



Kan man känna igen sin kollega på en hostning?

Ett rättsfonetiskt röstidentifikationsexperiment



Anna Ericsson

***Examensarbete 5 poäng (7.5 hp), VT 2008
Logonomprogrammet, SMI
Handledare: Owe Ander, Margareta Thalén***

INNEHÅLLSFÖRTECKNING

1. Inledning	1
2. Bakgrund	2
2.1. Lingvistik och fonetik inom rättsväsendet	2
2.2. Historik	3
2.3. Problem inom rättsfonetiken	5
2.4. Tidigare forskning	5
3. Metod	8
3.1. Inspelning	8
3.1.1. "Talmaterial"	8
3.2. Tester	9
3.2.1. Test 1 Identifiering	9
3.2.2. Test 2 Kön, ålder, längdbedömning	9
4. Resultat	10
4.1. Resultat test 1	10
4.1.1. Ett uppföljningstest - Kontroll av hur väl lyssnarna tycker sig känna försökspersonerna	12
4.2. Resultat test 2	14
4.2.1. Resultat könsbedömning	14
4.2.2. Resultat åldersbedömning	15
4.2.3. Resultat längdbedömning	16
5. Diskussion & Slutsatser	18
6. Sammanfattning	20
7. Litteratur och källor	22

1. Inledning

Det är fullt möjligt att känna igen en bekant röst på bara ett "hej" i telefonluren, det har nog de flesta erfarit. Men är det möjligt att identifiera personer på deras skratt, hostning eller s.k. feedback-ljud, bekräftningsljud, som /m/ och /a/?

Evolutionen har gett människan förmågan att snabbt identifiera olika arter. Vi kan på ett ögonblick avgöra om det är en björn eller en häst vi just hört. Denna förmåga har varit viktig för att hinna fly i tid från en annalkande fara. Vi är också skickliga på att urskilja och känna igen olika individer, ex. familjemedlemmar och vänner, och kan därmed lätt skilja dem från eventuella fiender. Men en del personers röster kan vara olika svåra att känna igen och därmed skilja från andras. Det finns personer med särskilt karaktäristiska röstegenskaper och andra med röster som inte alls är speciellt karaktäristiska. Det går att föreställa sig vissa skådespelares röster som det t.o.m. går att höra inom sig utan att de är i närheten, eftersom de är så utpräglat personliga. Samtidigt kan man lika lätt föreställa sig några vars röster det inte alls är lika lätt att påminna sig. På liknande vis kan man ofta lätt skilja mellan olika sångstilar som t.ex. mellan klassisk sång eller afrosång. Dessutom går det att utan svårighet skilja på olika sångares röster inom en och samma genre.

Anledningen till att vi minns olika röster olika bra och att vi dessutom lätt kan skilja röster från varandra är just att röster kan vara så olika. Olikheterna beror på flera saker. En av orsakerna är anatomin, hur själva talorganet eller röstinstrumentet är uppbyggt, det är t.ex. främst p.g.a. att män har större struphuvud och därmed längre och tjockare stämläppar (stämband) än barn och kvinnor som de har ett lägre taltonläge eller grundtonsfrekvens (F_0). Olikheterna beror också på hur vi använder oss av vårt instrument, hur mycket vi varierar tonen, hur fort vi talar, hur många och långa pauser vi gör, hur vi betonar m.m. Dessutom beror det också på vårt personliga språk, vilka ord och uttryck vi använder, om vi har en brytning, dialekt eller kanske någon sociolekt som återspeglas i talet.

Följande studie kan kallas rättsfonetisk. Inom rättsfonetik, även kallad forensisk fonetik, sysslar man bl.a. med identifikation av röster. Rättsfonetik och rättslingvistik berör områden inom lingvistik och fonetik som kan användas i rättsliga sammanhang. I denna studie kommer även den forensiska fonetiken att beskrivas närmare (se vidare under rubriken bakgrund).

Metoden i denna studie gick ut på att lyssnare fick försöka identifiera sina kollegor utifrån deras hostningar, skratt och bekräftningsljud. Frågan var här om man kan identifiera bekanta personer utifrån dessa minimala lingvistiska yttranden. Med vetenskapen om hur lätt man känner igen någon över telefonen skulle ett förväntat resultat vara att lyssnarna förmodligen klarar av identifieringen bra. Men troligen spelar det in hur väl lyssnarna känner dem som hostat, skrattat, sagt /m/ och /a/ (se vidare under rubriken metod).

I ett andra test fick en annan lyssnargrupp, utifrån samma material istället bedöma "talarnas" kön, ålder och längd. Denna lyssnargrupp kände ej försökspersonerna.

Det vore intressant om det går att se vilket som är lättare eller svårare att känna igen personerna på: host, skratt, /a/, eller /m/, det vore också intressant att se om det visar sig finnas någon slags samstämmighet bland lyssnarna.

Själva experimentet som ligger till grund för denna uppsats har ursprungligen utförts för en kurs i forensisk fonetik och i och med den har även hjälp med de statistiska beräkningarna erhållits.

2. Bakgrund

Här ges en beskrivning av forensisk fonetik och lingvistik. En historisk tillbakablick och läget kring tidigare forskning som berör denna studie kommer också att beskrivas liksom en del av de svårigheter som den forensiska fonetiken har att tampas med.

2.1. Lingvistik och fonetik inom rättsväsendet

Rättsfonetiken eller den forensiska fonetiken (fortsättningsvis kallad rättsfonetik) är något som används både utomlands och i Sverige, t.ex. för att avslöja brottslingar genom att identifiera deras röst. Vid ett brott är det inte ovanligt att gärningsmannen själv ringer till larmcentralen och rapporterar brottet. Sådana samtal kan användas i forensiska sammanhang då polisen vill veta om det är den person de gripit som är den gärningsman som ringt till larmcentralen. Det kan också handla om ett samtal från en avlyssnad telefon. Genom att analysera rösten i telefonsamtalet och jämföra den med gärningsmannens röst från förhørsinspelningar försöker man komma fram till om de två rösterna tillhör samma person. Ibland finns bara en vittnesuppgift om hur en gärningsmans röst lät eller vittnets minne av dennes röst eller sätt att tala. Då kan man skapa en s.k. "voice lineup" som kan jämföras med vittneskonfrontation men där det handlar om ett antal röster att lyssna på istället för personer att titta på. Öronvittnet som hört gärningsmannens röst får försöka identifiera gärningsmannens röst, om den återfinns bland ett antal röster, så kallade figuranter.

Med ett vidare begrepp kan man kalla detta för *rättslingvistik* eller *forensisk lingvistik* och i många fall skulle man kunna påstå att det är ett mer korrekt begrepp eftersom en undersökning ofta handlar om en bedömning av såväl grammatiska och andra språkliga egenheter som specifika röstegenskaper. Det kan röra sig om det taltonläge man använder (grundtonen, eller om man så vill F_0) och dess variation – själva melodin i talet. Det kan också vara vilken typ av röstkvalité och röst användning personen har, rytmen i språket, dialektala drag o.s.v. Vidare kan det vara sociolingvistiska språkliga kännetecken, personliga uttryck, speciella talesätt och särskilda artikulatoriska egenheter m.m. Philip Rose gör en tydlig beskrivning i sin bok *Forensic Speaker Identification*¹:

Forensic speaker identification is a part of forensic phonetics. Forensic phonetics is in turn an application of the subject of phonetics. [...] Phonetics is concerned primarily with speech: it studies especially how people speak, how the speech is transmitted acoustically, and how it is perceived.

As well as forensic speaker identification, forensic phonetics includes areas such as speaker profiling (in the absence of a suspect, saying something about the regional or socioeconomic accent of the offender's voice(s); the construction of voice line-ups; content identification (determining what was said when recordings are of bad quality, or when the voice is pathological or has a foreign accent); and tape authentication (determining whether a tape has been tampered with) (French 1994:170; 182- 4; Nolan 1997: 746).

¹ Rose 2002 s.2

2.2. Historik

Här följer en beskrivning av rättsfonetikens historia. Där inte annat anges är informationen hämtad från Tutorial on Forensic Speech Science av Anders Eriksson, professor i fonetik vid Göteborgs Universitet².

