

Test och jämförelse av appar för tal-till-text



Innehåll

Inledning.....	2
Process.....	2
Material	4
Resultat.....	5
Slutsatser	8
Diskussion	9

Inledning

Omvandling av tal till skriven text blir allt mer relevant inom området elektronisk kommunikation och det finns en mängd av programvaror som erbjuder funktionen tal-till-text. I och med mängden av programvaror som finns, både förskrivningsbara och konsumentprodukter, kan det vara svårt för förskrivare att veta vilka programvaror som finns och hur väl de fungerar. Att individuellt undersöka en mängd olika produkter är ett stort arbete. Centrum för Elektronisk Kommunikation (CFEK) har gjort detta test av ett antal vanligt använda programvaror för tal-till-text för att underlätta för landets förskrivare i deras urval av programvaror att informera om, ge råd kring eller förskriva.

CFEK tar inte ställning för eller emot någon av programvarorna som förekommer i testet. Vi har inte heller för avsikt att främja eller avråda från användning av någon specifik produkt. Vi uppmuntrar er att betrakta detta test som en utgångspunkt för reflektion och diskussion i ert fortsatta arbete kring tal-till-text.

Process

Urvalet av vilka programvaror för tal-till-text som användes i testet har gjordes genom insamling av vanligt använda och testade programvaror av förskrivare, tekniker och andra yrkesverksamma i CFEK Nätverksgrupp.

Programvaror för tal-till-text kan fungera på olika sätt och kallas till exempel diktering, transkribering eller automatisk textning. De programvaror som användes i testet och hädanefter benämns ”tal-till-text” är programvaror som automatiskt omvandlar tal till skriven text i realtid. Dessa programvaror finns som appar till mobil eller surfplatta. Textningen aktiveras vid appstart eller med manuell aktivering i appen, och avslutas vid appens avslut, manuell avstängning eller då appen inte uppfattar någon auditiv input.

Testet har inte tagit med programvaror som konverterar inspelat ljud till text i efterhand, eller funktioner där manuella insatser kontrollerar textningen, till exempel skrivtolkning. Testet inkluderar ej heller automatisk liveundertext i till exempel Teams.

Andra aspekter i val av programvara för tal-till-text kan vara exempelvis:

- Hur snabbt textningen sker, där kortast möjliga processtid är önskvärd
- Textningens formatering och styckesindelning
- Möjlighet att anpassa färg, typsning, textstorlek etc.
- Appens användarvänlighet

Dessa och ytterligare aspekter är viktiga och relevanta perspektiv i val av programvara. För att avgränsa testet har vi valt att i vår poängsättning endast jämföra textningens kvalitet. Ytterligare aspekter utöver textningens kvalitet skiljer sig åt mellan programvarorna och är inte alltid jämförbara, men detta tas upp vidare under ”Diskussion”.

Det är viktigt att poängtera att detta test är gjort under ”normala” förutsättningar, där ljudmiljön har varit ett vanligt tyst utprovningssrum, med hårdvaror som kan köpas i handeln, samtal inspelade av kollegor osv. Vi har gjort vårt bästa för att skapa goda men realistiska

förutsättningar för testet, och har inte använt någon specialutrustning, ljudisolerat rum etc. Testet bör kunna återskapas på andra alternativ telefoni-verksamheter.

För att genomföra testet har vi skrivit manus på korta samtal, ca 1-1,5 minut. Samtalen är tänkta att vara ”vanliga samtal” som är troliga att kunna ske med hjälp av tal-till-text, med en bredd av ”vanligt tal”, talesätt, namn, datum och tid etc. Texterna har talats in eller lästs upp av olika personer, manliga och kvinnliga röster, med olika dialekter och talhastigheter.

I förberedelse för testet spelade vi upp förinspelade ljudklipp för de olika programvarorna och upplevde ett medelmåttigt resultat. Vi såg därför över vilka hårdvaror vi använde och bytte till en lämpligare mikrofon och köpte in en ny högtalare. I samband med detta tog vi även beslutet att genomföra testet i två faser, en fas med förinspelat ljud och en fas med uppläst manus i realtid, ”live”.

