



LUND UNIVERSITY

Faktorer som påverkar vår bedömning av kvinnliga och manliga talares ålder

Schötz, Susanne

Published in:
Vetenskapssocieteten i Lunds. Årsbok 2008

2008

[Link to publication](#)

Citation for published version (APA):

Schötz, S. (2008). Faktorer som påverkar vår bedömning av kvinnliga och manliga talares ålder. I V. Molnár, B. Jonsson, & G. Andersson (Red.), *Vetenskapssocieteten i Lunds. Årsbok 2008* (s. 75-84). (Årsbok). Vetenskapssocieteten i Lund.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

FAKTORER SOM PÅVERKAR VÅR BEDÖMNING AV KVINNliga OCH MANliga TALARES ÅLDER

Susanne Schötz

Inledning

Åldrandet är en komplex process som påverkar alla levande varelser på en mängd olika vis. Hos oss människor gäller detta även sättet vi talar på. Våra röster och hur vi talar förändras från tidig barndom till hög ålder, och följaktligen avspeglas vår ålder i talet. De flesta människor är ganska bra på att höra hur gammal en okänd – och icke synlig – person är (Shipp och Hollien, 1969; Linville, 2001). Men hur nära kan vi med gissningar komma en okänd talares faktiska, kronologiska, ålder enbart genom att lyssna på ett talprov, till exempel en ljudinspelning med någon som läser en mening? Vilka ledtrådar använder vi och kan vi använda samma slags ledtrådar vid bedömning av alla slags talare eller finns det skillnader mellan olika talares talmönster när de åldras? Den här artikeln handlar om hur mycket och vilken slags information i tal vi behöver för att kunna göra en så bra åldersbedömning som möjligt av talare.

Bakgrund och tidigare forskning

När vi bedömer en talares ålder beror vår noggrannhet bland annat på vilka precisionskrav vi har, det vill säga hur nära talarens exakta ålder vi vill komma, och även på talprovets längd och typ, till exempel en utdragen vokal, en läst mening eller en spontan monolog. Ju mindre information det finns i talprovet, desto svårare blir uppgiften. Men även om informationen är mycket knapp är lyssnaren inte helt utlämnad åt gissningar. Noggrannheten kan även påverkas av karaktärsdrag hos både talaren och lyssnaren (Ramig och Ringel, 1983; Linville, 2001), såsom kön, åldersgrupp, talarens fysiska och fysiologiska tillstånd samt lyssnarens förtroenhet med liknande talare (exempelvis dialekt, åldersgrupp). Följaktligen kan en del talare vara svårare att bedöma än andra.

Det har bedrivits en hel del forskning om hur mänskliga lyssnare uppfattar talares ålder (Ptacek och Sander, 1966; Huntley m.fl., 1987; Braun och Cerrato, 1999; Linville, 2001; Brückl och Sendlmeier, 2003). Tyvärr är dessa studier ofta svåra att jämföra på grund av skillnader både i talprover (stimuli) och metod. Skillnaderna finns i:

1. talarnas språk
2. talprovernans längd (stimulusduration)
3. talets typ (utdragna vokaler, viskade vokaler, enstaka ord, läst tal (uppspelat rätt eller baklänges) spontantal o.s.v.)
4. ljudkvalitet (HiFi, telefonöverfört tal o.s.v.)
5. talarnas kön och ålder
6. lyssnarnas kön och ålder
7. bedömningsuppgiften (klassificering i 2/3/7 åldersgrupper, exakt i antal år o.s.v.)
8. mått för resultaten (korrelation, absolut medelfel, % rätt o.s.v.)

Det finns ytterligare faktorer att ta i beaktande: tydliga fonetiska skillnader mellan kvinnliga och manliga talare, och mellan äldre och yngre, i såväl struphuvudet och svalget som näs- och munhålorna (Hollien, 1987; Klatt & Klatt, 1990; Biemans & Van Bezooijen, 1999; Linville, 2001). Dessa skillnader kan härledas till fysiska och fysiologiska faktorer, d.v.s. talorganens storlek och form. Kvinnor och män utvecklas dessutom olika (Higgins & Saxman, 1991). Ett exempel är i puberteten, under vilken pojkars röstlägen sjunker markant, medan flickors

