer behaglig [än AnnMarie m. lexikon]. Har mer verklighetstrogen intonation [än Ingmar]. Läste mer rätt [än Emma]. Jämnare, bättre uttal [än Emma]. Jämnare, talet flöt bättre [än Ingrid], inga hack. Bättre uttal på ”observerbara”, rösten är mer klangfull [än AnnMarie]. Jämnare hastighet [än den mänskliga rösten]. Jämnare läsning [än Ingrid]. Uttalade namnen rätt, lät bättre [än Ingmar].(R7)
- Saknar betoning, monotont läst. Läser utan paus, tonhöjning i slutet av meningen. Utan betoning på rätt ord. Låter inte bekvämt. Det låter monotont. Entonigt till slutet då tonhöjden går upp istället för att gå ner, som jag vill.(R8)

Mänsklig röst

- För låg ljudnivå (R1)
- Lugnare, behagligare, samma tempo hela tiden. Mumlade för mycket. Mindre burkig [än AnnMarie m. lexikon]. Har bra röst och uttal men dålig pausering. Lugn och stabil. Har skön, stadig och mjuk röst. Bra(R4)
- Låter som om den pratar dialekt från västsverige, lite svårare. Låter mer cool och ung, som dr Alban, bra för generell lyssning, Ej bra för akademiska texter, lite svår att förstå, slutför ej fraser och ord. Låg volym, lämplig för lyssnare som har svenska som modersmål, mer begränsad inflection. Har ojämn hastighet. Tolkade texten bäst, sämre volym och hastighet [än AnnMarie]. För låg volym, bättre röstkvalitet . Låter överdramatisk. Ok för långa texter, låter som göteborgska, gillar inte. Mer mänsklig, några ord är avkapade i slutet, bra för akademiska texter. Lite snabb, texten är bättre, men jag gillar inte att lyssna på den. Sämre uttal [än Ingmar]. Borde varit mer långsamt och ha högre volym. Kräver mer av lyssnaren då hastigheten är snabbare, ok röst för akademiska texter om man är van att lyssna på rösten, ok röst(R5)
- Lättförståeligt Läste bättre [än Ingrid], det flöt bättre, hade sämre röst [än Ingrid]. Lät klart bäst, men pauserna gjorde att det blev oflyt i texten. För långsam, jag hade tryckt upp hastigheten på DAISY. Sätt denna röst på den andres [AnnMaries] läsning. Mest lättförståelig. Behagligast. Låter minst datorgjord.(R6)
- Bättre intonation. Mer verklighetstrogen intonation. Hade onödiga pauser, gick för långsamt med för långa pauser, rösten var behaglig. Pausade på rätt ställe. Mer realistisk. Tog bra pauser. Ändrade hastighet, hade dålig volym.(R7)

- Har slö röst. Låter bättre än alla andra röster jag hört hittills. Slö. Den sväljer ord. Har för låg röst, inte problem med volym, jag hör inte alla orden. Snabbare läsning, men sväljer ord. Läser utan energi, jag måste anstränga mig för att förstå. Har alltför låg röst, inte problem med volym. Har bra pauser och betoning på rätt ställen.(R8)