Fördelen med förinspelade ljudklipp är att alla programvaror får samma förutsättningar. Det kan också liknas vid att få automatisk textning av digitala möten, podcasts eller annat material som spelas upp digitalt. Nackdelen är att olika programvaror verkar ha olika möjligheter att ta in digitalt ljud, troligtvis på grund av att uppspelning, även med en bra högtalare, inte har samma omfång som talad röst i verkligheten.

Det vanligaste användningsområdet för tal-till-text är att få talet textat för personen man pratar med som befinner sig i samma rum, vilket efterliknas bäst med uppläsning av manus ”live”. Fordelen är även att flera av programvarorna är gjorda med detta syfte och verkar fungera bäst på detta sätt. Nackdelen är att det inte går att säkerställa att alla programvaror får samma förutsättningar, då uppläsningen kan variera i talhastighet, volym, betoning etc., oavsett talarnas bästa insats till att tala lika varje uppläsning.

Testet genomfördes i två faser. I fas ett användes förinspelade ljudklipp, inspelade vid ett tidigare tillfälle med testets mikrofon. Ljudklippen spelades upp från en dator med hjälp av högtalare, till de olika programvarorna inkopplade med mikrofon. Programvarorna har testats en i taget, i olika enheter men inkopplade till samma mikrofon. Rummet som använts i alla testerna har inga utstickande bakgrundsljud och en bakgrundsvolym på runt 30 dB. Ljudklippen har spelats upp till en ljudvolym som vi i rummet upplevt som klar och stark, på en ljudvolym strax över 60dB, vilket räknas som en vanlig samtalsvolym.

De inspelade ljudklippen var Biobesök, Telefonförsäljare och Biluthyrning (se bilaga för originaltexterna). Biobesök lästes upp av två kvinnliga röster, medelhastighet. Telefonförsäljare lästes upp av en manlig röst, hög hastighet och en kvinnlig röst, låg hastighet. Biluthyrning lästes upp av en kvinnlig röst, låg hastighet.

I fas två kopplade vi upp så många programvaror samtidigt som möjligt, i de enheter som fanns tillgängliga för oss, för att undvika att varje samtal behövde läsas upp flera gånger. Vi kunde ha fyra programvaror igång samtidigt, vilket innebar att varje manus endast behövde läsas upp två gånger. Manus lästes upp i realtid av olika personer, med hög och klar röst och med samma avstånd till mikrofonen.

Manus till uppläsningen var Samtal med chefen, Mobilförsäljare och Biluthyrningen (se bilaga för originalmanus). Samtal med chefen lästes upp av en manlig och en kvinnlig röst, hög hastighet. Mobilförsäljare lästes upp av två kvinnliga röster, låg och medelhastighet. Biluthyrningen lästes upp av en kvinnlig röst, låg hastighet. Manus för

Biluthyrningen användes till både fas 1 och 2 då samma uppläsare var tillgänglig vid både tillfällena vilket gav möjlighet att jämföra textningen av detta manus i båda faserna.

Textningarna har förts över till dator och styckeindelats på samma sätt, utan att någon text har förändrats. För att kunna jämföra textningarna har vi delat upp originaltexterna mening för mening. Textningarna har delats upp på samma sätt, mening för mening.

Varje mening i originaltexten har poängsatts. Poäng har getts för de ord och fraser som vi bedömt som ”viktiga”, de meningsbärande ord och fraser som gör att samtalet går att följa och förstå, totalt 30 poäng per samtal. Poäng har sedan satts på textningarna, om dessa viktiga ord och fraser har textats korrekt eller ej.

Bedömning om poäng har gjorts enligt följande: Om de poängsatta orden och fraserna har textats med samma innebörd ges poäng. Om bortfall skett men innebörden kvarstår ges poäng. Exempelvis om originaltexten anger ”Ja, det vill jag” kan både ”Ja, det vill jag”, ”Ja” eller ”Det vill jag” ge poäng.

Stavfel har godkänts, om innebörden blir densamma. Stavfel som ger ordet eller frasen en annan innebörd har inte godkänts.

