

TaLiS

- Taluppfattning och Listening Span-test

*Mats Andréén, Hanna Björk, Lars Petter Björkman, Anna Hillertz,
Edvard Johansson, Carola Markström, Sanna Åsberg*

Linköpings universitet, juni 2003

Denna studie var tänkt att undersöka om taluppfattning och arbetsminne kan testas i en klinisk miljö. Kognitiva förmågor påverkar människans taluppfattning. Personer med samma typ och grad av hörselnedsättning behöver inte få samma resultat i ett taluppfattningstest, då de kan ha olika kognitiva förutsättningar. Ett kliniskt test skulle fungera som ett komplement till andra metoder vid utprovning och anpassning av hörapparater där hänsyn tas till individens kognitiva förmågor. Ett stort antal meningar och frågor på meningarnas innehåll skapades. Målen var att undvika redundans samt att få meningar och frågor lika svåra. Meningarna utvärderades kvalitativt, såväl som kvantitativt och därefter testades meningarnas grad av svårighet. Det gjordes även försök att minska spridningen av meningarnas svårighetsgrad. Sedan var meningarna klara för pilottestet, som bland annat fungerade som ett listening span-test. För att undersöka eventuella inlärningseffekter, testades tio försökspersoner två gånger. Där fanns en korrelation som visade att signal/brus-förhållandet sänktes i andra omgången, vilket kan tydas som en inlärningseffekt. Testet gav upphov till takeffekter, men materialet kommer att användas för fortsatta studier.

Inledning

Forskning har visat att taluppfattning skiljer sig från hur människan uppfattar ljud i allmänhet och därför inte kan beskrivas med samma principer som gäller uppfattning av andra ljud. Ett flertal kognitionspsykologiska modeller har påverkat (den traditionellt sett fysiologiskt inriktade) forskningen kring hörsel och hörselskador då man har sett att enbart ett audiogram inte är tillräckligt för att beskriva hur man uppfattar tal, utan att de individuella kognitiva förmågorna också har betydelse för detta (Andersson, 2001).

Tidigare forskning på INR (Institutionen för nervsystem och rörelseorgan) tillsammans med IBV (Institutionen för beteendevetenskap) här i Linköping har arbetat för att ta fram metoder för att testa personers, främst hörselskadades, taluppfattningsförmåga i samband med utprovningen av hörapparater. Dagens hörapparater har en rad olika funktioner, vissa med avancerade algoritmer som alltefter omgivningens akustiska egenskaper varierar bearbetningen av de akustiska signalerna.

Forskning har visat att personer med goda högre kognitiva funktioner lättare kan tillgodogöra sig informationen från dessa dynamiska hörapparater, medan andra istället blir störda av den information som dessa hörapparater kan tillhandahålla (Gatehouse et al., under tryckning). Utifrån detta ville vi utforma ett kliniskt användbart test för att mäta både taluppfattningsförmåga och arbetsminne. Grundtanken var att undersöka hur arbetsminnet är relaterat till taluppfattning i brus, men den huvudsakliga frågeställningen begränsades till hur arbetsminnet går att mäta i klinisk miljö. I förlängningen skulle ett sådant test kunna användas som ett komplement till de traditionella hörseltesten och bidra till att försöka anpassa en hörapparat utifrån varje enskild individs förutsättningar.

Teoretisk bakgrund

Taluppfattning och talperception

Det finns en skillnad mellan taluppfattning och talperception. Talperception innebär att ta till sig, att varsebli, det som sägs, vilket inte måste innebära att man förstår det som sägs.

Taluppfattning däremot är den mer kognitiva delen av språkförståelse, och innebär bearbetning av det som sägs och därmed förståelse för det som har tagits in. Talperception och läsande är ganska lika i vissa avseenden, men skiljer sig åt på några punkter (Eysenck, 2000). Vid läsande kan orden ses som enheter, medan tal består av ett kontinuerligt flöde av skiftande ljud med korta avbrott, vilket innebär att ett segmenteringsproblem uppstår. Dock ger talat språk mottagaren fler ledtrådar (tempo, intonation, timing) än skriftspråk om meningsstrukturen och den avsedda betydelsen.