Redan romaren Quintilianus (Marcus Fabius Quintilianus 35-96 f.Kr.) menade att en människas röst är lika lätt att urskilja för örat som ett ansikte för ögat³. Denna egenskap har utnyttjats i rättsliga sammanhang genom talaridentifikation och använts som bevis så tidigt som på 1600-talet. Ett känt fall från 1660 är rättegången där William Hulet anklagades för att ha mördat Kung Charles I. Ett vittne hade hört bödeln och kände igen honom som Hulet "by his speech"⁴. Hulet dömdes då till döden, eftersom han inte var den som skulle ha utfört avrättningen. Senare erkände den ordinarie bödeln att det var han som utfört avrättningen och Hulet friades. I rättsliga sammanhang idag är det inte ovanligt att förövaren på liknande vis blivit hörd men inte sedd.

Ett annat känt fall där röstigenkänning har spelat en stor roll var då Charles Lindbergs son kidnappades 1932. Vid överlämnandet av lösensumman hörde Lindberg kidnapparens röst. 29 månader senare identifierade Lindberg den misstänktes röst som kidnapparen efter att ha hört honom upprepa samma fras. Med tanke på hur minnet fungerar är det ganska anmärkningsvärt att han tyckte sig känna igen en röst efter så lång tid (se vidare under rubriken tidigare forskning).

I och med att man uppfann telefonen och inspelningsutrustning öppnades nya dörrar för forensisk fonetik. Verktyg för att analysera tal utvecklades och spektrografen uppfanns som gjorde det möjligt att få fram en bild av tal. Mycket av utvecklingen skedde på *Bell Telephone Laboratories* under sent 30-tal och framåt. En senare variant av spektrografen utvecklades av *Kay Elemetrics* och såldes i kommersiellt syfte under namnet *Sonagraf*. Spektrografen utvecklades från början i syftet att användas för fonetisk forskning av tal och var tänkt att användas i uttalsträning för döva eller studenter med brytning. Den användes egentligen aldrig riktigt på det viset men spektrografen har haft stor betydelse för den fonetiska forskningen. Men eftersom denna tidiga forskning i USA räknades som ett krigsprojekt har ingenting av dessa tidiga studier publicerats. Först 1945 skriver Potter, en av forskarna vid Bell Labs, i en av de första publikationerna efter kriget:

The work here described was begun before the war. Because of related war interests it was given official rating as a war project, and has progressed far enough during the war period to justify its being brought to public attention.

Det har spekulerats kring vad militären egentligen ville använda denna forskning till, ville de i själva verket ta fram en pålitlig metod att identifiera röster med⁵? Tidigt började termen "voiceprint" användas i dessa sammanhang för de spektrogram eller röstavbilder man fick från spektrografen. Begreppet voiceprint ger intryck av att det handlar om ett röstavtryck motsvarande fingeravtrycket. Denna term användes också när man senare sysslade med talaridentifikation utifrån spektrogram, vilket starkt har ifrågasatts⁶. Dels är termen voiceprint kontroversiell eftersom ett spektrogram av en röst inte alls är lika oföränderligt som ett fingeravtryck, dels är talarigenkänning

² Eriksson 2005

³ Quintilianus Den fulländade talaren 2002 Wahlström & Widstrand Stockholm

⁴ Eriksson 2005 s 2

⁵ Eriksson 2005 s 3 refererar till Meuwly 2003, Grey & Kopp 1944

⁶ Dessa experiment startades av Lawrence Kersta, tidigare forskare vid Bell Labs.

utifrån spektrogram en verksamhet man kan ifrågasätta just därför att talet är så varierat. Denna voiceprintdebatt tas inte upp mer ingående här men en rolig detalj att nämna i sammanhanget är vad som hände då FBI-chefen var på Sverigebesök 1970. Dåvarande FBI-chefen J Edgar Hoover var på Stockholmsbesök och intervjuades av DN. Det resulterade i en debatt mellan den svenske talteknologen och forskaren Gunnar Fant⁷ och J Edgar Hoover. Debatten handlade om voiceprint och dess tillförlitlighet. Hoover uttalade sig om förträffligheten med detta voiceprint. Han hävdade att en röst är lika karaktäristisk som ett fingeravtryck och att man därför också lika exakt skulle kunna identifiera en person genom dennes röstavtryck. Fant svarade att ett voiceprint inte alls kan jämföras med ett fingeravtryck i fråga om tillförlitlighet, utan att rösten är betydligt mer föränderlig. Han hamnade då på bild på förstasidan i DN bredvid J Edgar Hoover, med bildtexten "Gunnar Fant - FBI enemy number one"⁸.

Denna debatt pågår egentligen alltjämt, inte minst för att man återkommande luras genom olika TV-serier och filmer att tro att en person enkelt kan identifieras på rösten, ofta genom att man skickar in en inspelning i någon magisk apparat som genast talar om vem det är på inspelningen. Det kan förekomma i kriminalserier som "CSI" och liknande. Det finns också charlataner som säljer vad de påstår är lögn-detektorer som kan laddas ner till mobiltelefonen. Med hjälp av den ska det gå att avgöra om någon är förälskad i en man undanhåller det eller ljuger om något annat⁹.

I Sverige utfördes tidigare rättsfonetiska analyser och talaridentifikationer av SKL (Statens kriminaltekniska laboratorium). Numera görs inte längre sådana analyser på SKL men här kan fortfarande tvättning/rensning av inspelningar utföras. Det kan t.ex. behövas om en inspelning innehåller brus eller andra störningar. En inspelning kan ibland bli betydligt tydligare efter en sådan rensning.

Eftersom den här typen av analyser främst används inom rättsliga sammanhang är det ofta myndigheter som beställer analysen (ex. migrationsverket, militären, säkerhetstjänsten, polis, åklagare, advokater eller försäkringsbolag¹⁰). Men även privata kunder eller "deckare" förekommer¹¹. Det kan t.ex. handla om att identifiera en person på en inspelning.

Det finns fortfarande personer som utför rättsfonetiska analyser i Sverige men de är få. Några av dem finns vid Göteborgs Universitet, där de annars bedriver forskning inom fonetik och rättsfonetik¹². Det finns även ett fåtal företag som utför rättsfonetiska analyser.

Den första utbildningen i forensisk fonetik i Sverige gavs våren 2004 vid Göteborgs universitet som en doktorandkurs för fonetiker eller lingvister. Nu ges liknande kurser även vid Stockholms universitet, både på fortsättningsnivå i lingvistik och som en mer populärvetenskaplig kurs¹³.

⁷ Gunnar Fant är professor emeritus vid Kungliga tekniska högskolan (KTH) i Stockholm. Han var tidigare professor i talöverföring vid KTH, 1966–1987

⁸ Fant 2004 s.30

⁹ Eriksson & Lacerda, 2007

¹⁰ Enligt handouts från föreläsning med Anders Eriksson "introduktion till rättsfonetik" 080213

¹¹ Enligt handouts från föreläsning med Anders Eriksson "introduktion till rättsfonetik" 080213

¹² Institutionen för lingvistik hemsida i Göteborg: <http://hum.gu.se/institutioner/lingvistik>, 080403

¹³ Institutionen för lingvistik hemsida i Stockholm, fristående kurser: http://www.ling.su.se/pub/jsp/polopoly.jsp?d=6931_080403, 15.45

2.3. Problem inom rättsfonetiken

Man måste inom den forensiska fonetiken handskas med problem som "the telephone effect", den effekt som påverkar talets tydlighet och identifikation av rösten genom telefonens begränsade återgivning. Det måste fonetikern ha i åtanke om en telefoninspelning jämförs med en annan sorts inspelning. Det är inte sällan som just kvalitén på inspelningen gör det svårt att utföra en noggrannare fonetisk analys.

Inom "voicelineups" finns också problem där det största är tidsaspekten. En konfrontation måste hållas så snart som möjligt eftersom minnet är vanskligt och därför påverkar resultatet. Minnet av en röst försvagas med tiden och därför får man en bättre identifikation ju tidigare man kan utföra "öronvittneskonfrontationen". Ett annat problem är att ordna en så bra "lineup" som möjligt. Man minns nämligen olika sorters röster olika bra och därför måste man ha lagom karaktäristiska röster att jämföra med den misstänktes röst.

I rätten får ett rättsfonetiskt utlåtande inte användas som bevis utan bara som indicium, d.v.s. kan inte fälla en person utan andra bevis eller indicier. Detta är visserligen inte något som rättsfonetikern tar ställning till men som denne kan behöva påpeka i sin rapport så att resultatet i den rättsfonetiska undersökningen inte används på fel sätt.

2.4. Tidigare forskning

Det finns en hel del tidigare forskningsresultat inom rättsfonetiken och en del av dessa är relevant att ta upp i anslutning till denna studie. Här redovisas några.