påverkas i mer måttlig grad (Lindblad, 1992). Det finns även en mängd sociala faktorer (kultur, miljö, vanor) att beakta. Därför använder kanske lyssnare olika strategier när de bedömer olika talargruppers ålder. I en studie av automatisk (dator-) bedömning av äldre talares ålder, lyckades Müller m.fl. (2003) framgångsrikt bygga automatiska könsspecifika ålders-klassificerare. Författaren (2005) fann skillnader hos både mänsklig och automatisk bedömning av kvinnors och mäns åldrar i en studie med enbart enstaka ord som stimuli. Medan röstläget (grundtonsfrekvens, F_0) utgjorde en bättre ledtråd för kvinnlig ålder, var spektrala akustiska mönster som beror på talrörets längd och form (främst formantfrekvenser) bättre ledtrådar för bedömning av manlig ålder. En möjlig förklaring till detta är att kvinnliga röster verkar uppfattas som mer komplexa än manliga (Murry och Singh, 1980). Detta skulle innebära att lyssnare behöver en delvis annorlunda och kanske även större uppsättning ledtrådar när de ska bedöma kvinnlig talarålder.

För att få svar på några av de återstående frågorna rörande perception av såväl kvinnlig som manlig talarålder genomfördes en perceptionsstudie som jämförde lyssnares åldersbedömningar från tal av olika typ och längd uttalat av både kvinnliga och manliga talare. Studien ingår även som en del av kapitel 2 i Schötz (2006).

Studiens syfte, mål, frågor och hypoteser

Syftet med undersökningen var att bestämma i vilken utsträckning lyssnares bedömning av såväl kvinnlig som manlig talarålder påverkas av 1) olika typer av stimuli (talprover bestående av enstaka ord samt spontantal) och 2) olika stimulusduration. Målet var att ytterligare belysa hur mycket information i tal som behövs för noggranna åldersbedömningar.

Följande forskningsfrågor ställdes:

1. Hur påverkar stimulustyp och -duration precisionen i lyssnares bedömningar av ålder?
2. Finns det skillnader i hur vi skattar kvinnlig och manlig talarålder vad avser stimulusduration och -typ?

Hypoteserna var att stimuli med spontantal och längre duration skulle leda till högre bedömningsprecision och att det skulle finnas skillnader i åldersbedömning av kvinnliga och manliga talare.

Material och metod

Talproverna till studien hämtades från den taldatabas som samlades in för det svenska dialektprojektet SweDia 2000 (Bruce m.fl., 1999). Den innehåller talprover (bl.a. ordlistor med enstaka ord och spontana intervjuer) av över 600 talare från drygt 100 orter i Sverige och de svensktalande delarna av Finland. Sex talare från vardera fyra olika grupper – äldre kvinnor (63–82 år), äldre män (60–75 år), yngre kvinnor (24–32 år) och yngre män (63–82 år) valdes slumpartat ut bland talarna från Götaland. För var och en av de 24 utvalda talarna valdes fyra talprover ut. Dessa signalbehandlades först för att utjämna eventuella ljudnivåskillnader mellan talproven (intensitetsnormalisering) och användes sedan som stimuli i fyra olika perceptionstester, där lyssnare ombads bedöma talarnas ålder:

Test 1: ca 10 sekunder spontan monolog

Test 2: ca 3 sekunder spontan monolog

Test 3: 6 enstaka ord spelade efter varandra: *käke, saker, själen, sot, typ* och *tack*; ca 4 sek.

Test 4: ett enstaka ord: *rasa*; ca 0.65 sek.

De fyra perceptionstesterna – ett för varje uppsättning stimuli – genomfördes med olika lyssnargrupper. Två lyssnargrupper deltog i var sitt test, medan en tredje grupp utsattes för två av testerna. Kön- och åldersfördelningen för de tre lyssnargrupperna samt vilken grupp som deltog i vilket test visas i Tabell 1. Samtliga försökspersoner var fonetikstudenter vid Lunds universitet. Uppgiften bestod i att lyssna på 24 stimuli som spelades upp från en dator (Apple PowerBook G4 med Harman Kardon SoundSticks högtalare) i slumpmässig ordning och därvid göra så exakta åldersbedömningar av talarna som möjligt. Lyssnarna fick höra på varje stimulus bara en gång och bedömningen skulle baseras på deras första intryck av talarens ålder. Efter själva bedömningsuppgiften ombads de att skriva ner ledtrådar som skulle kunna ha påverkat deras bedömningar.

