All Ears:
Adults' and Children's Earwitness Testimony

Lisa Öhman

UNIVERSITY OF GOTHENBURG
DEPT OF PSYCHOLOGY

Printed in Sweden by Ale Tryckteam, Bohus
Gothenburg, 2013
ISSN: 1101-718X
ISBN: 978-91-628-8623-3
ISRN: GU/PSYK/AVH--272--SE

For the e-published version of this thesis go to: http://hdl.handle.net/2077/32014

Abstract

Öhman, L. (2013). *All Ears: Adults' and Children's Earwitness Testimony*. Department of Psychology, University of Gothenburg, Sweden

Many crimes are committed under conditions of darkness, by masked perpetrators or over a phone. In such cases the witnesses' auditory observations may have a vital role in the investigative phase and in court. Nevertheless, earwitness testimony is a neglected research area. The present thesis investigated earwitnesses' (i) identification performance for an unfamiliar voice, (ii) memory for the perpetrator's statement, and (iii) ability to describe the voice. All four studies used the same general setup; exposure to an unfamiliar voice for 40 seconds, and an interview including a seven-voice lineup after a two week delay. High ecological validity was a specific aim across all studies. **Study I** explored the performance of children aged 7–9 (*N* = 95), 11–13 (*N* = 78), and adults (*N* = 91). Half were exposed to a Target-Present lineup (TP), and half to a Target-Absent lineup (TA). For both types of lineups the participants performed poorly. In the TP condition only the 11–13-year olds (27 % correct) performed above chance level. Furthermore, in the TA condition, all age-groups showed a high willingness to make an identification. **Study II** investigated the influence of presentation format (direct vs. mobile phone recorded voices) on voice recognition accuracy. The participating adults (*N* = 165) were assigned randomly to one of the four conditions (Initial exposure: direct vs. mobile phone recorded voice; Lineup presentation: direct vs. mobile phone recorded voices). The overall accuracy for correct identification was 13%, which is expected by chance. Further, the results did not reveal any significant effect of presentation format or lineup format. **Study III** compared three types of interviews intended to enhance witnesses' voice memory, as well as content recall. Additionally, an interview protocol developed by the Swedish Security Service, for questioning people that have only heard the perpetrator, was evaluated. After exposure, 11–13-year-olds (*N* = 119) and adults (*N* = 93) were interviewed, and returned after two weeks for an additional interview and a lineup. Overall performance for correct identifications was poor (children: 20%, adults: 19%), and an interview shortly after the witnessed event did not seem to help. The Cognitive Interview (vs. the Swedish Security Service protocol) was found to be beneficial for recalling the content of a brief conversation. **Study IV** investigated the effect of the perpetrator's *tone of voice* and *time delay* on voice recognition accuracy. Further, two types of voice description interviews intended to strengthen the encoding of the voice, were tested. Adults (*N* = 148) and 11–13-year-olds (*N* = 160) either heard the perpetrator speak in a normal tone both at encoding and in the lineup, or in an angry tone at encoding and in a normal tone in the lineup. Witnesses were then interviewed about the voice, either with global questions, or by rating voice characteristics. Half of the witnesses were presented with a lineup shortly after the interview and the others after two weeks. Overall, neither age-group performed above chance level (children: 13%, adults: 10%) and only time delay affected accuracy significantly. Children tested immediately performed better (21% correct) compared to those children tested after two weeks (9% correct). Further, voice descriptions were found to be poor. In sum, after testing a total of 949 witnesses under a number of different conditions, the message is clear; voice identification under reasonably realistic conditions is a highly difficult task. Actors in the legal system should therefore treat voice identification evidence with caution. For earwitnesses to be really useful we must find ways of improving their performance for voice identification, content recall, and voice descriptions.

Keywords: Earwitnesses, voice identification, content memory, voice descriptions, children

Lisa Öhman, Department of Psychology, University of Gothenburg, Box 500, SE-405 30 Gothenburg, Sweden. Phone: +46 31 786 19 34, E-mail: lisa.ohman@psy.gu.se

ISSN: 1101-718X ISBN: 978-91-628-8623-3 ISRN: GU/PSYK/AVH--272--SE

Svensk sammanfattning

Vid vissa brott kan offrets eller vittnets minne av gärningsmannens röst och andra auditiva observationer vara en viktig ledtråd och spela en väsentlig roll både i utredningsfasen och i domstolen. Exempel på sådana situationer är brott som begås i mörker eller av maskerade gärningsmän, som vid överfallsvåldtäkter och olika varianter av rån. En annan kategori är brott som begås över telefon såsom obscena samtal, bedrägeriförsök och andra hotfulla samtal. Visuella iakttagelser är här begränsat men vittnet kan likväl göra viktiga auditiva observationer. Trots att öronvittnens observationer inte är sällan förekommande så är vittnens minne för röster samt minne för vad som sades kraftigt eftersatta forskningsområden. Ett öronvittne kan framförallt bidra med information av tre olika slag; minne för gärningsmannens röst, minne för vad gärningsmannen sa, samt beskrivning av själva rösten. Hur bra ett öronvittne minns auditiv information av detta slag beror på en mängd olika faktorer. I denna avhandling undersöks flera sådana faktorer; vittnets ålder (Studie I, III, & IV) presentationsformatets betydelse (Studie II), retentionsintervallets längd (Studie IV), samt gärningsmannens tonfall (Studie IV). Vidare har olika intervjumetoder undersökts i syfte att försöka förbättra vittnets minne för rösten (Studie III & IV), för vad gärningsmannen sa (Studie III), samt vittnets beskrivning av rösten (Studie IV).

Avhandlingen baseras på fyra experimentella studier utförda med samma grundupplägg, men med vissa variationer beroende på respektive studies specifika syfte. Upplägget syftade till att försöka simulera en verklig situation och omfattade två tillfällen. Vid det första tillfället blev deltagarna exponerade för en okänd röst i 40 sekunder. För att skapa en realistisk situation fick deltagarna föreställa sig att de var i en klädbutik och att de väntade på sin tur utanför en provhytt. Ett skynke hade hängts från taket för att skapa känslan av en riktig provhytt och deltagarna ombads placera sig framför det. De instruerades att de skulle får höra något inifrån provhytten, men inte specifikt att det var en röst. Högtalare var placerade bakom skynket och uppspelningen startade med en mobiltelefonsignal som följdes av en man som svarade och talade med en annan (som ej hördes) angående ett planerat brott. Efteråt blev deltagarna ombedda att återkomma om två veckor för en intervju gällande händelsen som de precis hade bevittnat. De fick dock ingen information om vilken aspekt av händelsen som den kommande intervjun skulle fokusera på. Vid andra tillfället, två veckor senare, tog deltagarna del av en konfrontation innehållande sju inspelade röster. De blev informerade om att rösten de hörde för två veckor sedan kan finnas med bland rösterna, men att det också är möjligt att den inte finns med. Först blev deltagarna ombedda att lyssna noga på alla de sju rösterna (22–26 sekunder per röst) utan att ta något beslut. Därefter fick de höra rösterna en gång till, dock kortare röstfragment (11–14 sekunder per röst). Denna gång ombads deltagarna att göra sitt slutgiltiga val, det vill säga om de ansåg att gärningsmannens röst fanns med bland rösterna och i så fall ange vilket nummer, eller avstå från att identifiera någon röst. Efter konfrontationen fick deltagarna berätta allt de mindes av vad gärningsmannen hade sagt vid observationstillfället.

Människor i alla åldrar kan falla offer för eller bli vittne till brott och det är därför viktigt att undersöka hur bra olika åldersgrupper är på att identifiera röster. **Studie I** undersökte förmågan hos barn i åldrarna 7–9 (*N* = 95) och 11–13 (*N* = 78), samt vuxna (*N* = 91). Variationen från grunduppläget var att hälften av deltagarna tog del av en konfrontation där målrösten var inkluderad, och den andra hälften tog del av en konfrontation där målrösten inte var inkluderad. Prestationen, för båda varianterna av konfrontation, var dålig. När målrösten fanns med var det endast 11–13-åringarna (med 27% korrekta identifikationer) som presterade bättre än slumpen (12.5%, 8 alternativ). När målrösten inte var inkluderad så uppvisade samtliga åldersgrupper en stark benägenhet att göra ett utpekande, det vill säga ett felaktigt utpekande (totalt medelvärde = 53%). För båda barngrupperna samvarierade röstidentifieringen med talhastighet och grundtonsnivå. Ingen av dessa faktorer korrelerade signifikant med de vuxnas identifieringar.

Det frekventa användandet av mobiltelefoner speglas av det höga antalet brott där mobiltelefoner används. Det är därför av stor vikt att veta hur korrektheten vid en röstkonfrontation påverkas av att rösten har hörts via en mobiltelefon. I **Studie II** undersöktes hur presentationsformatet (direkt-inspelade vs. mobiltelefoninspelade röster) påverkar röstidentifieringars korrekthet. De vuxna deltagarna (*N* = 165) fördelades slumpmässigt mellan fyra olika betingelser (Initial exponering: direkt vs. mobilinspelad röst; Konfrontationen: direkt vs. mobilinspelade röster). Totalt sett gjorde 13% av deltagarna ett korrekt utpekande, vilket betyder att de presterade på slumpnivå (12.5%), samt mer än hälften av deltagarna (57%) gjorde ett felaktigt utpekande. Resultaten uppvisade inga signifikanta effekter av presentationsformat eller konfrontationsformat. Dessa resultat indikerar att effekten av mobiltelefoners sämre ljudkvalité inte är speciellt stor för röstigenkänning. Resultaten antyder även att det inte är någon fördel att använda sig av mobiltelefoninspelade konfrontationer i de fall där rösten initialt har hörts via en mobiltelefon.

Eftersom forskning har visat att vittnen är relativt dåliga på att känna igen och identifiera en okänd röst syftade **Studie III** till att försöka förbättra öronvittnens prestation vad gäller röstidentifiering samt minne för vad gärningsmannen sa. I studien jämfördes tre olika intervjumetoder. Ett ytterligare syfte med studien var att utvärdera ett intervjuprotokoll utformat av Svenska Säkerhetspolisen för att användas vid förhör av personer som har utsatts för ett brott där de endast har hört gärningsmannen tala. Avvikelsen från grundupplägget var att efter exponeringen av rösten fick deltagarna föreställa sig att de kontaktade polisen för att anmäla vad de precis bevittnat. Deltagarna blev slumpmässigt fördelade till en av de tre intervjumetoderna. Återigen deltog både 11–13-åringar (*N* = 119) och vuxna (*N* = 93). Totalt gjorde 20% av barnen och 19% av de vuxna ett korrekt utpekande och det var ingen signifikant skillnad mellan de olika intervjumetoderna. Däremot visade sig den kognitiva intervjun vara fördelaktig för de vuxna vad gäller minne för vad gärningsmannen sa. Vidare visade det sig att de vuxna återgav mer korrekt information gällande vad gärningsmannen sa jämfört med barnen. Svenska Säkerhetspolisens intervjuformulär visade sig varken vara fördelaktig för röstidentifiering eller för vad som sades. Snarare, de som intervjuades utifrån detta formulär gjorde fler felaktiga utpekanden än de som intervjuades med en ”standard” intervju, samt rapporterade färre korrekta detaljer jämfört med de som blev intervjuade med den kognitiva intervjun. Slutligen, vittnenas beskrivningar av gärningsmannens röst visade sig vara få och generella.

I **Studie IV** undersöktes två vanligt förekommande faktorer som kan påverka vittnets minne, nämligen gärningsmannens tonfall vid brottstillfället samt effekten av den tid som hinner passera mellan bevittnandet av brottet och röstkonfrontationen. Det är rimligt att anta att en gärningsman ofta talar i ett argt tonfall vid brottstillfället och i ett normalt tonfall vid en eventuell konfrontation. Det är därmed viktigt att veta hur detta kan påverka igenkänningsförmågan hos vittnet. I verkliga fall går det alltid en tid mellan brottstillfället och en möjlig röstkonfrontation. Kunskap om tidslängdens betydelse är därför av stor vikt. Vidare, att öronvittnen är relativt svaga på att minnas okända röster kan till viss del vara ett resultat av dålig inkodning av rösten. Därför jämfördes två olika röstbeskrivningsintervjuer som syftade till att förstärka minnet av rösten. Deltagarna, 11–13-åringar (*N* = 160) samt vuxna (*N* = 148), hörde antingen gärningsmannen tala i ett argt tonfall vid exponeringen och i ett normalt tonfall i konfrontationen (inkongruent), eller i ett normalt tonfall vid båda tillfällena (kongruent). En annan avvikelse från grundupplägget var att alla deltagare, kort efter exponeringen, blev intervjuade om rösten, antingen genom globala öppna frågor om rösten eller genom att skatta olika röstegenskaper på en skala. Båda intervjuerna avslutades med frågan om deltagarna trodde att de skulle kunna känna igen rösten om de fick chansen att höra den igen. Hälften av deltagarna tog del av konfrontationen kort efter den första intervju, medan resterande återkom två veckor senare. Totalt gjorde endast 13% av barnen och 10% av de vuxna ett korrekt utpekande. Varken gärningsmannens tonfall eller intervjumetod visade sig ha en signifikant effekt på korrekthetsnivån. Däremot visade resultaten att de barn som tog del av röstkonfrontationen direkt presterade signifikant bättre (21% korrekt) jämfört med de barn som tog del av röstkonfrontationen efter två veckor (9% korrekt). Mest överraskande var den dåliga prestationen av de som testades under de mest fördelaktiga förhållandena, dvs kongruent tonfall och tog del av konfrontationen direkt. Av dessa gjorde endast 25% av barnen och 19% av de vuxna ett korrekt utpekande. Vidare, majoriteten barn (86%) och vuxna (63%) trodde att de skulle kunna känna igen gärningsmannens röst vid ett senare tillfälle, dock var det endast 13% av dessa barn och 4% av dessa vuxna som faktiskt gjorde ett korrekt utpekande. De fria beskrivningarna av rösten var få och generella och utgjordes till stor del av situationsspecifika beskrivningar (t ex. stressad, arg, nervös) som inte har med själva röstens karaktär att göra.

Sammantaget visar studierna i doktorsavhandlingen att både barn och vuxna presterar dåligt när de ställs inför uppgiften att identifiera en okänd röst som de har hört under realistiska förhållanden. Avhandlingen visar också att 11–13-åringar presterar på samma nivå, eller i vissa fall bättre, än vuxna. Det innebär att *om* rättsväsendet är beredd att använda sig av röstkonfrontationer för att testa en hypotes i utredningsfasen och/eller som bevis i rätten, så bör det gälla även om vittnet tillhör åldersgruppen 11–13 år. Trots den dåliga prestationen verkar dock vittnen vara av den uppfattning att de presterar bättre än vad de i själva verket gör. Sådan överkonfidens kan vara ett problem då det kan missleda rättsliga aktörer såväl i utredningsfasen som i domstolen. Vidare, det intervjuformulär som används i nuläget av Svenska Säkerhetspolisen visade sig varken vara fördelaktigt för röstigenkänning eller för vad gärningsmannen sade. Således, efter att ha testat 949 personer under ett antal olika betingelser är avhandlingens slutsats tydlig: aktörer i rättsväsendet bör behandla öronvittnens utsagor med stor försiktighet. För att öronvittnen skall vara användbara måste metoder som förbättrar öronvittnens röstidentifieringar, minne för vad gärningsmannen sa, samt röstbeskrivningar utvecklas.

Acknowledgements

I would like to express my sincere gratitude to:

My supervisor, Professor *Pär Anders Granhag,* and my second supervisor, Professor *Anders Eriksson,* both for their patience with me and for making me believe that I can do this. I feel privileged to have had the chance to work with them and to have benefited from their great knowledge. I also want to thank them for all of the inspiring discussions.

My collaborators and friends in the research unit for Criminal, Legal and Investigative Psychology (CLIP): *Erik Adolfsson*, *Helen Alfredsson*, Associate professor *Karl Ask*, *Sebastian Cancino Montecino, Franziska Clemens*, *Ivar Fahsing, Angelica Hagsand, Malin Karlén, Melanie Knieps*, Dr *Sara Landström*, *Linda Lindén, Eric Mac Giolla, Simon Moberg Oleszkiewicz, Anna Rebelius*, Dr *Emma Roos af Hjelmsäter* (my morning buddy), *Tuule Sooniste,* Associate Professor *Leif Strömwall* (my statistical guru), *Sara Svedlund, Petra Valej*, *Rebecca Willén, Ann Witte* and *Olof Wrede*. Franziska my friend, you are one of a kind and I am glad to have shared this journey with you!

I wish to thank all my colleagues and friends at the Department of Psychology for creating a nice working atmosphere and for all of the laughter we have shared together. Special thanks to *Linnéa Almqvist* for becoming such a good friend.

Many thanks to all of the participants and to all of you who have helped me with the thesis, including all kinds of practical details, just to mention some; *Ann Backlund*, *Linda Lindén,* Professor *Björn Lyxell*, *Elaine Mc Hugh*, and *Karin Sjöö Åkeblom*.

I also want to thank my friends outside of academia who support and encourage me when it is most required. Special thanks to *Frida Johansson* and *Karin Sjöö Åkeblom* for patiently listening to my ups and downs.

