

**Röstigenkänning och Ansiktsigenkänning  
Kan yrkeserfarenhet påverka prestation och  
konfidensbedömningar?**

C-uppsats i Psykologi

Carina Borgström

Vårterminen 2007

Handledare: Mats Dahl

## Abstrakt

Syftet med detta studium var att undersöka om det fanns skillnader i förmågan till identifikation och konfidensbedömningar för röstigenkänning och ansiktsigenkänning relaterat till yrkeserfarenhet och erfarenhet av identifikationsuppgifter. Undersökning genomfördes som ett experiment med två grupper. Experimentgruppen (N=27) bestod av yrkesverksamma poliser, väktare, larmoperatörer och kriminalvårdare. Kontrollgruppen (N=27) utgjordes av försökspersoner som saknade yrkeserfarenhet av identifikationsuppgifter. Resultatet visade att det fanns en signifikant interaktionseffekt mellan de två identifikationsuppgifterna (röstigenkänning vs. ansiktsigenkänning) och de två grupperna. Experimentgruppen visade sig ha signifikant bättre identifikationsförmåga för röster. Experimentgruppen var också signifikant mer underkonfidenta för röstigenkänning jämfört med kontrollgruppen.

Nyckelord: Röstigenkänning, ansiktsigenkänning, öronvittnen, ögonvittnen, konfidensbedömningar

## **Röstigenkänning och Ansiktsigenkänning Kan yrkeserfarenhet påverka prestation och konfidensbedömningar?**

Bruno Richard Hauptmann avrättades den 3 april 1936 i Trenton State Prison, USA, för kidnappning och mord på den världsberömde piloten Charles Lindberghs 19-månader gamle son. Sonen kidnappades i mars 1932 och återfanns i ett dike, mördad i maj samma år. Lindbergh identifierade vid rättegången Hauptmanns röst som kidnapparens. Lindbergh hade hört rösten på sin sons kidnappare nästan tre år tidigare. I hopp om att sonen fortfarande var vid liv befann sig då Lindbergh, tillsammans med en följeslagare, på en kyrkogård för att överlämna den lösensumma på \$ 50 000 som kidnapparen hade begärt. I mörkret hörde Lindbergh en man med tysk brytning uttala orden: *"Hey doctor! Here doctor, over here"*. Under rättegången vittnade Lindbergh under ed och försäkrade att Hauptmann var den man han hört på kyrkogården. Förutom denna öronvittnesutsaga fanns ytterst få bevis mot Hauptmann. Han nekade själv till alla anklagelser men dömdes till döden (Solan & Tiersma, 2003).

Till skillnad från det omfattande intresse som forskningen har visat för psykologin i ögonvittnesidentifikationer är den empiriska forskningen på öronvittnesidentifikationer mycket mer begränsad (Yarmey, 1995). Den allmänna uppfattningen att vi människor, med säkerhet kan identifiera andra genom att höra deras röst verkar ha funnits sen urminnes tider. Modern vetenskaplig forskning i och kring röstidentifikation visar att denna tro till viss del är sann, men att sådana identifikationer inte är helt felfria (Read & Craik, 1995). I samband med vissa brott är det enda bevis som finns att tillgå för en domstol en målsägande, eller ett vittne som har hört en röst. Om alla de brott som begås i mörker läggs till dem som begås av maskerade gärningsmän, inser man det betydelsefulla i att studera öronvittnens tillförlitlighet. Tillfogas dessutom alla hot och obscena samtal som sker över telefon så framstår studier av öronvittne som än mer motiverat (Clifford, 1980; Yarmey, 1995).

Frågan hur väl ett vittne som Lindbergh kan identifiera en röst väcks återkommande i rättsliga sammanhang. Vid tiden för Hauptmanns rättegång fanns ingen kompetens för att bedöma tillförlitligheten i Lindberghs vittnesmål. Faktum är att det vetenskapliga intresset för hur väl vi kan identifiera andra, genom att höra deras röst, initierades av Hauptmann rättegången. Ett år efter avrättningen på Hauptmann inledde Frances McGehee, professor i psykologi vid University of Illinois, USA, en serie experiment för att undersöka hur väl en grupp studenter kunde identifiera en okänd röst de hört. Hon lät studenterna lyssna till en person som satt bakom en skärm och läste en text omfattande 56 ord. Studenternas förmåga till igenkänning testades sedan efter olika tidsintervall. Uppgiften bestod av att identifiera den ursprungliga rösten i en grupp av fem röster. Efter två dygn blev resultatet 83 % korrekta identifikationer, efter två veckor var resultatet 68 % korrekta identifikationer, efter tre månader 35 %, och efter fem månader hade de korrekta identifikationerna sjunkit till 13 % - vilket är sämre än ren chanstagning (Clifford, 1980). Stora delar av McGehee's resultat har bekräftats av senare forskning. Vår förmåga att identifiera främmande röster kvarstår ett tag, men försämras avsevärt över tid (Read & Craik, 1995). Forskningen visar också att alla

individer inte presterar likvärdiga resultat. Individer varierar dramatiskt i sin förmåga att identifiera röster, något som det inte finns någon känd förklaring till (Solan & Tiersma, 2003).

### *Fundamental Röstperception*

Vi människor föds med förmågan att skilja på mänskliga röster och andra ljud. Ljudvågorna i talat språk är mycket komplexa. Ljudet som alstras av den mänskliga rösten uppvisar ett harmoniskt spektrum med ett i princip oändligt antal deltoner (McAllister, 1998). Den social-kognitiva forskningen om öronvittne har inte riktat in sig på örats fysik eller den neuropsykologiska aspekten av ljud. Istället har denna forskning rört de faktorer som bidrar till att vi i allmänhet kan urskilja och förnimma tonhöjd, styrka och varaktighet i en röst. Dessa perceptuella dimensioner av röstigenkänning har ett samband med fysiska dimensioner av akustiska signaler. Frekvens (Hertz) har ett samband med tonhöjd. Intensitetsnivå (Decibel) har ett samband med röststyrka. Varaktighet (Millisekunder) är ett mått på längd. Frekvensen i en röst beror på graden av vibrationer i stämbanden. En hög frekvens av vibrationer uppfattas oftast som ett högt tonläge av lyssnaren (Yarmey, 1995).

Mansröster är typiskt ”djupare” eller ”mörkare” än kvinnoröster. Den viktigaste skillnaden ligger i tonhöjden – kvinnoröster har överlag högre tonhöjd än mansröster. Detta beror på att kvinnoröster har högre frekvens – stämläpparna vibrerar snabbare i kvinnoröster än i mansröster. Anledningen till detta är att mäns struphuvud och stämläppar som regel är större än kvinnors. Av samma anledning ligger barnröster högre än vuxnas röster. I mansröster svänger stämbanden runt 100-200 gånger per sekund (dvs. med 100-200 Hz). I kvinnoröster brukar man räkna med ungefär det dubbla. Hos små barn går svängningsfrekvensen ofta upp till 400-500 Hz (Engstrand, 2004).

I normal samtalston producerar vi människor i genomsnitt tio ljud per sekund. Den exakta varaktigheten i dessa ljud måste mätas i millisekunder, vilket gör det svårt för oss att uppskatta deras längd. Denna begränsning gör det svårt att urskilja ljud, i en för oss, obekant röst. Vår förmåga att överhuvudtaget kunna identifiera en röst vi hört vid ett eller flera tillfällen beror istället på förmågan att upptäcka och identifiera språkvariationer. På grund av de biologiska skillnaderna i strukturen av röstmekanismer är skillnaderna i röstkvalité och de utmärkande dragen mellan olika talare större än skillnaden mellan en och samma talares röst vid olika tillfällen. Regionala, sociala och kulturella skillnader märks i accent, rytm och inlärd språkvanor som tonfall och pauseringar. Även speciellt utmärkande uttal av vissa ord och fraser och även talrubbningar bidrar till språkvariationer (Yarmey, 1995).

Det finns en mängd kända faktorer som påverkar ett vittnes prestation och möjligheter att identifiera en röst. Dessa kan organiseras i tre olika kategorier: röstspecifika faktorer, vittnesspecifika faktorer och situationsspecifika faktorer (Yarmey, 1995).