Redan 1937 gjorde McGehee studier där hon lät lyssnare identifiera olika typer av röster efter olika lång tid mellan första exponeringstillfället och identifieringen (1, 2, 3 dagar, 1, 2, 3, veckor, 1, 2, 3, mån)¹⁴. Hon fann flera saker. Identifikationen är mer korrekt ju tidigare man gör identifikationstestet efter exponering av rösten (korrekt identifikation ökar med kortare tidsintervall). Korrekt identifikation minskar ju fler röster man har att välja på vid identifikationen och korrekt identifikation minskar också om grundtonen är olika vid presentation och identifikation. Hon fann också att mans- respektive kvinnoröster identifierades ungefär lika bra. Några år senare testade McGehee identifikation av inspelade röster¹⁵ istället för att ha talaren i samma rum bakom ett draperi. Hon testade även här tidsaspekten (identifikation efter 2 dagar, 2 veckor, 1, 2 mån) och även vilken röst som stack ut mest i karaktär jämfört med de andra. Hon undersökte hur tilltalande rösterna tycktes vara och vilken ålder, längd och vikt de uppskattades ha. Hon hade dessutom olika lyssnargrupper med både taltränade personer, musiker och personer utan speciell träning. Hon fann att man med fördel kan använda sig av inspelat material, att det inte var någon generell överensstämmelse för vilken röst som var mest olik, förutom för lägst grundton (men ej högst). Ingen överensstämmelse fanns heller för vilken röst som var mest tilltalande. Ålder, längd och vikt verkade svårt att bedöma men en tendens fanns till att bedöma storlek efter tonhöjd (F_0) och taltempo. Den röst som valdes som mest tilltalande efter en lyssning var inte densamma som senare föredrogs. Papcun m.fl.¹⁶ fann i sina experiment en trend till ökad korrekthet med ökad säkerhet på den egna bedömningen. De fann också en trend till färre korrekta svar som en funktion av fördröjning (mellan

¹⁴ McGehee 1937

¹⁵ McGehee 1944

¹⁶ Papcun, Kreiman, Davis 1989

presentation och identifikation). I ytterligare en annan studie¹⁷ fann man inte någon signifikant skillnad mellan identifiering direkt och fördröjt men det kan förklaras med att deras fördröjning bara var 24 timmar. Såpass kort fördröjning påverkar alltså inte identifieringsresultaten enligt den studien.

Vad det gäller identifikation av röster påpekar Rose¹⁸ att vissa röster är lättare att identifiera än andra (Rose refererar här till Papcun m.fl. 1989 och Rose & Duncan 1995). Rose menar att man därför kan anta att vissa röster innehåller mer individuell information som lyssnare använder sig av då de identifierar än vad andra röster gör. Hollien m.fl.¹⁹ gjorde en studie där lyssnare fick identifiera förställda, stressade och normala röster och fann att de lyssnare som var bekanta med rösterna lyckades bättre än andra lyssnare, vilket senare har bekräftats i liknande experiment.

Vad det gäller könsbestämning av röster har man i tidigare studier²⁰ funnit att det går det bra att identifiera vilket kön en röst tillhör, utifrån endast korta yttranden. Lyssnare fick i denna studie könsklassifiera 2500 /ä/-vokaler. Utifrån uppmätta formantfrekvenser och dess amplitud (ljudstyrka) räknades grundtonsfrekvens (F_0 , för att illustrera larynx-storlek) och ansatsrörlängd (motsvarande kroppsstorlek) ut med hjälp av en formel från Lieberman & Blumstein 1993²¹. Då båda dessa variabler användes tillsammans var bestämningen näst intill helt korrekt över lag.

Könsbestämning utifrån minimala lingvistiska yttranden, som isolerade viskade vokaler²² (/i/ och /a/) och tonlösa frikativor²³ (alla olika varianter av tonlösa frikativor enligt IPA's alfabet²⁴), har också studerats. Dessa studier visar att man kan bedöma könstillhörighet även utan information från grundtonsfrekvensen. Könsbestämning utifrån viskade vokaler var nära perfekt. Alla /a/-vokaler var korrekt identifierade och bland /i/-vokalerna gjordes bara fyra misstag på 80st exempelljud. Utifrån tonlösa frikativor var könsidentifikationen bättre än chansning för alla frikativor (från 54 % som lägst för /f/) och upp till 91 % korrekt för /h/ och 73-93 % för /s/ och sje-ljud.

I tidiga studier med åldersbestämning²⁵ fick 10 lyssnare skilja yngre vuxnas röster från äldre. De lyckades förvånadsvärt bra vad gällde tre olika stimuli: förlängda vokaler, baklänggestal och vanligt tal. Medelvärdet korrekta identifikationer sträckte sig mellan 78 % och 99 %. Lyssnarna sade sig bedöma utifrån fraserings, talhastighet, intensitet, tveknings, röstkvalitet och röstbrott (voice breaks) s.k. tupper. Efterföljande experiment har generellt visat på liknande resultat som dessa tidiga undersökningar. I en sammanfattning av sju studier utförda mellan 1969 och 1998²⁶ har korrelationer mellan uppfattad och kronologisk ålder hamnat mellan 0,68 och 0,88.

Hur man lyckas uppskatta kroppsstorlek (främst längd och vikt) har testats i många studier²⁷. Generellt har man lyckats bra. Lyssnare verkar använda sig mest av

¹⁷ Saslove & Yarmey 1980

¹⁸ Rose, 2002 s.101

¹⁹ Hollien, Majewski, Doherty 1982

²⁰ Bachorowski och Owren, 1999

²¹ Lieberman och Blumsteins formel för uträkning av ansatsrörets längd ($VTL=Vocal\ Tract\ Length$)
 $VTL = (2k + 1)c/4F_k + 1$. Lieberman Blumstein 1993.

²² Schwartz 1968

²³ Ingemann 1968, Schwartz 1968

²⁴ IPA: International Phonetic Association's alphabet. Fonetisk förteckning över världens alla språkljud.

²⁵ Ptacek och Sander 1966

²⁶ Cerrato *et al* 2000

²⁷ Dommelen 1993, Dommelen 1995, Krauss, Freyberg & Morsella 2002

grundtonsfrekvens (F_0) som en indikator på kroppsstorlek. Vilket visserligen känns logiskt rent intuitivt, de allra flesta skulle nog förvänta sig att en röst med låg F_0 tillhör en stor person. Sådana intuitioner märks speciellt i de egenskaper vi tilldelar syntetiserat tal (datoriserat tal, skapat för att likna mänskligt tal). En av DECTalk's syntetiserade röster, kallad "Huge Harry", är ett exempel på där den låga grundtonsfrekvensen tolkas som rösten hos en stor människa. Till artikeln "Review of text-to-speech conversions for English" av H. Klatt²⁸, som är en sammanställning av många talsyntessystem, finns en tillhörande internetsida där man kan lyssna på hur alla talsynteserna låter²⁹. I verkligheten är det dock inte alltid så att stora människor har låg F_0 . Flera studier³⁰ har visat svaga korrelationer mellan talarens F_0 och längd och korrelationerna har inte varit signifikanta. Om nu lyssnare ändå kopplar samman grundton med längd kan man i så fall förvänta sig att de inte lyckas speciellt bra i att uppskatta längd utifrån röstsampel. Det har också visat sig i ett flertal studier att så är fallet³¹. Van Dommelen och Moxness har i experiment undersökt flera olika parametrar men inte funnit några korrelationer med talarens längd eller vikt, med enda undantaget för taltempo hos manliga talare. Manliga talare med långsamt taltempo tolkas som långa och/eller tunga. Deras test innehöll även en perceptionsdel där de fann en korrelation när manliga lyssnare bedömde manliga talare, mellan verklig och uppskattad längd och vikt. I en annan studie där man parade ihop bilder med röster³² fick man fram måttligt signifikanta korrelationer mellan uppskattad och verklig längd.

Vad det gäller åldersbedömning har det visat sig att det är normalt med en felmarginal på ca 10 år³³. Vad det gäller könsbestämningen har tidigare studier visat på god könsidentifikation, t.o.m. enbart utifrån tonlösa frikativor³⁴ därför kan den förväntas gå bra även i förestående experiment.