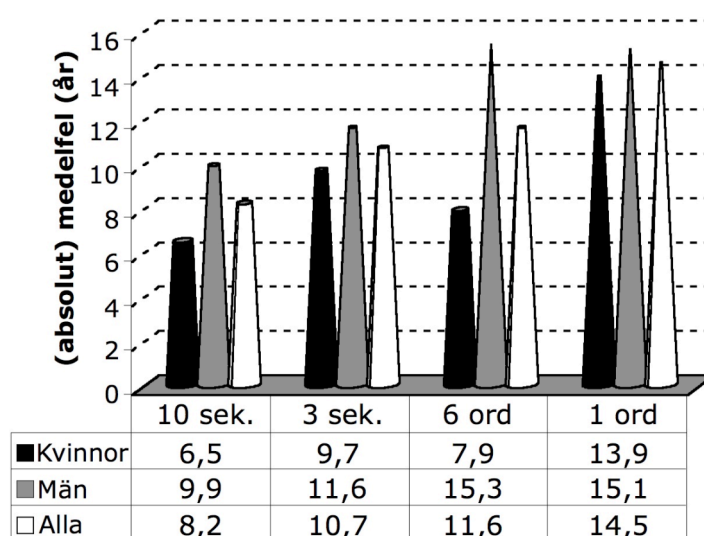
Tabell 1. De fyra testernas stimulustyp, antal försökspersoner samt köns- och åldersfördelning (*N* = antal, *K* = kvinnor, *M* = män).

Test (stimuli)	<i>N</i>	<i>K</i>	<i>M</i>	Åldersfördelning (medelvärde/median)
1 (10 sek. spontantal)	31	18	11	19–65 (27/23)
2 (3 sek. spontantal)	33	22	11	19–57 (25/23)
3 (6 ord)	37	33	4	19–55 (26/24)
4 (1 ord)	37	33	4	19–55 (26/24)

Resultat

Precision

Figur 1 visar det absoluta medelfelet, det vill säga medelvärdet på den absoluta skillnaden mellan lyssnarnas åldersbedömningar och talarnas egentliga kronologiska ålder i antal år, för kvinnliga, manliga samt för alla talare (båda könen) i de fyra testerna. I de tre första var lyssnarnas bedömningar mer än dubbelt så exakta som bedömningarna från en så kallad baslinje, där samtliga 24 stimuli skattades till 47,5 år (= talarnas medelålder, se Tabell 2). Resultaten för Test 4 med de kortaste talproverna nådde bara halvvägs mellan baslinjen och de övriga testerna.



Figur 1. Absolut medelfel för de fyra stimulityperna för kvinnliga, manliga och alla talare.

I samtliga fyra tester var lyssnarnas bedömningar av kvinnor bättre än av män. Det allra lägsta absoluta medelfelet – och alltså det bästa resultatet – uppmättes för kvinnliga stimuli med 10-

sekunders spontantal (6,5 år), medan det högsta felet uppmättes för manliga 6-ords-stimuli (15.3 år). Summan, medelvärden och medianvärden för samtliga 24 talare i de fyra perceptionstesterna och även för baslinjen visas i Tabell 2.

Tabell 2: Summa, medelvärde och medianvärde av det absoluta medelfelet för samtliga talare i de fyra testerna samt för baslinjen (beräknad till att alltid gissa på 47,5 år, d.v.s. medelvärdet för samtliga talars ålder).

<i>Test</i>	<i>1 (10 sek.)</i>	<i>2 (3 sek.)</i>	<i>3 (6 ord)</i>	<i>4 (1 ord)</i>	<i>Baslinje</i>
summa	196,5	256,1	277,6	348,7	497,0
medelvärde	8,2	10,7	11,6	14,5	20,7
median	7,2	10,0	10,0	16,7	19,5

Köns- och stimuluseffekter

För att testa den statistiska signifikansen utsattes testresultaten (medelvärdet av lyssnarnas absoluta medelfel) för två olika variansanalyser (ANOVA). Talarnas *kön* utgjorde ”within-subject factor” i båda, medan ”between-subjects factor” i den första analysen var *stimulusduration* (kort (1 ord), medium (6 ord och 3 sek.) eller lång (10 sek.)) och i den andra *stimulustyp* (spontantal eller ord).