I normalt tal överlappar fonemen varandra och påverkar således hur påföljande ord uttalas, detta fenomen kallas för koartikulation och är tydligast vid konsonanter vars uttal ofta är beroende av vilken vokal som följer (Eysenck, 2000).

Arbetsminne

Baddeley och Hitch utvecklade på 1970-talet en arbetsminnesmodell (Eysenck, 2000) bestående av tre olika komponenter som kallas för centrallexekutiven, fonologiska loopen och visuo-spatiala skissblocket. Centrallexekutiven är nyckelkomponent i arbetsminnesmodellen och påminner om uppmärksamhet. Den fonologiska loopen innehåller information i talbaserad form och visuo-spatiala skissblocket är specialiserat för spatial och/eller visuell kodning. Komponenterna är oberoende av varandra. Detta innebär att om två uppgifter ska utföras samtidigt och båda kräver att samma komponent används, kan inte båda uppgifterna utföras på ett bra sätt. Om uppgifterna däremot kräver användning av olika komponenter är det möjligt att med bra resultat utföra båda uppgifterna samtidigt.

Arbetsminnet är begränsat och kan ses som en samling mentala processer som låter information hållas temporärt i ett tillgängligt tillstånd för att tjäna en mental uppgift (Cowan, 1998). Vid övningar i talförståelse framgår ofta inte den avsiktliga betydelsen förrän man har hört efterföljande ord, alltså måste ordet kunna kvarhållas i minnet. Det är först när hela meningen har hörts som ordets betydelse kan fastställas. Detta är en förutsättning för att man ska kunna förstå tal.

Daneman och Carpenter utvecklade 1980 ett Reading Span-test som skulle mäta individers arbetsminneskapacitet (Baddeley och Gathercole, 1993). Reading span-testet skiljer sig från konventionella minnesomfattningstest då det både ser till en individs förmåga till att processa och lagra lingvistiskt material.

Testet går till så att en försöksperson (FP) får en lista med meningar att läsa och efter varje läst mening blir ombedd att i följd återge det sista ordet från varje tidigare läst mening. Ju fler meningar en person har hunnit läsa innan den inte längre klarar av att återge alla tidigare meningars sista ord, desto bättre minnesomfattning har individen. Denna metod överfördes även på auditivt presenterade meningar, det så kallade listening span-testet. Starka kopplingar mellan minneskapaciteten och resultat i språkförståelsetest har påvisats i ett flertal test. Daneman och Carpenter antog att skillnader i arbetsminnet mellan individer främst beror på hastigheten och effektiviteten med vilka kognitiva operationer utförs.

Tidigare studier

Hagermans meningsmaterial

I början på åttiotalet utformade Björn Hagerman, Karolinska Institutet i Stockholm, ett test för att mäta taluppfattningströskeln (speech reception thresholds, SRT), den nivå där man nätt och jämnt uppfattar en bestämd procent (50 % i Hagerman 1984) av orden i uppspelade meningar. Denna nivå ställs in genom att ändra signal/brus-förhållandet, d.v.s. ljudnivån på meningarna och bruset.

Ett material bestående av 50 ord som kombineras till meningar med låg redundans, d.v.s. få kontextuella ledtrådar, skapades för att sedan kombineras slumpvis. Eftersom det bara var 10 förnamn, 10 verb, 10 siffror, 10 adjektiv och 10 substantiv, blev variationerna av meningarna begränsade och inlärningseffekter uppstod (Nilsson, Soli & Sullivan, 1994).

Hearing In Noise Test (HINT)

HINT är en amerikansk studie och mätte, liksom Hagermans test, taluppfattningströskeln genom att meningar spelades upp i brus varpå volymen justerades tills FP nätt och jämnt kunde uppfatta

en viss procent av orden. De utgick ifrån ett brittiskt textmaterial baserat på transkriptioner av barns tal framtaget 1979 av Bench och Bamford (Nilsson, Soli & Sullivan, 1994).