Dear parents and sisters, thank you for believing in me and encouraging me. You are simply the best!

Last but not least, *Per* and *Albin* “my boys”, thank you for all the love and support, and for showing me what is important in life. Because of you, I never forget that there is a world outside of research.

This research has been financially supported by the Crime Victim Compensation and Support Authority.

Lisa Öhman
Gothenburg, March 2013

List of Publications

This thesis consists of a summary and the following four papers, which are referred to by their roman numerals:

I. Öhman, L., Eriksson, A., & Granhag, P.A. (2011). Overhearing the planning of a crime: Do adults outperform children as earwitnesses? *Journal of Police and Criminal Psychology, 26,* 118–127. doi: 10.1007/s11896-010-9076-5

II. Öhman, L., Eriksson, A., & Granhag, P.A. (2010). Mobile phone quality vs. direct quality: How the presentation format affects earwitness identification accuracy. *The European Journal of Psychology Applied to Legal Context, 2,* 161–182.

III. Öhman, L., Eriksson, A., & Granhag, P.A. (Available online: 27 Feb 2012). Enhancing adults' and children's earwitness memory: Examining three types of interviews. *Psychiatry, Psychology and Law*. doi: 10.1080/13218719.2012.658205

IV. Öhman, L., Eriksson, A., & Granhag, P.A. (2013). Angry voices from the past and present: Effects on adults' and children's earwitness memory. *Journal of Investigative Psychology and Offender Profiling, 10,* 57–70. doi:10.1002/jip.1381

Table of Contents

Introduction

On January 17, 2005, Fabian Bengtsson, the Chief Executive of one of the major Swedish electronics companies, SIBA, was kidnapped under gun threat by two men who kept him for seventeen days in a narrow wooden case. The purpose: to blackmail his family. Fabian Bengtsson never saw his kidnappers, however he could hear them speak and he also made other auditory observations. The kidnapping received full media attention and fortunately ended well. After his release, Fabian Bengtsson's thorough observations of, for example, what time the delivery car of a well-known ice cream company (playing a characteristic tune to attract customers' attention) passed by outside, enabled the police to find the apartment where he had been held. In the apartment they found one of the kidnappers as well as evidence of the crime, such as DNA traces which indicated that Fabian Bengtsson had been in the apartment. Although it took much effort to close this case and to convict the perpetrators, had it not been for Fabian's auditory observations the kidnappers might not have been identified. Other examples of cases in which witnesses' auditory observations have played a key role are the classic "Charles Lindbergh case" (1935) and the more recent high-profile cases of "Amanda Knox" (2007) and "Trayvon Martin" (2012).

Observations made by victims and witnesses are the most frequent and often the most important evidence in criminal cases (Kebbell & Milne, 1998). In most cases the victim has seen the perpetrator. In other cases, however, the voice or other auditory information can be an important clue. An earwitness is a witness or a victim who has heard, but not seen, the perpetrator for different reasons. Fortunately, kidnapping cases are rare (at least in Sweden), but there are a number of other more common situations where the perpetrator is only heard. Examples of such situations are crimes committed under conditions of darkness or by disguised perpetrators, such as hooded rape or robbery. There are also cases where the victim has been blindfolded. Yet another category is crimes committed over the phone, such as obscene phone calls, frauds, ransom demands and other threatening calls.

In a case like the mentioned kidnapping of Fabian Bengtsson, the first impression might be that one would definitely remember (and never be able to forget) the voice of the kidnapper. People's experiences of easily recognizing familiar voices, such as the voices of relatives, friends, politicians and actors have created the notion that voice recognition is often very accurate (Hammersley & Read, 1996; Yarmey, 1995). However, empirical studies have shown mixed results concerning the recognition of familiar voices (Bartholomeus, 1973; Hollien, Majewski, & Doherty, 1982; Read & Craik, 1995) and it has even been found that we are not always able to identify the voices of our own family members (McClelland, 2008). There seems, unfortunately, to be limited awareness about the reliability of earwitnesses' voice identification performance in the judicial system (Solan & Tiersma, 2003). Further, research has shown that potential jurors (undergraduate students) hold inaccurate beliefs about people's ability to correctly identify familiar voices (Yarmey, Yarmey, Yarmey, & Parliament, 2001), as well as unfamiliar voices (van Wallendael, Surace, Hall-Parsons, & Brown, 1994; Yarmey, 1995).

The first documented case of voice identification used in court of law dates back to as early as 1660 (Deffenbacher et al., 1989). Voice identification is still treated as direct evidence of identity in modern law enforcement (Stern, Mullennix, Corneille, & Huart, 2007) and occurs all over the world (Hollien, 2012). Nevertheless, victims' and witnesses' memory for voices is, compared to eyewitness identification, a neglected research area (Wilding, Cook, & Davis, 2000). As it is shown that earwitness identification does not mirror eyewitness identification (Hollien, 2002; Hollien, Bennett, & Gelfer, 1983), it is important to conduct research within this area. The present thesis has a psycho-legal approach with a special focus on ecological validity. The general aim of the present thesis is to explore earwitnesses' identification performance for an unfamiliar voice heard under conditions that bear a reasonable resemblance to a real life criminal situation. Further, an earwitness is often initially required to create a voice and speech profile of the perpetrator (Broeders & Rietveld, 1995). Therefore, an additional aim is to investigate how good witnesses are at describing voices.

Another important aspect is earwitness memory for what the perpetrator said. Imagine that a witness overhears a terrorist talking to an accomplice about the planning of an attack. It would be of great value if the witness could accurately remember and report to the police what was said. In spite of being a very important topic, witnesses' memory for criminal conversations is a much understudied area (Davis & Friedman, 2006). Therefore, a further aim with the present thesis is to investigate earwitnesses' memory for the content of a perpetrator's account.

The thesis is organized as follows: First, I define some basic earwitness terminology and explain a few acoustic features. Second, I introduce the three main domains of study concerning earwitness testimony, followed by a review of basic memory processes and voice memory. In the following three sections, I provide a general overview of previous empirical work within each of the three areas of special interest in the present thesis, namely, memory for content, voice descriptions, and voice identification. The last section ends with an evaluation of past research. Finally, I summarize the empirical studies of the present thesis and conclude with a general discussion of the main results, as well as some legal implications and directions for future research.

Defining Earwitness Terminology

As a start, it may be of use to define some terms that are commonly used within psycho-legal earwitness research. The definition of *Earwitness identification evidence* chosen for this thesis is provided by Yarmey (1995):

> The process of a witness hearing the voice of a target person or persons, retaining that information in memory, retrieving that information later when called to identify the suspect(s) either in a 1-person voice lineup or a many-person voice lineup, and finally, testifying or communicating this decision to a police investigator, trial judge, or jury (p. 795).

Voice lineup refers to when a witness is presented with a number of voices (usually five to eight) in an attempt to identify an earlier heard voice. Basically, there are two types of lineups; *target-present* and *target-absent* lineups. As the name reveals, a target-present lineup is a lineup in which the target voice (perpetrator's voice) is present, conversely in a target-absent lineup the target (perpetrator's) voice is not present. However, it should be noted that it is only in controlled experiments that it is possible to distinguish between these two types of lineups. In a real investigation it is not known if the suspect is the perpetrator. The other persons in the lineup (who are known not be the perpetrator) are called *foils*.

In a target-present lineup the witness can make four types of responses. The witness can correctly identify the target (*correct identification*), select a foil (*false identification*), report that the perpetrator's voice is not present *(false rejection)*, or respond "I don't know". In a target-absent lineup there are three possible outcomes; the target is reported to not be present (*correct rejection)*, the selection of a foil (*false identification),* or an "I don't know" response.

There is always some time delay between the initial exposure to the target and a possible subsequent voice lineup. In the literature, this delay is often referred to as a *retention interval* (or *time delay).* Further, the amount of time that the witness is exposed to the perpetrator's voice is commonly called *duration.*

Acoustic Features of a Voice

Since the present thesis has a focus on voices, a few basic voice features need to be explained. Three acoustic cues that have been suggested as important in voice similarity judgements are *articulation rate, pitch variation* and *pitch level* (Petrini & Tagliapietra, 2008). The definitions of these cues are: *articulation rate* is the speaking rate excluding pauses expressed in syllables per second; *pitch variation* is the standard deviation of the fundamental frequency divided by the mean; and *pitch level* is the fundamental frequency base line (see Lindh & Eriksson, 2007). These cues will be further explained below.

Articulation rate is a way of quantifying how fast a speaker is talking between pauses. Produced rate of speech is, although not perfectly, correlated with perceived rate of speech, and therefore a speaker with a high articulation rate is also perceived as talking fast.

Voice pitch is dependent on the vibration of the vocal cords; the higher the frequency of the vibration, the higher the pitch. For example, men have longer and thicker vocal cords than women, which results in lower rates of vibration compared to women. Pitch can be described by two measurements, namely, level *(pitch level)* and variation *(pitch variation)*. Pitch level is often described by the mean, although the base line actually provides a better description. The base line of vocal cord vibration can be seen as a relaxed position, the frequency to which a speaker continuously returns when modulating their voice (e.g., for the intonation of words and phrase structure, such as marking the end of a phrase) (Traunmüller & Eriksson, 1995). This base line is relatively stable for a given individual at normal vocal effort levels. When engaging in a conversation, people seldom talk monotonously. Instead, people are most likely to modulate their voice; pronouncing some things vividly or intensely and others more calmly. This implies that the mean of the frequency can vary markedly in a conversation, whereas the base line stays approximately the same.

Pitch level and *pitch variation* translates to the perception of speech in that people with a low pitch level are often perceived as having a deep voice (and vice versa), and speakers with a great variation in pitch are often perceived as talking in a lively manner (and vice versa).

Differences between speakers (inter-speaker variability), as well as differences within the same voice on different occasions (intra-speaker variability) might affect how well a voice is remembered. Biological differences such as uniqueness and voice quality are examples of inter-speaker variability. Differences within the same voice at different occasions can, for example, be a result of the situation, emotional state, intention, and health status. Depending on the situation the voice may be altered and thereby affect, for example, the articulation rate, pitch level and pitch variation.

Referring to earwitness voice identification, Yarmey (2007) pointed out that we have to assume that earwitnesses will base their decision more on inter-speaker variability (foil vs. suspect) than intra-speaker variability (the suspect's voice on different occasions). Unfortunately, research indicates that intra-speaker variability plays an important role for witnesses' decisions in a voice lineup situation (see below the section on the effect of tone of voice).

Although focusing on voices, the present thesis has a legal-psychology approach. Therefore, it is beyond the scope of this thesis to provide a more detailed description of the "voice".

Earwitness Testimony: Three Main Domains of Study

Testimonies by victims and witnesses play a significant role in criminal investigations (Kebbell & Milne, 1998); the type of information that witnesses might contribute with can be categorized into three main domains of study (see Figure 1). Information gathering is an important part of a criminal investigation, and the aim here is to elicit as much accurate and detailed information about the crime as possible. Therefore, an investigation often starts with the police interviewing the victim and the witnesses about their observations (event recall). The police might further ask the witnesses to describe the perpetrator's appearance (e.g., face description). If there is a suspect in the case, the witness might be confronted with a lineup in an attempt to identify the suspect (face recognition).

These three main domains of study are also applicable to earwitness testimony. First, the memory of what the perpetrator said is one important domain (content recall). The witness may have spoken directly to the perpetrator or merely overheard critical information. The police would most likely start by gathering information about the content of the conversation. A second source of information is the memory of the voice per se (voice description). In order to narrow down the number of suspects, a witness may be asked to describe the *perpetrator* based on information obtained from the voice (e.g., sex, age, dialect, accent). In addition, the witness might be asked to describe the perpetrator's *voice* (e.g., speech rate, pitch level). Thirdly, if there is a suspect, a voice lineup may be conducted (voice recognition).

	Recall		Recognition
Eyewitness testimony	Event recall →	Face description →	Face lineup (Face recognition)
Earwitness testimony	Content recall →	Voice description →	Voice lineup (Voice recognition)

Figure 1. Three main domains of study of a witness testimony.

All three domains may be important in criminal investigations and each domain has attracted research attention, although in varying degrees. In this section I will therefore discuss previous findings with respect to; (a) memory for content, (b) voice description and (c) voice recognition. However, first I will introduce three basic memory processes that are important when discussing earwitnesses testimony and briefly discuss voice memory from a theoretical perspective.

Basic Memory Processes

Our memory is characterized by three main processes; *encoding, storage* and *retrieval* (e.g., Reisberg, 2010). *Encoding* is the process that takes place when new information is acquired, and the information that surrounds us needs to be converted into a form that can be stored. For example, visual coding is used when we are forming memories of people's faces, whereas memories for auditory information are encoded acoustically (Nevid, 2003). However, mere exposure to a stimulus will not result in a high quality memory, therefore attention plays a critical role. Attending to a stimulus during the encoding process enhances future memory of that stimulus (Mulligan & Brown, 2003). Further, elaborative rehearsal and deep processing might be needed to effectively encode information into long-term memory (Reisberg, 2010). *Storage* refers to the process of retaining the encoded information in memory, and furthermore, *retrieval* is the process of accessing stored information at a later occasion. However, these three processes should not be viewed as separate stages. For example, previously stored knowledge affects how well we encode new information. Further, how the information is stored affects the retrieval process. That is, information needs to be appropriately indexed and organized to enable retrieval in future situations (Reisberg, 2010). Hence, it is evident that these basic memory processes are intertwined and they are found to be important for all domains of study within witness testimony. For each process there are a number of factors that might affect the quality of the memory. Further, these different factors are (more or less) applicable to each of the three main domains of study within earwitness testimony (i.e., content recall, voice description and recognition). All factors that are mentioned below will be discussed in more depth in later sections of this thesis.

Encoding. How well an earwitness will encode the voice and the content of a conversation may depend on, for example, the age of the witness, the duration of the conversation, in what tone the perpetrator spoke and further, whether the voice was heard live or via a mobile phone. Other relevant aspects are to what extent the witness was prepared to memorize the voice and what was said, and if the witness was both seeing and hearing the perpetrator.

Storage. After the encoding some time might elapse before the witness will be questioned about the event. Meanwhile, the information needs to be retained in memory. One obvious factor that might affect how well the voice and content are retained in memory is the length of the retention interval.

Retrieval. At the retrieval phase, factors like type of interview technique and lineup procedure may play an important role for memory performance. Further, if the initial voice is heard via a mobile phone, the retrieval process may be enhanced if the voices in the lineup are recorded via a mobile phone.

Memory Models of Voices

A well-known memory model that is important for remembering is the multi-component system model often termed as "working memory" (e.g., Baddeley, 1990). The working memory is characterized by an attention-controlling central executive and three sub-systems; the visuo-spatial sketchpad, the phonological loop, and a more recently added episodic buffer (e.g., Baddeley, 2012). The system of most interest for the processing of voices is the phonological loop, as it is the subsystem that processes and encodes auditory information. The phonological loop comprises of two main components; a passive phonological store for memory traces and an articulatory rehearsal component where the memory trace needs to be rehearsed, otherwise the trace will decay (e.g., Baddeley, 2000). The loop can be illustrated as the inner voice that we, for example, can use to repeat items in our head that we need to remember, such as a phone number. Besides its capacity for remembering digits and unrelated words for a short period of time, it has been questioned why the loop should be a feature of human cognition (e.g., Baddeley, Gathercole, & Papagno, 1998). Thus, research has shown that the primary function of the phonological loop is to temporarily store new words while more stable long-term phonological representations are being constructed (Baddeley et al., 1998; Baddeley, Papagano, & Vallar, 1988). That is, the phonological loop is a system for supporting language learning. It has been acknowledged that not much is known with respect to whether or not the same system is used for non-linguistic auditory information, such as environmental sounds and music (Baddeley, 2012). As the present thesis focuses on voices, it is necessary to acknowledge the seemingly neglected relation between the phonological loop and voice processing. A picture or a face can be retained and rehearsed in the visual sketchpad and a number, word, or a sentence can be retained by repeating it auditorily in the phonological loop. But what about a voice, where is the voice rehearsed and how is it consolidated into long-term memory? The role of long-term memory for remembering voices is as central as the working memory. Unfortunately, there is relatively little knowledge about how listeners perceive and remember unfamiliar voices (Kreiman & Papcun, 1991). One memory model suggests that voices are remembered in terms of a "prototype" – an average voice – and a set of deviations from that prototype (Kreiman & Papcun, 1991; Papcun, Kreiman, & Davis, 1989). Though, the deviations are found to be forgotten as time passes. The prototype model explains the difference between familiar and unfamiliar voices by suggesting that the recognition of unfamiliar voices relies on the prototype plus deviations. Conversely, when it comes to a familiar voice, people learn the specific features of that particular voice and therefore no longer use the prototype; instead they only use features that deviate from the prototype (Papcun et al., 1989). The stronger the deviations are, the easier the voice is to identify (Lavner, Rosenhouse, & Gath, 2001).

In line with this, studies using functional neuroimaging have shown that different brain regions are found to be activated when processing familiar and non-familiar voices. Voice recognition activates both the posterior and the anterior Superior Temporal Sulcus (STS) with a right hemispheric dominance (e.g., Belin & Zatorre, 2003; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003; von Kriegstein & Giraud, 2004). However, in contrast to recognizing familiar voices, when processing non-familiar voices a higher activation is found in the right posterior STS (von Kriegstein & Giraud, 2004). Further, the recognition of non-familiar voices shows a bilateral activation and involves more areas in the brain compared to recognition of familiar voices, which is suggested to be related to the difficulty of recognizing non-familiar voices (von Kriegstein & Giraud, 2004). An important question in relation to this is when an unfamiliar voice becomes familiar.