Stimulusduration

Längre stimulusduration resulterade generellt i signifikant högre åldersbedömningsprecision ($F(2,100)=71.059$, $p<.05$). Resultaten för kvinnliga och manliga talare skiljde sig dock åt. Precisionen förbättrades mer med längre stimulusdurationer för kvinnliga talare än för manliga. För kvinnorna observerades ett lägre medelfel för 10-sekundersstimuli (6,5) jämfört med 3-sekundersstimuli (9,7). Dessutom var felet för 6-ordsstimuli (7,9) lägre än för 1-ordsstimuli (13,9). En betydligt mindre precisionsskillnad mellan längre och kortare stimulusdurationer observerades för de manliga talarna. Det absoluta medelfelet var 9,9 för de längsta talproverna (10 sek.), ett högre medelfel uppmättes för de två medellånga (3-sekunder: 11,6 och 6-ord: 15,3), och för de kortaste 1-ordstalproverna låg medelfelet (15,1) mellan medelfelen för de båda medellånga. Den statistiska så kallade interaktionen mellan talarnas kön och stimulusduration var dock generellt inte signifikant ($F(2,100) = 2.171$, ns).

Stimulustyp

I den andra analysen var interaktionen mellan talarnas kön och stimulustyp signifikant ($F(1,68) = 39.296$, $p<.05$). Lyssnarnas bedömningar för manliga talare var bättre för spontana stimuli (medelfel: 9,9 och 11,6) än för ordstimuli (medelfel: 15,3 och 15,1). För de kvinnliga talarna var däremot medelfelet för sexordsstimuli (7,9) lägre än för spontana trestekundersstimuli (9,7), men något högre jämfört med spontana tiosekundersstimuli (6,5).

Lyssnarledtrådar

Många av lyssnarna uppgav flera ledtrådar som de ansåg hade påverkat deras åldersbedömningar. De vanligaste redovisas i Tabell 3. Talarnas dialekt, röstläge och röstkvalitet verkar ha påverkat lyssnarnas bedömningar i samtliga lyssnartester, medan talprovernas lingvistiska (semantiska) innehåll endast gäller testerna med spontantal. En vanlig kommentar från lyssnare som bedömt spontana stimuli rörde talare som pratar om något som hänt tidigare – vare sig det var för fem eller femtio år sedan. Dessa talare bedömdes ofta som äldre, oavsett övriga ledtrådar som taltempo och dialekt.

Tabell 3: Ledtrådar, som enligt lyssnarna påverkat deras bedömningar (totalt antal lyssnare).

Test (stimuli)	Dialekt	Röstläge	Röstkvalitet	Semantiskt innehåll
1 (10 sek. spontantal)	24	23	25	22
2 (3 sek. spontantal)	25	26	32	17
3 & 4 (6 resp. 1 ord)	33	31	34	0

Diskussion

Även om endast ett begränsat antal stimulustyper och -durationer undersöktes i denna studie, observerades flera intressanta resultat. Dessa diskuteras nedan tillsammans med några förslag på vidare undersökningar.

Precision

Lyssnarna visade sig vara avsevärt bättre (ungefär dubbelt så bra) på att bedöma talarålder jämfört med baslinjen i tre av testerna, vilket stämmer bra överens med tidigare studier. Frågan kvarstår dock hur hög precision vi bör kunna förvänta oss. När vi har att göra med precision för bedömd ålder, måste även skillnader i talarnas kronologiska ålder tas med i beräkningen. Ett absolut medelfel på 10 år kan uppfattas som mindre precist om talarens kronologiska ålder är 30 år (en bedömning på 20 år kan beräknas som $20/30 = 66,7\%$ korrekt) än om den är 80 (en bedömning på 70 år kan beräknas som $70/80 = 87,5\%$ korrekt). Vi behöver uppenbarligen ett bättre mått på precision för den här typen av bedömningar.