Material

Programvaror:

- Anteckningar, dikteringsfunktion på Apples tangentbord - Apple
- OneNote, dikteringsfunktion - Microsoft Corporation
- Transkribering och aviseringar - Research at Google
- Group Transcribe - Microsoft Corporation
- Rogervoice - RogerVoice
- TERA-Mobile och TERA-Touch, Local och i telefonsamtal - Europea i Malmö AB
- Transett T-Talk - AB Transistor

Hårdvaror:

- Apple iPhone 11
- Apple iPad 10,2 (9th gen)
- Apple iPad Pro 12,9”
- Samsung Tab S5e
- Samsung Galaxy A52
- ASUS Zenfone 9
- Sennheiser SP 20 ML
- Logitech Z313

Manus: (för full text, se bilaga)

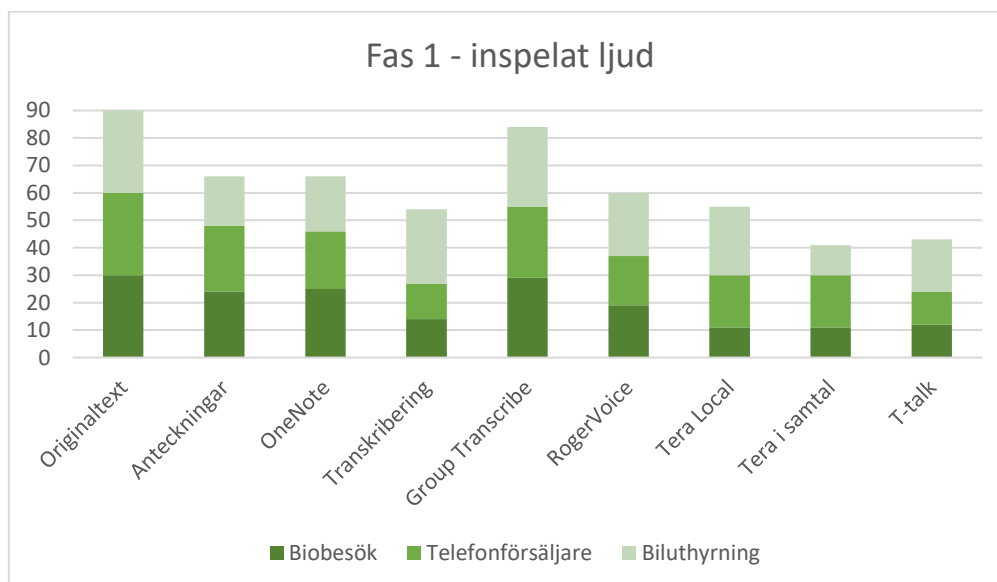
- Biobesök
- Telefonförsäljare
- Biluthyrning
- Samtal med chefen
- Mobilförsäljare

Resultat

Nedan följer resultat i poäng från fas 1, fas 2 och totala poäng. Resultatet finns i textform och i diagramform, för att underlätta tolkning av siffrorna. Full originaltext, textningar och poäng kan läsas i bilagan.