The aim of this section was not to give an extensive review of memory models and the neurological structure that underlies voice processing. Instead, my intention was to highlight the fact that the human brain contains regions that are strongly selective to voices and that different processes underlie the recognition of familiar and non-familiar voices.

Memory for Content

Many civil and criminal cases involve testimony regarding statements or content of specific conversations. Furthermore, there are “language crimes” (e.g., verbal sexual harassment, fraud) where the witness’s memory of a conversation is the only available evidence (Campos & Alonso-Quecuty, 2006). Nonetheless, this area has been largely neglected by psycho-legal research (Davis & Friedman, 2006).

There are many aspects of oral communication that can be of legal relevance, such as *who* said it, to *whom* it was said, *when* it was said, the *sequence* of communication etc. I do not intend to cover the full range of these (for a review see Davis & Friedman, 2006), instead I will focus on the aspect most relevant for the scope of this thesis, namely *what was said*.

People’s memory for content has been tested both by recall tests (e.g., Miller, deWinstanley, & Carey, 1996; Stafford & Daly, 1984) and recognition tests (e.g., Bates, Masling, & Kintsch, 1978; MacWhinney, Keenan, &

Reinke, 1982). Research indicates that people’s recognition memory is better (e.g., MacWhinney et al., 1982) than their recall memory (e.g., Stafford & Daly, 1984). In a forensic context there is often no knowledge of what was actually said and the outcome of free recall is therefore of most relevance for the present thesis. A free recall can consist of two types of memory traces, namely *gist* memory and *verbatim* memory. Gist memory refers to the kernel of the meaning, that is, the content of the to-be-remembered without specific details. The verbatim memory refers to a more detailed memory, such as the memory of actual wordings and syntactic form. The two types of memory traces are suggested to be independent, that is, both types are encoded in parallel (Brainerd & Reyna, 1993).

Research within this area has mainly focused on memory for mundane conversations (e.g., MacWhinney et al., 1982; Stafford & Daly, 1984). Though, it has been found that *what* is said may affect how well it is remembered. As an example, adult participants who heard a recorded conversation between a man and a woman, recalled sexual content better than neutral content (Pezdek & Prull, 1993). After a five week delay, the meaning of sexual utterances was better recalled than neutral utterances, however, the verbatim memory for both types of utterances was rather poor. Further, in a case study, children’s (age 8–16) memory of a self-experienced obscene phone call was examined (Leander, Granhag, & Christianson, 2005). It was found that the children were quite accurate in their reports, however, they omitted almost all of the sexual and sensitive information. The fact that they remembered more of the neutral information indicated that they probably also remembered the sexual information, although they chose not to report it. A possible reason, suggested to explain the finding, was that the children experienced shame and embarrassment. This finding of omitting information is noteworthy and an important aspect that needs to be considered when interviewing victims of a crime where what was said is crucial. Conversations with criminal content often contain attention attracting details not present in other types of conversations, for example, like the previously mentioned sexual accounts, brutal violence, threats etc. Hence, it might not be possible to generalise findings about memory for everyday conversations to memory for conversations with criminal content.

Though, in line with research on memory for everyday conversations (e.g., Miller et al., 1996), research on memory for criminal conversations using free recall as a memory test has shown that witnesses’ statements contain mostly gist memory and that verbatim memory is very poor (Campos & Alonso-Quecuty, 2006; Neisser, 1981; Pezdek & Prull, 1993). Further, memory is found to decay with time, both for criminal (e.g., Campos & Alonso-Quecuty, 2006; Yarmey, 1992) and non-criminal content (e.g., Stafford, Burggraf, & Sharkey, 1987). In addition, verbatim memory is found to decline faster than gist memory (Reyna & Brainerd, 1995). Rehearsal has been found to be beneficial. Recall accuracy for what a perpetrator said, tested after a week delay, was higher for participants who rehearsed (vs. those who did not rehearse) by freely recalling everything that they remembered very shortly after the event (Boydell & Read, 2011). In line with eyewitnesses (e.g., Ibabe & Sporer, 2004), it has been found that adults remember more central than peripheral details from a perpetrator's account (Boydell & Read, 2011). The same pattern is found for children, though tested on non-criminal content (Gibbons, Anderson, Smith, Field, & Fischer, 1986). Although earwitnesses' recall of a criminal account (both seen and heard) can be rather accurate, confidence is suggested not to be a reliable predictor of accuracy (Boydell & Read, 2011). Though, an earwitness's level of confidence has been found to be more reliable when reported shortly after the event compared to when stated at trial.

A stable finding in eyewitness research is that children spontaneously recall less information than adults when asked to describe an event (e.g., Cole & Loftus, 1987). The same type of age-related difference has also been found for memory of content. That is, children's recall of the content of a criminal conversation has been found to be less detailed than that of adults' (Ling & Coombe, 2005; Saywitz, 1987). Although the overall recall for a novel conversation was rather poor, children aged 11–16 performed even more poorly compared to the adults (Ling & Coombe, 2005). Further, young children (age 8–9) were found to remember significantly less in their free recall of a heard mock-crime compared to older children (age 11–12 and 14–16) (Saywitz, 1987). In brief, these studies suggest that children's recall of the content of a heard criminal conversation is less detailed than that of adults'.

Another aspect to consider is if the witness only heard the perpetrator speak or both saw and heard the perpetrator. Research has shown that participants in an auditory-only condition report less correct information (Campos & Alonso-Quecuty, 2006) and show a greater decrement in memory performance after a delay than participants in an audio-visual condition (Campos & Alonso-Quecuty, 2006; Toglia, Shlechter, & Chevalier, 1992). In line with adults, young children's (4–7-years old) memory for a story was found to be poorer in an auditory-only condition compared to an audio-visual condition (Gibbons et al., 1986; Ricci & Beal, 2002).

It should be acknowledged that earwitness research has focused mainly on verbal stimuli. However, earwitnesses may also pick up on important nonverbal auditory information. For example, in the previously mentioned kidnapping case, Fabian Bengtsson's observation about what exact time the ice cream car passed outside were essential to the investigation. Other examples of such nonverbal stimuli are the number of gun shots and what direction a particular sound came from. Memory for the number of gunshots (nonverbal auditory stimuli) heard in a criminal context (mixed with other modality stimuli) has been found be less well remembered than verbal and visual stimuli (Experiment 2, Huss & Weaver, 1996). Further, in car-accidents, a witness's estimation of vehicle speed and direction may be important. It has been found that children (5, 8, & 11 years) are poor at identifying vehicle sounds, but that this ability increases with age (Pfeffer & Barnecutt, 1996). When comparing auditory, visual, and audio-visual estimations of traffic speed, it has been found that adults in an auditory mode tend to make more errors compared to the other two conditions (Barnecutt, Pfeffer, & Creswell, 1999).

To sum up, memory for criminal content and other non-verbal stimuli that might have legal relevance has been investigated to some extent, but not nearly as much as voice recognition. This is noteworthy since, in real-life investigations, it is much less common that witnesses are asked to identify a previously heard voice. Witness statements about what was said are far more frequently used (Davis & Friedman, 2006).

Voice Description

Yet another important aspect of earwitness testimony is the description of the perpetrator's voice. Voice descriptions may serve at least two purposes. First, accurate and detailed descriptions may allow the police to narrow their search for potential suspects. Secondly, it has been suggested that the selection of foils for lineups should be based on the witness's description of the perpetrator (Wells et al., 2000). However, voice descriptions are usually too limited to provide adequate information needed for the selection of foils (Broeders & Rietveld, 1995). Nevertheless, the quality of earwitnesses' voice descriptions is a neglected research area.

For speaker profiling, it would be helpful if earwitnesses could accurately estimate person characteristics such as sex, age, height and weight based solely on the voice. Though, is that possible? It is suggested, from an evolutionary perspective, that humans distinguish voices according to gender. After knowing that someone is a person, to determine the person's gender has been of utmost importance due to reproduction (Nass & Gong, 2000). Therefore, it is not that surprising that listeners are found to be skilled at determining the sex of adult speakers (e.g., Cerrato, Falcone, & Paoloni, 2000), even from non-verbal sounds such as a cough or laughter (Eriksson, 2008). Studies examining listeners' judgments of speakers' height and weight have shown mixed results. Some have found that listeners are able to accurately estimate such characteristics (e.g., Krauss, Freyberg, & Morsella, 2002; Lass, Barry, Reed, Walsh, & Amuso, 1979), whereas others have found this to hold true only for male speakers (van Dommelen & Moxness, 1995). Other studies, however, have found no significant correlation between actual and estimated weight and height (e.g., Yarmey, 1992) and some of the earlier studies (e.g., Lass et al., 1979) have been criticised for only using overall means which might overstate the estimation accuracy. A re-analysis of those studies (using a different method) showed, in contrast, that listeners are not skilled at estimating speakers' weight and height based solely on their voice (Gonzalez, 2003). The pitch of a voice depends primarily on the length of the vocal folds and the timbre of the voice on the shape and size of the vocal tract. The acoustic measures correlated with pitch and timbre have repeatedly been shown not to correlate with body size in any useful way, and in a recent study by Hatano et al. (2012) the physical size of the vocal tract was shown not to correlate with body height. Poor estimations based solely on the voice may therefore not be that surprising. As for age estimations, the same mixed pattern is found; while some studies show that listeners can reliably estimate the age of a speaker (e.g., Krauss et al., 2002), others do not (e.g., Yarmey, 1992). It is suggested that for forensic situations, age should only be broadly classified as the speaker being "young", "adult", or "old" (Cerrato et al., 2000).

The few studies focusing on describing the *voice*, which is the focus of the current thesis, have found that voices are hard to describe. The numbers of described dimensions are often few and most of them general and non-specific, such as the sex of the speaker (note that this is rather a *person* description) and pitch (Yarmey, 2001, 2003). Further, it is reasonable to assume that witnesses who are better at describing the voice should also be better at recognizing it. Contrary to intuition, the quality or number of reported voice descriptions has not been found to have a significant association with identification accuracy (Yarmey, 2001).

A possible solution to the problem with insufficient descriptions is to ask the witnesses to rate different voice characteristics on a scale. When using such a method, witnesses are prompted to think about features that otherwise might be omitted or not thought of. Studies using this method have shown that ratings of distinctive voices are more reliable over time than ratings of non-distinctive voices (Yarmey, 1991a), and a discussion between two witnesses does not seem to influence rated descriptions of the perpetrator's voice (Yarmey, 1992). These studies have focused on the mean ratings of various voice characteristics. Another interesting aspect is the level of agreement between witnesses' rated descriptions. Such knowledge is highly relevant in cases where there are several witnesses. Unfortunately, this aspect has received little attention.

In sum, vague descriptions of a voice may have negative consequences for the construction of a lineup, and as a result the composition of the lineup might need to be based on information other than the description made by the earwitness. Further, not much is known about the agreement between different witnesses' rated descriptions.

Voice Recognition

Identification of a suspect is often considered as strong evidence in court. However, eyewitness research and the introduction of evidence, such as DNA, have shown that mistaken eyewitness identification is the largest single contributing factor to the conviction of innocent people (Wells et al., 2000; Wells & Olson, 2003). Therefore it is not surprising that much eyewitness research has been devoted to face recognition. The same pattern is found in earwitness research, where most research has been on the identification of a voice. Hence, some of the more important factors affecting voice recognition performance will be discussed below.

Research on Voice Recognition: An Overview

As mentioned, studies have shown that recognizing familiar and unfamiliar voices are two independent abilities (e.g., von Kriegstein & Giraud, 2004). This implicates that findings within one area cannot be generalized to the other. Hence, it needs to be clarified that the focus of the present thesis is on unfamiliar voices. Though, the definition of an unfamiliar voice may not be that straightforward as there might be different levels of (un)familiarity. A voice heard a couple of times for a short amount of time, like a neighbour, might be judged as more familiar than a once-heard voice, but less familiar than the voice of a family member. A lineup is not recommended if the witness claims that the voice of a perpetrator belongs to a highly familiar person (Broeders & Rietveld, 1995), whereas a voice lineup could be used if, for example, the neighbour is accused. Though, the definition of an unfamiliar voice used in the present thesis is a voice heard only on one occasion.

There are numerous variables that might affect recognition of an unfamiliar voice, and in the following section I will give an overview of some of these variables. I will end this section with a brief discussion of voice identification in Sweden.

Earwitness as well as Eyewitness

In many situations, the witness both sees and hears the perpetrator. Though, few scholars have investigated how the two modalities might interact with and affect each other. The first study to compare the ability of subjects to make accurate auditory and visual identifications from the same event, found that visual identifications were far more accurate than auditory identifications (Hollien et al., 1983). In another early study, where the effect of both hearing and seeing the perpetrator was tested, greater attention to the voice was expected as the lightning deteriorated (Yarmey, 1986). However, the results contradicted the researcher's expectation. Subjects in four different illumination conditions did not differ in terms of their voice identification accuracy. More recent studies have found a face overshadowing effect. To both see and hear the perpetrator has been found to impair the processing of the voice and result in lower voice identification accuracy, compared to only hearing the perpetrator (Cook & Wilding, 1997b, 2001; McAllister, Dale, Bregman, McCabe, & Cotton, 1993; Stevenage, Howland, & Tippelt, 2011, though see Armstrong & McKelvie, 1996; Legge, Grosmann, & Pieper, 1984, for an opposite result when using a two-alternative forced-choice recognition test). It is suggested that when the face of a perpetrator is exposed, the attention to the voice is primarily focused on emotions and what is being said, rather than information useful for voice recognition (Yarmey, 2007). In contrast, hearing the perpetrator has not been found to impair face identification (Stevenage et al., 2011). Rather, a bimodal lineup (when both hearing and seeing the perpetrator) has been found to result in a higher number of correct face identifications compared to a visual lineup only (Melara, DeWitt-Rickards, & O'Brien, 1989).

Recent research has shown that both hearing and seeing the perpetrator can affect other tasks than identification, such as person descriptions and memory for conversation. Although not affecting photo lineup accuracy, poorer descriptions of the perpetrator's physical appearance and poorer memory for the perpetrator's message have been found when the perpetrator is speaking with a foreign-accent compared to no accent (Pickel & Staller, 2011). Further, the presence of a weapon (visual information) has been found to worsen memory for the perpetrator's statements without affecting voice descriptions and voice identification (Pickel, French, & Betts, 2003). While the former study shows that auditory information about a perpetrator can have a negative effect on visual memory, the latter shows that visual information can impair auditory memory.

The general conclusion to be drawn from this review is that auditory and visual information can interfere with each other. This in turn shows the importance of clearly distinguishing between situations where the earwitness is presented with both visual and auditory information and auditory only. It implicates that research findings cannot be generalized between the different situations. As the present thesis concerns earwitnesses in the absence of visual information, I will hereafter exclusively focus on situations where the perpetrator is only heard.

Exposure Time

One factor that has attracted much attention is the effect of speech duration. How long the witness is exposed to the voice is a factor that is likely to affect voice identification accuracy and it is suggested that the longer the exposure, the better the identification performance (e.g., Legge et al., 1984; Orchard & Yarmey, 1995; Yarmey & Matthys, 1992). Though, research has shown that it is possible to recognize a *familiar voice* from a vowel segment of 25 ms in duration (Compton, 1963). For unfamiliar voices, there seems to be a tendency for longer durations to produce more hits in the target-present lineups, whereas the result for the target-absent condition is mixed. While some studies have found that the advantage of longer duration was partly counteracted by high degrees of false alarms (Yarmey, 1991b; Yarmey & Matthys, 1992), other studies have not shown an increased number of false identifications (Kerstholt, Jansen, van Amelsvoort, & Broeders, 2004; Orchard & Yarmey, 1995).

There may, however, not be a very straightforward relationship between exposure time and accuracy. The number of heard vowel sounds has, for example, been found to moderate this relationship, at least for relatively short utterances (Pollack, Pickett, & Sumby, 1954; Roebuck & Wilding, 1993). To be exposed to a larger repertoire of the perpetrator's voice has been found to result in higher identification accuracy, whereas increased sentence length as such did not have an effect on the performance (Pollack et al., 1954; Roebuck & Wilding, 1993). Though, Cook and Wilding (1997a) replicated the Roebuck and Wilding study (both with an immediate test and with a one week delay) and found the opposite pattern; that the length rather than vowel variety had a positive effect on identification accuracy.

Furthermore, there are studies showing that the identification accuracy is superior when hearing the same voice at two or three occasions compared to hearing the voice for the same length of time for one massed trial (Goldstein & Chance, 1985 in Deffenbacher et al., 1989; Yarmey & Matthys, 1992, though see Procter & Yarmey, 2003, for no effect of distributed learning for whispered voices). This advantage of distributed learning over massed practice has been found for different types of tasks (for a review see, Donovan & Radosevich, 1999).

In sum, the only conclusion that can be drawn is that the effect of length, vowel variety and distribution on voice identification accuracy is at present not fully understood.