### *Röstspecifika faktorer*

*Bekanta och obekanta röster.* I vardagen känner vi lätt igen röster på familjemedlemmar, vänner och arbetskamrater. Vi kan även med lätthet identifiera rösten på många radio- och TV-personligheter, politiker och skådespelare. Denna

erfarenhet har gett upphov till den existerande myten att all röstigenkänning är korrekt och tillförlitlig. Det finns dock en viktig skillnad i kommunikationen mellan främmande människor och den där vänner talar med varandra eller när vi hör bekanta röster. Identifikationen av bekanta röster underlättas av en mängd faktorer som t.ex. användandet av hälsningsfraser, favorituttryck, smeknamn, stereotypa fraser och annan information som har med sammanhanget att göra. Bekanta röster verkar bestå av ett unikt mönster av perceptuella faktorer (Yarmey, 1995). Mindre familjära röster är svårare att känna igen, därför att de saknar information som går att härleda till kontext. Graden av familjaritet har större betydelse än vi tror. I en studie av Yarmey visade resultatet att människor kan identifiera rösterna på dem som står en mycket nära som t.ex. familjemedlemmar med 89 % säkerhet vid en röstkonfrontation, men antalet korrekta identifikationer sjunker till 69 % för övriga bekantskaper som grannar och arbetskamrater (Solan & Tiersma, 2003). Även Hollien (2002) betonar att prestationsförmågan för att identifiera bekanta röster, skiljer sig markant mellan mycket bekanta röster och bekanta röster. För att en röst överhuvudtaget skall räknas som bekant för en lyssnare menar Hollien att lyssnaren måste ha god hörsel och ha hört rösten någorlunda regelbundet under minst två års tid.

Hollien, Majewski och Doherty (1982) utförde en serie experiment för att jämföra skillnaden i prestation mellan försökspersoner som kände den som talade, och de som inte kände den som talade. Vid normala röstförhållande gjorde försökspersonerna, som var bekanta med talaren 98 % korrekta identifikationer, medan de som inte var bekanta med talaren endast gjorde 33 % korrekta identifikationer. Det finns emellertid anledning att infoga en varning för röstigenkänning för bekanta röster. En bekant röst förknippas ofta med den speciella miljö där man brukar höra den. Detta leder till att då en bekant röst lyfts från sitt naturliga sammanhang, minskar möjligheten att identifiera den (Van Lancker, Kreiman & Emmorey, 1985).

*Främmande språk, brytning och dialekter.* Forskning om ”cross-racial effekter” dvs. vilken effekt främmande språk eller accenter har på förmågan att identifiera röster är starkt eftersatt jämfört med samma effekt för ansiktsigenkänning. Den forskning som gjorts talar för att minnet för olikheter i t.ex. accent är mer beroende av olikheter i själva lyssnandet än i det uttalade. Röster som talar ett främmande språk uppfattas som homogena eller likadana, medan vi kan urskilja olikheter i röster som talar med samma accent som oss själva (Yarmey, 1995). Det finns alltså forskning som talar för att öronvittnen har lättare för att känna igen röster talade på det egna språket, än röster talade på ett främmande språk. De frågor som rör hur t.ex. en gärningsmans dialekt, brytning m.m. påverkar öronvittnens tillförlitlighet är mycket angelägna. Trots detta är området näst intill outforskat (Goggin, Thompson, Strube & Simental, 1991).

*Faktorer som påverkar den som talar.* Den som talar kan också vara utsatt för påverkan som ger störningar. Rädsla, ångest eller andra stresstillstånd är vanligt förekommande när en gärningsman talar under tiden ett brott begås. Sådana tillstånd kan markant påverka en eller flera parametrar i rösten (Hollien, 2002). Det är inte heller ovanligt att en gärningsman skriker åt ett vittne i samband med ett brott. Forskning visar att om gärningsmannen skrikit och vid en senare vittneskonfrontation talar i normalt tonfall, kommer detta att försvåra igenkänningen av hans röst (Granhag, 2001). Även droger och alkohol, och till och med tillfälliga hälsotillstånd som förkylning kan ge tillfälliga förändringar i rösten (Hollien, 2002).

*Röstförvrängning.* Uppkomst av mänskligt tal innebär komplex motorisk aktivitet som resulterar i en stor mängd rörelser och gester som är tidsmässigt samordnade med stor precision (McAllister, 1998). Enligt Yarmey (1995) förändras de akustiska beståndsdelarna i en röst när den som talar väljer att förställa sin röst. Även röstförändringar som att alternera tonerna från ett normalt till ett förargat röstläge ger signifikant lägre korrekt röstigenkänning.

En gärningsman i ett brottmål har minst två chanser att förställa sin röst:

1. Då öronvittnet hör honom för första gången.
2. Då han talar in sitt "bidrag" till den röstkonfrontation han tvingas medverka i.

Forskningen visar att även mycket subtila förvrängningar kan ställa till med stora problem (Yarmey, 1995).

Ett effektivt sätt att förställa en röst är att viska. En viskande röst saknar viktiga karaktärsdrag som tonhöjd, tonfall och intonation. I en studie av Orchard och Yarmey (1995) fann man att en röst som först talat i ett normalt tonläge, var mycket svårare att identifiera för de försökspersonerna som senare fick höra samma röst viska än för de personer som fick höra rösten i samma normala tonläge som vid första tillfället.

Clifford (1980) refererar till en undersökning av Pollack, Pickett och Sunby, där bekanta röster endast kunde identifieras till 30 % när den som talade viskade, jämfört med 95% korrekta identifikationer om samma talare använde ett normalt tonläge.

Att medvetet förställa rösten är också ett effektivt sätt att försvåra identifiering. En penna kan t.ex. bli ett effektivt sätt att maskera en röst. Det finns rapporter om att brasilianska kidnappare har placerat en penna mellan tänderna under tiden de framfört krav på lösensumman. Detta enkla trick påverkar tungans och käkarnas rörelser på ett sätt som gör rösten mycket svårare att identifiera (Solan & Tiersma, 2003).

### *Vitnessspecifika faktorer*

*Ålder.* Det har hävdats att kronologisk ålder kan påverka förmågan till röstigenkänning (Hollien, 2002). Förutom att äldre öronvittnes generella förmåga att uppfatta ljud kan påverka, finns andra åldersrelaterade påverkansfaktorer. Även om barn är bra på att känna igen en röst de har hört, är de inte lika bra som vuxna på att känna igen en röst de bara hört vid ett enda tillfälle. Forskningen visar att barn först någon gång mellan 10 och 15 års ålder börja komma i kapp vuxna vad gäller förmågan att identifiera obekanta röster (Granhag, 2001). Enligt Bull och Clifford (1984) är vår förmåga att identifiera röster generellt som bäst mellan 16-40 års ålder. Men eftersom det finns individuella skillnader i minnesförmågan går det inte med säkerhet att säga att en individ över 40 år är sämre på röstigenkänning än en yngre individ.

*Kön.* I McGehee's (1937) experiment visade resultaten att de manliga försökspersonerna var markant bättre än de kvinnliga på röstigenkänning (refererat i Hollien, 2002). Nyare forskning har inte visat att det finns några könsskillnader vad gäller förmågan till röstidentifikation (Hollien, 2002). Däremot visar studier att vi generellt är bättre på att identifiera röster av det egna könet än röster av det motsatta könet (Saslove & Yarmey, 1980).

*Stress och arousal.* När röstidentifikation används av vittnen i rättssammanhang är det viktigt att uppmärksamma tänkbar stress, rädsla och andra emotioner som händelsen i sig kan ha väckt. Yarmey (1995) menar att det ännu inte är fastställt hur effekten av sådana trauman påverkar minnet, om de främjar eller försvårar minnesåtergivning. Vissa studier pekar på att minnet påverkas negativt av stress och rädsla, medan andra studier visar det motsatta – stress underlättar minnesprocessen. Yarmey (1995) påpekar även att problemet med dessa studier är att det är näst intill omöjligt att i kontrollerade experiment väcka samma känslor som ett vittne känner vid ett verkligt brott.