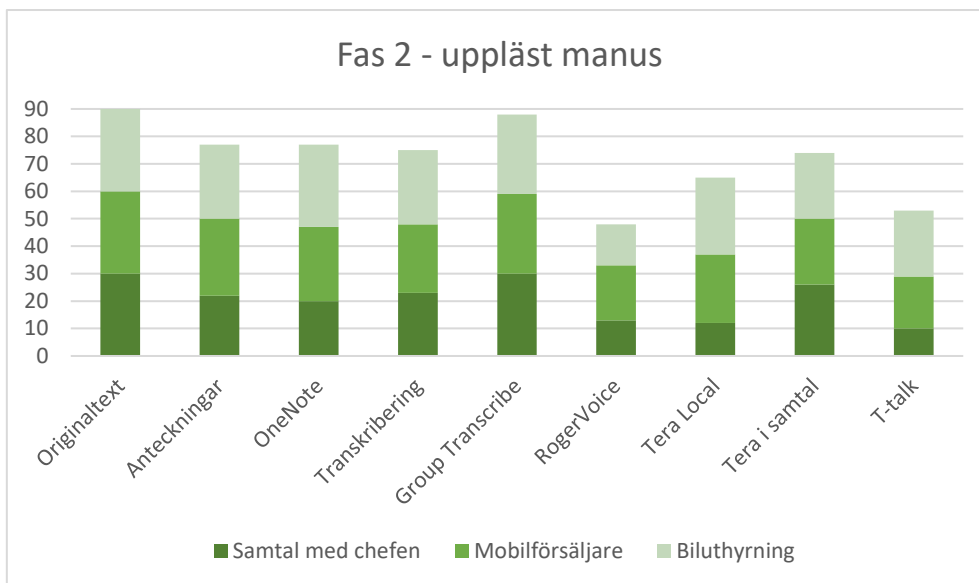
Fas 1 – inspelat ljud, 2023-08-30

	<u>Biobesök</u>	<u>Telefonförs.</u>	<u>Biluthyrning</u>	<u>Totalt</u>
Originaltext	30	30	30	90
Anteckningar	24	24	18	66
OneNote	25	21	20	66
Transkribering	14	13	27	54
Group Transcribe	29	26	29	84
RogerVoice	19	18	23	60
TERA Local	11	19	25	55
TERA i samtal	11	19	11	41
T-Talk	12	12	19	43