²⁸ Klatt 1987

²⁹ För ljudexempel gå till: <http://festvox.org/history/klatt.html> eller <http://www.icsi.berkeley.edu/eecs225d/klatt.html>

³⁰ Lass och Brow 1978, Künzel 1989

³¹ Dommelen 1993, Dommelen 1995

³² Krauss m.fl. 2002

³³ Schötz 2001

³⁴ Ingeman 1968

3. Metod

Metoden var att låta några personer hosta, skratta och ge bekräftningsljud på en inspelning. Denna spelades sedan upp för personernas arbetskollegor som skulle försöka identifiera personerna bakom ljuden. Ett andra lyssnartest utfördes också där andra lyssnare som inte kände dem som producerat exempeljuden fick bedöma kön, ålder och längd hos försökspersonerna.

3.1. Inspelning

För att undersöka om lyssnare kan identifiera sina kollegor på dessas skratt, hostningar, /m/ och /a/, har 10 försökspersoner, 5 män och 5 kvinnor, spelats in. Alla är personer som arbetar på institutionen för lingvistik vid Stockholms universitet. De blev informerade vid inspelningstillfället om att det var önskvärt att få med just skratt, hostning och bekräftningsljuden: /m/ och /a/ i inspelningen. De spelades in i det s.k. "tysta rummet", det ekofria rummet på institutionen för lingvistik vid Stockholms universitet. Inspelningen skedde delvis under en dialog med försöksledaren, som pratade med försökspersonerna via en intercom (intertelefon). Men de fick också till sin hjälp med sig en vitsbok, några roliga bilder och ett antal journalgrodor in i "tysta rummet" för att lättare locka fram äkta skratt. Inspelningen startades direkt, tidigare än försökspersonen visste om, vilket gjorde att dialogen mellan försöksledare och försöksperson också spelades in under vad försökspersonen förmodligen trodde var kalibrering inför inspelningen. Från denna dialog kunde sedan spontana bekräftningsljud plockas och dessutom några av skratten. Hostningarna bedömdes inte behöva inspireras på något vis för att vara äkta.

3.1.1. "Talmaterial"

Från varje försökspersons inspelning plockades sedan ett skratt, en hostning, ett /m/ och ett /a/. Kriterierna för de exempelljud som valdes ut var att de skulle vara så naturliga som möjligt och inte sticka ut från mängden. Därför valdes t.ex. hostningar som var motsvarande ungefär lika långa och innehöll ungefär lika mycket grundtonsinformation, vilket var ganska lite F_0 -information. Om en hostning dock bedömdes innehålla något mer grundtonsinformation än de andra så gjordes den istället lite kortare. Eftersom identifikationen blir lättare ju mer F_0 -information yttrandet innehåller och ju längre det är gjordes urvalet, eller anpassades materialet, så att de olika exempelljuden uppskattades vara lika lätta eller svåra att identifiera utifrån. Om de uppskattades som olika anpassades materialets längd så att de skulle bli likvärdiga. På liknande sätt bedömdes och valdes även skratten och bekräftningsljuden ut. Det togs inte någon större hänsyn till om de utvalda exempelljuden var speciellt representativa för just de personerna eller inte, det viktiga bedömdes vara att få fram naturliga exempelljud. Sammanlagt blev det en stimuluslista med fyra exempelljud från varje försöksperson med ett skratt, en hostning, ett /a/ och ett /m/. Eftersom det var tio försökspersoner blev det totalt fyrtio exempelljud. Dessa lades i slumpmässig ordning på en CD som sedan användes vid de båda testerna.

3.2. Tester

TVå tester utformades med samma lyssningsmaterial. Det ena där lyssnare från samma institution, d.v.s. kollegor, personer bekanta med försökspersonernas röster fick lyssna och försöka identifiera försökspersonerna (test 1). Det andra där lyssnare som inte känner försökspersonerna fick lyssna och försöka bedöma försökspersonernas kön, ålder och längd. Dessa lyssnare var elever från grundkursen i lingvistik (test 2).

3.2.1. Test 1 Identifiering

Identifikationstestet utfördes av personer som arbetar på institutionen för lingvistik, väl bekanta med försökspersonerna och deras röster. Det var från början tänkt att både de inspelade kvinnorna och männen skulle delta tillsammans med en lika stor grupp män och kvinnor från institutionen som inte blivit inspelade, alltså sammanlagt tjugo personer. Nu blev det inte så eftersom det var svårt att få speciellt dem som redan blivit inspelade för studien att ställa upp en andra gång. Istället blev det fem personer som även spelats in för testet (två kvinnor, tre män) och elva personer som inte blivit inspelade (fem kvinnor och sex män). Totalt var det alltså sexton personer som känner försökspersonerna som lyssnade på materialet och gjorde identifikationstestet. Den enda information de fick i förväg om vilka personer som förekom bakom exempelljuden var att alla var deras arbetskollegor men de visste inte vilka kollegor som blivit inspelade för studien. På arbetsplatsen fanns vid inspelningstillfället ett 40-tal kollegor att välja bland. De som själva blivit inspelade för studien fick informationen att de kanske förekom i materialet och kanske inte. Lyssnarna fick ett svarsformulär och en CD med exempelljuden så de kunde göra testet på sitt eget rum i sin egen takt och kunde lyssna på ljuden flera gånger om de ville.

3.2.2. Test 2 Kön, ålder, längdbedömning

Det andra testet innebar bedömning av ålder, längd och könstillhörighet utifrån samma talmaterial. Totalt 20 personer, 4 män och resten kvinnor, från grundkursen i lingvistik, bedömde alla exempelljud. De fick göra testet i grupper om fyra eller fem lyssnare åt gången som lyssnade samtidigt. Alla fick var sitt par hörlurar som de kunde reglera volymen på. De kunde också påverka hur snabbt exempelljuden presenterades (så att alla hann med att skriva) och de fick höra exempelljuden fler än en gång om de ville. Alla fick ett svarsformulär där de fyllde i uppfattad ålder, längd och kön på varje exempelljud.

4. Resultat

Generellt verkar det för test 1 som om det var ganska svårt att identifiera försökspersonerna (fp) utifrån exempelljuden, åtminstone för vissa talare och för vissa exempelljud. För andra försökspersoner och exempelljud var däremot korrekt identifiering desto högre. För test 2 kan man generellt säga att männen verkar vara lättare att könsbestämma utifrån exempelljuden. Dessutom att försökspersonernas ålder bedömdes med en felmarginal på strax under 13 år i genomsnitt, vilket är bra och som förväntat utifrån tidigare studier. Över lag verkar det också som om försökspersonerna bedömdes vara yngre än de egentligen var i nästan alla fall. Vad det gäller längden bedömdes den med en felmarginal på bara 2 cm i genomsnitt. Här ser det också ut som om kvinnorna har bedömt längden lite närmare försökspersonernas riktiga längd än vad männen gjort.

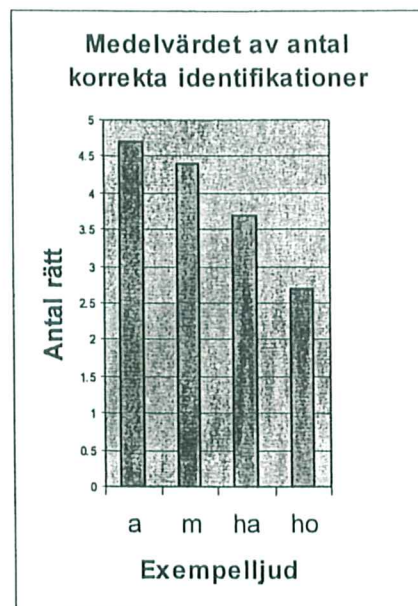
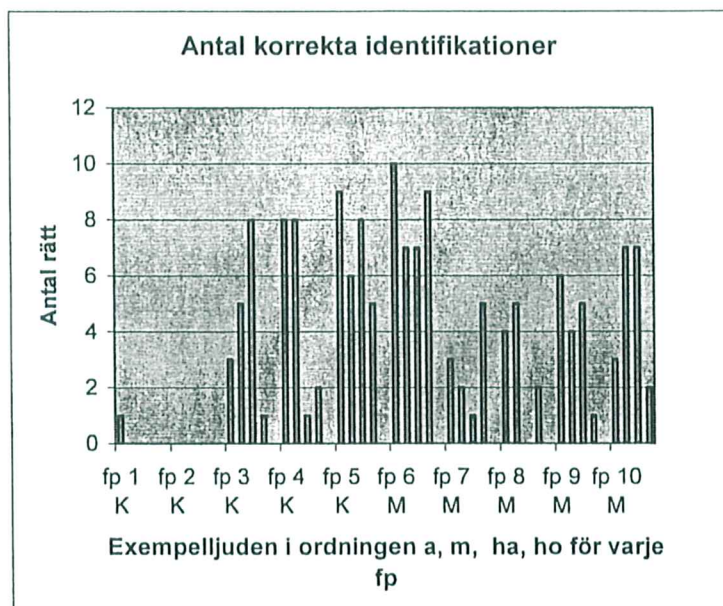
4.1. Resultat test 1

I det första testet där lyssnare som känner försökspersonerna fick försöka identifiera dem utifrån exempelljuden har resultaten delats upp och presenteras för två lyssnargrupper var för sig. Den ena gruppen utgörs av resultaten från de lyssnare som inte själva blivit inspelade för studiens syfte. Den andra visar resultaten från de lyssnare som även blivit inspelade.