*Vittnets karaktär.* Få forskningsprojekt har fokuserat på direkta karaktärsdrag som påverkar att vissa personer är konstant bättre på uppgifter som röstigenkänning. En förståelse för sådana faktorer skulle underlätta förståelsen för ett öronvittnes prestation. Hollien (2002) refererar dock till en studie utförd av DeJong. DeJongs studie utfördes i syfte att undersöka om en individs minne och medfödda auditiva förmåga och musikaliska begåvning påverkade förmågan till röstigenkänning. DeJong rapporterade att intelligens, i form av kognitiva processer, var en bättre prediktor på subjektiv förmåga för röstigenkänning jämfört med auditiv minnesförmåga. Hon rapporterade dock att en lyssnare med en hög grad av musikalisk begåvning kan förväntas ha en högre förmåga till röstigenkänning än den som saknar musikalisk begåvning.

### *Situationsspecifika faktorer*

*Tidsmässig fördröjning.* Hollien (2002) menar att det är påtagligt att möjligheten för ett vittne att identifiera en okänd röst försämras efter en tid. Detta bekräftas genom flera studier, men det går inte alltid att förutsäga ett exakt mönster för denna glömska (Clifford, Rathborn & Bull, 1981; Read & Craik, 1995). Det enda som kan sägas med säkerhet är att korrekta röstidentifikationer har ett samband med tidsaspekten, men att glömskekurvan ser olika ut för olika individer och att det finns många fler variabler än tid som kan påverka glömskeprocessen. En viktig men okontrollerad variabel i röstigenkänning är alltså vittnens minne. Ett komplicerande faktum är att hörselminnet är ganska olika synminnet. Det har t.ex. visat sig att ögonvittnes minne inte försämras signifikant för korrekta utpekande även efter så lång tid som flera månader, medan samma *fördröjning* för hörselstimuli ger markanta försämringar i röstigenkänning (Hollien, 2002).

För att summera de studier som gjorts kan man konstatera att minnet för en främmande röst avtar mycket lite eller inte alls under de första 24 timmarna. Med två till fyra veckors fördröjning sjunker antalet korrekta identifikationer. Med anledning av detta kan vi alltså skilja på kort fördröjning (timmar) och lång fördröjning (dagar och veckor) (Clifford, 1980). I en egen studie konstaterade Clifford (1980) att antalet korrekta identifikationer var 50 % efter en vecka, 43 % efter två veckor och 9 % efter tre veckor. Det bör dock påpekas att röstidentifikationer med endast några minuters fördröjning inte nödvändigtvis leder till en hög frekvens av korrekta identifikationer (Yarmey, 1994).

*Initial exponeringstid.* Tiden (varaktigheten) ett vittne lyssnar till en röst påverkar möjligheten till korrekta identifikationer. Under optimala observationsförhållande kan röster identifieras vid röstprov på en längd av två sekunder eller mindre (Yarmey, 1995). I de flesta experiment har man låtit vittnet lyssna på korta röstfragment – ofta med en varaktighet av 20 sekunder eller därunder. Detta upplägg stämmer ganska väl

med vad som händer i verkliga situationer – en person som ringer in ett hot håller sällan några längre tal (Granhag, 2001). En mängd studier visar emellertid att ju längre tid en försöksperson får möjlighet att lyssna till en röst desto större blir säkerheten i identifikationerna. Korrekta identifikationer är överlagset större vid en lång exponeringstid - 8 minuter i förhållande till kort exponeringstid - 30 sekunder (Yarmey, 1995). Ett problem är att med ökad exponeringstid följer också fler felaktiga identifikationer. Man kan naturligtvis fråga sig hur det både kan bli fler korrekta, och fler inkorrekta identifikationer vid lång exponeringstid. Svaret är att vittnen som inledningsvis fått lyssna på ett relativt långvarigt röstfragment gör fler identifikationer. Eftersom de vittnen som hört ett längre röstfragment gör fler identifikationer än de som lyssnat till ett kortare fragment, har de utrymme för att göra både rätt och fel oftare. Det är till och med klarlagt att längden på exponeringstider påverkar antalet felaktiga utpekande mer än det påverkar antalet korrekta utpekande vad gäller röstigenkänning (Yarmey, 1994).

*Upprepad exponering.* Det finns studier som visar att längden på ett röstprov är av mindre betydelse än om försökspersonen har möjlighet att höra rösten flera gånger under kortare tid (Yarmey, 1995). Forskning visar att det är lättare att känna igen en röst man hört vid fyra tillfällen á 15 sekunder, än en röst man hört vid endast ett längre tillfälle á 60 sekunder (Clifford, 1980). Samma goda identifikationsförmåga gäller för försökspersoner som hört en röst tre gånger i 20 sekunder, jämfört med att ha hört en röst en gång i 60 sekunder (Solan & Tiersma, 2003). Ett annat intressant forskningsresultat är att vårt minne är klart bättre om vi haft en aktiv konversation med personen i fråga, än om vi under en lika lång tidsperiod passivt lyssnat till rösten (Granhag, 2001).

*Planerad/oplanerad inkodning.* I undersökningssituationer förvarnas ofta vittnet om att hon eller han strax kommer att få höra ett röstfragment. Detta rimmar dåligt med vad som ofta händer i verkliga situationer. Inte överraskande har man funnit att förvarnade öronvittnen är bättre på att komma ihåg röster än icke förvarnade. Få eller om ens några vittnen är förberedda på att bli vittnen i samband med ett brott. Minnen från ”verkliga livet” inkodas och lagras oftast automatiskt. Clifford (1980) refererar till en undersökning av Gaiselman och Bellezza som kunde konstatera att åtminstone talarens kön inkodas med automatik i lyssnarens minne. I en undersökning av Saslove och Yarmey (1980) fann man att förvarnade försökspersoner klarade att identifiera en främmande röst med 70 % korrekthet. Försökspersoner som inte förvarnats klarade motsvarande identifikation med 62 % korrekthet.

*Teknisk påverkan.* I samband med röstigenkänning finns flera faktorer av teknisk art som kan ge upphov till problem. Ett vittnes förmåga till korrekta identifikationer påverkas av ljudkvaliteten på själva rösten. Försämring av ljudkvaliteten sker när någon pratar i telefon, när det är dålig kvalitet på bandupptagningar eller när inspelningar har skett med mikrofoner med begränsad kapacitet. Detta gör att viktig information om den som talar går förlorad och denna kan aldrig ersättas. Bandinspelningar kan också inkludera störande bakgrundsljud som vind, motorljud, musik eller andra röster (Hollien, 2002).

*Träning.* En relevant fråga i sammanhanget är om det går att träna upp sin förmåga till bättre röstigenkänning. Hollien (2002) menar att en summering av forskningen visar



att t.ex. tränade och professionella fonetiker klarar uppgiften att identifiera röster bättre än otränade individer. Generellt klarar välutbildade fonetiker uppgiften bättre under följande förhållande:

1. Minimal forensisk träning.
2. Rösten är obekant.
3. Oförberedda på uppgiften.
4. Presenterade för enbart korta röstprov.

Däremot finns inga belägg för att generell träning i fonetik ger signifikant bättre prestationsförmåga för röstigenkänning (Hollien, 2002). Mot detta talar att det finns forskning som visar att blinda öronvittnen presterar bättre än seende öronvittnen i igenkänningstest (Bull, Rathborn & Clifford, 1983). Clifford (1980) menar att den individuella förmågan till inkodning är en troligt avgörande faktor för om ett vittne vid ett senare tillfälle skall kunna identifiera en främmande röst. Forskningen visar alltså att det är svårt att träna människor att bli bättre öronvittnen. Däremot verkar det rimligt att en lyssnare som har god förmåga till uppmärksamhet kommer att lyckas bättre än ett öronvittne som saknar samma goda förmåga (Clifford, 1980).