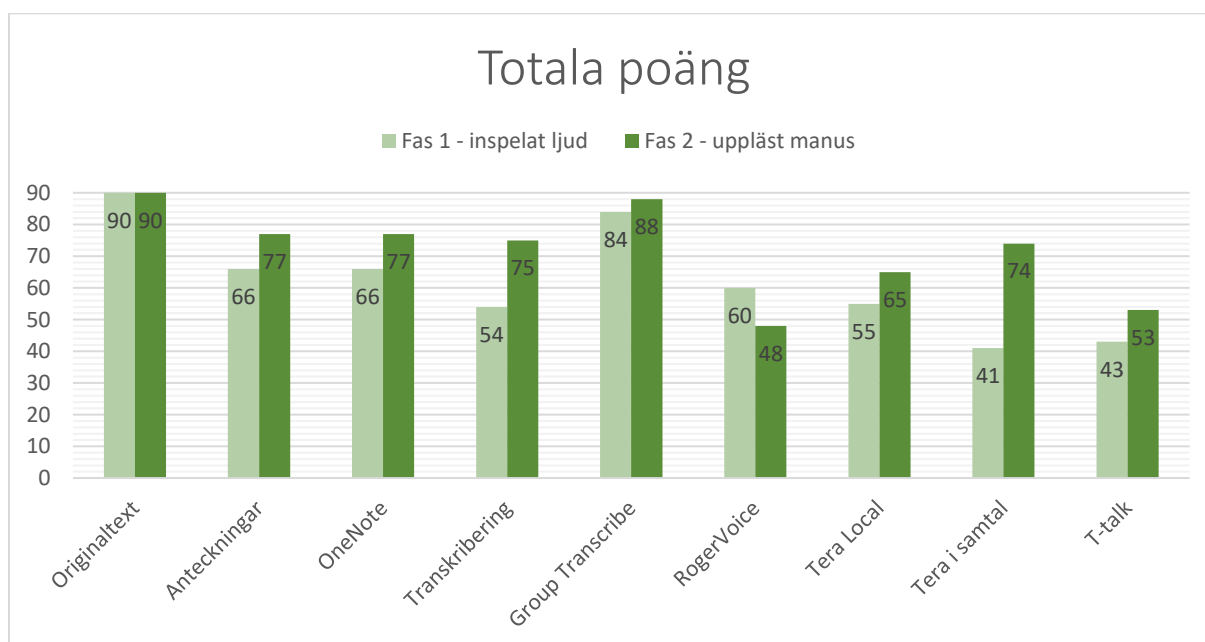
Fas 2 – uppläst manus, 2023-10-06

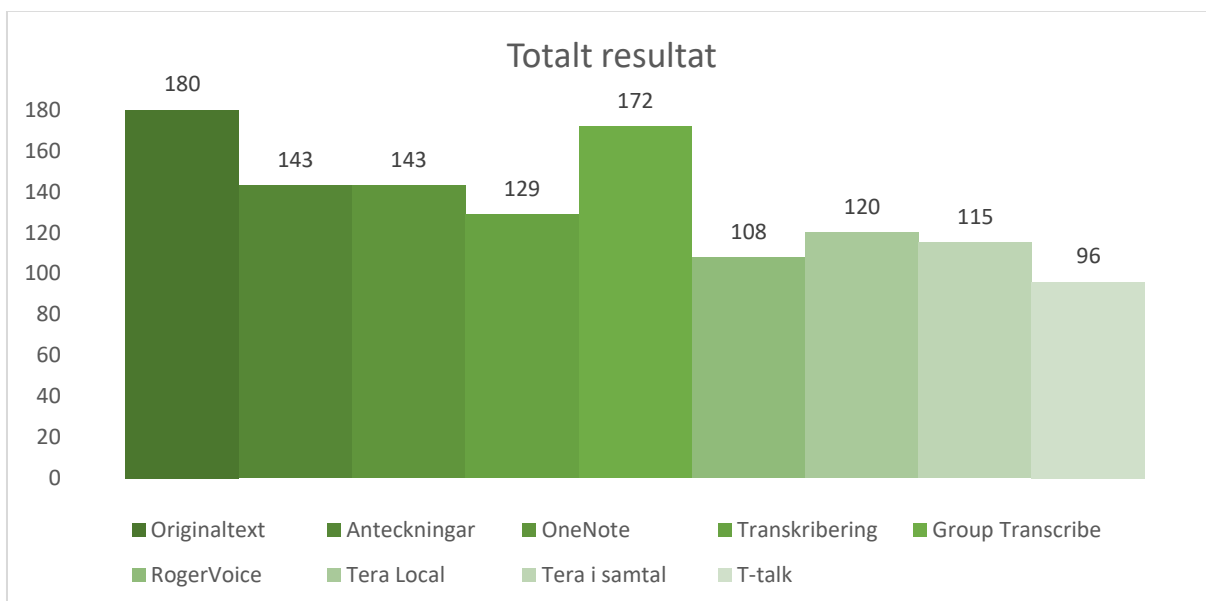
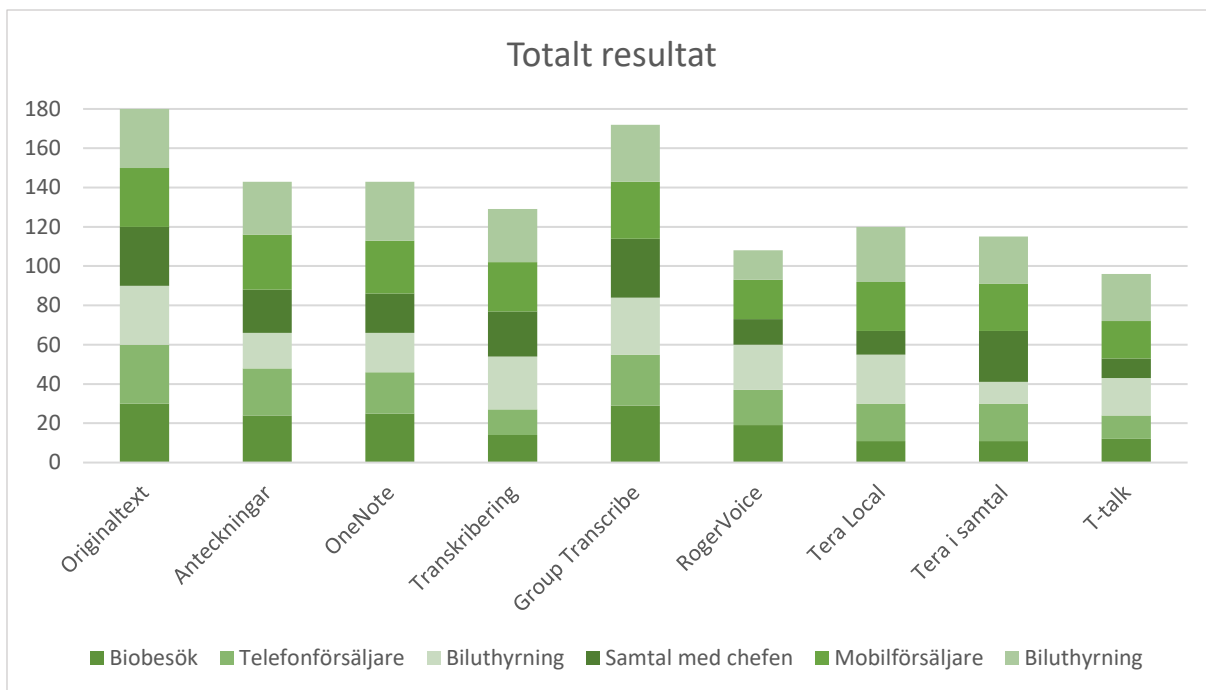
	<u>Med chefen</u>	<u>Mobilförs.</u>	<u>Biluthyrning</u>	<u>Totalt</u>
Originaltext	30	30	30	90
Anteckningar	22	28	27	77
OneNote	20	27	30	77
Transkribering	23	25	27	75
Group Transcribe	30	29	29	88
RogerVoice	13	20	15	48
TERA Local	12	25	28	65
TERA i samtal	26	24	24	74
T-Talk	10	19	24	53