I de fall lyssnarna (båda grupperna) absolut inte kunde komma på vem de hört ombads de att istället ange ålder, längd och kön på dessa exempelljud. Dessa bedömningar har dock inte analyserats närmare, utan syns endast i figurerna som icke korrekta bedömningar.

Resultaten presenteras i stapeldiagram där identifieringsresultaten i antal korrekta identifikationer visas för varje försökspersons alla exempelljud. De kvinnliga försökspersonerna visas till vänster och männen till höger. Exempelljuden visas i ordningen: /a/, /m/, skratt, hostning.

De båda grupperna (inspelade och inte inspelade) presterade liknande resultat, med ungefär lika andelar korrekta identifikationer för fp4, fp5 och fp6 men de fördelar sig olika över de olika exempelljuden mellan grupperna. De sämsta identifieringsresultaten har båda grupperna för fp1 och fp2. För lyssnargruppen som inte spelats in visas antal korrekta identifikationer och medelvärdet för antalet korrekta identifikationer för varje exempelljud i figur 1a och 1b nedan. Resultaten för korrekta identifikationer och medelvärden för lyssnargruppen som blivit inspelade visas i figur 2a och 2b på nästa sida.



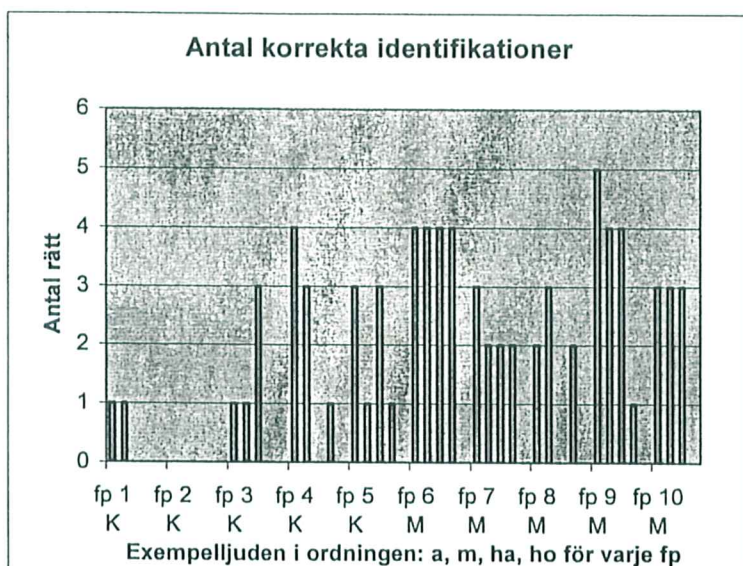
Figur 1a Antal korrekta identifikationer av försökspersonerna gjorda av lyssnare som känner dem men som inte är inspelade själva (11 st).

Figur 1b Medelvärdet av de korrekta identifikationerna för samma data som till vänster.

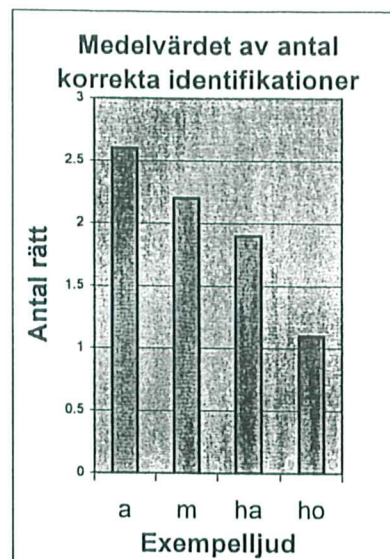
De lyssnare som själva blivit inspelade och som utförde identifikationstestet var endast fem stycken. Tre av de fem inspelade männen och två av de fem inspelade kvinnorna, (se figur 2a och 2b på nästa sida). En av de inspelade kvinnorna var dessutom försöksledaren själv, som visste vilka försökspersonerna var och vilka exempelljud som använts och därför inte kunde utföra identifikationstestet.

Vad det gäller identifikation av den egna "rösten" bland dessa så varierade resultatet kraftigt. För de tre männen hade en av dem prickat in sin egen röst i samtliga fall (fp7). En annan av männen lyckades identifiera alla förekomster av den egna rösten utom den som var hans skratt (fp8), han angav inte heller någon annan person som ägare till den rösten som skrattade. Den tredje mannen misslyckades med att känna igen sin egen röst vid samtliga förekomster (fp10). Två av dem, hans /m/ och /a/, bedömde han tillhöra en och samma person som har en liknande heshet i sin röst som hans egen men denne person fanns inte med i det inspelade materialet för denna studie. Sitt skratt bedömde samme man som en annan persons och hostningen namngavs inte.

Kvinnorna som både spelats in och som gjort identifikationstestet var bara två. Den ena av dem (fp4) identifierade sin egen röst korrekt i samtliga fall förutom skrattet som bedömdes tillhöra en annan. Men här kan tilläggas att de båda faktiskt skrattar mycket lika. Den andra kvinnan (fp1) identifierade sitt /m/ korrekt men de andra exempelljuden bedömde hon tillhöra andra. Hon påpekade dessutom att det var svårt att hålla isär tre röster: sin egen och två andras. Det visade sig också i resultaten att det också var de två andra rösterna hon nämnt som hon förväxlat sin egen röst med.



Figur 2a Antal korrekta identifikationer av försökspersonerna gjorda av de lyssnare som också blivit inspelade (5 st).



Figur 2b Medelvärde av de korrekta identifikationerna för samma data som till vänster.

4.1.1 Ett uppföljningstest - Kontroll av hur väl lyssnarna tycker sig känna försökspersonerna

Det verkar rimligt att anta att man känner igen rösten bättre på de personer som man faktiskt också känner bättre (jämför med liknelsen med snabb identifiering i ett telefonsamtal ovan). Det skulle kunna innebära att man inte känner igen människor på rösten så lätt, även om man delar arbetsplats med dem och pratar med dem då och då eller kanske t.o.m. ofta. Därför gjordes en liten kompletterande undersökning av hur väl lyssnarna tycker sig känna de personer som producerat exempelljuden i testet. Det skulle kunna visa om det är så att man identifierar dem man känner väl, även om det man får höra bara är en ynka hostning eller ett kort skratt.

Kontrolltestet utfördes så att alla som deltagit i identifikationstestet fick fylla i ett nytt formulär där de kryssade för hur väl de själva tyckte att de känner de inspelade personerna (här avslöjades alltså vilka försökspersonerna var). De fick fem olika grader på en skala att välja mellan från "inte alls" till "jättebra" däremellan fanns också "inte så väl", "ganska väl" och "bra", det framgick också av instruktionen att det skulle föreställa en skala och att "ganska väl" på så vis var svagare än "bra". Denna skala överfördes sedan till siffror från ett till fem där "inte alls" gav 1 poäng och "jättebra" 5 för att kunna redovisas och jämföras med de korrekta identifikationerna på ett överskådligt sätt.

På nästa sida redovisas resultatet av identifikationstestet för alla lyssnare tillsammans. Dessutom resultatet av det kompletterande kontrolltestet av hur väl lyssnarna tycker sig känna talarna. Höga staplar i identifikationstestet innebär korrekt identifikation gjord av många lyssnare. På samma vis innebär höga staplar i kontrolltestet eller höga "kändispoäng" att många lyssnare tycker sig känna försökspersonen bra. Försökspersonernas "kändispoäng" redovisas i samma ordning som i identifikationstestet.

Dessa stapeldiagram liknar inte varandra speciellt mycket eftersom de försökspersoner som fått god identifiering inte samtidigt fått höga "kändispoäng". Hur

4.2. Resultat test 2

I det andra testet där lyssnare som inte känner försökspersonerna skulle bedöma deras kön, längd och ålder har resultaten delats upp och redovisas i de tre kategorierna könsbedömning, åldersbedömning och längdbedömning var för sig.