### *Öronvittneskonfrontationer*

De flesta studier av röstigenkänning har utförts i form av experiment som har karaktären av en vittneskonfrontation. En öronvittneskonfrontation, eller en röstparad som det även kallas, är när en person som har hört, men inte sett, en målperson antas kunna identifiera samma målperson genom att enbart höra dennes röst. Målpersonens röst presenteras bland en mängd andra röster – vanligtvis fem till åtta – och vittnet ombeds identifiera vilken av rösterna som tillhör målpersonen (Hollien, 2002). Vittneskonfrontationer används ofta i syfte att vägleda Polisen i spanings- och utredningssammanhang eller för att man skall få bevismaterial i en rättegång. I dessa sammanhang är det lika viktigt att en oskyldig kan frias från misstankar som att den skyldige kan pekats ut. Värdet av en vittneskonfrontation är beroende av hur den genomförs och dokumenteras. Rätt genomförd kan en vittneskonfrontation vara en ovärderlig tillgång i en brottsprocess. En bristfällig vittneskonfrontation kan däremot helt sakna värde eller till och med försvåra en brottsutredning. Vittneskonfrontationer kräver med andra ord kunskap och noggrannhet (RPS, 2005). En öronvittneskonfrontation genomförs oftast under följande förutsättningar:

1. Målpersonen är inte känd för vittnet.
2. Det finns en fördröjning mellan den tidpunkt då vittnet hörde målpersonen första gången och tidpunkten när öronvittneskonfrontationen hålls – ibland ganska lång tid.
3. Vittnet har ofta mycket begränsad träning i röstigenkänning.
4. Processen är kontrollerad, trots detta kan emotionella och andra personliga variabler påverka.

Öronvittneskonfrontationer är inte lika vanliga som ögonvittneskonfrontationer, trots det förekommer de. Problem associerade med öronvittneskonfrontationer är av växande betydelse. Hollien (2002) anger tre skäl till den ökade betydelsen:

1. Denna form av röstigenkänning har blivit vanligare än det varit tidigare.

2. I vissa fall (troligen ganska många) är dessa vittneskonfrontationer konstruerade av individer med endast marginell kompetens.
3. Adekvata riktlinjer för denna form av vittneskonfrontationer finns ännu inte.

*Viktiga faktorer vid öronvittneskonfrontationer.* Precis som ögonvittnen är öronvittnen känsliga för olika typer av påverkan. Granhag (2001) nämner några saker som bör beaktas i samband med öronvittneskonfrontationer, för att undvika onödiga och förödande felkällor:

1. Människor är i allmänhet mycket sämre informerade om sin hörsel förmåga än t.ex. om sin synförmåga. I samband med en vittneskonfrontation bör därför en allmän test av vittnets hörsel genomföras.
2. Rösterna som förekommer i vittneskonfrontationen skall vara inspelade. Om man väljer att låta ett vittne ta del av rösterna ”live” exponerar man sig för flera felkällor. Ett öronvittne kan påverkas av så små saker som att en av figuranterna kommer av sig eller tittar bort.
3. Alla de röster som spelas upp skall naturligtvis låta relativt lika. Detta gäller t.ex. för kön, ålder och dialekt. Det är också givet att alla figuranter yttrar samma fras.

Ett område som har visat sig vara betydelsefullt är vittnets förväntningar att målpersonen skall finnas med i konfrontationen. Hollien (2002) refererar till flera undersökningar, där problem har uppstått för att vittnen antar att målpersonen skall finnas bland de röster som förekommer i konfrontationen. Dessa undersökningar visar att ytterst få vittnen använde sig av möjligheten att indikera att målpersonen inte fanns med i öronvittneskonfrontationen, trots att de *visste* att så var fallet. Vittnets förväntningar att målpersonen skall finnas med bland rösterna kan alltså påverka deras identifikation. I en vittneskonfrontation finns fyra möjliga resultat i identifikationsuppgiften.

Tabell 1.

*Fyra möjliga resultat av en identifieringsuppgift*

Vittnets identifikation	Rätt Målperson	Annan person
Vittnet gör ett utpekande	Sant Positivt	Falskt Positivt
Vittnet gör inget utpekande	Falskt Negativt	Sant Negativt

(Haber & Haber, 2000)

I två av fyra alternativ har vittnet gjort en korrekt identifikation (den översta vänstra och den nedersta högra cellen). Det finns även möjlighet att vittnet gör två felaktiga identifieringar. Dessa två misstag i identifieringen är inte ekvivalenta. När ett vittne i rättssammanhang felaktigt identifierar en person som gärningsmannen - *Falskt Positivt* - riskerar en oskyldig person att bli anklagad, och till och med dömd för ett brott. Detta måste betraktas som ett mycket allvarligare fel än - *Falskt Negativt* - där en skyldig gärningsman oriktigt utpekats som oskyldig. Undersökningsresultat visar att

ögonvittnesidentifikationer ofta har en Falskt Positivt grad som överstiger 25 % (Haber & Haber, 2000). Dessa misstag kan minskas men inte helt elimineras, genom att vittnet informeras om att målpersonen möjligen inte finns med i konfrontationen (Hollien, 2002). I en rapport från Interpol (2001) varnas emellertid för att en alltför ensidig strävan att eliminera antalet *Falskt Positiva* identifikationer, kan leda till oacceptabla frekvenser av *Falskt Negativa* identifikationer och/eller en orealistiskt låg nivå av korrekta identifikationer.

*Jämförelse mellan ögonvittnes/öronvittnes konfrontationer.* Det är konstaterat att det finns många likheter mellan de båda processerna och dessa är självklara, men det finns också olikheter. Det finns t.ex. skillnader i hur:

1. Auditiva och visuella minnet hanteras.
2. Hur en röst är strukturerad och därmed hur den kategoriseras eller analyseras jämfört med ett ansikte.
3. Rädsla, ilska eller arousal kan påverka de båda identifikationsprocesserna på olika sätt.
4. Endast ett kort röstprov begränsar oss mer än en snabb blick på en person.
5. Skillnader i medfödd förmåga och andra utmärkande drag som främjar gott minne för synintryck till skillnad från minne för röster och andra ljud.

Hollien (2002)

### *Konfidensbedömningar*

En viktig fråga för forskning inom området vittnespsykologi, och där med också för röstigenkänning, är vittnets objektiva förmåga till korrekta utpekande och hans/hennes subjektiva konfidens till denna korrekthet (Clifford, 1980). Den grad av konfidens en person känner kan uttryckas med en konfidensbedömning från "helt säker" till "ren gissning". Konfidensbedömningar går under den bredare beteckningen metakognitiva bedömningar. Begreppet metakognition handlar bland annat om människors reflexioner kring vad de tror sig veta och minnas. Graden av giltighet eller realism i personers konfidensbedömningar kan definieras som den grad konfidensbedömningarna överensstämmer med bedömarens faktiska minnesprestation (Johansson, 2004).

En av de främsta anledningarna till att studera relationen mellan (a) graden av säkerhet och (b) graden av korrekthet, är att man vet att juridiskt ansvariga tenderar använda just vittnets grad av säkerhet vid utvärdering av en utsagas tillförlitlighet (Granhag, 2001). Det verkar intuitivt mest troligt att en person gör en korrekt bedömning när hon/han är helt säker på att identifikationen är riktig (Clifford, 1980). Kalibreringsforskningen visar dock att människor ofta är säkrare på sin sak än de har fog för (Johansson, 2004). Uppgiftens svårighetsgrad har också visat sig ha betydelse för konfidensbedömningar. Överkonfidens är starkt relaterat till svåra igenkänningsuppgifter (Olsson, 2000). Det huvudsakliga antagandet är att missledas att tro att den som uppger sig vara säker på sin sak också antas vara mest trovärdig (Clifford, 1980).

Även vittnespsykologisk forskning visar att vittnen är överkonfidenta, det vill säga att de kopplar alltför höga säkerhetsbedömningar till sina utsagor. Ett stort antal studier visar att vittnen tror att de minns rätt i långt större utsträckning än vad som faktiskt är fallet (Allwood & Granhag, 1999). Detta gäller för såväl ögonvittnen som öronvittnen (Yarmey, 1995).