The Effect of Retention

In a real-life investigation there is often a time gap between the crime and a possible voice lineup. In fact, one of the first factors to be examined in earwitness research was the effect of time delay on voice identification accuracy. It was the kidnapping case of the famous aviator Charles Lindbergh's son that raised this question and inspired the pioneering studies examining to what extent it is possible to identify a one-time heard unfamiliar voice after a very long period of time. Lindbergh's positive identification of the suspect's voice three years after first hearing it was accepted in court as evidence and the defendant was sentenced to the death penalty (McGehee, 1937). Studies examining the effects of different retention intervals have shown mixed results. For short delays, up to 24 hours after exposure, little loss in voice recognition has been found (Saslove & Yarmey, 1980; Yarmey, 1991b). For longer (and more realistic) delays, some studies show no difference in performance between a 1 week and 2 week delay (van Wallendael et al., 1994), between a 1 week and an 8 week delay (Kerstholt, Jansen, van Amelsvoort, & Broeders, 2006) or between a 2 week and an 8 week delay (McGehee, 1944). Other studies, however, have shown that memory for voices decline over time (e.g., Clifford, Rathborn, & Bull, 1981), a significant drop in performance between week 2 and week 3 (Clifford & Denot cited in Bull & Clifford, 1984), and that the false alarm rate increases after a week (Yarmey & Matthys, 1992). This mixed pattern does not offer a clear prediction for the effect of retention interval.

Age-differences in Voice Recognition

One possible important factor for voice identification accuracy is the age of the witness. Developmental change and growth continues throughout life and it is more intense for children as they differ fundamentally from one developmental period to another (Sroufe, Cooper, & DeHart, 1992). Research on cognitive development has established that the brain undergoes significant change during the onset of puberty at around 11–12 years of age (Sroufe et al., 1992), for example a decrement in cognitive efficiency is found (McGivern, Andersen, Byrd, Mutter, & Reilly, 2002). Regarding adults, research on aging and memory has shown that older adults perform more poorly on long-term memory tasks compared to younger adults (e.g., Brickman & Stern, 2009). Further, hearing ability is found to decrease with increasing age (e.g., Baltes & Lindenberger, 1997). Hence it is clearly important to know how these age-changes affect voice recognition. Nonetheless, not much is known about different age-group's voice identification performance.

As for children, only a handful of empirical studies concerning voice-memory are found. Studies of children's ability to recognise *familiar* voices have shown promising results. Children aged four and older are suggested to have an adult-like ability to recognize and identify their classmates (Bartholomeus, 1973), and children from the age of three are impressively good at identifying cartoon characters (Spence, Rollins, & Jerger, 2002). However, findings concerning familiar voices have limited forensic relevance. In most criminal situations, where the testimony of an earwitness would be of interest, the heard voice is unfamiliar and it is not possible to generalise findings on familiar voices to the identification of unfamiliar voices (Cook & Wilding, 1997a; van Lancker & Kreiman, 1985).

The recognition of *unfamiliar* voices has been tested in children aged 6 to 16 years and in adults (Mann, Diamond, & Carey, 1979). The overall results show that the number of correct identifications increased dramatically from the age of 6 to the age of 10. The 6-year olds performed at chance level, whereas the 10-year olds performed on the same level as adults. There was a decrease in performance for 11– to 13-year olds, and a return to adult-level at the age of 14. Even though this study focused on unfamiliar voices, it still has modest forensic value. That is, the testing phase took place immediately after the listening phase, and the participants were presented with a forced choice test between two or four voices. Such a setup does not mirror what takes place in a real criminal investigation.

In a study using a setup that better reflects real-life situations, both children's and adults' voice memory for unfamiliar voices were tested

(Clifford & Toplis, 1996). The participants both saw and heard the targets, and were then confronted with two voice lineups (one female and one male). The results showed that voice identification was poor for all age-groups (5–6, 8–9, 11–12-year olds and adults), but that false positive errors were found to decrease with age. The 5– to 6-year olds evidenced the highest proportion of incorrect responses and were found to be relatively more prone to making false identifications. The two youngest age-groups were found to perform worse than adults, whereas the 11– to 12-year olds performed better than the adults. Although this study used a more realistic situation (i.e. a self-experienced event), it still differs from a real criminal situation because of the very short time-delay between exposure and test.

To investigate the effect of delay (24–48 hours or 3 to 4 weeks) and naturally occurring stress on children's voice memory, children aged 3 to 8 years were tested with respect to a dental appointment (Peters, 1987). The children were tested for their ability to identify the voice of the dentist as well as the dental assistant. It was found that the overall accuracy level did not differ significantly from chance and no effect of stress, retention interval or age was found.

Even less attention has been given to older age-groups. Though, research indicates that listeners over 40 years tend to perform poorer than younger adults (Bull & Clifford, 1984). To my knowledge, no study has tested the performance of elderly witnesses.

To sum up, when it comes to unfamiliar voices, children under the age of 10 generally seem to perform rather poorly (Clifford & Toplis, 1996; Mann et al., 1979; Peters, 1987). The results for children aged 11– to 13-years are mixed; one study found a decrease in performance (Mann et al., 1979), whereas another study found that this age group performed better than adults (Clifford & Toplis, 1996). Further, younger adults (21 to 40 years) seem to be more reliable than older adults (over 40). This review shows that it is not possible to draw any precise conclusions concerning children's voice identification ability. Not only are the available studies few, the variation in methodology between the studies is also considerable. Further, there are gaps to fill with respect to knowledge of the performance of older age-groups.

Presentation Format – The Effect of Telephones

Many earwitness cases may be a consequence of crimes committed over a phone such as obscene phone calls, extortion, frauds, ransom demands and other threatening phone calls. Today's widespread use of mobile phones is reflected in the number of crimes where mobile phones are used. The sound quality mediated by telephone, and mobile phones in particular, is vastly inferior to that of a direct recording of good quality. Hence, it is important to have knowledge about how phones affect voice identification accuracy.

The effect of the degraded quality on the acoustic analysis of speech for forensic purposes has been examined in some studies (e.g., Byrne & Foulkes, 2004; Künzel, 2001). In these studies the limited telephone bandwidth has been shown to affect the position of the lowest resonance (first formant) in vowels which may affect the reliability of acoustic speaker comparison. Landline phone and mobile phone transmissions have been found to influence the sound quality negatively and in partly the same way (e.g., limited bandwidth, transmission losses and the effect of usually poor microphone quality) but not necessarily in identical ways. The transmission speed is greater for landline phones (typically 64 kb/s) than for mobile phones (24.40 kb/s or less). Furthermore, the speed of the wireless transmission between the mobile phone and the mobile network may change many times during a single call. A lower transmission speed causes information loss and results in a change of voice quality information. It is therefore not reasonable to assume that landline phones and mobile phones would affect voice recognition in exactly the same way. However, it seems to be the case that no study has tested the effect of these differences for voice recognition or memory, but it has been shown to have a considerable effect in automatic speaker recognition (Brümmer & Strasheim, 2009). When reference samples and test samples differ in quality this is called a mismatch condition. The mismatch in the Brümmer and Strasheim study (2009) was landline recordings vs. mobile phone recordings. Mismatch conditions resulted in higher error rates. Further, there are also considerable differences between mobile calls and landline calls with respect to speaking style. For example, it has been found that people have a tendency to speak more loudly when using a mobile phone, and because of mobility, there is often more background noise where mobile phones are used (Byrne & Foulkes, 2004; Foulkes & Barron, 2000).

The available studies concerning the effect of telephones on voice identification made by lay people are few and the methodology varies considerably. In the first study that addressed the effect of landline telephones, listeners heard six paired voice samples, using the original direct recordings or bandpass filtered versions simulating telephone quality (Rothman, 1977). Their task was to decide if the speaker was the same for the two samples. It was found that the simulated telephone quality voices were more difficult to identify than the voices in the original recording. In a study slightly more forensically relevant, direct recordings and landline telephone recordings were used (Rathborn, Bull, & Clifford, 1981). The participants heard either a directly recorded voice or a telephone recorded voice, and were then immediately confronted with a six-voice target-present lineup consisting of either directly recorded or telephone recorded voices. Each participant did six such trials (three male and three female voices). It was found that subjects who were presented with a directly recorded target and then a directly recorded lineup, performed significantly better than the participants in the other three conditions. There were no significant differences between the conditions where telephone recordings were used. These studies imply that landline telephone quality speech decreases recognition performance in general. However, the use of pairwise comparison (Rothman, 1977) or several lineups per subject (Rathborn et al., 1981) is not comparable to the conditions typical for real life investigations.

In more recent studies where subjects have been presented with a single target voice and a subsequent lineup (after varying delays) no effect of landline telephone quality has been found (Perfect, Hunt, & Harris, 2002; Yarmey, 2003), or only by a small margin (Nolan, McDougall, & Hudson, 2009). No significant difference in identification accuracy was found when comparing participants who had heard a directly recorded voice and participants who had heard a voice recorded through a telephone (Perfect et al., 2002), or when comparing participants who spoke to the target over the telephone and participants who spoke to the target face-to-face (Yarmey, 2003). When comparing four combinations of quality (exposure: studio/telephone quality; lineup: studio/telephone quality), studio exposure and lineup was found to result in marginally more correct identifications compared to telephone exposure and lineup (Nolan et al., 2009). Notably, the mixed conditions produced even less correct identifications.

The only study, to my knowledge, that has tested the effect of a mobile phone used a quite unusual setup as the speakers did not use the mobile phone directly (Kerstholt et al., 2006). Instead recorded speech was presented over a loudspeaker and a mobile phone was held close to the loudspeaker. No significant differences in identification accuracy were found between the participants in the telephone condition and those who heard a directly recorded voice. The results in these studies thus imply that telephone quality does not have any clear effect on voice identification accuracy, in contrast to the earlier cited studies.

Yet another aspect to consider when a phone is involved is how to conduct the lineup. It is suggested that the police share the commonsense belief that voice recognition will be enhanced if the test takes place under the same circumstances as the initial hearing, meaning that when a voice is initially heard over a telephone it is desirable to conduct the lineup using voices recorded over a telephone (Rathborn et al., 1981). However, research does not support the belief of the police (though see Nolan et al., 2009, for an opposite opinion). Rather, studies have shown that using a telephone recorded lineup when the voice is originally heard over a telephone will not result in greater accuracy (Kerstholt et al., 2006; Rathborn et al., 1981). If it could be established that mobile phone recorded lineups do not improve earwitness accuracy, then directly recorded voice samples could be used for the lineups without causing any negative effect on recognition. That would, to some degree, facilitate the work of the police and phonetic experts.

In sum, whereas the results of earlier work imply that landline telephone quality speech decreases recognition performance in general (Rathborn et al., 1981; Rothman, 1977), the results of more recent studies imply that telephone quality does not have any clear effect on voice identification accuracy (Kerstholt et al., 2006; Perfect et al., 2002; Yarmey, 2003). In addition, only one previous study has tested the effect of using what is now the most common type of telephone, the mobile phone.

The Effect of Tone of Voice

A shared problem for ear- and eyewitnesses is that the perpetrator can intentionally alter their voice/appearance between the crime and later occasions. The perpetrator may, for example, disguise their voice during the crime. As expected, research has shown that disguise lowers the number of correct voice identifications (Bull & Clifford, 1984; Hollien et al., 1982). Another example is whispering, which conceals some of the most important vocal characteristics, such as speech prosody (the melody of speech) and (at least partly) the timbre of the voice. There are also marked effects on vowel quality (Ito, Takeda, & Itakura, 2005; Petrushin, Tsirulnik, & Makarova, 2010). It is therefore only logical that research has shown that whispered voices, both familiar and unfamiliar, are more difficult to identify compared to normal-tone voices (e.g., Procter & Yarmey, 2003; Yarmey et al., 2001).

A more frequent problem for earwitnesses is unintentional change. Differences within the same voice at different occasions can for example be a result of the situation, emotional state, intention, and health status of the individual. Depending on the circumstances the voice may be altered and thereby affect the acoustical components of speech, like the articulation rate and pitch level. Further, it is reasonable to assume that it is common to sound upset or angry when committing a crime. Research has shown that such a relatively simple change in speaking style between presentation and lineup may reduce voice identification to chance level (e.g., Read & Craik, 1995;

Saslove & Yarmey, 1980). Even a smaller alteration, such as shifting from a casual speaking style to a formal speaking style, has shown to have a negative effect on voice identification accuracy (Bahr & Pass, 1996). Thus, the effect of tone of voice is an important aspect of voice identification and combined with the slim literature available this speaks for the need of more research. Further, to my knowledge, there is no previous research on the effect of tone of voice on child witnesses.

Interviewing Earwitnesses

Researchers have paid much attention to developing interview techniques aimed at helping eyewitnesses to accurately remember the experienced event. Much recent research has focused on the Cognitive Interview (e.g., Memon, Meissner, & Fraser, 2010). The Cognitive Interview (CI) was developed to enhance witnesses' memory and elicit as much correct information as possible, by using several cognitive techniques (e.g., Fisher & Geiselman, 1992). In short, the technique is based on two well-known memory principles; the multi-component view of memory (e.g., Bower, 1967) and the "encoding-specificity principle" (e.g., Tulving & Thomson, 1973). From these two principles, four mnemonics were derived: (1) a "report everything" instruction, which encourages the interviewee to report as many details as possible; (2) a "mental reinstatement of context" instruction, where the interviewee is asked to mentally reinstate both the internal (e.g., feelings) and external (e.g., physical surroundings) context of the event; (3) a "reverse order recall" instruction, that encourages the interviewee to recall the event in a reverse order, starting with the end of the most memorable part of the event; and (4) a "change perspective" instruction, which asks the interviewee to recall the event from the perspective of another person who was present (if relevant). The CI has shown to be beneficial for *recall* as it has been found to considerably improve the number of correct details reported by eyewitnesses (for a recent meta-analysis see Memon et al., 2010) and moreover, the positive effect has also been found for children of different ages (e.g., Larsson, Granhag, & Spjut, 2003; McCauley & Fisher, 1995).

When considering the effect of the CI on *recognition,* the results are more ambiguous. Mental reinstatement of context and variations of that mnemonic have shown to be effective for facial recognition (Krafka & Penrod, 1985; Malpass & Devine, 1981; Shapiro & Penrod, 1986). In a meta-analysis, where context reinstatement was defined as "whether or not context was reinstated with the use of cues previously associated with the targets or the incident at the study phase" (p. 141, in Shapiro & Penrod, 1986), it was found that the context reinstatement variable yielded the largest positive effect on hits, but also a negative effect on false alarms. In a study were the mental reinstatement of the witnessing context was induced by the interviewer providing information about the event, it was found that recognition accuracy was greater for the participants in the "guided memory" interview compared to a control group (Malpass & Devine, 1981). In a study that was more ecologically valid, the mental context reinstatement was done while the witness inspected physical evidence (Krafka & Penrod, 1985). The results showed that the context reinstatement significantly increased accurate identifications (compared to a control group). Though, it is not possible to determine whether that result was due to the context reinstatement or the psychical evidence. On the contrary, studies using the CI (instead of only context reinstatement) have shown a negative effect (Finger & Pezdek, 1999, Experiment 1) or no difference (Gwyer & Clifford, 1997) in terms of identification accuracy. In the "context reinstatement" studies that show a positive effect, the participants did not verbally describe the perpetrator as elaborately as they did in the CI studies (Finger & Pezdek, 1999). It has been demonstrated that verbally describing a stimulus may impair subsequent recognition performance, an effect termed "verbal overshadowing" (Schooler & Engstler-Schooler, 1990). Finger and Pezdek (1999) showed, however, that the CI only reduced face identification accuracy when the description took place immediately before the identification task. When a time delay was inserted (1h in Experiment 2, and 24 min in Experiment 3) the verbal overshadowing effect was eliminated.

Surprisingly few researchers have tried to investigate whether the impressive results of the CI also hold true for earwitnesses, or the effect of any type of interview for that matter. One exception is a study by Memon and Yarmey (1999) that compared the CI with the Structured Interview (SI) with respect to identification performance in a voice lineup. No significant differences were found between the two interview types for voice identification accuracy. As is the case for eyewitnesses, earwitness research has also shown a verbal overshadowing effect (Perfect et al., 2002; Vanags, Carroll, & Perfect, 2005). In these studies, however, the interviews were conducted immediately before the lineup, and in the study by Memon and Yarmey (1999) the interviews took place after a 2-day delay and immediately before the voice lineup. In a real criminal situation there is most often a time gap between the witness statement and a possible voice lineup. It is not known if the finding that the verbal overshadowing effect for eyewitnesses may be eliminated after a delay (Finger & Pezdek, 1999) would also hold true for voice identification.

Considering the effect of the CI on memory for conversation, the few previous studies show promising results. In the study by Memon and Yarmey (1999) the CI resulted in 24% more correct details compared to the SI. The difference was not significant, however. Campos and Alonso-Quecuty (2008) investigated the effect of the CI on witnesses' memory for a criminal conversation by comparing it to the Spanish Traditional Interview (STI). As predicted, participants in the CI condition showed a better recall of the content of the conversation without an increase in errors.