Sammanställning av poäng

	<u>Fas 1</u>	<u>Fas 2</u>	<u>Totalt</u>
Originaltext	90	90	180
Anteckningar	66	77	143
OneNote	66	77	143
Transkribering	54	75	129
Group Transcribe	84	88	172
RogerVoice	60	48	108
TERA Local	55	65	120
TERA i samtal	41	74	115
T-Talk	43	53	96





Slutsatser

Vi har dragit några slutsatser utifrån poängsättningen som visas i resultatet.

Group Transcribe gör överlägset mest korrekta textningar i detta test, med en pricksäkerhet på 95,5%. Diktering i Anteckningar och OneNote ligger på en gemensam andra plats med en pricksäkerhet på 79,4%, vilket trots ett bra resultat ändå är 16% sämre än Group Transcribe. Lägst resultat har T-talk, med endast 53,3% korrekta textningar.

De flesta appar har fler korrekta textningar med ett uppläst manus än med inspelat ljud, vilket är förväntat då naturligt talad röst har ett större omfång än digitalt uppspelat ljud. Undantaget är RogerVoice, som har fått högre poäng med inspelat ljud än uppläst manus.

Anteckningar och OneNote har fått lika många poäng totalt. Antal poäng i de individuella samtalen skiljer dig åt, endast totalsumman i respektive fas är detsamma. Apparna använder dessutom olika textningsmotorer (Apple respektive Microsoft), så detta bör vara ett slumpmässigt sammanfallande.

De programvaror som fått högst poäng (Group Transcribe, Anteckningar, OneNote och Transkribering) är alla utvecklade med stora företag i ryggen; Google, Microsoft och Apple. Ekonomi och resurser kan påverka vilka programvaror som har möjlighet att utvecklas mest och på flest språk, och därmed producera mer korrekta textningar.

TERA Local och TERA i telefonsamtal har fått olika antal poäng. Detta trots att det är samma programvara och därmed textningsmotor, och att programvarorna har fått samma förutsättningar i möjligaste mån. Detta kan tyda på att små skillnader kan påverka mer än vad vi tror, eller att det finns någon annan teknisk aspekt som kan förklara skillnaden i resultatet.

Diskussion

Utifrån mitt arbete som förskrivare av Alternativ Telefoni i Region Örebro Län och som representant för Centrum för Elektronisk kommunikation kommer jag att väga in mina erfarenheter och tankar, och dessutom andra aspekter av användningen av ovanstående programvaror, och dela med mig av några ytterligare tankar som förhoppningsvis kan vara till nytta för dig i fortsatta reflektioner, diskussioner och framtida arbete kring tal-till-text. Ha detta i bakhuvudet när du läser följande diskussion, och dra alltid dina egna slutsatser!

Group Transcribe har gjort överlägset bäst ifrån sig i detta test. Group Transcribe formaterar även textningen på ett naturligt sätt, vilket de flesta andra appar inte gör. Ord för ord tas in och textas, när meningen är färdig formateras den med korrekta versaler, gemener, interpunktion och styckesindelning. Detta gör det enkelt att följa med i texten. En annan fördel med Group Transcribe är att flera enheter kan koppla upp sig till samma samtal via Bluetooth, QR-kod, länk eller mötes-ID, där varje enhet fungerar som en separat mikrofon. Detta innebär att appen även kan informera om vem som säger vad. Group Transcribe finns bara till iOS och kan därför bara användas av iPhone- och iPadanvändare. En nackdel med Group Transcribe är att det inte finns några inställningar att göra för att ändra textstorlek, färg, typsnitt etc, förutom de inställningar du kan göra i din enhet.