### *Ögon- och öronvittnens konfidensbedömningar och prestation*

Det finns ett problem i det objektiva förhållandet mellan korrekthet och konfidens, nämligen att detta förhållande inte alltid visat sig vara positivt. Detta gäller framförallt i ögonvittnesundersökningar. Det är också mycket svårt att jämföra det stora antal studier som gjorts på ögonvittnen. Vissa av dessa har använt sig av verkliga händelser, vissa av filmer eller videomaterial och andra har endast använt ansiktsfotografier. Ett annat problem är att sättet som använts för att beräkna den relevanta konfidensen varierar mellan de olika studierna (Clifford, 1983). En metastudie av 24 röstigenkänningsundersökningar visar en låg positiv korrelation ( $r = .25$ ) mellan korrekta röstidentifikationer och konfidens. Undersökningar visar också att en felaktig identifikation oftast görs med hög konfidens (Yarmey, 1994).

I en studie av Olsson, Juslin och Winman (1998) konstaterades att prestationen och kalibreringen var mycket sämre för öronvittnen jämfört med ögonvittnen. I studien visade öronvittnena att de klart överskattade sin förmåga till korrekta identifikationer – de var alltså överkonfidenta, jämfört med ögonvittnen. Detta kan delvis förklaras av att öronvittnenas uppgift i denna studie kunde ansågs vara svårare än motsvarande uppgift för ögonvittnen. Ett fenomen som kan uppstå när öronvittnen testas på röster och ögonvittnen testas på ansiktsfotografier. Ett dilemma i denna typ av jämförelser är alltså att uppnå samma svårighetsgrad för de båda uppgifterna. Olsson, Juslin och Winman (1998) utförde därför ytterligare en studie där syftet var att mäta kalibreringen för röst- och ansiktsgenkänning med identiska processer. I praktiken innebär det att försökspersonerna testades på röst- och ansiktsgenkänning med samma figuranter. Även denna studie visade lägre nivå för korrekta identifikationer för röstigenkänning jämfört med ansiktsgenkänning. Resultaten visar också på att försökspersonerna var starkt överkonfidenta i sina identifikationer för röster.

Vad är det då för möjliga skillnader i uppgiften att identifiera röster och ansikten?

Olsson, Juslin och Winman (1998) anger att en möjlig förklaring kan vara att det auditiva stimulimaterialet lider mer av den fysiska reproduktionen i experimentsituationen än det visuella stimulimaterialet. En annan förklaring kan vara att det föränderliga i en främmande röst gör att identifikationen är beroende av tidsfaktorn, medan all motsvarande information är ögonblickligt tillgänglig när vi ser på ett ansikte. Olsson, Juslin och Winman (1998) ger också en möjlig förklaring till varför försökspersonerna i deras studie, var mer överkonfidenta för röstigenkänning än för ansiktsgenkänning. I det dagliga livet upplever sig människor ofta med lätthet kunna identifiera röster, vilket ger upphov till en föreställning att denna typ av identifikation är en tämligen enkel uppgift. Att identifiera främmande röster ingår dock mycket sällan i det vardagliga livet. I en experimentsituation med främmande röster skiljer sig därför förutsättningar drastiskt från dagliga erfarenheter. Enligt Olsson, Juslin och Winman (1998) finns det en möjlighet att försökspersonerna håller fast vid sin uppfattning att röstigenkänning är en tämligen enkel uppgift. Detta kan bidra till höga värden i konfidensbedömningarna och därmed också leda till att försökspersonerna blir starkt överkonfidenta. Denna generalisering av god prestationsförmåga i en situation, som förs över till en annan situation med betydligt sämre prestationsförmåga, kan vara en av förklaringarna till den höga överkonfidens som har observerats för röstigenkänning.

Olsson, Juslin och Winman (1998) menar att studier ger belägg för att det forensiska systemet skall vara försiktiga när man utvärderar resultatet från röstigenkänningstest.

Detta med anledning av att korrektheten i röstigenkänningstest för okända röster oftast är låg. Det finns även ett bristande samband mellan konfidens och korrekthet för sådana röstigenkänningstest. I liknande jämförelsetest, bör därför rättssystemet sätta större tilltro till ögonvittnen än till öronvittnens konfidensbedömningar. Det är mycket troligt att ögonvittnen är överkonfidenta, men det är tyvärr troligt att öronvittnen, i än högre grad är överkonfidenta.

### *Studiens syfte och hypoteser*

Syftet med denna undersökning är att undersöka om det finns skillnader i förmågan till identifikation och konfidensbedömningar för röstigenkänning och ansiktsigenkänning som är beroende av de deltagande gruppernas yrkeserfarenhet. En grupp, kontrollgruppen, har liten eller ingen träning, utöver den som man får i vardagslivet. Den andra gruppen, experimentgruppen, har erfarenhet i identifikationsuppgifter. Till den andra gruppen räknas bl.a. anställda i kriminalvården, larmoperatörer, poliser och väktare. Dock har denna grupp ingen eller ytterst ringa formell träning i röstidentifiering, det rör sig istället om erfarenhet förvärvad genom yrkeslivet.

De konkreta hypoteserna i denna undersökning är:

H<sub>1</sub> = Yrkesmässig erfarenhet av röstidentifikationsuppgifter ger bättre förmåga till korrekta identifikationer av röster.

H<sub>2</sub> = Yrkesmässig erfarenhet av röstidentifikationsuppgifter ger bättre kalibrering i konfidensbedömningarna av röster.

## Metod

### *Undersökningsdeltagare*

I undersökningen deltog sammanlagt 54 försökspersoner fördelade på två grupper. Försökspersonerna i kontrollgruppen var utvalda genom bekvämlighetsurval. Samtliga saknade dock yrkesmässig erfarenhet av identifikationsuppgifter. I kontrollgruppen deltog totalt 27 försökspersoner, 11 män och 16 kvinnor, i åldern 16 -70 år (M=36.6 år). I experimentgruppen deltog 27 yrkesverksamma väktare, poliser, kriminalvårdare eller larmoperatörer, med minst ett års yrkeserfarenhet (M = 13 års yrkeserfarenhet). Av försökspersonerna i experimentgruppen var 20 män och 7 kvinnor, alla i åldern 20-60 år (M=31.4 år). 52 av försökspersonerna hade svenska som modersmål, en finlandssvenska och en försöksperson hade polska som modersmål. Ingen av försökspersonerna erhöll ersättning för sin medverkan.

### *Undersökningsdesign*

Samtliga försökspersoner testades på ansiktsigenkänning och röstigenkänning vid samma tillfälle. Experimentet förekom i sex olika versioner där filmsekvenserna och ljudklippen presenterades i olika ordning. För ansiktsigenkänning användes korta filmsekvenser och ansiktsfotografier. För röstigenkänning användes röster i form av inspelade ljudklipp. Vid skapande av stimulimaterialet eftersträvades att svårighetsgraden i identifikation av röster och ansikten skulle vara snarlika.

### *Material*

Undersökningsmaterialet utgjordes av 152 bilder i form av ett bildspel. Bildspelet var producerat i programvaran Microsoft PowerPoint. Bildspelet innehöll förutom instruktioner, totalt tio filmer, 60 ansiktsfotografier, 65 ljudklipp och 20 distraktionsuppgifter. För att genomföra experimentet användes en bärbar dator. Datorn var utrustad med ett grafikkort av märket ATI Radion Mobility 16 Mb AGP. Vidare användes två externa högtalare av märket Ozaki, AMP output power 2 W, 55 dB, 140Hz~16KHz.

*Filmer och ansiktsbilder.* De tio filmerna som användes i experimentet var inspelade med en digitalkamera av märket Minolta med 3.3 megapixels (2056 x 1544). Vid inspelningarna användes ett stativ. Samtliga filmer saknade ljud, visades i färg och var redigerade i programvaran Windows Movie Maker. Filmerna visades i bildspelet med bildstorleken 8 x 10,65 cm. I var och en av filmerna förekom en person som rörde sig framför kameran på ett avstånd av 2-6 meter, sysselsatt med någon vardaglig aktivitet. Personen fanns hela tiden i bild och visades framifrån eller från sidan. I fem av filmerna förekom en kvinna och i de övriga fem filmerna en man. Samtliga var mellan 25 och 65 år. Det förekom inga störande moment i filmerna, i form av andra personer eller händelser i bakgrunden. Varje film var 15 sekunder lång. 14 av de 60 ansiktsbilderna, som hörde till filmerna och användes i undersökningen, var tagna med samma kamera som användes för filminspelningarna. Övriga bilder var hämtade från sökmotorn Google Sverige med hjälp av bildsökning. 30 av bilderna föreställde kvinnor och 30 föreställde män. Generellt för bilderna var att de var tagna på samma avstånd, (ca 1,5 meter) och att ljussättningen var likvärdig. Bilderna visades i bildspelet med bildstorleken 6,35 x 4,75 cm. Samtliga bilder var i färg.