Meningarna anpassades till amerikansk engelska och likställdes med avseende på naturlighet, längd och hur lätta de var att uppfatta. Tio personer ombads skatta meningarnas naturlighet på en skala mellan ett och sju där ett var mycket konstlad och sju helt naturlig. De meningar som inte klarade sig omarbetades och testades igen. Meningarna lästes in av en man och justerades så att de fick samma medelamplitud. Ett brus med samma genomsnittliga frekvensspektrum som meningarna skapades och lades på, sedan utfördes ytterligare ett flertal test med försökspersoner (FP) för att justera signal/brusförhållandet för varje mening tills FP presterade ungefär lika bra på alla meningar. Listor med samma genomsnittliga fonemdistribution skapades. Slutresultatet blev ett stabilt material av likvärdiga meningar som utgjorde en bra grund för framtida test i taluppfattning.

Listening Span and Speech Recognition (LiSSpeR)

LiSSpeR (Hällgren, 2002) var ett försök att skapa ett svenskt, kliniskt användbart, auditivt test som undersöker de kognitiva förmågor som är relevanta för talförståelse.

Hagermanmeningarna fick ligga till grund för testet eftersom detta meningsmaterial var väl beprövat. I testet spelades meningsmaterialet upp i brus, som innehöll samma frekvensspektrum som meningarna, och FP ombads att efter varje mening repetera vad de hade hört. Detta gav ett värde på individens förmåga att uppfatta tal i brus. Detta jämfördes sedan med prestationen på ett auditivt kognitivt test för att se om det fanns någon korrelation. Resultaten blev inte signifikanta, men pekade ändå åt rätt håll. Försökspersonerna som deltog var alla hörselskadade och de testades med och utan hörapparat. Det auditivt kognitiva testet bestod av en del med frågor och en del med en upprepningssuppgift. När FP upprepade meningarna korrekt, så fick de upprepa ytterligare en mening, innan de fick återge (upprepningssdelen) eller besvara en fråga (frågedelen) på den tidigare korrekt återgivna

meningen. Det visade sig att testet gav stora inlärningseffekter när FP fick upprepa det, vilket gjorde att det inte var kliniskt användbart.

Metod

Det första förberedande steget inför pilotstudien var att skapa ett meningsmaterial med meningar samt frågor på meningarnas innehåll. Därefter spelades meningarna in och deras volym korrigerades för att kompensera för svårigheter att uppfatta meningarna. Detta gjordes genom ett test som utfördes i tallabbet på Linköpings universitet. Meningsmaterialet fördelades sedan på tolv listor efter vissa kriterier, innan pilotstudien genomfördes.

Framtagning av meningsmaterialet

I framtagandet av meningarna följdes till stor del den metod som använts i HINT (Nilsson, Soli & Sullivan, 1994). Vid utformningen av meningarna tog vi hänsyn till att de skulle presenteras auditivt och inte visuellt genom att de formulerades som talspråk, snarare än skriftspråk. Efter en jämförelse med LiSSpeR (Hällgren, 2002) beräknades att pilottestet skulle kräva 120 meningar. För att ha marginal att kunna sälla ut avvikande meningar ur materialet togs 150 meningar fram utifrån några huvudkriterier. Målet var att meningarna skulle ligga på samma svårighetsgrad och kännas naturliga. I skapandet av meningarna användes ett större antal olika substantiv, verb, objekt etc., än vad som finns i Hagermanmeningarna. Alla meningar inleddes med subjekt och predikat då denna ordföljd inte gav någon synlig inlärningseffekt i HINT. Dessutom hade vissa bara just subjekt – predikat – objekt, medan andra dessutom hade adjektiv, tids- eller rumsadverb. Detta, i kombination med det stora antal ord som våra meningar bestod av, ansåg vi vara tillräckligt för att undvika en inlärningseffekt.