Based on eyewitness research, is has been suggested that the CI may only be beneficial when the witnessed event is relatively rich in details, for example, both seeing and hearing an event, or a situation in which a number of things occur simultaneously (Geiselman, Fisher, MacKinnon, & Holland, 1985). For the earwitness studies that have found a positive effect the critical events have been rather long in duration, seven minutes (Memon & Yarmey, 1999) and 15 minutes respectively (Campos & Alonso-Quecuty, 2008). Therefore, it is not known whether the positive effect of CI applies to criminal accounts with shorter durations.

Most police interviews are not as open and interviewee driven as the CI, however. Instead, police interviews are characterized by being organized around a pre-determined set of questions that are asked in a pre-determined order, which lends the interviewee a rather passive role (Fisher, Ross, & Cahill, 2010). In accordance with this, the Swedish Security Service (SÄPO) has developed an interview protocol for earwitnesses in the form of a check-list. This check-list is used for questioning people that have been exposed to a crime and/or threats and where they have heard – but not seen – the perpetrator; for example a receptionist that has received a bomb threat over the telephone or a bank clerk that has been robbed by masked perpetrators. Because the check-list is used by the Swedish Security Service and there are a growing number of terrorist oriented incidents, it is therefore important to know what effect, if any, this check-list may have on earwitness memory and voice descriptions. To the best of my knowledge, this check-list has not been scientifically evaluated.

In sum, finding ways to enhance earwitnesses' memory for voices heard once is of utmost importance and one possible approach is to develop an appropriate interview technique for questioning earwitnesses. Further, it would be of great interest to scientifically evaluate the interview procedures presently used by the police and to contrast these against more theoretically based interviews like the CI.

Voice Identification in Sweden

Although voice identification occurs all over the world (Hollien, 2012), in Sweden there is no tradition of using voice confrontations for identification purposes. This is not that surprising considering that in Sweden there are no established methods with respect to how to conduct a voice identification test with a witness. In the report "Vittneskonfrontation" (Witness confrontation) published by The National Police Agency (RPS Rapport, 2005:2) it is recommended that a voice lineup should be conducted with the help of a phonetic expert. If, however, the police should decide to conduct such lineups themselves, the available guidelines are few. The suggested principle for selecting voices (foils) is that, for example, sex, age and dialect should be fairly similar to the suspect's voice. Further, the voices should be tape-recorded to avoid unexpected incidents (the foils or the suspect may hesitate or refuse to talk) and to enable repeated listening. Finally, it is recommended that all the voices in the lineup should pronounce the same phrases or sentences. If the witness has heard the perpetrator speaking in a normal tone, all voices should repeat the exact phrase uttered by the perpetrator.

These guidelines are not very detailed and not in complete agreement with results from earwitness research. For example, research has not found lineups where the foils and the suspect are uttering identical phrases or the exact words of the perpetrator (said at the initial exposure) to result in more correct identifications or less false identifications (Hammersley & Read, 1996; Yarmey, 2001). Further, guidelines in other countries, for example in The Netherlands (Broeders & van Amelsvoort, 1999) and in The United States (Hollien, 1996), are more detailed. Hence, it is evident that the Swedish guidelines need to be improved.

Evaluation of Past Research

By now it should be very clear that earwitness research is a neglected area. Most of the variables studied have been examined to a limited extent and the outcome is often mixed. Combined this makes it difficult to draw any firm conclusions on how the examined variables affect voice identification accuracy. The diverse findings are most likely due to the large variation in methodology. To illustrate, the effect of retention interval depends on several factors like attention, tone of voice, and duration. Further, the performance depends on how the memory is tested (lineup vs. forced-choice). Different researchers have tested variables under different conditions and, as pointed out by Hollien (2012), the area suffers from the lack of robust structuring and adequate standards.

In addition to the fact that the area suffers from a lack of common ground, there are some problematic gaps in the literature. In my opinion there are at least four important gaps that merit attention. First, the performance of different age-groups has received relatively little attention. Almost all studies have used participants in the age range 20 to 40 years, and have neglected children and older adults. This is a problem as both general memory and hearing ability are likely to change throughout the life-span and people at all ages can become victims of or witnesses to a crime. Secondly, a common drawback for past research is that it is often inadequate in terms of theory and short in terms of explanations for *why* earwitnesses perform poorly. The area would benefit from focusing on basic research and investigating for example; how we remember voices, how a voice becomes familiar, and what cues are used to recognize a voice. Thirdly, in line with the finding of earwitnesses' poor memory performance, it is noteworthy that ways to enhance earwitness memory have received very little attention (see Memon & Yarmey, 1999, for an exception). This is true for all three types of information that an earwitness might contribute (voice identification, memory for content and voice descriptions). Fourth, the lineup procedure itself would benefit from more research. It would be of value if a standard method could be adopted as this, for example, would facilitate comparison between studies. In addition, the development of new procedures might enhance identification accuracy.

From a psycho-legal perspective, parts of past research may be viewed as lacking legal relevance. As an example, some studies have used a forced-choice test, which does not mirror how an identification task would be administrated in a real investigation. Further, many studies have used an immediate recognition test, which would not be possible in a real criminal case. However, it is not to say that such research is invalid, as examining such conditions can contribute to knowledge on how voice memory works. I would, however, prefer a clearer distinction and expressed awareness between research that has a legal perspective with an applied focus and research with a more basic perspective.

Summary of the Empirical Studies

In some criminal cases the victim's or witness's memory of the perpetrator's voice may be an important clue and therefore have a vital role in the investigative phase and in court. Nevertheless, compared to eyewitness identification, earwitness identification is a much neglected area (e.g., Wilding et al., 2000). The empirical studies in the current thesis sought to explore how well earwitnesses perform at identifying an unfamiliar voice that has been heard under rather realistic criminal conditions. A total of 949 participants have been tested under a number of frequently present variables that might affect the accuracy level, namely; *age* (Study I, III, & IV), *presentation format* (Study II), *tone of voice* (Study IV), and *time delay* (Study IV). An additional aim was to find ways to enhance earwitnesses' memory for voices (Study III & IV) and content recall of a criminal conversation (Study III). Further, witnesses' voice descriptions have been found to be very vague (e.g., Yarmey, 2003). Therefore, a further attempt was made to find a more effective way of interviewing witnesses about the voice (Study IV).

The voice samples used were chosen from a set of 16 recordings where people who spoke an educated form of the regional dialect were recorded in a semi spontaneous dialogue and also reading a mock incriminating call from a manuscript. One specific person that sounded reasonably "involved" and not as though he was reading was selected as the mock perpetrator. The seven foils needed for the lineups were selected based on the outcome of a perceptual evaluation test taken by two groups of undergraduate students (37 altogether). Following suggestions by Hollien (2002), we selected, two speakers who were perceived as quite similar, two rather dissimilar and the remaining three in the middle. Three Target-Present (TP) lineups were constructed, each included the same voices but presented in different orders. Further, for Study I, three Target-Absent (TA) lineups were constructed which were identical to the TP lineups, except that the perpetrator's voice had been replaced by a foil whose voice had been judged to be most similar to the perpetrator's voice. Each voice was simultaneously recorded via a mobile phone so that three identical mobile phone recorded TP-lineups could be constructed for Study II.

As the thesis has an applied focus, the chosen method was aimed to mirror what could take place in real-life, both at encoding and in the lineup. That is, a method that maximizes possible differences between the manipulated variables was not chosen. All four studies in the thesis used the same general experimental setup, with some variations depending on the major focus of the study (see Table 1). This was motivated by two basic reasons. First, the use of the same general setup enabled comparisons between the four studies. Secondly, as the earwitness research area suffers from a lack of structure, the current thesis sought to contribute with more systematic knowledge. The setup consisted of two phases.

In the first phase, the witnesses were exposed to an unfamiliar voice. To simulate a realistic situation, the listeners were instructed to imagine that they were in a shop to buy cloths and that they were standing in front of a dressing room waiting for their turn. To increase realism, a curtain had been hung from the ceiling and participants were instructed to place themselves in front of it. They were told that they would overhear something taking place behind the curtain, but not that they would hear a voice. Loudspeakers presenting the recorded target were placed behind the curtain. The presentation started with a mobile phone signal followed by a person behind the curtain answering the call and talking to someone (not heard) for about 40 seconds about the planning of a crime. After listening to the critical event the participants were told that they would be interviewed about the event two weeks later; however they were not informed about what aspects of the phone call the interview would concern.

In the second phase, which took place two weeks later (except for half of the participants in Study IV) each participant was presented with a recorded voice lineup consisting of seven voices. The participants were informed that the person they had heard speaking two weeks earlier *may* or *may not* be present in the lineup. First the participants were instructed to listen carefully to all seven voices (sample length 22–26 seconds) without making any decision. After hearing all voices once, the participants were instructed that they would now hear the voices once again, but shorter voice samples (11–14 seconds per voice). This time the participants were asked to report *if* they recognised the voice they had heard earlier and if so, which of the numbered voices it was. If they thought that the voice was not present they were instructed to simply say so. The participants were also asked to report how confident they were that their decision was correct. After the lineup, the participants were asked to report everything they could remember of what the perpetrator had said. Finally, the participants were asked for personal background information such as age, place of birth and years of life spent in Sweden (when relevant).

Due to the fact that the main dependent variable (lineup decision) is dichotomous (i.e. correct or incorrect), Chi-square analyses were used to test possible effects of the manipulated variables. As the two remaining dependent variables (content recall and voice descriptions) require a different statistical treatment, they were coded and analysed for accuracy and agreement.

Table 1

Overview of the four empirical studies.

	Phase one			Phase two		
	Event	Interview	Retention interval	7-voice lineup	Manipulated variables	Dependent variables
Study I	Hearing a voice for 40 s	-	2 weeks	TP/TA	Age (7–9 vs. 11–13 vs. adults)	Voice identification (Memory for content)
Study II	Hearing a voice for 40 s	-	2 weeks	TP	Presentation format (direct vs. mobile phone)	Voice identification
Study III	Hearing a voice for 40 s	Yes	2 weeks	TP	Age (11–13 vs. adults) Interview type (CI vs. BI vs. SSSI)	Voice identification Memory for content Voice descriptions
Study IV	Hearing a voice for 40 s	Yes	Immediate/ 2 weeks	TP	Age (11–13 vs. adults) Tone of voice (normal-normal vs. angry-normal) Time delay (immediate vs. 2 weeks) Interview (global questions vs. scale ratings)	Voice identification Voice descriptions

Study I

It is not unusual that children are asked to provide testimony as plaintiffs or as witnesses (e.g., Gordon, Baker-Ward, & Ornstein, 2001). However, there are only a handful of studies that have tested children's ability to recognize unfamiliar voices. Hence, the aim of Study I was to explore the performance of children aged 7–9 and 11–13, and adults.

A total of 264 subjects (95 7–9-year-olds, 78 11–13-year-olds, and 91 adults) participated in the study. In accordance with the general setup, the witnesses were exposed to an unfamiliar voice for 40 seconds and, after a two week delay, they were presented with a seven-voice lineup. The variation was that half of the participants were exposed to a TP lineup, and the other half to a TA lineup.

We had three main predictions. First, we predicted that the adults would make more correct identifications and more correct rejections than the youngest children. Previous research for the older children is too ambiguous to permit any precise hypothesis regarding their performance level compared to that of adults. Second, based on eyewitness research, we predicted that both groups of children would make more false identifications than the adults (e.g., King & Yuille, 1987). Third, irrespective of age, we expected a weak relationship between accuracy and confidence.

Results

The overall mean performance level (referring to the total number of participants that correctly identified the target in a TP lineup or correctly rejected a TA lineup) across all age-groups and conditions was 33.8%, which was significantly above the level of chance (12.5%, 8 possible alternatives). Broken down for each age-group and condition, all age-groups performed above chance level for the TA condition. For the TP condition the older children performed significantly above chance level (27% correct), whereas the younger children (14% correct) and the adults (20%) did not.

Across all three age-groups, 53% of the participants made a false identification in the TA condition, and 44% of the participants selected a foil in the TP condition. Contrary to our prediction, the results showed that the adults had the highest percentage of false identifications for both TP and TA, compared to both the younger and the older children. In order to investigate why some of the speakers were chosen more often than others, three acoustic cues (important for voice similarity judgements) were explored: *Articulation rate, Pitch variation* and *Pitch level.* All three of these cues were found to correlate with false identifications. For the youngest children there were highly significant correlations between the number of false identifications per foil and all three acoustic cues, whereas only *Articulation rate* and *Pitch level* correlated significantly with false identifications for the older children. None of the acoustic cues correlated significantly with false identifications for the adults. The direction of dependence was that foils with higher articulation rates and greater pitch variation were chosen more often. For pitch level the trend was in the opposite direction; foils with lower pitch were chosen more often.

Possible age-differences were tested for all types of responses, however, the only significant difference was that the youngest children made significantly more false rejections in the TP condition compared to the older children and the adults (Chi-square, $p < .05$). The older children and the adults did not differ significantly from each other with respect to false rejections. In line with our prediction, the overall confidence-accuracy correlation was not significant.

Conclusions

The results of this study confirm previous research showing that witnesses, regardless of age, are not very good at recognizing an unfamiliar voice heard under realistic circumstances. Further, there was high level of false identifications overall, with almost half of the participants identifying an innocent person. On the positive side one can note that the older group of children performed at the same level as the adults and in some respects even better. However, this should be interpreted with some caution until more research is carried out involving this particular age-group.

An interesting observation in connection with the analysis of false identifications was that the children were significantly influenced by the pitch and rate cues, factors which did not affect the judgements of the adults. A lesson to be learned from the present study is that linguistic factors like speaking rate may distort the fairness of the lineup.

Study II

Today it is very common that mobile phones are used during crimes. However, there is so far only one study that has tested whether or not mobile phone quality affects voice identification accuracy (Kerstholt et al., 2006). Hence, the aim of Study II was to investigate if presentation format (direct recording vs. mobile phone recording) affects voice recognition accuracy. A further aim was to investigate the usefulness of conducting a lineup with mobile phone recorded voices for a situation where the target voice was originally heard over a mobile phone.

A 2 (Initial exposure: direct vs. mobile phone) x 2 (Lineup presentation: direct vs. mobile phone) between-group design was employed. A total of 165 adult participants were assigned randomly to one of the four conditions. The procedure was almost identical to the general setup. The only difference in phase one was that half of the participants heard exactly the same conversation recorded through a mobile phone, and in phase two; half of the participants were exposed to a mobile phone recorded lineup which was exactly the same as the direct recordings. For this study only TP lineups were used.

We predicted that participants who heard a directly recorded voice would perform better at voice recognition than participants who heard a voice recorded through a mobile phone. Further, we predicted that a mobile phone recorded lineup would not improve voice recognition accuracy in cases where the target was recorded via a mobile phone. In addition, as for Study I, we expected no or only a weak relationship between accuracy and confidence.

Results

The overall mean for correct identification across all conditions was 12.7%, which is exactly what would be expected by chance alone (12.5%, 8 possible alternatives). The number of correct identifications was somewhat lower for the directly recorded lineups compared to the mobile phone recorded lineups, which is in direct contradiction of our prediction, but none of the differences were significant.

Across all conditions, 57% of the participants made a false identification. There was little variation in the number of false identifications as a function of presentation format or lineup format.

Of the participants who made a false identification, 54% selected one particular foil. The number of times this particular foil was chosen was significantly above chance level for all conditions. The proportion of false rejections was almost the same for all four conditions. As predicted, the overall confidence-accuracy correlation was not significant.

Conclusions

The results showed no difference in voice identification accuracy between a mobile phone quality voice and a direct quality voice, and the idea that mobile phone lineups should be used if the target voice is originally heard over a telephone receives little support. The overall performance level was at chance level and the witnesses showed a strong tendency to make false identifications. Why some foils may attract more attention than others, although selected after a perceptual evaluation test, is something that needs to be examined more closely in future studies.

Study III

As research has shown that earwitnesses are poor at remembering and recognizing an unfamiliar voice heard under realistic circumstances, the aim of Study III was to find ways to enhance earwitnesses' memory for voices as well as content recall of a criminal conversation. For that purpose, three types of interviews were compared. An additional aim was to evaluate an interview protocol developed by the Swedish Security Service for questioning people that have been exposed to a crime and/or threats, where they have only heard the perpetrator.

The Cognitive Interview (CI) has been proven to have strong positive effects for eyewitnesses recall accuracy (e.g., Memon et al., 2010). Surprisingly few have tried to investigate whether the impressive results of the CI also holds true for earwitnesses. Hence, this study investigated if a CI, with questions about the voice and content, employed immediately after the witnessed event would enhance earwitness memory at a later stage. Though, research has shown that verbally describing a voice may impair subsequent voice recognition (Perfect et al., 2002; Vanags et al., 2005), an effect termed "verbal overshadowing" (Schooler & Engstler-Schooler, 1990). With potential verbal overshadowing in mind, we contrasted interviews with no questions about the voice (Baseline interview), specific questions about the voice (Swedish Security Service checklist) and open-ended questions together with more specific questions about the voice (CI).

To our knowledge, there is no previous study focusing on the CI and children as earwitnesses. This is noteworthy since it is not unusual for children to be victims or witnesses of a crime. Therefore, both 11–13-year-olds (*n* = 119) and adults (*n* = 93) participated in the study. The variation from the general setup was that after hearing the voice, the participants were told to imagine that they had decided to report to the police what they had just heard. After waiting 5–10 minutes, they were interviewed using one of the three interview techniques. Hence, all witnesses were interviewed twice (in the first phase as well as in the second phase). One further difference was that the most commonly identified foil in Study I and II was replaced in the lineup by the voice that was judged as most similar to the perpetrator (used in the TA lineup in Study I).