*Ljudklipp.* Samtliga ljudklipp var inspelade med en MP3 spelare av märket T.Sonic TM 630. Vid inspelningarna användes en mindre extern så kallad slipsmikrofon, MICTC2. Av de 65 ljudklipp som användes i undersökningen var 33 inspelade kvinnor i åldern 16-65 år och 32 ljudklipp var inspelade män i åldern 20-65 år. Alla inspelade personer hade svenska som modersmål och talade dialekter från södra Sverige. I samtliga ljudklipp läste personerna i normal samtalston en text bestående av 41 ord. Texten löd:

*Människor är i grunden inte särskilt olika, men i ett krisläge finns det dels de som blir hysteriska och bara gråter, dels de som omedelbart börjar agera för att försöka lösa problemen. Det gemensamma för båda sorterna är att återfå kontrollen (Werkelin & Svanberg, 2006, sid. 128).*

De 65 ljudklippen varierade i tid mellan 12–23 sekunder, beroende på inläsningshastighet (M=16 sekunder). Samtliga ljudklipp var redigerade i programvaran Audacity.

*Distraktionsuppgifter.* I experimentet förekom 20 distraktionsuppgifter i form av logiska uppgifter. Bilderna föreställde en figur där en del av figuren saknades och var markerad med ett frågetecken. Ovanför figuren fanns texten – *Vilken del saknas?* Under figuren fanns fem svarsalternativ numrerade 1-5. 15 av distraktionsuppgifterna visades i färg och de övriga fem i svartvitt.

### *Procedur*

Samtliga försökspersoner genomförde experimentet under liknande villkor. Undersökningen genomfördes på försökspersonernas arbetsplatser eller i hemmiljö. Experimentet genomfördes individuellt eller med två försökspersoner samtidigt. Experimentet genomfördes utan tidsgräns och tog i snitt ca 45 minuter att utföra. Före själva undersökningen informerades försökspersonen om att hon/han strax skulle få se ett antal filmer och höra ett antal ljudklipp. Det var endast tillåtet att se filmsekvenserna och höra ljudklippen en gång. Uppgiften bestod i att, efter varje film eller ljudklipp, försöka identifiera personen som försökspersonen sett eller hört bland sex ansiktsbilder (ansiktsgenkänning) eller sex ljudklipp (röstigenkänning). Försökspersonen informerades om att det inte var säkert att den person som förekommit i filmen eller hörts i ljudklippet, fanns med bland de sex ansiktsbilderna eller de sex ljudklippen. Varje identifikationsuppgift hade således sju svarsalternativ. I uppgiften ingick även att göra en bedömning av hur säker försökspersonen var på sitt svar, uttryckt i procent (konfidensbedömning) efter varje identifikation. Svaren lämnades på en förtryckt svarsblankett. På svarsblanketten ombads försökspersonen även att lämna uppgift om ålder, kön, yrkeserfarenhet, ev. musikalitet och de personliga strategier som använts för ansiktsgenkänning respektive röstigenkänning. Experimentet började med att försökspersonerna fick höra ett ljudklipp som inte tillhörde identifikationsuppgifterna. Detta ljudklipp användes för att försökspersonerna skulle kunna göra en personlig inställning av ljudnivån. Därefter startades bildspelet med stimulimaterialet. Försökspersonerna fick omväxlande se en av de tio filmsekvenserna och höra ett av de tio ljudklippen med rösterna. Efter varje film eller ljudklipp visades en distraktionsuppgift på skärmen. Försökspersonerna fick under ca 90 sekunder försöka lösa den logiska uppgiften. Efter distraktionsuppgiften fick sedan försökspersonerna, se eller höra de sex ansiktsbilderna eller de sex ljudklippen som hörde till uppgiften. Försökspersonerna antecknade sina svar och konfidensbedömningar på svarsblanketten efter varje uppgift. På svarsblanketten fanns angivet att en konfidensbedömning på 16 % eller lägre, räknades som ren gissning.

I hälften av de totalt 20 identifikationsuppgifterna saknades målpersonen bland de sex ansiktsbilderna eller de sex ljudklippen som hörde till uppgiften (Target-absent Lineup). I andra hälften av uppgifterna förekom målpersonen bland de sex ansiktsbilderna eller ljudklippen (Target-present Lineup).

## Resultat

### *Identifikationsprestation och respons*

Totalt utförde de 54 försökspersonerna 540 identifikationsuppgifter för ansiktsgenkänning och 540 identifikationsuppgifter för röstigenkänning. För ansiktsgenkänning gjordes totalt 261 (48,3 %) korrekta utpekanden eller avvisanden. För röstigenkänning gjordes totalt 314 (58,1 %) korrekta utpekanden eller avvisanden. Som figur 1 visar var det ingen signifikant skillnad i korrekt respons för ansikten och röster vad kontrollgruppen beträffar. Tabell 2 och 3 visar frekvensdata för de olika responskategorierna och konfrontationsvillkoren för ansikts- och röstigenkänning.

Tabell 2.

*Frekvens av identifikationsrespons för ansiktsigenkänning och röstigenkänning, där målpersonen fanns med, fördelat på de båda grupperna i undersökningen.*

Identifikationsrespons: Målpersonen finns med i identifikationsuppgiften (Target-present)							
Experimentvillkor:	Korrekta		Felaktiga		Inkorrekta avvisande		Totalt N
	N	%	N	%	N	%	
<b>Ansiktsigenkänning</b>							
Experimentgruppen	63	46,7	33	24,4	39	28,9	135
Kontrollgruppen	70	51,9	31	22,9	34	25,2	135
Totalt	133	49,3	64	23,7	73	27,0	270
<b>Röstigenkänning</b>							
Experimentgruppen	97	71,9	18	13,3	20	14,8	135
Kontrollgruppen	92	68,1	26	19,3	17	12,6	135
Totalt	189	70,0	44	16,3	37	13,7	270

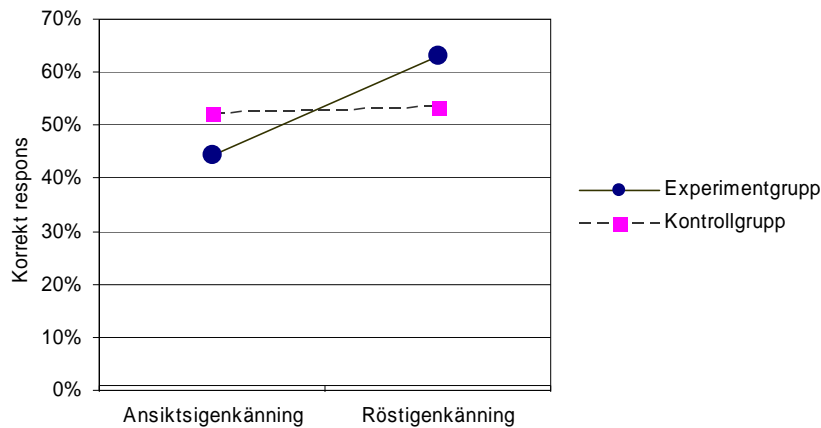
Tabell 3.