För att kontrollera att meningarna inte kändes alltför konstlade utvärderades de kvantitativt genom att ett antal personer fick betygsätta meningarna efter hur naturliga de upplevdes (samma procedur genomfördes i HINT). Skalan gick från ett till sju, där sju var en helt naturlig mening och ett var en väldigt konstlad mening.

Om de inte godkändes i första omgången gjordes de om och utvärderades igen. Varje mening utvärderades av 15 personer och för att godkännas krävdes efter avrundning ett medelbetyg på sex. På grund av tidsbrist hann endast 142 meningar godkännas innan de skulle spelas in, då hade 165 personer gett sitt omdöme. Då ett tillräckligt antal meningar hade godkänts specificerades tre typer av frågor: på meningens subjekt, på meningens predikat samt en fråga gällande antingen meningens objekt eller adverbial. Frågorna specificerades sedan individuellt till varje mening. Då problem med svårighetsgraden uppstod vid specificeringen av predikatfrågan, ströks denna typ av frågor ur materialet.

Likställande av svårighetsgrad

Nästa steg i processen var att få så jämn svårighetsgrad som möjligt på meningarna. Även detta moment var inspirerat av HINT. Alla godkända meningar spelades in (en kvinnlig röst användes) och samma brus som använts till Hagermanmeningarna lades på.

Bruset var ett så kallat spektralt anpassat brus, vilket innebär att man filtrerar vitt brus som innehåller alla frekvenser och erhåller ett brus med samma genomsnittliga frekvensspektrum som de inspelade meningarna (Hagermanmeningarna). Då vi på grund av tidsbrist inte kunde generera ett eget brus så utgick vi från att Hagermanbruset har ungefär samma frekvensspektrum som vårt eget generade brus skulle ha fått. Detta eftersom en kvinnlig röst använts för inspelningen av meningarna i både vårt eget meningsmaterial och till Hagermanmeningarna.

Försöket gjordes i tallabbet på Institutionen för datavetenskap vid Linköpings universitet. Totalt 45 FP delades upp i fyra grupper och fick, en i taget, lyssna på de uppspelade meningarna och upprepa vad de hört, eller trodde sig ha hört. Efter det att samtliga i en grupp hade genomfört testet beräknades medelvärdet på svårighetsgraden ut genom ett poängsystem för antal korrekt återgivna ord i varje mening. Varje enskild menings avvikelse från medelvärdet användes sedan för att justera meningens ljudnivå, ett bestämt antal decibels höjning eller

sänkning av ljudnivån per 15 % avvikelse från det totala medelvärdet, inför nästa grupps testomgång. Brusets ljudnivå hölls konstant under hela försöket. Efter den sista gruppen FP hade meningarnas svårighetsgrad utjämnats så att standardavvikelsen på de enskilda meningarnas ljudnivåjusteringar, mätt i decibel, var 0,20.

Under experimenten i tallabbet användes ett eget datorprogram som vi skrev i Matlab, programmet skulle sköta det rent praktiska genomförandet av testet. Programmet skötte uppspelningen av meningarna, ställde in ljudnivån på meningarna och mixade meningarna med brus. Programmet styrdes i sin tur av en försöksledare som kryssade i vilka ord som försökspersonerna svarade rätt på i meningarna. Programmet sparade även ner resultaten.

Meningarna fördelades preliminärt, efter grad av justering, på tolv listor. Sedan likställdes listornas svårighetsgrad även avseende de lingvistiska aspekterna. Hänsyn togs till fördelningen av återkommande subjekt, fördelningen av andrafrågans variationer i frågestrukturen (gällande objektet eller tidsrumsadverbial) samt fonemfördelningen i de olika listorna.