We had three main hypotheses. First, an interview that makes the participants reflect on the voice shortly after the exposure will increase the ability to make a correct identification, and even more so when having been interviewed with the CI. Second, irrespective of age, participants in the CI condition would recall more information about the content of the criminal conversation compared with the other two conditions. Third, irrespective of interview type, the adults would recall more of the content than the children.

Results

The overall mean performance for correct identifications was 20.2% for the children and 19.4% for the adults, which was significantly above the level of chance (12.5%). However, broken down for each age group and interview condition, only children in the baseline condition performed above chance level. Across conditions, no significant differences were found between children and adults for any response type.

There was no significant difference in correct identifications between the three interview conditions, either for the children or the adults. Though, for the adults, there was a significant difference in the number of false identifications found; adults in the baseline condition made significantly fewer false identifications compared with the adults in the other two conditions.

As partial support for our hypothesis, adults in the CI condition (compared to the other two conditions) reported significantly more correct information both immediately after witnessing the event and at the second interview. Unexpectedly, this was not the case for the children. Though, as predicted, the adults reported significantly more correct information than the children in all conditions.

The overall agreement between the witnesses for the questions in the Swedish Security Service checklist was moderately high. When asked to freely describe the voice, both children and adults gave few descriptions, and the adults gave significantly more descriptions than the children.

Conclusions

Both children and adults were found to perform poorly in terms of voice identification and an interview shortly after the witnessed event did not seem to help. However, the study advances previous knowledge on earwitnesses' performance in two ways. First, the CI, in the case of adults, was found to be beneficial for recalling the content of a brief conversation. The finding that the CI did not seem to help children's content memory makes it an urgent task to develop interview techniques that also work for children. Second, the checklist used by the Swedish Security Service had no positive effect on either recognition or recall. On the contrary, the checklist was found to produce more false identifications compared to a "standard" police interview, and less recalled information compared to the CI. In addition, the voice descriptions elicited by this checklist were not found to be very useful. The combined evidence shows the importance of properly evaluating the methods used by the police.

Study IV

Although research has shown that earwitnesses perform poorly (e.g., Read & Craik, 1995; Studies I, II, and III in the present thesis), there are reasons to believe that the reliability of voice lineups might be higher under certain circumstances. If this assumption is correct this would be helpful when deciding if a lineup should be conducted, and further, when assessing the diagnostic value of earwitness evidence. Study IV investigated two frequently present variables in voice identification cases that might affect voice recognition accuracy, namely; the effect of the perpetrator's *tone of voice* and *time delay*. Poor voice encoding is an important general factor that might contribute to the low accuracy levels found. Hence, an additional aim was to examine two types of voice description interviews intended to strengthen the encoding of the voice.

It is reasonable to assume that it is common to sound upset or angry when committing a crime. Research has shown that a relatively simple change in speaking style between first presentation (angry tone) and lineup (normal tone) markedly reduces voice identification accuracy (Read & Craik, 1995; Saslove & Yarmey, 1980). Although it is not unusual for children to be victims or witnesses of a crime, to our knowledge, there is no previous research on the effect of tone of voice on child witnesses. Therefore, both 11–13-year-olds (*n* = 160) and adults (*n* = 148) participated in the study. The perpetrator either spoke in a normal tone both at encoding and in the lineup (*congruent*), or in an angry tone at encoding and in a normal tone in the lineup (*incongruent*). Witnesses were then interviewed about the voice; either with six global questions about the voice (e.g., Can you describe the voice? Was there anything unusual about the voice?), or by rating the perpetrator's voice on a 6-point scale for eleven voice characteristics (e.g., timbre, harshness). Both interviews ended with a question asking if the witnesses believed that they would recognize the voice if they had the opportunity to hear it again. Half of the witnesses were presented with a lineup shortly after the interview (*immediate*) and the remaining after two weeks (*delayed*).

First, we expected an effect of time such that witnesses tested immediately would perform better than witnesses tested after two weeks. Secondly, we expected an effect of tone of voice. That is, witnesses in the congruent condition were expected to outperform witnesses in the incongruent condition in terms of identification accuracy. Thirdly, we expected an association effect such that witnesses in the *congruent-immediate* condition would perform better than the witnesses in the other conditions. Furthermore, two exploratory questions were addressed; (1) are global open-ended questions about the voice more beneficial in terms of correct identifications and elicited information, compared to ratings of voice characteristics?, (2) what is the level of agreement between witnesses when asked to rate voice characteristics on a scale?

Results

Overall mean performance was 13.4% correct identifications for the children and 9.6% for the adults. Thus, neither group performed above the level of chance (12.5%). Across conditions, no significant differences were found between children and adults for any response type.

For the children, there was a reliable effect of time for correct identifications, such that children tested immediately (21% correct) performed better compared to children tested after a two week delay (9% correct). This was not the case for adults. Further, interview type and (in)congruency between tone of voice did not significantly affect voice identification accuracy, either for the children or the adults. Though, as predicted, witnesses in the *congruent-immediate* condition performed the best, with 25% of the children and 19% of the adults making a correct identification.

Although 86% of the children and 63% of the adults believed that they would recognize the perpetrator's voice if they had the opportunity to hear it again, only 13% of these children and 4% of these adults actually made a correct identification.

The witnesses reported few descriptions when asked global questions about the voice, and there was no significant association between the number of reported descriptions and identification accuracy. Further, only 45% of the descriptions could be categorized as an actual description of the voice. Agreement between the participants for the voice characteristic ratings was high in the neutral tone condition as well as in the angry tone condition.

Conclusions

Contrary to our prediction and past research (e.g., Saslove & Yarmey, 1980), there was no difference in terms of voice identification accuracy between witnesses who heard the perpetrator speak in an angry tone at encoding and in a normal tone in the lineup, and those who heard him speak in a normal tone on both occasions. At first sight this may be viewed as a positive result, as it suggests that a voice lineup may still be recommended even though a perpetrator has spoken in an angry tone during the witnessed event. However, it becomes less encouraging when considering the low accuracy level for both the congruent and incongruent condition.

As partial support for our prediction, time was found to affect voice lineup accuracy for children, such that children tested shortly after exposure made more correct identifications compared to those tested after a two week delay. Unexpectedly, this was not the case for the adults.

No difference in terms of identification accuracy was found between the two interview conditions. On a positive note, a relatively high agreement was found between the witnesses' perception of the voice in the scale-rating interview condition. If voice perceptions do not vary too much between witnesses, it means that the use of scale based descriptions may play an important role for speaker profiling. Consistent with previous research (Yarmey, 2003), witnesses generated few descriptions when asked to verbally describe the voice. Thus, global open-ended questions did not seem to help the witness to give a richer description; the reported descriptions were very vague and less than half of the descriptions could be categorized as an actual description of the voice. Hence it is evident that better ways for eliciting more detailed descriptions are needed.

In sum, the most important finding was the overall low accuracy level and especially the poor performance by those tested under what would seem to be rather favourable conditions (*congruent-immediate*). This, in conjunction with the finding that the majority of witnesses thought that they would be able to recognize the voice, makes it appropriate to advice legal actors to treat voice identification with great caution.

General Discussion

The major aim of this thesis was to explore earwitnesses' identification performance for an unfamiliar voice that has been heard under rather realistic criminal conditions. This question was addressed in four studies and, taken together, these four studies focused on several forensically relevant variables that might affect the accuracy level. A further aim was to try to enhance voice recognition by comparing different interview techniques intended to strengthen voice memory. Additional aims were to investigate people's ability to describe voices and their memory for what the perpetrator said.

In some criminal cases verbal information is the most critical clue, such as crimes committed in darkness, over the phone, or by masked perpetrators. Nevertheless, victims' and witnesses' memory for voices – and for what the perpetrator said – are neglected research areas. In brief, the most important findings derived from the present thesis are as follows. First, regardless of age, earwitnesses are poor at recognizing an unfamiliar voice when tested under rather realistic criminal conditions (see Table 2). However, earwitnesses seem to believe that they are better than they really are. Such expressed overconfidence is a problem as it can mislead legal practitioners in their decisions, both in the investigative stage and at the trial stage. Thirdly, the method currently used by the Swedish Security Service for questioning people that have been exposed to a crime and where they have heard – but not seen – the perpetrator, was found to produce more false identifications compared to a "standard" police interview, and less recalled information compared with the Cognitive Interview. Hence, the overall conclusions that can be derived from the present thesis are that actors in the legal system should treat voice identification evidence with great caution and that better methods for gathering information from earwitnesses need to be developed.

In the following sections I will expand on these and other findings and discuss some possible explanations for the overall low accuracy level, as well as some limitations. Finally, I will present some legal implications and future directions.

Table 2

A summary of all studies showing the overall percentages for each possible response type.

	Study I			Study II	Study III		Study IV		TOT
	7–9	11–13	Adults	Adults	11–13	Adults	11–13	Adults	
	%	%	%	%	%	%	%	%	%
TP									
Correct id	14%	27%	20%	13%	20%	19%	13%	10%	17%
False id	36%	46%	50%	57%	35%	43%	48%	52%	46%
False rej	50%	27%	30%	30%	45%	38%	39%	38%	37%
TA									
Correct rej	51%	51%	40%						
False id	49%	49%	60%						

Do Adults Outperform Children at Voice Identification?

Although people of all ages can become witnesses, very few earwitness studies have examined children's performance. Therefore, three of the four studies in the present thesis included both children and adults.

The effect of age was the principal focus of Study I and the main finding was that there was almost no difference in performance between the different age-groups (7–9, 11–13, & adults). The older children performed better than chance at correct identifications, whereas the younger children and the adults did not. This finding is in line with what Clifford and Toplis (1996) found, but contradicts the decrease in performance found for this age group by Mann et al. (1979). Further, the results of Study III and IV showed that the differences in performance level between adults and children aged 11–13 were small and mostly not significant. In line with Study I, although not significant, the tendency was that the older children performed better than the adults. The rapid cognitive development that older children undergo (Blakemore & Choudhury, 2006) may be one possible explanation for the ambiguous pattern of results, and underlines the need for more research involving this age-group.

In line with previous findings (Clifford & Toplis, 1996; Mann et al., 1979; Peters, 1987), the youngest children did not perform above chance level in the TP condition. It seems as if the cognitive demands of the lineup situation are too high for children under the age of 10. It has been found that knowledge of *when* it is necessary to remember undergoes a rapid growth in the first 12 years of life (Davies, 1996). Therefore, it may be the case that the younger children did not think that the original event was something that they needed to memorize. However, these explanations need to be further investigated and cannot account for the adults' weak performance.

All age-groups performed well above the level of chance in the TA condition (Study I). The two groups of children even outperformed the adults concerning correct rejections, although not significantly so. However, the "positive" result for the TA condition must be interpreted with caution. It is more difficult to respond correctly in a TP lineup because the only way to be correct is to identify the target (a rejection or the selection of a foil would be incorrect). In contrast, in a TA lineup a participant is correct if he or she merely states that the voice is not present. However, it is not possible to say with any reasonable degree of certainty to what extent the "not-present" answers represent real awareness that the perpetrator's voice was not present, or if this answer was given only because the participants found the identification task too difficult. The degree of reported certainty might have been a guide, but since there was no significant correlation between certainty and accuracy this question will remain unanswered.

It is noteworthy that it was the adults who made most false identifications for both the TP and the TA lineups in Study I. This is contrary to what is usually found in eyewitness studies. In eyewitness research, children have been found to make false identifications more often than adults, especially in TA lineups (e.g., King & Yuille, 1987; Lindsay, Pozzulo, Craig, Lee, & Corber, 1997; Parker & Carranza, 1989). However, eyewitness research showing children to make relatively more false identifications has often used a simultaneous lineup procedure. Irrespective of age, simultaneous lineups have, compared to sequential lineups, been found to result in markedly more false identifications, and especially so when the target is absent from the lineup (Steblay, 2007). The explanation for this is that simultaneous lineups encourage a relative judgment, whereas the sequential lineup encourages an absolute judgment. All voice lineups are sequential since voices must necessarily be presented one at a time. This fact may reduce the children's guessing tendency by demanding a relatively high cognitive ability to remember all of the voices and afterwards choose one (if any). If we assume that adults are better at retaining the line of voices in memory, this should make them more likely to make a relative judgment and experience a greater pressure to make an identification.

The only significant age difference was that the youngest children made more false rejections in TP lineups than the other two age-groups. The results that the youngest children made least false identifications (although not significantly so) and made most false rejections may not be that surprising. When combining these results, a possible interpretation is that the youngest children simply avoided making an identification in both types of lineups, possibly because they perceived the task to be too difficult. This may also explain why the youngest children performed better than the adults in the TA condition (although not significantly so). However, this finding is contrary to what has been found in eyewitness research, and shows that it is not always possible to generalize findings from eyewitness research to earwitness research.

In conclusion, although all age-groups perform poorly, a voice identification test seems to be too demanding for children under the age of 10. However, the older group of children generally performed at the same level as the adults and in some respects even better. From a practical, forensic perspective, this suggests that *if* we are prepared to accept adults as earwitnesses, there is no reason why we should not also view children (in the age range 11–13 years) as potentially useful earwitnesses.

Does Presentation Format Matter?

As has been pointed out above, mobile phone transmission may alter the sound quality quite substantially compared to directly recorded speech or speech heard in face-to-face exchanges. The perceived timbre of a voice depends on resonances in the vocal tract. The acoustic correlates of these resonances are called formants. Due primarily to limited bandwidth, the formant information may be distorted in telephone transmission. This has been observed in several studies (e.g., Künzel, 2001). In the present thesis, differences in measured formant values for the first two formants on the order of 150 Hz on average were found between the mobile recordings and the direct recordings. When listening to one of the directly recorded voices and afterwards listening to the exact same passage recorded via a mobile phone, one clearly hears a degradation of the sound. Due to the poorer sound quality in mobile phones, it was expected that the number of correct identifications would be higher in the direct condition compared to the mobile phone condition. The overall mean for correct identification across conditions was equal to what would be expected by chance (12.7%) and contrary to the expectation, the number of correct identifications was somewhat lower for the directly recorded lineups (Direct/Direct and Mobile/Direct) compared to the mobile phone recorded lineups (Mobile/Mobile and Direct/Mobile). However, none of these differences were significant. One might speculate whether there could still be an influence of mobile phone quality but that the effect was not detected due to the low number of correct identifications. Though, the analyses failed to produce a significant effect of mobile phones, not just for correct identifications, but for any of the obtained results, including the significant number of false identifications of one particular foil. Taken together, the results strongly suggest that the detrimental effect on voice recognition suggested by the poorer sound quality of mobile phone recordings is minimal, if any at all. This is in line with findings in past similar studies (Kerstholt et al., 2006; Perfect et al., 2002; Yarmey, 2003). Further, the results also imply that using a mobile phone recorded voice lineup when the voice is originally heard over a mobile phone is not likely to improve identification accuracy (nor does it seem to impair the identification accuracy).

The absence of a detrimental effect of mobile phone quality on any response type is somewhat surprising considering that there was a clear difference in sound quality. It is difficult to find an explanation for this lack of effect. In a study where human listeners were compared to automatic speaker verification for their performance on telephone speech recognition, it was found that both performed worse when the quality was degraded by background noise and poor transmission conditions (Schmidt-Nielsen & Crystal, 1998). However, the humans generally outperformed the automatic system. A suggested partial explanation was that humans depend greatly on speech habits (pronunciation, characteristic laughs etc.) when recognizing speakers.

The present study aimed to test if mobile phone sound quality caused by technical factors has an impact on voice identification accuracy, and did not include other factors connected with mobile phone speech like speaking style and background noise. For the current study, the direct and mobile phone speech samples were recorded simultaneously, which most likely meant that the speakers did not feel as if they were speaking in a mobile phone and therefore did not adjust their speech the way they might normally have done. It has been found that a speaker's fundamental frequency (F0) might increase when talking in a mobile phone (Byrne & Foulkes, 2004). This potential aggravation for speaker recognition was thus not a problem in the present thesis. Further, the witnesses heard the voice in a silent room. This means that background noise that might otherwise interfere with witnesses' ability to clearly hear and attend to the voice was not present. Although this might partly explain the absence of a difference between the two conditions, it is not sufficient to explain why the poorer quality did not have an effect.

Can an Interview Improve Earwitnesses' Performance?

A primary aim of the present thesis was to try to enhance witnesses' memory for voices and content. Poor performance may be the result of failure in retrieval as well as poor encoding. In an attempt to enhance performance, different types of interviews were examined. In Study III the well-known Cognitive Interview was contrasted to a "baseline" interview (in which the interviewees were simply asked to report everything they remembered without any questions about the voice), and a check-list developed by the Swedish Security Service (containing specific questions about the voice). Study IV focused on encoding and addressed the question of whether open-ended global questions about the voice would strengthen memory and be more beneficial in terms of correct identifications compared to rating voice characteristics on a scale.