4.2.1. Resultat könsbedömning

Generellt verkar det vara enklare att könsbestämman männens exempelljud korrekt än kvinnornas, vilket framgår av de båda tabellerna tabell 1 och tabell 2 på nästa sida. Eftersom testet utfördes av 20 lyssnare så innebär siffran 20 i tabellerna 100 % korrekt bedömning. Där det förekommer förvirring i könsbestämningen är det endast för en av männen (fp9) och bara för hans hostning. För kvinnorna förekommer olika mycket förvirring i könsbestämningen beroende på vilket exempelljud det är – svårast verkar hostning vara. Den innehöll heller inte så mycket F_0 -information vilket förmodligen är förklaringen till detta. Tidigare studier³⁵ har visat att F_0 är den starkaste ledtråden man använder vid könsidentifiering så det vore en rimlig förklaring. Även skratt innebar felaktigheter i könsbestämningen av två försökspersoner där den ena (fp 5) bedömdes fel av en lyssnare och den andra (fp 3) av fyra lyssnare. Just detta skratt innehöll även det ganska lite F_0 -information vilket förstås kan vara förklaringen till detta. Vidare blev det också lite hopblandning av /a/ hos två försökspersoner. I dessa två fall hade dock den ena kvinnan en lite lägre tonhöjd än vad som är normalt för henne och för den andra var det bara en lyssnare som tog fel. Kvinnornas /m/ blev dock aldrig felaktigt tolkade som män.

Kvinnor	Fp 1	Fp 2	Fp 3	Fp 4	Fp 5	Medel
a	14	20	20	20	19	18.6
m	20	20	20	20	20	20.0
ha	20	20	16	20	19	19.0
host	20	18	18	16	20	18.4

Tabell 1 Antal korrekta bedömningar av könstillhörighet för kvinnornas exempelljud och medelvärdet av dessa, bedömt av 20 st lyssnare som ej känner fp.

Män	Fp 6	Fp 7	Fp 8	Fp 9	Fp 10	Medel
a	20	20	20	20	20	20.0
m	20	20	20	20	20	20.0
ha	20	20	20	20	20	20.0
host	20	20	20	16	20	19.0

Tabell 2 Antal korrekta bedömningar av könstillhörighet för männens exempelljud och medelvärdet av dessa, bedömt av 20 st lyssnare som ej känner fp.

³⁵ Bachorowski och Owren 1999, Hollien 1987

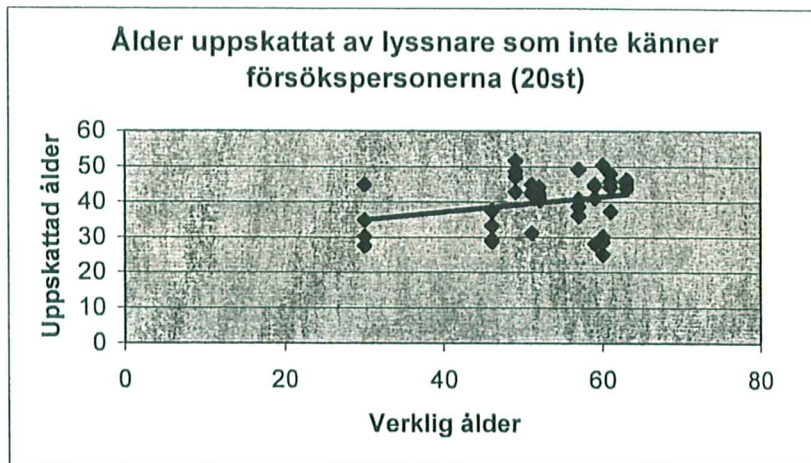
Nedan visas alla lyssnarnas sammanslagna bedömningar i ett stapeldiagram. Eftersom 20 lyssnare deltog så innebär 20 antal rätt alltså 100 % korrekt bedömning även här. Som synes har lyssnarna över lag könsbestämt försökspersonerna riktigt bra.



Figur 5 Antal korrekta könsbedömningar för varje försöksperson och varje exempelljud bedömda av alla 20 lyssnare som inte känner försökspersonerna. Kvinnliga försökspersoner syns till vänster och manliga till höger.

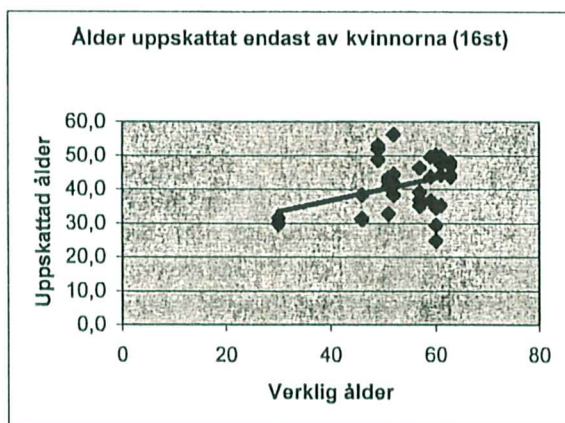
4.2.2. Resultat åldersbedömning

Bedömningen av försökspersonernas ålder visade att lyssnarna bedömde deras ålder ganska bra, eller ungefär som man kan vänta sig utifrån tidigare studier. Lyssnarna prickade in den rätta åldern med en felmarginal på strax under 13 år (12.86). Vissa var förstås betydligt närmare än så men samtidigt var det andra lyssnare som inte lyckades så bra. Över lag bedömer dessa lyssnare försökspersonerna som yngre än de egentligen är i nästan alla förekomster. Den yngsta talaren bedöms dock vid ett tillfälle som äldre men annars som väldigt nära den verkliga åldern. Medelvärden av åldersbedömningen för alla stimuli visas i figur 6. Korrelationen mellan verklig ålder och uppskattad ålder är starkt signifikant ($P < 0.001$) men den förklarade variansen är inte så stor, vilket tyder på att denna faktor visserligen är signifikant men att det kan finnas andra faktorer som spelar in.

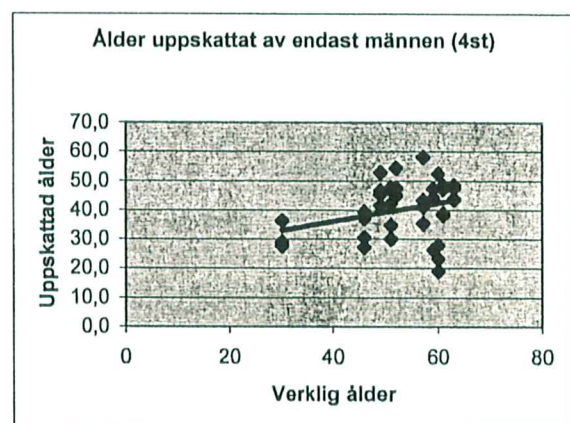


Figur 6 Korrelationen mellan verklig ålder och medelvärdet av uppskattad ålder bedömd av alla 20 lyssnare som inte känner försökspersonerna. (Förklarad varians 16.4%, $R^2=0.1636$)

Det verkar som om kvinnorna bedömer åldern på försökspersonerna som producerat de olika exempelljuden lite närmare den verkliga åldern än vad männen gör men denna skillnad ser inte ut att vara speciellt stor, se figur 7 och 8 nedan.