Pilotstudien

Pilotstudien följde upplägget i LiSSpeR, och genomfördes i en ljudisolerad studio på Universitetssjukhuset i Linköping. Till testet användes det menings- och frågematerial som tagits fram tidigare under projektets gång. En något modifierad version av det program vi skrev i Matlab, till det tidigare nämnda testet i tallabbet på Linköpings universitet, användes även här för att ta hand om de praktiska detaljerna i testet. Försökspersoner var 20 studenter i åldern 20-35. För att ta reda på om vårt test, trots mer varierade meningar och frågor än i LiSSpeR, ledde till de inlärningseffekter som detta tidigare test haft problem med, fick hälften av FP utföra testet två gånger i rad med exakt samma betingelser. Pilottestet var uppdelat i tre delar. I första delen fick varje försöksperson lyssna på 30 meningar och efter varje mening återge vad de hört. Signal/brus-förhållandet

ställdes under dessa meningar in så att försökspersonen till slut hörde i genomsnitt 80 % av orden i meningarna. Utgångsvolyten vid testets början, för både signal (meningarna) och brus, låg på 65 dB SPL (Sound Pressure Level), vilket ungefär motsvarar ljudstyrkan på normal samtalston.. Bruset låg på tre sekunder före och en sekund efter meningens. Ordningen mellan deltest två och tre, *frågedelen* och *upprepningsdelen*, varierade beroende på om det var en FP med ett jämnt eller ojämnt ordningsnummer. Ordningen mellan listorna och den inbördes ordningen mellan meningarna på varje lista slumpades fram under varje testomgång. De FP som gjorde testet två gånger fick de två deltesten i samma ordning vid varje testtillfälle för att förutsättningarna skulle vara desamma. Det var bara på meningar som repeterades helt korrekt som en fråga eller upprepningsuppgift gavs. Förutsättningarna var sådana att upprepningsdelen och frågedelen skulle genomföras 12 gånger vardera av FP. Detta innebar att FP var tvungen att repetera sammanlagt 24 meningar korrekt, på det återstående meningsmaterialet, för att testet skulle fungera. Vilken frågetyp som kom till en mening slumpades. Dock fick varje försöksperson sex stycken av varje frågetyp. Då sex av de ursprungliga FP inte klarade av att repetera 24 meningar korrekt innan meningarna tog slut (vilket gjorde deras resultat oanvändbara), så testade vi sammanlagt 26 personer i pilotstudien.

Resultat

Sammanlagt deltog 236 försökspersoner i studien. Vi uppnådde inte syftet med vår studie då det slutliga pilottestet visade sig vara för enkelt och kraftiga takeffekter uppkom. I hela 37 % av testomgångarna fick FP alla rätt på frågedelen och i 17 % av testomgångarna hade de alla rätt på upprepningsdelen. De sämsta prestationerna på frågedelen respektive upprepningsdelen var åtta av tolv respektive sex av tolv, och någon risk för golveffekter verkar därmed inte föreligga i testets nuvarande utformning. Det visade sig också vara skillnad i svårighetsgraden på de två frågetyperna. Förstafrågan hade i genomsnitt 86 % korrekta

svar medan andrafrågan i genomsnitt hade hela 97 % korrekta svar.

Inga signifikanta inlärningseffekter vad gäller prestation på frågedelen och repetitionsdelen mellan testomgångarna påvisades hos de FP som upprepade testet. Det är svårt att säga om det beror på att takeffekterna i sig försvårar en förbättring, eller om testets utformning förhindrar en sådan inlärning. Däremot fanns en tydligt signifikant sänkning av signal/brus-förhållandet mellan testomgångarna för de FP som gjorde testet två gånger. Samtliga FP hade lägre signal/brus-förhållande i andra omgången och måste därför ha presterat bättre på de första 30 meningarna under testomgång två än under testomgång ett. Detta påvisar i sin tur att en viss inlärningseffekt förekom.

Diskussion

Kognitiva funktioner som spelar in vid taluppfattning är arbetsminne, uppmärksamhet och informations- bearbetning. Ju bättre förmågor, desto lättare att förstå tal och desto mindre handikapp får man av en hörselnedsättning.