The present thesis lends further support to the CI as an effective memory enhancing tool as it resulted in significantly more correct information (both immediately after the witnessing event and at the second interview) compared to both the Swedish Security Service check-list (SSSI) and the baseline interview (BI). It is worth noting that this result occurred even though the to-be-remembered situation contained only auditory information and was not particularly rich or complex in nature. Research has shown that witnesses in an auditory-only condition (compared to audio-visual condition) tend to report less correct information (e.g., Campos & Alonso-Quecuty, 2006; Gibbons et al., 1986) and show a greater decrement in memory performance (e.g., Toglia et al., 1992). This makes the finding that the CI elicited relatively more correct recall and resulted in relatively less forgetting even more important. The positive effect was, however, only significant for the adults. Unexpectedly, the result suggests that the CI may not be beneficial for enhancing children's content recall. This is surprising because previous research on eyewitness testimony has shown the CI to be effective for children (e.g., Larsson et al., 2003). Further, the children reported less correct information than the adults regardless of interview condition. Hence, it is important to develop interview techniques that also work for children as earwitnesses. On the positive side, one may note that neither the children nor the adults reported many fabrications. This might be due to the fact that they were explicitly told (in the CI) not to guess and that it was okay to say "I don't know" or "I don't remember".

Unfortunately, none of the interview types enhanced voice recognition. In both Study III and IV, the overall performance was poor and there was no significant difference between the interviews in terms of the number of correct identifications. The adults interviewed with the CI and the SSSI made slightly fewer accurate identifications compared with adults in the BI condition. This result may lend some support to previous observations that describing a voice produces a verbal overshadowing effect and results in fewer correct identifications (Perfect et al., 2002; Vanags et al., 2005). A time delay between the voice description and the recognition task did not seem to prevent this effect.

Although the number of correct identifications seemed relatively unaffected by the type of interview, the interviews with questions about the voice seem to have had effects in other ways. The adults in the BI condition who did not answer any questions about the voice made significantly fewer false identifications compared with witnesses in the other two interview conditions (where they were asked to give a description of the voice). It may be interpreted that instructions to describe the voice may increase the willingness to make an identification, although without an increase in accuracy. Perhaps describing the voice makes the participants think that they have a better memory of the voice than they really have. It should be acknowledged, however, that a TA condition was not included, so whether the same tendencies would also appear in such situations is yet to be examined.

In sum, interview type was found to affect earwitness recall more than voice recognition. Regardless of interview type, both children and adults performed poorly in terms of voice identification and an interview shortly after the witnessed event did not seem to help strengthen the memory. The findings advance previous knowledge on earwitnesses' performance in two ways. First, the CI, in the case of the adults, was found to be beneficial for recalling the content of a brief conversation. Second, the checklist used by the Swedish Security Service had no positive effects on either recognition or recall. On the contrary, the checklist was found to produce more false identifications compared to a "standard" police interview, and less recalled information compared with the CI. The combined evidence shows the importance of scientifically evaluating the methods that are currently used by the police. Further, it is evident that there is a need to develop an interview technique specially designed for earwitnesses, with the goal to enhance both child and adult witnesses' voice recognition performance.

Is Performance Better under Certain Conditions?

As the overall performance in Study I, II, & III was so poor and an interview did not seem to improve the outcome, the aim of Study IV was to find out if the reliability of voice lineups might be higher under certain other conditions. Therefore, the effects of the perpetrator's tone of voice and time delay on voice recognition accuracy were tested.

Change of Tone

The perpetrator used in the first three studies altered his tone of voice during the telephone call (encoding). Initially he was somewhat angry, but later on he spoke more quietly in order to avoid being heard by others. He may also have been perceived as stressed since he wanted to finish the conversation as quickly as possible. It was a deliberate choice to use a conversation in which the perpetrator altered the tone of the voice, because one may assume that it is rather common to sound upset or angry when engaging in a conversation on criminal matters. When the participants returned two weeks later, and were confronted with a voice lineup, several of the participants reported that they thought it was difficult to recognize the voice in the lineup since the voice they remembered had talked in a different tone of voice. Hence, this might be a factor contributing to the overall low performance level. Therefore, in Study IV we manipulated the perpetrator's tone of voice and made two new recordings (with the same speaker and exact same content). To ensure differences between the two versions, the actor was instructed to sound angrier than in the version previously used, and to sound as neutral as possible in the other condition.

Contrary to the prediction and past research (e.g., Saslove & Yarmey, 1980), there was no difference in terms of voice identification accuracy between witnesses who heard the perpetrator speak in an angry tone at the to-be-remembered-event and in a normal tone in the lineup and those who heard him speak in a normal tone on both occasions. At first sight this may be viewed as a positive result, as it suggests that a voice lineup may still be recommended even though a perpetrator has spoken in an angry tone during the witnessed event. However, it becomes less encouraging when considering the low accuracy level for both the congruent and incongruent condition. Put differently, the finding that the witnesses performed at chance level when the tone of voice had changed from angry to normal is in fact in accordance with past research on adult witnesses (e.g., Saslove & Yarmey, 1980). The difference is that those who heard the perpetrator talk in a normal tone on both occasions also performed at chance level, which is lower than usually found in previously studies (e.g., Saslove & Yarmey, 1980; Yarmey, 2001).

The Effect of Delay

In real-life investigations it is not possible to conduct a voice lineup immediately after the critical event. However, controlled experimental studies may give insight into how rapidly the memory of a voice decays. Therefore, an immediate condition was included in Study IV and it was expected that witnesses tested immediately would perform better than those tested after two weeks. As partial support for the prediction, time was found to affect voice lineup accuracy for children, such that children tested shortly after exposure made more correct identifications compared to those tested after a two week delay. The same effect was however not found for adult witnesses. The fact that adults performed worse than chance level, although tested almost immediately after hearing the voice, was very much unexpected.

Past research has shown that time delay may have a negative effect on the number of false identifications (Yarmey & Matthys, 1992), however, this was not supported by Study IV. Instead the trend was, although not significant, that both children and adults were more likely to reject the lineup if tested after a delay. This finding makes sense considering that a witness is more uncertain of their memory when some time has passed and therefore is more reluctant to make an identification. However, it should be acknowledged that a TA condition was not included in Study IV, so it is unknown whether a time delay would have had an effect on false identifications when the target was not present.

A Combination Effect?

Although the witnesses performed poorly, regardless of being tested immediately or with a congruent tone of voice, as expected, both the children and the adults in the *congruent-immediate* condition performed the best. However, they did not perform significantly better than the witnesses in the other conditions, and only 25% of the children and 19% of the adults made a correct identification. These accuracy levels are remarkably low considering the circumstances, and much lower than shown in previous research (e.g., Saslove & Yarmey, 1980; Yarmey, 1991b). It should be noted that the pattern in terms of correct identifications across conditions for the children follows a logical pattern, as those in the *congruent-immediate* condition performed best, followed by the *incongruent-immediate* condition, and those in the delayed conditions performed worst. In contrast, for the adults we found a rather random pattern, as those in the *incongruent-delayed* condition performed second best and were not significantly different from the *congruent-immediate* condition, and none in the *incongruent-immediate* condition could accurately identify the perpetrator.

In conclusion, although somewhat mixed results, it is probably safe to say that memory for unfamiliar voices will decline over time. Though importantly, an immediate test will not ensure high accuracy. Unexpectedly, a change of tone between the witnessed event and the lineup did not result in a negative effect. However, the most important finding was the overall low accuracy level and especially the poor performance by those tested under what would seem a rather favourable condition (*congruent-immediate*).

Few Differences – A Matter of Methodology?

It is now clear that almost none of the manipulated variables significantly affected identification accuracy in the present thesis. Performance was at chance level regardless of condition. How should this be interpreted? Does this reflect people's voice recognition capacity or is it a matter of methodology?

The goal with the present thesis was to use an ecologically valid setup that could mirror the procedure of a real-life investigation, both at encoding stage and in the lineup situation. Hence, it could be argued that the method used made it impossible to find significant differences. For example, retention intervals shorter than two weeks have shown better results. However, a shorter retention would not be possible in real-life investigations. Though, an immediate test was used in Study IV, and although it resulted in higher accuracy compared to after two weeks, the accuracy levels were not very high. It should be noted though that the test was not precisely immediate, because the participants were first interviewed about the voice before they were presented with the lineup. This is relevant because describing the voice has been found to possibly impair a subsequent recognition test (e.g., Perfect et al., 2002; Vanags et al., 2005).

In the present thesis the witnesses only overheard a criminal conversation, i.e., the conversation was not directed to the listeners and they did not actively speak to the perpetrator. Such a setup was used because children were tested. To simulate a realistic overhearing situation, the perpetrator's voice was heard for a rather short amount of time and the witnesses were unprepared for a later voice recognition task. A longer duration and prepared witnesses actively speaking to the perpetrator might have resulted in higher accuracy and perhaps more differences.

For the selection of foils, a voice similarity test was conducted with a separate group of listeners. Following recommendations, we selected two voices that were perceived to be quite similar, two moderately similar and two rather dissimilar (Hollien, 2002). Higher accuracy could possibly have been found if the foils were more dissimilar. Further, as only one person served as the perpetrator it might be that the chosen perpetrator had a voice that was too usual or non-distinctive, which made recognition very difficult. It would have been interesting to compare the results to a second voice. However, the distinctiveness of voices is impossible to control in real cases.

Hence, it might be argued that a setup with more facilitating conditions would have resulted in higher accuracy levels and allowing more differences to emerge between the manipulated variables. As the aim with the present thesis was to mirror a real-life situation, differences that do not pertain to those tested under ecologically valid conditions were not of interest. From this perspective, the answer to my initial question is that it seems as if people are not simply better when tested under reasonably realistic conditions. Considering the poor performance, I believe that the next step within this area should be to study more closely how the voice memory works without an applied perspective. The aim should be to investigate if the poor voice memory is due to a failure of encoding, storing, retrieval, or a mix of these stages. Such understanding could then contribute to the development of methods that could enhance voice identification accuracy.

Analyzing False Identifications

Besides correct identifications, it is important to consider the selection of foils, i.e., false identifications. In the following analysis a distinction will be made between the "target", referring to the target voice in the mock incriminating phone conversation (at encoding), and the "suspect", referring to the target voice in the lineup.

People perceive speech through different acoustic cues that they weigh together (e.g., Nittrouer, Manning, & Meyer, 1993). In Study I, significant correlations between the acoustic cues *Articulation rate, Pitch level* and *Pitch variation* and false identifications were found for the children. The youngest children were more frequently found to choose foils that spoke with a higher articulation rate, greater pitch variation and lower pitch level. Similarly, for the older children articulation rate and pitch level were found to correlate with false identifications (with the same direction of dependence). In contrast, the adults' false identifications did not correlate with any of these acoustic cues (neither in Study I, nor in Study II). These results parallel findings in studies of speaker similarity, where the weight of such cues have been found to vary with age, referred to as the Developmental Weighting Shift (Nittrouer, Crowther, & Miller, 1998; Nittrouer et al., 1993). The pitch cue has been found to be particularly strong for the youngest children (Petrini & Tagliapietra, 2008). It is suggested that this age-difference is a result of developmental changes in speech perception where children, as they become more familiar with their native language, change the weight they assign to the various acoustic properties (Nittrouer & Miller, 1997). Younger children tend to attach more weight to global cues like speaking rate and intonation than older children and adults (e.g., Fowler, 1991; Nittrouer, 2006). An analysis of the target's voice showed that the articulation rate was higher than that of any of the foils (the suspect’s articulation rate was also among the fastest), but for the other two cues the target (and the suspect) were in the middle of the total range. Both articulation rate and pitch are relevant factors when assessing speaker similarity. To rely on them in the rather stereotypical manner that the children seem to have done is, however, not an efficient strategy. In fact, to rely heavily on these cues might have hindered performance, as the target spoke in a different tone at the encoding situation compared to the tone of the suspect in the lineup (except Study IV).

Although none of the acoustic cues correlated significantly with false identifications for the adults, one particular foil was selected much more often than the others. This foil was present in three of the studies and when present, more than half of the adults who made false identifications selected this particular foil (50% in Study I, and 54% in Study II & IV). Why this voice attracted so much attention is an important question, since such a bias may severely impair recognition accuracy. Great care was taken to ensure a fair foil composition, and the voices for the lineup were chosen after a perceptual evaluation test taken by two separate groups of students. In this test, this particular foil was only evaluated as moderately similar to the target voice. However, the students compared the voices with a sample representative of the speaking style used in the lineup (where he spoke in a normal conversational tone). At the encoding situation he talked in a more stressed and upset manner. An analysis of the voice of the most often selected foil showed that his voice was most similar to the target with respect to articulation rate and pausing.

The group of listeners that heard the suspect’s voice (at the voice similarity evaluation test) estimated instead that a speaker with a relatively low articulation rate sounded most similar. This foil replaced the target voice in the TA condition in Study I, and in Study III he replaced the most identified foil in Study I & II. However, this foil was only chosen by 7% of the participants in Study I. As mentioned, the answer to this discrepancy between the voice similarity evaluations test and foil selections might be found by looking at intra-speaker differences, namely the two groups compared the target voice from two different contexts and emotional states. In line with this, only 15% identified this foil in Study III. Somewhat puzzling though was that another foil, judged as second most similar to the suspect now attracted no less than 45% of the false identifications. This might be due to the exclusion of the otherwise most identified voice, but why this foil now attracted more attention, rather than the voice judged as most similar to the suspect, is not clear.

As intra-speaker variability seems to be of importance, it could be that foil selection differs between the witnesses in Study IV that heard the perpetrator speak in an angry tone and those who heard him speak in a neutral tone. When hearing the angry tone, most children and adults chose this particular fast talking foil (children 38%, and adults 40%, respectively). However, even more adults chose the fast talking foil (67%) in the neutral tone condition, while most children chose the foil judged as sounding most similar to the suspect (37%) (though, note that this is not the same as the most similar voice that replaced the suspects voice in the TA condition in Study I). For this I have no explanation.

Since the target voice spoke with the highest articulation rate, one can't help wondering if this may be one explanation why both groups of children more often chose speakers with a high articulation rate, and why the adults most often chose the speaker with the highest articulation rate. Put differently, the reason for this might be that the participants remembered the target as a fast talker. If this reasoning holds true, it would give more credibility to the participants' ability as earwitnesses. In line with our finding, Zetterholm, Sarwar, and Allwood (2009) found (using a TA lineup), that the foil most similar to the target voice with respect to articulation rate and pitch, also turned out to be the foil that was chosen most often. However, it is equally possible that the participants in the present studies, as well as the earwitnesses in the study by Zetterholm et al. (2009) chose those speakers because the higher articulation rate attracted their attention in general. Such factors, which may severely distort a recognition task, should be given more attention in future research. Some recent studies have tested other voice characteristics that may affect earwitness memory, such as high-pitch vs. low-pitch voices (Mullennix et al., 2010) and high-typical vs. low-typical voices (Mullennix et al., 2011). It has been found that high-typical foils and high-typical targets are more often confused with each other than low-typical voices. It has therefore been suggested that the perceived typicality of the suspect's voice and the voices of the foils should be taken into account when constructing lineups (Mullennix et al., 2011).

Taken together, it seems as if intra-speaker variability is an important factor, but that inter-speaker variability also plays an important role. Further, the composition of voices in the lineup might affect foil selections as well as the specific voice to-be-remembered (for a further discussion, see the section about Voice distinctiveness). It is evident from the present thesis that articulation rate is an important factor to consider when conducting voice lineups. Furthermore, a practical implication from Study I is that for lineups intended for child witnesses one should make sure that pitch level, pitch variation and articulation rate are reasonably similar between the voices in the lineup.

Real Witnesses’ Identification Tendency

The results of the present thesis and previous research have shown that, when confronted with a lineup, people are prone to make an identification. However, not much is known about how well these laboratory-based findings mirror the behaviour of real witnesses. In studies examining real eyewitness cases, it has been found that about 40% of the witnesses identified the suspect, around 20% of the witnesses identified a foil and further, 40% of the witnesses did not make an identification (Valentine, Pickering, & Darling, 2003; Wright & McDaid, 1996). These results concern facial lineups and, to my knowledge, real earwitness cases have not been examined. However, these numbers give at least an indication of the potential outcomes of real-life investigations. It seems as if real witnesses are more reluctant to make an identification when confronted with a lineup compared to mock witnesses in experiments. Awareness of the possible severe consequences a false identification may have in a real case may be an explanation for this discrepancy. This suggests that it is necessary for future research to try to create experimental high-stake situations to better reflect the conditions of real life cases.

Can We Trust Earwitnesses Subjective Confidence?

If an earwitness identifies the suspect, a crucial question is if it is possible to establish whether the identification is correct or not. To answer that question, researchers have mostly focused on witnesses’ subjective confidence in their decision. Most research has shown a low correspondence between earwitnesses’ subjective confidence and identification accuracy (e.g., Olsson, Juslin, & Winman, 1998; Yarmey, 2001), and have highlighted that judges and jurors often rely heavily on witnesses’ confidence when assessing the reliability of a testimony (e.g., Cutler, Penrod, & Stuve, 1988; Solan & Tiersma, 2003). Further, in line with eyewitness research, post-identification feedback has been found to inflate earwitnesses’ confidence in their decision (Quinlivan et al., 2009). This thesis is no exception, as the confidence-accuracy relationship was found to be weak for all four studies and across conditions. However, the weak relationship is not a particularly surprising result. A very low accuracy rate (as was the case in all studies) leaves little room for a high correspondence between accuracy and confidence. Though, confidence assessed after an identification does not seem to be a good predictor of accuracy.