Figur 7 Korrelationen mellan verklig ålder och medelvärdet av uppskattad ålder bedömd av endast de 16 kvinnorna. (Förklarad varians 10%)

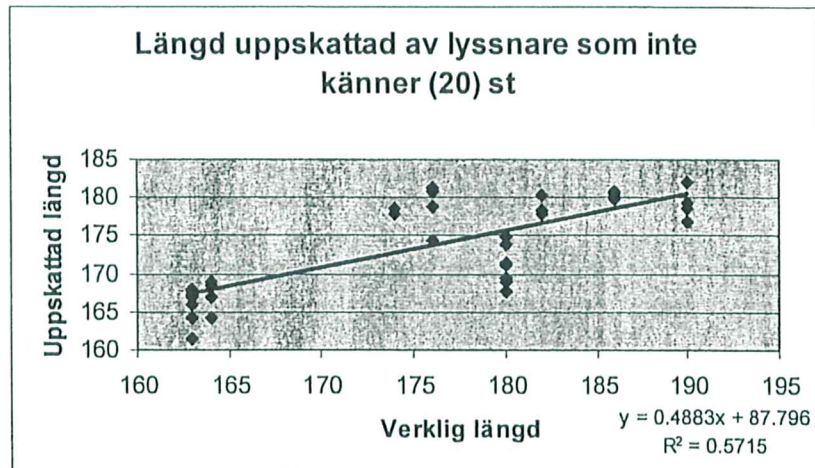


Figur 8 Korrelationen mellan verklig ålder och medelvärdet av uppskattad ålder bedömd av endast de 4 männen. (Förklarad varians 17%)

4.2.3. Resultat längdbedömning

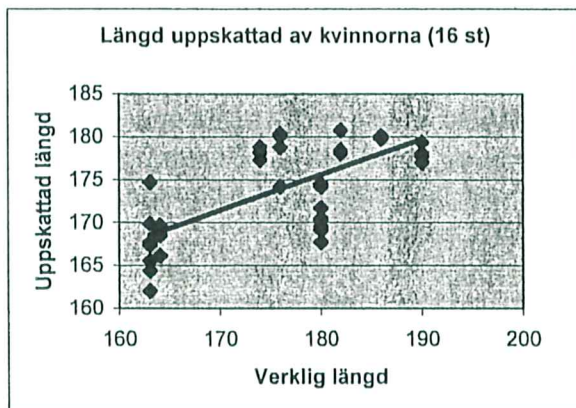
Av de två variablerna längd och ålder verkar det vara längd som är lättast att bedöma. Längden bedömdes med en felmarginal på bara 2 cm i snitt för lyssnarnas samlade bedömning, se figur nedan. Den förklarade variansen (se figur 9) är betydligt högre för längduppskattningen än för åldersuppskattningen. Det finns dock en liten fallgrop här som jag återkommer till.

Eftersom kvinnor i genomsnitt är kortare än män kan det inte uteslutas att talarens kön spelat en roll vid uppskattning av längden. Dessutom bedömer man som lyssnare generellt mer korrekt när den verkliga längden eller åldern hos försökspersonen är närmre lyssnarens egen längd eller ålder. Därför kan bedömningen ha påverkats av lyssnarnas egen ålder och längd även i denna undersökning.

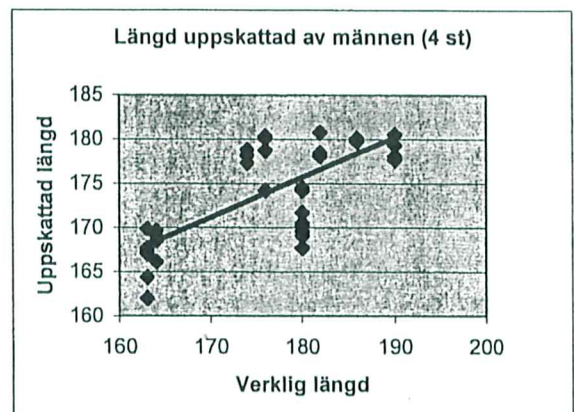


Figur 9 Korrelationen mellan verklig längd och medelvärdet av uppskattad längd bedömd av alla 20 lyssnare som inte känner försökspersonerna.

I det här fallet finns det inga större skillnader mellan de kvinnliga och manliga bedömarens uppskattningar. De kvinnliga försökspersonerna har bedömt längden med något mindre felmarginal men skillnaden är inte signifikant.



Figur 10 Korrelationen mellan verklig längd och medelvärdet av uppskattad längd bedömd av endast de 16 kvinnorna. (Förklarad varians 49%)



Figur 11 Korrelationen mellan verklig längd och medelvärdet av uppskattad längd bedömd av endast de 4 männen. (Förklarad varians 56%)

Som sagts ovan kan man misstänka att talarens uppfattade kön kan ha påverkat uppskattningen av talarens längd. En närmare analys visar att detta var en riktig gissning. Uppskattad längd är signifikant korrelerad med både talarens kön och längd. I själva verket har talarens kön ett större förklaringsvärde än den verkliga längden. En intressant skillnad är att lyssnarna varit bättre på att uppskatta de kvinnliga talarnas längd än mäns. Korrelationen mellan verklig längd och uppskattad längd är starkt signifikant för de kvinnliga talarna ($p < 0.001$) men endast marginellt signifikant för de manliga ($p = 0.13$). Förklaringsvärdet är dock lågt för bägge grupperna.

5. Diskussion & Slutsatser

En glad överraskning var att det inte var så svårt att få försökspersonerna att skratta vid inspelningen som befarat. Flera spontana skratt kom av de roliga journalgrodorna och andra skrattade åt den konstiga situationen att skratta på befallning. Endast en person medgav att hans skratt kanske var lite konstlat, trots det blev detta skratt korrekt identifierat som hans av sju av elva lyssnare som känner honom. Det som mot förmodan var svårast att locka fram vid inspelningen var spontana bekräftningsljud men här var det en stor hjälp att försöksledaren kunde prata med försökspersonen under inspelningen och därigenom få fram naturliga feedbackljud.

En kommentar från en av lyssnarna som kan vara värd att nämna är hans motivering när han fick reda på att han identifierat en av försökspersonernas hostning som tillhörande en person som inte fanns med i materialet. Lyssnaren förklarade att han endast identifierat hostningen som västerbottnisk och råkade sedan välja fel västerbottning på institutionen.

En annan punkt att ta upp är att en av männen flera gånger misstagits för att vara kvinna. Det var visserligen endast på hans hostning som innehöll förhållandevis lite grundtonsinformation. Detta kan bero på att han eventuellt har en lite högre grundton än de andra manliga försökspersonerna (inga sådana mätningar har gjorts men rent perceptuellt kan det uppfattas så). Hans röstkvalité är ganska hes och innehåller mycket s.k. jitter och shimmer (kan beskrivas som skrovligheter och ojämnheter i röstklngen). Denne försöksperson är rökare och kanske kan denna speciella röstklng vara en följd av långvarig rökning och att rökningen i så fall bidragit till de felaktiga könsbestämningar som gjorts på hans hostning. Å andra sidan brukar grundtonen påverkas så att den blir lägre av rökning, i så fall den motsatta mot vad som uppfattas här, men att söka svar på hur rökning påverkar rösten är inte syftet med denna studie. En annan förklaring skulle kunna vara det faktum att mäns röster får högre grundton med hög ålder³⁶. Denne man (rökaren) är dock inte nämnvärt äldre än de andra men man åldras olika fort och även röstens åldrande sker individuellt.

Försökspersonerna bedömdes över lag vara yngre än de egentligen var. Det kan bero på att lyssnarna i de flesta fall var yngre än alla försökspersonerna. Det har tidigare visats att det är lättare att korrekt bedöma personers ålder som är mer jämgamla med en själv. Det finns ett samband mellan talarens och lyssnarens ålder³⁷. Detta kan ha påverkat lyssnarna, samtidigt som man också får beakta att man tenderar att bedöma äldre som yngre än de är och tvärtom.

Vidare skulle det vara önskvärt att titta närmare på resultaten från identifikationstestet för att kunna säga något om bättre eller sämre identifikation av försökspersonerna gjorda av de lyssnare som kände dem för att i och med detta kanske kunna säga något mer om de personer som de valt istället för den som producerat exempelljudet då de inte identifierat rätt. T.ex. hade ju en lyssnare gissat på en västerbottning som visserligen är en av kollegorna men inte den västerbottning som blivit inspelad för studien. På liknande sätt kanske det finns flera bland försökspersonerna som kan förväxlas med andra kollegor, t.ex. p.g.a. röstlikhet. Rent intuitivt verkar det som om det skulle kunna vara så, flera av exempelljuden låter faktiskt som om de skulle kunna vara så, flera av exempelljuden låter faktiskt som om de skulle kunna komma från en annan person på institutionen och flera lyssnare har också gissat på andra personer som inte spelats in. I en eventuell

³⁶ Hartman 1979.

³⁷ Hartman 1979, Hollien & Tolhurst 1987, Huntley m.fl. 1980, Kreiman & Papcun 1985.

fortsättning av denna studie skulle det vara intressant att se om lyssnarna följer något sådant mönster.

Vad det gäller resultaten av hur väl lyssnarna tycker sig känna sina kollegor, kan det vara missvisande att "kändispoängen" redovisas sammanslagna. Om man istället tittade på person för person, skulle det kunna se annorlunda ut. En enskild person kanske känner en av försökspersonerna mycket väl och därför identifierar denne väl, samtidigt som andra lyssnare kanske inte tycker sig känna denne så väl och därför inte heller identifierar denne så väl. Då allas "kändispoäng" visas sammanslagna här, kan den personens resultat som identifierat någon person bra försvinna i mängden. Detta vore också spännande att titta närmare på i framtiden.