*Frekvens av identifikationsrespons för ansiktsigenkänning och röstigenkänning, där målpersonen saknades, fördelat på de båda grupperna i undersökningen.*

Identifikationsrespons: Målpersonen saknas i identifikationsuppgiften (Target-absent)					
Totalt	Korrekta avvisande		Felaktiga identifikationer		N
	N	%	N	%	
<b>Ansiktsigenkänning</b>					
Experimentgruppen	57	42,2	78	57,8	135
Kontrollgruppen	71	52,6	64	47,4	135
Totalt	128	47,4	142	52,6	270
<b>Röstigenkänning</b>					
Experimentgruppen	73	54,1	62	45,9	135
Kontrollgruppen	52	38,5	83	61,5	135
Totalt	125	46,3	145	53,7	270

En Mixed Model ANOVA användes med identifikationsuppgift (två nivåer: ansiktsigenkänning vs. röstigenkänning) som inomgruppsvariabel och grupp (två nivåer: experimentgrupp vs. kontrollgrupp) som mellangruppsvariabel och korrekt respons som beroende variabel. Analysen visar på en signifikant interaktionseffekt,  $F(1,52) = 9.58$ ,  $p < .05$ ,  $\eta^2 = .16$ , mellan de två identifikationsuppgifterna och de två grupperna.





*Figur 1.*

*Interaktionseffekt mellan de två identifikationsuppgifterna och de två grupperna. M och SD visas i tabell 4.*

I undersökningens första hypotes antogs att yrkesmässig erfarenhet av röstidentifikationsuppgifter skulle ge bättre förmåga till korrekta identifikationer av röster. En one-way between-groups ANOVA utfördes för att undersöka om det fanns några skillnader mellan grupperna för frekvensen av korrekt respons för ansikts- och röstigenkänning. Resultatet visade att det inte fanns någon signifikant skillnad,  $F(1,52) = 3.68$ ,  $p = .06$ ,  $\eta^2 = 0.07$ , mellan experimentgruppen och kontrollgruppen för ansiktsgenkänning. Däremot var experimentgruppen signifikant bättre,  $F(1, 52) = 4.95$ ,  $p = .03$ ,  $\eta^2 = 0.09$ , än kontrollgruppen vad gäller röstigenkänning.

### *Kalibrering*

Sambandet mellan försökspersonernas konfidensbedömningar och korrekt respons (Confidence-Accuracy Relationship) har i denna undersökning beräknats med den procedur som kallas kalibrering. Kalibrering kan ses som ett mått på hur väl angivna konfidensbedömningar matchar med korrekt respons. Formeln för beräkning av kalibrering är:

$$C = \frac{1}{N} \sum_1^T n_t(x_t - c_t)^2.$$

$N$  är det totala antalet besvarade frågor,  $T$  är antalet konfidensnivåer som använts,  $c_t$  är andelen korrekta svar för alla item med konfidensnivån  $x_t$  och  $n_t$  är antalet gånger konfidensnivån  $x_t$  användes. Graden av kalibrering kan variera från 0 (perfekt kalibrering) till 1 (total avsaknad av kalibrering). Graden av över/underkonfidens (O/U) kan variera från  $-1$  till  $+1$ . Mätvärdet visar i hur stor utsträckning försökspersonerna i sina konfidensbedömningar, underskattat (negativt värde) eller överskattat (positivt värde) graden av egen korrekt respons.

I undersökningens andra hypotes antogs att yrkesmässig erfarenhet av röstidentifikationsuppgifter skulle ge bättre kalibrering i konfidensbedömningarna av röster. Inga signifikanta skillnader för kalibrering kunde påvisas mellan grupperna eller mellan identifikationsuppgifterna. För medelvärden och standardavvikelse se tabell 4.

För över/underkonfidens visade analysen att experimentgruppen var signifikant mer underkonfident än kontrollgruppen för röstigenkänning,  $F(1, 52) = 13.56$ ,  $p < .01$ .  $\eta^2 = .21$ . Inga andra signifikanta skillnader kunde påvisas. För medelvärden och standardavvikelse se tabell 4.

Tabell 4.

*Medelvärde och standardavvikelse för Korrekt respons, Kalibrering och O/U fördelat på de båda identifikationsuppgifterna för respektive grupp.*

	N	Korrekt respons		Kalibrering		O/U	
		M	SD	M	SD	M	SD
<b>Ansiktsigenkänning</b>							
Experimentgruppen	27	4.44	1.48	0.14	0.10	0.03	0.22
Kontrollgruppen	27	5.22	1.50	0.15	0.09	0.07	0.22
<b>Röstigenkänning</b>							
Experimentgruppen	27	6.30	1.64	0.12	0.10	-0.07	0.17
Kontrollgruppen	27	5.33	1.54	0.14	0.10	0.10	0.17

## Diskussion

Ett av huvudsyftena med denna undersökning var att undersöka om det fanns skillnader i förmågan till korrekta identifikationer för röstigenkänning och ansiktsigenkänning som är beroende av gruppernas yrkeserfarenhet. Resultatet av denna jämförelse visar att det fanns en signifikant interaktionseffekt mellan de två identifikationsuppgifterna och de två grupperna. Experimentgruppen visade sig vara signifikant bättre än kontrollgruppen vad gäller röstigenkänning. Detta ger stöd för undersökningens första hypotes, att yrkesmässig erfarenhet av identifikationsuppgifter ger bättre förmåga till korrekta identifikationer av röster. Liknande resultat har inte gått att finna i tidigare forskning. Resultatet av denna undersökning får därför ses som något överraskande och det krävs ytterligare undersökningar av liknande slag för att verifiera resultatet. Som tidigare nämnts finns inga belegg för att generell träning i fonetik ger bättre prestationsförmåga (Hollien, 2002), men att det samtidigt också finns forskning som visar att blinda öronvittnen presterar bättre än seende öronvittnen i igenkänningstest (Bull, Rathborn & Clifford, 1983). I den forskning som finns kring effekten av träning lämnas flera motstridiga förklaringar. Utebliven effekt av träning har bl.a. förklarats med att det auditiva minnessystemet saknar semantisk struktur, liknande det som finns i det visuella minnessystemet. Men då saknas å andra sidan förklaringar till hur och varför vi kan minnas röster som vi känner igen (Clifford, 1980). En tänkbar förklaring, till den bättre prestationen, kan vara att experimentgruppen gör betydligt fler identifikationer av röster jämfört med kontrollgruppen. En högre frekvens av feedback för röstigenkänning skulle då eventuellt kunna ge någon form av implicit kunskap, något som kontrollgruppen saknar. Om så skulle vara fallet finns det antagligen många fler yrkeskategorier som har likvärdig erfarenhet av röstidentifikation och därmed god förmåga till röstigenkänning, t.ex. telefonister och flygledare. En fråga som kan vara väl värd att utreda ytterligare. Det måste också ifrågasättas om det är själva yrkeserfarenheten i sig som ger upphov till skillnaden i prestationen för röstigenkänning. De mekanismer som påverkar

prestationen måste kartläggas ytterligare för att de skall gå att dra några egentliga slutsatser och att andra möjligheter skall kunna uteslutas. I denna undersökning användes bekvämlighetsurval i urvalet av försökspersoner till kontrollgruppen. Detta faktum ger en osäkerhet som bör elimineras i framtida undersökningar.

I undersökningen användes korta filmsekvenser istället för ansiktsfotografier vid inkodningstillfället för ansikten. Detta för att öka den ekologiska validiteten. Syftet med denna metod har varit att efterlikna verkliga situationer, där målpersonen rör sig i en miljö och sedan skall identifieras med hjälp av ett ansiktsfotografi från en annan miljö. Att närmare bestämma en undersöknings ekologiska validitet är en mycket svår uppgift. Det kan alltid hävdas att verkliga ögonvittnen är bättre, sämre eller på annat sätt annorlunda från dem som används i ögonvittnesexperiment (Wells, 1993). För röstigenkänning användes samma röstinspelningar av målpersonerna vid själva inkodningstillfällena som vid själva konfrontationstillfällena. Detta kan eventuellt ha påverkat identifikationen i de fall målpersonen fanns med i konfrontationsgruppen. I undersökningar liknande denna bör därför olika inspelningar av målpersonen användas vid inkodnings- och identifikationstillfället. Tidigare forskning har dragit slutsatsen att röstigenkänning är svårare än ansiktsigenkänning (Olsson, 2000; Olsson, Juslin & Winman, 1998) beroende på prestationsförmåga. Denna undersökning visade att prestationen var bättre för röstigenkänning. Om det finns några verkliga skillnader i förmågan att identifiera röst och ansikte går inte att konstatera så länge det inte finns utarbetade metoder för att mäta skillnaden i svårighet att identifiera röst jämfört med ansikte.