Då det inte gjordes någon kontroll av lyssnargruppernas samstämmighet, kan man inte veta om det faktum att tre olika lyssnargrupper deltog i studiens fyra tester kan ha påverkat resultaten. I samtliga fyra tester var lyssnarnas åldersbedömningar av kvinnor mer precisa än av män, kanske för att majoriteten av lyssnarna själva var kvinnor. Det är dock fortfarande oklart hur lyssnarnas kön påverkar precisionen i åldersbedömningsuppgifter, något som måste undersökas vidare. Även om de flesta studier inte rapporterat några precisionsskillnader mellan kvinnliga och manliga lyssnare, har en del forskare rapporterat att kvinnor är bättre än män på att bedöma talarålder, medan andra menar att män är något bättre än kvinnor (Braun och Cerrato, 1999). En annan möjlig förklaring till den här studiens resultat är att den manliga talargruppen bestod av fler svårbedömda talare än den kvinnliga. Shipp och Hollien (1969) fann att talare vars ålder var svår att bedöma ledde till stor spridning i perceptionstesternas bedömningsresultat – med standardavvikelser på nio år eller högre. Eventuellt skulle ett sådant mått kunna användas för att bedöma om talare är representativa för sin kronologiska ålder eller ej.

Stimuluseffekter

I den här studien ledde längre stimulusdurationer oftast till högre precision i lyssnarnas åldersbedömningar. Detta väcker frågan om vilken som är den optimala talprovs längden i åldersperceptionstester. När leder en ytterligare durationsökning inte längre till en högre precision i bedömningarna? Fler experiment med större och mer systematiska variationer i stimulusdurationer behövs för att besvara denna fråga.

Signifikanta effekter observerades både för precision och kön i de två stimulustyperna i denna undersökning. Enstaka ord och spontantal kan dock vara svåra att jämföra i en studie av talarålder. Flera av lyssnarna angav att de påverkades av det lingvistiska (semantiska) innehållet i spontana stimuli, vilket skulle kunna förklara varför manliga spontana stimuli uppnådde högre precision än motsvarande ordstimuli. Spontantal ger inte bara mer information om talaren (dialekt, ordval med mera), utan innehåller dessutom mer prosodisk och spektral variation än enstaka ord. Den lägre precisionen för trestekundersstimuli jämfört

med sexordsstimuli för kvinnliga talare kan dock inte förklaras enbart som en effekt av stimulustyp. Det skulle vara intressant att jämföra ett större antal stimulustyper för att identifiera de lämpligaste typerna för bedömning av såväl kvinnlig som manlig talarålder. Framtida studier inkluderar undersökningar där ett flertal olika stimulustyper jämförs och varieras mer systematiskt med avseende på fonetiskt innehåll och kvalitet såväl som fonetisk variation och dynamik.

Könseffekter

Skillnaderna mellan kvinnor och män när det gällde vilken typ av stimulus som resulterade i en högre åldersbedömningsprecision har redan nämnts. En möjlig förklaring till dessa skillnader är att lyssnare använder olika ledtrådar och strategier när de uppskattar kvinnlig och manlig ålder. Författaren (2005) föreslår att vi som lyssnare möjligen använder mer prosodiska ledtrådar (framför allt röstläge) när vi ska bedöma kvinnliga talare, men att vi föredrar spektrala ledtrådar (formanter, spektral balans med mera) när vi ska bedöma manliga. Resultaten från studien som beskrivs i denna artikel pekar i samma riktning: manligt spontantal verkar förse lyssnare med mer spektral information – och därmed bättre manliga åldersledtrådar – medan längre stimulus-durationer innehåller mer prosodisk information, som verkar behövas för att bättre kunna bedöma kvinnlig ålder. Det krävs dock fler studier för att belysa skillnaderna mellan perception av kvinnlig och manlig ålder, och talarnas kön måste beaktas i framtida undersökningar av perceptoriska ledtrådar såväl som akustiska korrelerat till talarålder.

Slutsatser

Även om det behövs mer forskning för att verifiera och bygga på resultaten från denna studie, går det att dra två försiktiga slutsatser:

1. Stimuli med längre durationer och med spontantal verkar förbättra lyssnares precision vid bedömning av talarålder.
2. Det finns skillnader mellan perception av kvinnlig och manlig ålder, vilka kan förklaras med motsvarande skillnader i lyssnarstrategier och val av perceptoriska ledtrådar.