Att ta fram meningar till testet var inte helt enkelt, eftersom vi ville ha meningar som på samma gång var naturliga och lågredundanta. Flertalet FP hävdade att det var svårt att lista ut delar av meningens som de inte hörde, vilket tyder på att meningsmaterialet var lågredundant. En högre kvalitet på meningsmaterialet skulle antagligen ha uppnåtts om vi hade haft längre tid på oss i och med att vi då hade kunnat göra fler deltester vid framtagandet av meningensfrågematerialet.

Bedömningarna av meningarna var subjektiva eftersom det bland gemene man råder delade meningar om vad som är en naturlig mening. Det kvantitativa testet hade kanske blivit mindre subjektivt om vi hade försökt att ge försökspersonerna en bättre definition på vad vi menar med en naturlig mening. Vår tanke med utvärderingen var att mäta den spontana reaktionen på meningarna, men denna försvann i de fall då personerna som utvärderade meningarna analyserade och letade fel.

Vi tog bara hänsyn till ljudstyrka när vi skulle likställa meningarnas svårighetsgrad. Denna

metod var grovhuggen, men bättre än att inte kompensera alls.

Den struktur frågorna vi ställde hade, kan ha medverkat till den takeffekt som uppkom i pilottestet. Anledningen kan vara att frågorna inte tvingade FP att bearbeta innehållet, det räckte att återkalla orden i meningen. FP kan även ha skapat strategier och en tänkbar sådan är att särskilt koncentrera sig på att minnas en mening som man kände sig säker på, och därför inte koncentrera sig maximalt på nästföljande. En bidragande faktor till uppkomsten av en sådan strategi beror möjligen på att FP i instruktionerna fick reda på att fråge- och upprepningssuppgiften endast gällde de meningar som repeterats korrekt. Om FP följde en sådan strategi kan det ha inneburit att FP inte stördes av nästföljande mening, som var tänkt som ett störningsmoment.

Ju längre tid som går desto troligare är det att FP hinner glömma meningen. Till skillnad från LiSSper hade vi inte en bestämd tid mellan meningarna, vilket hade varit bra eftersom samspelet mellan FP och försöksledare ska ha så lite betydelse som möjligt.

Slutsats

Testet visade sig vara för enkelt och gav upphov till kraftiga takeffekter. Meningsmaterialet var tänkt att motverka inlärningseffekter, men pilottestet gav inte tillräckligt med underlag för att avgöra om vi hade gjort det. Vi ser tre lösningar till problemet med takeffekter. Att ställa fler frågor under frågedelen och att förlänga repetitionsdelen är den första, att göra om frågorna den andra, och slutligen att ändra upplägget och strukturen på testet.

Även om man lyckas komma tillrätta med takeffekterna så kvarstår inlärningseffekterna, samt att visa att testet verkligen kan användas för att påvisa individuella skillnader i arbetsminnet hos människor kopplat till deras taluppfattningsförmåga.

Referenser

Andersson, U. (2001). *Cognitive deafness*. Institutionen för beteendevetenskap, Linköping.

Cowan, N. (1998). Visual and auditory working memory capacity. I: *Trends in Cognitive Sciences*, nr 2, 3:e upplagan s 77 <<http://www.sciencedirect.com>>. Hämtat 12/2 2003.

Eysenck, M. W. & Keane, M. T. (2000). *Cognitive psychology*. Psychology Press Ltd.

Gatehouse, S., Elberling, C. & Naylor, G. (under tryckning) Benefits From Hearing Aids in Relation to the Interaction between the User and the Environment. Överlämnad till: *International Journal of Audiology*

Gathercole, S. E. & Baddeley, A. D. (1993). *Working memory and language*. Lawrence Erlbaum Associates Ltd.

Hagerman, B. (1984). *Some aspects of methodology in speech audiometry*. Avdelningen för teknisk audiologi, Karolinska Institutet, Stockholm.

Hällgren, M. (2002). *Listening span and speech recognition*. (manuskript) Avdelningen för teknisk audiologi, Linköpings universitet.

Nilsson, Michael, Soli, Sigfrid D. & Sullivan, Jean A. (1994) Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. I: *Journal of the Acoustic Society of America* nr 2, s. 1085-1098.