Diktering i Anteckningar och OneNote har båda fått höga poäng. De har dessutom fått samma poäng, trots att vi har testat dem i olika enheter. Det kan vara så att eftersom diktering är en funktion som används av många, inte bara vår målgrupp, är det en funktion som har fått resurser till att utvecklas. Diktering kan egentligen användas överallt där det finns ett textfält, och inte bara i dessa appar. Det är att rekommendera att använda en anteckningsapp, men du skulle likväl kunna diktera in i t.ex. din webbläsares sökfält, vilket såklart blir väldigt obekvämt att läsa. Då diktering används i en programvara tas ofta en stor del av skärmen upp av en ”inkommande ljud”-ruta, vilket gör att det blir mindre plats kvar att läsa textningen. Det kan dock finnas en del inställningar att göra för textstorlek, färg och typsnitt, beroende på vilken app man använder diktering i. Något som kan vara en nackdel med funktionen diktering är att den generellt stänger av sig när den inte ”hör” något inkommande tal. Det kan göra att man behöver starta upp dikteringen flera gånger, och det finns risk att man missar en del av det som sägs.

Det är även viktigt att notera om man använder diktering från en app eller från tangentbordet. I detta test har diktering i Anteckningar gjorts med Apples tangentbord, medan diktering i OneNote gjorts med den inbyggda dikteringsfunktionen i OneNote-appen. Vill man exempelvis använda Googles dikteringsmotor i valfri app kan man ladda hem Googles tangentbord Gboard, och därmed använda tangentbordets dikteringsfunktion. Att kontrollera om man använder en inbyggd dikteringsfunktion i en app eller en dikteringsfunktion på tangentbordet är något att vara vaksam på om man upplever ojämna resultat i textningen.

Transkribering är en app som har många fördelar i form av smidighet. Då den är utvecklad som en så kallad tillgänglighetsapp av Google finns den idag redan inlagd i många nya Android-mobiler och surfplattor. Ibland är den dold i enhetens applista, även om den finns nedladdad. Den går då att hitta under Inställningar och sedan Tillgänglighet. Här kan olika Android-versioners inställningsmenyer se olika ut, men appen hittas under t.ex. Hörförbättringar eller Transkribering. Här finns även möjlighet att lägga till Transkribering

som en tillgänglighetsgenväg. Det innebär att det finns en snabbgenväg till appen som man kan komma åt enkelt oavsett vilka andra appar som är öppna på enheten.

Textningen startar direkt då appen öppnas, den behöver inte aktiveras manuellt. Det går enkelt att ändra storlek på texten, byta språk och några andra enkla inställningar via kugghjulsikonen, eller öppna Fler inställningar där det går att ändra mellan ljust och mörkt läge, lägga till ord i en anpassad ordlista osv.

Transkribering är enkel att hantera även vid låg teknisk kompetens. Appen fungerar bra att använda utan att göra några personliga inställningar. Vill man göra ytterligare inställningar eller lägga till Transkribering som en tillgänglighetsgenväg behövs det endast göras en gång, och appen kan sedan fortsätta användas som vanligt utan att behöva göra ytterligare val.

TERA är den app av de vi har testat som har bäst möjligheter till att göra synanpassningar, det vill säga att ändra färger, textstorlek och typsnitt. Detta kan göras i viss mån i anteckningsappar, men inte i samma utsträckning som i TERA. Att kunna göra dessa justeringar är en stor fördel till exempel vid synnedsättning.

TERA har även fördelen att den kan texta både lokala samtal och telefonsamtal, och är den enda appen vi har testat som har båda dessa funktioner i samma app.

TERA är ett hjälpmedel som är MDR-klassat. TERA går att få via förskrivning från exempelvis Hörcentral, Habilitering eller verksamhet för Alternativ Telefoni. Sortiment och riktlinjer kan skilja sig åt i olika regioner, vilket kan innebära att TERA i vissa regioner endast finns tillgänglig för vissa målgrupper och i andra regioner inte alls.

Rogervoice är den enda gratisappen i testet som textar telefonsamtal. Även med ett abonnemang saknas vissa funktioner som finns i andra länder då man köper ett abonnemang, till exempel att få ett eget telefonnummer till appen.