Sammanfattning

Flera tidigare undersökningar har visat att mänskliga lyssnare är ganska bra på att gissa en okänd (och icke synlig) talares ålder. Studien som beskrivits i denna artikel tar upp några av de frågor om mänsklig åldersbedömning som ännu inte besvarats av tidigare forskning. Perception av talarålder undersöktes i fyra lyssningstester med stimuli av olika längd (0,65, 3, 4 och 10 sekunder) och typ (enstaka ord och spontantal). Resultaten pekar på att lyssnare använder olika ledtrådar och strategier vid bedömning av kvinnlig och manlig talarålder. Längre stimulusduration (oavsett typ) innehåller mer prosodisk information som verkar vara viktig vid kvinnlig åldersbedömning, medan stimulustypen spontantal (oavsett längd) ger mer spektral information som förefaller viktigare när vi bedömer manlig talarålder. Den här studiens resultat kommer att användas som utgångspunkt i ytterligare perceptionsstudier av talarålder med mer systematisk variation av stimulusduration och -typ.

Litteratur

- Biemans M & Van Bezooijen. (1999) Biological gender and social gender in relation to voice quality. Proceedings of ICPhS 99 (San Francisco), 1249–1252.
- Braun A & Cerrato L. (1999) Estimating speaker age across languages. Proceedings of ICPhS 99 (San Francisco), 1369–1372.

- Bruce G, Elert C-C, Engstrand O & Eriksson A. (1999) Phonetics and phonology of the Swedish dialects – a project presentation and a database demonstrator. Proceedings of ICPHS 99 (San Francisco), 321–324.
- Brückl M och Sendlmeier W. (2003) Aging female voices: An acoustic and perceptive analysis. Proceedings of VOQUAL' 03 (Geneva), 163–168.
- Cerrato L, Falcone M & Paoloni A. (1998) Age estimation of telephonic voices. Proceedings of the RLA2C conference (Avignon), 20–24.
- Higgins M B & Saxman J H. (1991) A comparison of selected phonatory behaviours of healthy aged and young adults. Journal of Speech and Hearing Research 13, 1000–1010.
- Hollien H. (1987) “Old Voices”: What Do We Really Know About Them? Journal of Voice 1(1), 2–13.
- Huntley R, Hollien H & Shipp T. (1987) Influences of listener characteristics on perceived age estimations. Journal of Voice 1, 49–52.
- Klatt D & Klatt L. (1990) Analysis, Synthesis, and Perception of Voice Quality Variations Among Female and Male Listeners. JASA. 87(2): 820–857.
- Lindblad P. (1992) Röst. Lund: Studentlitteratur.
- Linville S E. (2001) Vocal Aging. San Diego: Singular Thomson Learning.
- Müller C, Wittig F & Baus J. (2003) Exploiting speech for recognizing elderly users to respond to their special needs. Proceedings of Eurospeech 2003 (Geneva), 1305–1308.
- Murry T & Singh S. (1980) Multidimensional analysis of male and female voices. JASA 68 (5), 1294–1300.
- Ptacek P H & Sander E K. (1966) Age recognition from voice. Journal of Speech and Hearing Research 9, 273–277.
- Ramig L A & Ringel R L. (1983) Effects of physiological aging on selected acoustic features. Journal of Speech and Hearing Research 26. 22–30.
- Schötz S. (2005) Prosodic cues in human and machine estimation of female and male speaker age. In G. Bruce & M. Horne (Eds.) Nordic Prosody: Proceedings of the IXth Conference, Lund, 2004. Frankfurt am Main: P. Lang, 215–223.
- Schötz, S. (2006) Perception, Analysis and Synthesis of Speaker Age. PhD thesis, Travaux de l'Institut de linguistique de Lund 47. Lund: Dept. of Linguistics and Phonetics, Lund University.
- Shipp T & Hollien H. (1969) Perception of the aging male voice. Journal of Speech and Hearing Research 12, 703–710.

Presentation av min forskning

Susanne Schötz disputerade i fonetik 2006 med en avhandling om perception, analys och syntes av talarålder. Hon var under sin doktorandtid även ansluten till Sveriges nationella forskarskola i språkteknologi (GSLT). Sedan 2007 forskar hon i projektet SIMULEKT (Simulering av svenska dialekter), som är ett samarbete mellan Lunds universitet och Kungliga tekniska högskolan (KTH) i Stockholm finansierat av Vetenskapsrådet. Hon är dessutom IT-pedagog i Humanistlaboratoriet på Språk- och Litteraturcentrum, där hon undervisar, utvecklar läromedel samt ansvarar för den fonetiska utrustningen.