Vad det gäller kommentaren i inledningen om ett hej över telefonen, så innehåller ett "hej" betydligt mer formantinformation och därmed mer information om talaren och dennes personliga röstegenskaper än vad en kort hostning eller skratt gör och förmodligen också mer information än ett så kort /m/ eller /a/ som lyssnarna fått identifiera utifrån i denna studie. Därför är det betydligt enklare att känna igen en person över telefonen på det viset. Dessutom har man ett någorlunda begränsat antal personer att välja mellan, då man vet ungefär vilka personer man kan förvänta sig telefonsamtal ifrån samtidigt som man förmodligen är mer bekant med dessas röster än en del av sina kollegors röster. Den uppgift lyssnarna fått i identifieringstestet är med andra ord inte helt enkel och med det i åtanke kan man ändå tycka att de lyckats ganska bra med sina identifieringar, även om resultaten varierar. Jämför man med vad man funnit i tidigare liknande studier stämmer resultaten i denna studie väl överens med dem.

6. Sammanfattning

Kan man identifiera personer som man känner på bara en F_0 -fattig hostning, ett /m/ eller /a/ eller på enbart deras skratt?

För att undersöka detta har 10 personer som arbetar på samma arbetsplats (5 kvinnor, 5 män) spelats in. Deras hostningar, /m/, /a/ och skratt har sedan presenterats för en lyssnargrupp på 16 arbetskolligor som borde vara väl bekanta med dessa personer och deras röster (test 1). Lyssnarna är dels 5 st av de personer som blivit inspelade och dels 11 st andra från arbetsplatsen. Detta möjliggör också att se om de lyssnare som också spelats in kan identifiera sina egna hostningar, skratt, /m/ och /a/. Lyssnarnas uppgift var att försöka identifiera personerna utifrån exempelljuden (försökspersonernas hostningar, skratt, m och a). Den enda förhandsinformation de fick var att personerna som producerat dem var personer från deras arbetsplats. Ytterligare en lyssnargrupp användes som inte var bekant med de inspelade personerna (test 2). Denna lyssnargrupp fick istället bedöma vilket kön, ålder och längd de trodde personerna som producerat exempelljuden hade.

Resultaten för test 1 visar att det verkar som om det var ganska svårt att identifiera försökspersonerna utifrån exempelljuden, åtminstone för vissa talare och för vissa exempelljud men det varierar också mellan de båda lyssnargrupperna. För test 2 kan man generellt säga att männen verkar vara lättare att könsbestämma utifrån exempelljuden än kvinnorna. Försökspersonernas ålder bedömdes med en felmarginal på strax under 13 år i genomsnitt, vilket är ett förväntat resultat om man ser till tidigare studier. Dessa lyssnare bedömde försökspersonerna över lag som yngre än de egentligen var i nästan alla fall med undantag för en yngre talare som bedömdes äldre än hon egentligen var. Detta att yngre bedöms vara äldre än de är och äldre bedöms yngre än de är, har observerats i tidigare studier. Vad det gäller längden bedömdes den med en felmarginal på bara 2 cm i genomsnitt, det ser också ut som om kvinnorna har bedömt även längden lite närmare försökspersonernas riktiga längd än vad männen gjort.

7. Litteratur och källor

- Bachorowski, J.-A. & M. J. Owren (1999). "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech." *Journal of the Acoustical Society of America* **106**: 1054–1062.
- Cerrato, L., Falcone, M. & Paoloni, A. (2000). "Subjective age estimation of telephonic voices." *Speech Communication* **31**: 107–112.
- Dommelen, W. A. (1993). "Speaker height and weight identification: a re-evaluation of some old data." *Journal of Phonetics* **21**: 337–341.
- Dommelen, W. A. (1995). "Speaker and listener sex for speaker height and weight identification." *Proceedings XIIIth ICPPhS*. Stockholm, KTH, Vol. 2: 738–741.
- Dommelen, W. A. & Moxness, B. H. (1995). "Acoustic parameters in speaker height and weight identification: sex-specific behavior." *Language and Speech* **38**: 267–287
- Eriksson, A. (2005). *Tutorial on Forensic Speech Science* Interspeech, Lisboa
- Eriksson, A & Lacerda, F. (2007). "Charlatanry in forensic speech science: A problem to be taken seriously". *International Journal of Speech Language and the Law* **14**: 2, in press.
- Fant, G. (2004). "Phonetics and phonology in the last 50 years." *Speech Research in a Historical Perspective* Sound to Sense: 11-13 juni, 20-41 MIT
- Handouts från föreläsning med Anders Eriksson "introduktion till rättsfonetik" 080213, 18-20.
- Hartman, D. (1979). "The perceptual identity and characteristics of aging in normal male adult speakers." *Journal of Communication Disorders* **12**: 53–61
- Hollien H., Majewski, W. & Doherty, E. T. (1982). "Perceptual identification of voices under normal, stressed disguise speaking conditions." *Journal of Phonetics* **10**: 139–148
- Hollien, H. (1987). "Old Voices: What do we really know about them?" *Journal of Voice* **1**: 2–17
- Hollien, H. & Tolhurst, G. (1987). "The aging voice." I Weinberg, B, (red.), Transcripts of the seventh symposium on care of the professional voice, part II: life span changes in the human voice. *The voice Foundation*. 67-73
- Huntley, R., Hollien, H. & Shipp, T. (1980). "Influences of listener characteristics on perceived age estimations." *Journal of Voice* **1**: 49-52
- Ingemann, F. (1968). "Identification of the speaker's sex from voiceless fricatives." *Journal of the Acoustical Society of America* **44**: 1142–1144.
- Institutionen för lingvistik hemsida i Göteborg:
<http://hum.gu.se/institutioner/lingvistik>. (080403) 14.30
- Institutionen för lingvistik hemsida i Stockholm, fristående kurser:
<http://www.ling.su.se/pub/jsp/polopoly.jsp?d=6931>. (080403) 15.45
- Klatt H. (1987). "Review of text-to-speech conversions for English." *Journal of the Acoustical Society of America* **82**: 737–793.
- Klatt H. (1987). Ljudexempel till artikeln "Review of text-to-speech conversions..."
http://www.acoustics.hut.fi/publications/files/theses/lemmetty_mst/appa.html
(080304) 11.30.

- Krauss, R. M., Freyberg, R. & Morsella, E. (2002). "Inferring speakers' physical attributes from their voices." *Journal of Experimental Social Psychology* **38**: 618–625.
- Kreiman, J. & Papcun, G. (1985). "Voice discrimination by two listener populations." Paper presented at *ASA Acoustical Society of America*, Austin TX.
- Künzel, H. (1989) "How well does average fundamental frequency correlate with speaker height and weight?" *Phonetica* **46**: 117–125.
- Lass, N. J. & Brown W. S. (1978). "Correlational study of speakers' heights, weights, body surface areas, and speaking fundamental frequency." *Journal of the Acoustical Society of America* **63**:1218–1220.
- Lieberman P. & Blumstein, S. E. (1993). *Speech Physiology, Speech Perception and Acoustic Phonetics*. University Press, Cambridge, MA
- McGehee, F. (1944). "An experimental study of voice recognition." *Journal of General Psychology* **31**: 53–65.
- Papcun, G., J. Kreiman, *et al.* (1989). "Long-term memory for unfamiliar voices." *Journal of the Acoustical Society of America* **85**: 913–925.
- Ptacek, P. H. & Sander, E. K. (1966). "Age recognition from voice." *Journal of Speech Hearing and Research* **9**:273–277.
- Quintilianus, M., F. *Den fulländade talaren* (2002). Wahlström & Widstrand Stockholm
- Rose, P. (2002). *Forensic Speaker Identification*. London, Forensic Science SERIES Taylor & Francis, Samtidigt utgiven i USA och Kanada av: Taylor & Francis Inc New York
- Saslove, H. & A. D. Yarmey (1980). "Long-term auditory memory: speaker identification." *Journal of Applied Psychology* **65**: 111–116.
- Schwartz, M. F. (1968). "Identification of speaker sex from isolated, voiceless fricatives." *Journal of the Acoustical Society of America* **43**: 1178–1179.
- Schwartz, M. F. & Rine, H. E. (1968). "Identification of speaker sex from isolated, whispered vowels." *Journal of the Acoustical Society of America* **44**: 1736–1737.
- Schötz, S. (2001). "A perceptual study of speaker age." *Working Papers* **49**: 136–139. Lund: Lund University, Department of Linguistics.