En fördel med att använda Rogervoice och ringa app till app är att textningen syns för båda personerna. Man har därför möjlighet att se hur det man sagt har blivit textat, och har möjlighet att själv korrigera om något har blivit fel.

T-talk har fått lägst poäng i vårt test. Det är möjligt att T-talk skulle uppnå bättre resultat om den t.ex. användes i kombination med andra hjälpmedel eller enheter. T-talk har fler funktioner än de som ingick i testet, till exempel att texta inkommande ljud via kabel till en annan enhet, till exempel telefonsamtal eller videomöten. T-talk kan även kopplas samman med andra hörselhjälpmedel, t.ex. Phonak Roger Select. Dessa funktioner har vi inte tagit med i testet då de inte finns i någon av de andra apparna som vi testat. Det är något som kan vara värt att testa själv om T-talk är ett aktuellt alternativ.

T-talk har dessutom en funktion där appen uppskattar textningens korrekthet. Man kan ställa in en gräns på 70, 80 eller 90 procent, och om appen bedömer textningen som mindre korrekt än den procent man ställt in markeras texten med gul bakgrund. Det är en funktion som jag upplever skulle vara önskvärd även i andra program, för att som användare få stöd i att bedöma om textningen har blivit rätt. Problemet som vi stötte på just i det här testet var att all textning i T-talk tyvärr blev gulmarkerad, även på den lägsta inställningen 70% korrekthet.

Vi har både testat appar som är gratis att ladda ner, gratis men med abonnemang och förskrivningsbara appar som kan vara gratis för användaren men innebär en kostnad för den förskrivande regionen.

Rogervoice är gratis i sitt grundläge, där man kan ringa app till app. För att ringa till andra telefoner som inte har Rogervoice installerat krävs ett abonnemang i appen.

TERA och T-talk är förskrivningsbara appar som för den förskrivande regionen innebär en kostnad vid inköp och ytterligare kostnader vid förlängning av licens. Om hjälpmedel är kostnadsfritt eller ej för patienten varierar mellan regionerna.

Övriga appar i testet är appar som är gratis att ladda ner och använda.

Oavsett vilken app man använder är det skillnad på ett uppstyrt test och verklig användning. Exempelvis, om apparna i testet har slutat texta under några meningar, innebär det förlorade poäng. Om appen skulle sluta texta i en verklig situation skulle man troligtvis fråga om eller be motparten att upprepa sig. Detsamma gäller faktorer som vi inte har fått med i detta test, såsom bakgrundsljud, fler dialekter, brytningar osv. Även användarna av tal-till-text skiljer sig såklart åt i till exempel synförmåga, teknikvana, kognitiv förmåga, drivenhet, intresse och så vidare, faktorer som också påverkar valet av programvara.

Som summering så har alla appar sina fördelar och sina olika användningsområden. Oavsett poäng i detta test har alla appar något att ge. Olika appar kan behövas beroende på situation, och det gäller att hitta vad som fungerar bäst för varje person. Detta test kan användas som ett verktyg för detta, och vi på CFEK kan agera bollplank vid frågor och tankar.

Tekniken kring tal-till-text är relevant och utvecklas fort. Utseende, funktioner och tekniken innanför skalet kan förändras, förhoppningsvis till det bättre. Apparna i det här testet kan ha hunnit uppdateras bara från det första testet tills att jag skriver dessa ord. Ha i bakhuvudet att teknik är en färskvara och det är viktigt att göra en egen bedömning av appens nuvarande status innan en rekommendation eller förskrivning görs. Genom att använda dessa programvaror, jämföra dem och tycka till, kan vi leda utvecklingen i rätt riktning.

Vi kan även se fram emot en framtid med nya tekniska lösningar. Automatisk textning både av generellt ljud och av telefonsamtal finns redan i bland annat Google Pixel-telefoner och är på väg till Samsung, och säkert kommer fler att följa efter. Då behövs ingen extra app för att få textade telefonsamtal, och kan användas av alla. Problemet i dagsläget är att funktionen inte finns på svenska, vilket vi såklart hoppas att de kommer lägga till.

Att hålla sig uppdaterad kring all ny teknik är inte alltid en enkel sak på ett brett fält. Vi på CFEK tar gärna emot både frågeställningar och tips på nyheter inom området elektronisk kommunikation. Hör gärna av er till oss!