Resultatet visar inga signifikanta skillnader mellan grupperna vad gäller kalibreringen för röstigenkänning eller ansiktsigenkänning. Undersökningens andra hypotes, att experimentgruppen skulle vara bättre kalibrerad för röstigenkänning får alltså inget stöd. Experimentgruppen visade sig dock vara signifikant mer underkonfident än kontrollgruppen för röstidentifikationsuppgifterna. Detta resultat är något överraskande. Forskning rörande konfidensbedömningar i arbetslivet visar att de flesta yrkesgrupper är överkonfidenta i sina bedömningar gällande professionella beslut (Allwood & Granhag, 1999). Det finns dock en mängd faktorer som kan påverka professionella konfidensbedömningar, bland annat yrkesgruppens vana och träning i att fatta snabba beslut. Möjligheten att själv påverka utgången av uppgiften eller tidsintervallet mellan uppgift och feedback tillsammans med graden av erfarenhet, nämns som andra faktorer som kan påverka yrkesmässiga konfidensbedömningar (Allwood & Granhag, 1999). Den forskning som idag finns om konfidensbedömningar i arbetslivet är begränsad och lämnar ingen tillfredställande förklaring till resultatet i denna undersökning.

Denna undersökning visar också att personer som arbetar med rättsvårdande uppgifter inte kan anses bättre, eller mer tillförlitliga i sina ögonvittnesutsagor än andra. Något som är värt att uppmärksamma. Poliser och väktare förekommer ofta som vittnen i våra tingsrätter och tillskrivs vanligtvis, med automatik, stor trovärdighet.

Den forskning som finns idag om röstigenkänning rör i huvudsak situationsspecifika faktorer som t.ex. exponeringstid och hur minnesförmågan påverkas av tidsmässig fördröjning. Forskning om röst- och vittnesspecifika faktorer lämnar mycket mer att önska. Enligt Rikspolisstyrelsen (RPS, 2005) finns det i Sverige idag ingen etablerad metod för att utföra röstkonfrontationer där vittnet skall identifiera en av flera röster.

Det svenska rättssystemet har helt enkelt valt bort röstidentifikation som ett verktyg med motiveringen att det är tidskrävande och förutsätter stor sakkunskap inom området (RPS, 2005). Detta är säkert ett klokt beslut med tanke på att forskningen i och kring röstigenkänning i dagsläget, får anses vara mycket begränsad och kan riskera rättssäkerheten. Men om vi vänder på myntet kan vi konstatera att om det finns ett ögonvittne till ett brott finns det i regel också ett öronvittne – nämligen en och samma person. Det kan alltså vara rimligt att anta att det skulle finnas lika många öronvittnen som ögonvittnen som kan lämna ett meningsfullt bidrag vid en eventuell vittneskonfrontation, om rättssystemet kunde hantera och ta till vara dessa vittnesmål. Det ena utesluter inte det andra och röstigenkänning kan aldrig ersätta ansiktsigenkänning i vittnessammanhang. Risken finns emellertid att rättssystemet för eller senare, måste förlita sig på ett enda ögonvittne med begränsad synförmåga som i mörker har bevitnat hur en gärningsman begått ett allvarligt brott. Det har gått mer än sjuttio år sedan Bruno Richard Hauptmann dömdes till döden och avrättades för ett brott han troligen inte hade begått (Solan & Tiersma, 2003). Detta misstag kan endast förklaras med dålig kunskap och att ett öronvittne med låg eller ingen tillförlitlighet tillskrevs en extremt hög trovärdighet. Tyvärr kan detta hända igen – forskningen kring röstigenkänning och öronvittnesproblematik har på dessa sjuttio år knappt hunnit börja.

## Referenser

- Allwood, C. M., & Granhag, P. A. (1999). Feelings of confidence and the realism of confidence judgements in everyday life. I P. Juslin & H. Montgomery (Eds.), *Judgment and decision making Neo-Brunswikian and process tracing approaches* (sid. 123-146). Hillsdale, N.J.: Lawrence Erlbaum Press.
- Bull, R., & Clifford, B. R. (1984). Earwitness voice recognition accuracy. I G. L Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (sid. 92-123). New York: Cambridge University Press.
- Bull, R., Rathborn, H., & Clifford, B. R. (1983). The voice-recognition accuracy of blind listeners. *Perception 12*, 223 – 226.
- Clifford, B. R. (1980). Voice Identification by Human Listeners: On Earwitness Reliability. *Law and Human Behavior, 4*, 373-395.
- Clifford, B. R. (1983). Memory for Voices: The Feasibility and Quality of Earwitness Evidence. I S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating Witness Evidence: recent psychological research and new perspectives* (sid.190-218). Chichester: Wiley.
- Clifford, B. R., Rathborn, H., & Bull, R. (1981). The effects of delay on voice recognition accuracy. *Law and Human Behavior, 5*, 201-208.
- Engstrand, O. (2004). *Fonetikens grunder*. Lund: Studentlitteratur.
- Granhag, P-A. (2001). *Vittnespsykologi*. Lund: Studentlitteratur.
- Goggin, J., Thompson, C., Strube, G., & Simental, L. (1991). The role of language familiarity in voice identification. *Memory and Cognition, 19*, 448-458.
- Haber, R. M., & Haber, L. (2000). Experiencing, remembering and reporting events. *Public Policy and Law Psychology, 4*, 1057-1097.
- Hollien, H. (2002). *Forensic Voice Identification*. San Diego: Academic Press.
- Hollien, H., Majewski, W., & Doherty, E.T. (1982). Perceptual identification of voices under normal, stress and disguise speaking conditions. *Journal of Phonetics, 10*, 139-148.
- Interpol, Forensic Science Symposium. (2001). *Forensic speech and audio analysis forensic linguistics*. Hämtad den 22 februari 2007 från [www.interpol.int/Public/Forensic](http://www.interpol.int/Public/Forensic).
- Johansson, M. (2004). *Realism in metacognitive judgments: effects of social factors*. Opublicerad doktorsavhandling. Lunds Universitet.

- McAllister, R. (1998). *Talkommunikation*. (Andra upplagan). Lund: Studentlitteratur.
- Olsson, N. (2000). *Realism of Confidence in Witness Identification of Faces and Voices*. Opublicerad doktorsavhandling. Uppsala Universitet.
- Olsson, N., Juslin, P., & Winman, P. (1998). Realism of Confidence in Earwitness Versus Eyewitness Identification. *Journal of Experimental Psychology: Applied*, 2, 101-118.
- Orchard, T., & Yarmey, D. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology*, 3, 249-260.
- Read, D., & Craik, F. I. M. (1995). Earwitness Identification: Some Influences on Voice Recognition. *Journal of Experimental Psychology: Applied*, 1, 6-18.
- Rikspolisstyrelsen. (2005). *RPS Rapport 2005:2 Vittneskonfrontation*. Stockholm: Rikspolisstyrelsen.
- Saslove, H., & Yarmey, A. Y. (1980). Long-term auditory memory: Speaker identification. *Journal of Applied Psychology*, 1, 111-116.
- Solan, M., & Tiersma, P. M. (2003). Falling on deaf ears. *Legal affairs*. November/december.
- Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters. Part I: Recognition of backwards voices. *Journal of Phonetics*, 13, 19-38.
- Wells, G.L. (1993). What do we know about eyewitness identification? *American Psychologist*, 5, 553-571.
- Werkelin, P., & Swanberg, K. (2006). *Klockan 10.31 på morgonen i Khao Lak*. Stockholm: Bonnier fakta.
- Yarmey, A. D. (1994). Earwitness evidence: Memory for a perpetrator's voice. I D. F. Ross, J. D. Read & M. P. Toglia (Eds.), *Adult Eyewitness Testimony: Current trends and developments* (sid. 101-124). Cambridge: Cambridge Univ. Press.
- Yarmey, A. D. (1995). Earwitness speaker identification. *Public Policy och Law Psychology*, 4, 792-816.