To further investigate to what extent witnesses are good at predicting their voice recognition ability, the participants in Study IV were asked if they thought that they would be able to recognize the voice if they had the opportunity to hear it again. It was found that when asked shortly after being exposed to the voice, the majority of both children and adults believed that they would be able to recognize the perpetrator's voice. However, only a small fraction of those optimistic witnesses could actually identify the perpetrator's voice in the lineup. Such unrealistic optimism can have important implications in real-life investigations. That is, a witness who expresses optimism in terms of being able to recognize the perpetrator's voice may have the result that the investigator decides to expose the witness to a lineup. If the memory of the voice is worse than predicted, and the data indicate that this is often the case, the witness may identify an innocent suspect or miss the actual perpetrator and this might set the investigation off on the wrong track. Thus, the finding extends previous research by showing that witnesses' wrongful estimations of their voice recognition ability may not only be a potential problem in court, but also in the investigative phase.

Why do Earwitnesses Perform so Poorly?

As already established, earwitnesses seem to perform poorly at identifying a once-heard voice when tested under rather realistic conditions. Though, I believe that it is important to clarify that this is not to say that humans are bad at voice recognition. From an evolutionary perspective, the recognition of *familiar* voices has been found to be very important as it has contributed to survival (Sidtis & Kreiman, 2012). As an example, infants are found to already be able to discriminate the voice of the mother at birth (e.g., DeCasper & Fifer, 1980; Mehler, Bertoncini, Barriere, & Jassik-Gerschenfeld, 1978). That ability is suggested to be essential for mother-infant bonding which is of utmost importance for the infant's well-being and survival (e.g., DeCasper & Fifer, 1980; Sidtis & Kreiman, 2012). On the contrary, in line with the present thesis, research on *unfamiliar* voices has shown that children under the age of 10 are poor at discriminating between unfamiliar voices (Clifford & Toplis, 1996; Mann et al., 1979; Peters, 1987). Further support for the superiority of familiar voices comes from the fact that the processing of familiar and unfamiliar voices are suggested to be independent. Studies on individuals with brain-damage have shown that an injury to the right hemisphere impairs the ability to recognize familiar voices, whereas an injury to either hemisphere impairs the recognition of unfamiliar voices (e.g., van Lancker & Kreiman, 1987). Therefore, individuals with a focal brain damage are able to recognize familiar voices, although they are unable to discriminate among unfamiliar voices. Hence, from an evolutionary perspective, a reasonable conclusion is that discriminating familiar voices from unfamiliar voice is of utmost importance for bonding, as well as for deciding if the person is a friend or foe. However, discriminating between different unfamiliar voices might not have been as crucial for our survival. Although the accuracy rates for familiar voices are not perfect, research supports the superiority of familiar voices as they are found to be far more accurately identified than unfamiliar voices (Yarmey, Yarmey, & Yarmey, 1994; Yarmey et al., 2001). As a voice also signals the emotional state of the speaker (e.g., Sidtis & Kreiman, 2012) it may be reasonable to speculate that the most important aspect to perceive when confronted with an unfamiliar voice is the speaker's emotions. To reflect on the emotional state is helpful when deciding if the person is friendly or hostile. That might be an explanation as to why most participants in the present thesis seemed to focus on the perpetrator's emotions (see the section Voice Descriptions).

Although the recognition of familiar voices has, from an evolutionary perspective, been of most importance, it is still interesting to discuss why we are so poor at recognizing unfamiliar voices. Except for the superiority of familiar voices, there are many possible explanations as to why it is so difficult to correctly identify an unfamiliar voice. In this section I will first discuss the differences between auditory and visual processing and then three possible explanations for the poor performance relevant for the present thesis.

Auditory Processing vs. Visual Processing

Pictures have been found to be much better remembered than words (Standing, 1973). A suggested explanation for this is that pictures (vs. words) are more likely to be encoded and stored both verbally and as mental pictures (Paivio, 1971). Further, pictures need to be meaningfully understood before they can activate a label, whereas words activate a label without requiring semantic processing (Nelson, Reed, & McEvoy, 1977). Semantic encoding requires a deeper processing and is found to be a more effective memory mnemonic (e.g., Craik & Tulving, 1975). The superiority of the visual modality is also found in studies where the memory for lists of words is tested. False memories of words are less likely to occur if the items studied are pictures instead of written words (Israel & Schacter, 1997), and the production of false words is higher when the words are presented auditorily compared to visually (Smith & Hunt, 1998). Relating this to voice memory, a voice may be very difficult to encode as a pictorial image and a lack of terminology may make verbal encoding difficult or even impossible for voices heard on one occasion. In line with this, it has been found that people are much better at recognizing faces than recognizing voices (Hanley, Smith, & Hadfield, 1998; Hollien et al., 1983; Stevenage et al., 2011; Yarmey et al., 1994). It is suggested that there is a selective interference effect towards prioritising visual input over acoustic input (Stevenage et al., 2011). In addition, even if the quality of visual observation condition declines, people do not seem to automatically shift attention from visual cues to auditory cues (Yarmey, 1986). Hence, people do not seem to be accustomed to using the voice as an identification cue when discriminating among unfamiliar voices. This may be one additional explanation as to why people are found to be poor at remembering a once-heard voice. One may then assume that blind people, who have not learned to rely on visual cues, should be particularly good at voice recognition. However, blind listeners as a group have not been found to be better at identifying unfamiliar voices, indicating that mere exposure to only voices may not be sufficient in order to perform better at voice identification (Elaad, Segev, & Tobin, 1998). Phonetic experts are found to be better at speaker identification than naïve listeners (Elaad et al., 1998; Schiller & Köster, 1998). It is suggested that professional voice identification experts are better than lay people because their extensive practice may have given them the ability to encode voices in a retrievable form in long-term memory, an ability that lay people lack (Elaad et al., 1998).

Level of Preparedness

A factor especially relevant for the present thesis that may explain the poor performance is the level of preparedness. Although the participants were told that they would overhear something behind the curtain, they were unprepared for a later voice identification task. In fact, many participants were surprised when they were confronted with a voice lineup, and several explicitly reported that they thought that the subsequent interview would concern the content of the conversation. None of the participants reported that they believed that they would be expected to remember the voice. Most previous research has found that unprepared (vs. prepared) witnesses perform more poorly in a voice identification task (Armstrong & McKelvie, 1996; Hollien, Huntley, Künzel, & Hollien, 1995; Saslove & Yarmey, 1980; Yarmey, 2003;

although see Perfect et al., 2002, for no effect of preparation). In relation to the basic memory processes; attention, rehearsal and deep processing are important (e.g., Mulligan & Brown, 2003; Reisberg, 2010). Being unprepared might have had the effect that the voice was not given enough attention and was therefore not properly encoded. Further, because the participants were unaware of the later voice recognition test, they most probably did not think of or rehearse the sound of the voice. As a consequence, the voice was perhaps never encoded into long-term memory. It should be noted that the setup in the present thesis reflects what is happening in most (but not all) real life situations; people who become victims or witnesses are seldom prepared (Clifford, 1980). Further, in most real cases the exposure to the voice may be too brief and occur too quickly for the listener to formulate an intention to memorize the voice (Read & Craik, 1995).

Passive Listening

Another possible explanation for the poor performance in the current thesis is that the conversation was overheard, i.e., not directed to the listener. The reason for choosing this particular setup was that children as young as 7–9-years served as participants in Study I. To expose young children to a more personal "threat" would be problematic for ethical reasons. Past research, though, suggests that actively talking to the person to be remembered leads to higher identification accuracy compared to passively listening to the perpetrator (Hammersley & Read, 1985). However, actively speaking to someone does not necessarily guarantee high voice recognition. In a study where witnesses spoke to the target in person or over the phone, less than one third were able to accurately identify the voice although tested after only two minutes (Yarmey, 2003). One explanation for the diverse findings might be the duration of conversation. The study that showed high identification accuracy when actively speaking to the target used a rather long conversation (five minutes) compared to the study that showed low accuracy levels for witnesses who had spoken to the target (30 seconds).

In addition, some recent studies have shown that listeners rarely notice if the voice is being replaced by another voice. Such a change deafness has been found under different levels of attention and cognitive demands, such as passively listening to the voice (Sauerland, Sagana, & Otgaar, 2012), repeating the words said by the voice (Vitevitch, 2003), and when actively engaged in a conversation with the speaker (Fenn et al., 2011). This suggests that listeners may not automatically encode the conversational content, as well as voice characteristics, when listening to speech. Hence, attention may be focused on what is being said. This might be an additional reason for poor voice recognition (see further discussion in the section Memory for Content).

Voice Distinctiveness

The low accuracy level may not only be explained by poor memory ability. The to-be-remembered voice *per se* may influence subsequent lineup decisions. Past research indicates that certain voices are more recognizable than others (e.g., Clifford et al., 1981), even when being assessed as equally distinctive (Philippon, Cherryman, Bull, & Vrij, 2007). The perpetrator's voice in the present thesis may fall into the category of less recognizable. Unfortunately, such individual variability in voices is impossible to control for in real-life criminal cases.

Research has shown that stereotypes about a person's occupation can be formed from voice alone and further, people can form the impression that a person is a criminal by only listening to their voice (Yarmey, 1993). In line with this reasoning, a suspect that sounds like a stereotypical criminal might run the risk of being chosen because of that and not because of a genuine memory for the voice. If a foil is instead perceived as having the voice of a criminal, this might cause the suspect to escape detection (Nolan & Grabe, 1996). It is therefore suggested that before the administration of a lineup, objective listeners should be asked to judge whether some of the voices deviate from the others, or sound like the voice of a criminal (Nolan et al., 2009). Although the selection of foils in the present thesis was based on the outcome of a similarity test completed by objective listeners, they were not asked to identify criminal sounding voices. Therefore, a possible reason why one particular foil was chosen so often in the present thesis (Study I, II, & IV) could be that he was perceived as sounding like a criminal (see also the section Analyzing False Identifications).

Memory for Content

The criminal conversation used in the present thesis was a telephone conversation where only one of the parties was heard. The duration was rather brief (40 s) and the content was not very rich or detailed in nature. Based on that, the content must be considered as rather abstract and hard to remember. Further, the memory was tested by free recall. Furthermore, no distinction was made between verbatim reports and reports rephrased in the subject's own words, but semantically equivalent to the original utterance. When taking all of the above into account, the overall memory performance could be viewed as rather reasonable. Though, the performance was highly influenced by age and delay.

As evident from the results of Study III and in accordance with past research (Ling & Coombe, 2005; Saywitz, 1987), children tend to remember and report less correct details of a heard conversation than adults. The same pattern was found in a study by Öhman, Eriksson, and Granhag (2012) that is based on the same data as Study I. The number of correctly reported details, after a two week delay, increased significantly with increasing age (7–9: 4% correct, 11–13: 11% correct, & adults: 23% correct, respectively). This parallels eyewitness findings where young children (under 10) have been found to have more difficulty in matching the performance of older children and adults when it comes to free recall with a high demand on detailed knowledge (e.g., Cole & Loftus, 1987; Saywitz, 1987). Basic memory research has shown that prior knowledge and understanding to a great extent determine what we can and cannot remember (Gordon et al., 2001; Reisberg, 2010). In essence, things we understand well are better remembered (Davies, 1996). The extremely poor memory shown by the youngest children (4% correct) might be explained by a low degree of comprehension. The abstract content may have resulted in that they did not completely understand what the perpetrator said. In fact, only 37% of the younger children expressed an awareness that the conversation was of a criminal nature, whereas 83% of the older children and all of the adults did so (Öhman, Eriksson et al., 2012). Further, although asked about what the man behind the curtain had said two weeks earlier, many of the younger children confused the to-be-remembered content with what the perpetrator and the foils said in the lineup. More positively, neither the children nor the adults reported many fabrications. This might be due to the fact that they were explicitly told not to guess and that it was okay to say "I don't know" or "I don't remember".

An interesting finding is that the older children performed better at the voice identification task than the adults (Study I), while the opposite was found for recall memory (Öhman, Eriksson et al., 2012). One might speculate that the adults focused more on the content of the conversation than on the voice, as they assumed that they would be asked questions about what the perpetrator said. Studies using functional neuroimaging have shown that different brain regions are activated when focusing on the speaker's voice compared to the verbal content (e.g., Relander & Rämä, 2009; von Kriegstein et al., 2003; von Kriegstein & Giraud, 2004). Both the content and the voice recognition task activated auditory language areas, but each task also activated regions that were not active during the other task. Further, the voice region was not activated during the verbal task, which suggests marginal analysis of vocal features when focusing on the content (von Kriegstein et al., 2003). This implies that in real-life investigations it might be of importance to ask earwitnesses if they focused on the content or the voice during the crime.

All interviews in Study III included questions about what the perpetrator said, as a police encounter shortly after the event would most certainly include such a question. However, the scenario in the Öhman, Eriksson et al., study (2012) was that the statements were not gathered until two weeks after the event. When comparing the results from that study with the baseline interview (BI) and the Swedish Security Service interview (SSSI), (which are identical to the questioning procedure for memory of content used in Öhman, Eriksson et al., 2012), it is evident that earwitnesses' statements should be collected in close connection to the crime. When asked immediately, the adults correctly remembered 39–40% (Study III) of the perpetrator's account, compared to 23% when the statement was taken two weeks later (Öhman, Eriksson et al., 2012). This is in line with previous research that has shown that adults in an auditory-only condition show a great decrement in memory performance when tested after a delay (Campos & Alonso-Quecuty, 2006; Toglia et al., 1992). The effect of delay was even more evident for the older children (11–13-year olds) as they could report 30–31% correct information when asked immediately (Study III) compared to only 11% when two weeks had elapsed (Öhman, Eriksson et al., 2012). As shown previously for both adults (Boydell & Read, 2011) and children (Ricci & Beal, 1998), the opportunity to give an immediate statement (Study III) resulted in better memory after two weeks (adults: 28–30% correct; children: 20–22% correct) compared to reports collected after two weeks only (Öhman, Eriksson et al., 2012, adults: 23% correct; children 11% correct). In addition, the younger children, who were only tested after a two week delay, merely remembered 4% of the content correctly.

To sum up, three main findings regarding memory for content can be derived from the present thesis; (1) adults recall and report more details of an perpetrator's account than children, (2) it is important that earwitnesses' statements about their verbal observations are gathered as soon as possible after the crime, and even more so for child witnesses, and (3) focusing on the content can impede voice recognition.

Voice Descriptions

Although person descriptions are thought of as vague and non-discriminative (Meissner, Sporer, & Schooler, 2007), eyewitnesses have been found to describe around 10 attributes of an offender (Sporer, 1996), which is around twice as many as usually found for earwitnesses (e.g., Yarmey, 2003). Therefore, in an attempt to enhance voice descriptions, the use of global open-ended questions was investigated in the present thesis (Study III & IV). Unfortunately, such questions did not seem to help witnesses to give a richer description as, in line with past research (e.g., Yarmey, 2003), the witnesses generated few and very vague descriptions (e.g., dark, normal, dialect). In addition, less than half of the descriptions could be categorized as an actual description of the voice. Although personal characteristics such as dialect, age and nationality may be important for establishing a perpetrator profile, it is noteworthy that these are the features that the witnesses report when explicitly asked to describe the *voice*. Furthermore, in both study III and IV, many of the descriptions could be categorized as situation dependent (e.g., angry, stressed, aroused) and may therefore be of limited value. This is in contrast with person descriptions as they are found to contain more permanent features (e.g., height, skin colour) than temporary features (clothing) (van Koppen & Lochun, 1997). A plausible explanation for the vague and meagre voice descriptions may be that we are not used to describing voices. Further, people are not familiar with the terms with which relevant features of the voice may be described (Broeders & Rietveld, 1995). Both Study III & IV showed that the adults reported significantly more descriptions than the children. The same age-related pattern is found for person descriptions, and a suggested reason is that children have a less developed linguistic ability and therefore a smaller vocabulary to describe people (Pozzulo, 2007). The same explanation will most likely hold true for voice descriptions.

An interesting finding was that, although using the exact same open-ended global questions, both adults and children reported fewer descriptions in Study III (children: M = 2.7; adults: M = 4.7) than in Study IV (children: M = 4.6; adults: M = 6.4). The finding is even more surprising when taking into account that the questions in Study III were embedded in a Cognitive Interview (CI), which has been found to generate more detailed descriptions from eyewitnesses (e.g., Memon et al., 2010). Another difference was that questions about what the perpetrator said preceded the description of the voice in Study III, whereas the situation was the opposite in Study IV. This might be a possible explanation as to why the CI did not have an enhancing effect on the voice memory in Study III; there may have been too much focus on the content before the participants were asked to concentrate on the voice. The order was therefore changed for Study IV. Although identification accuracy did not seem to be improved by the sequence of questions (see discussion of voice identification), it seems as if the order had an impact on the amount of reported descriptions. It would have been interesting to examine if memory for content also showed an order effect. Unfortunately, such a comparison cannot be made as the content memory for Study IV has not yet been transcribed and coded.

As for the use of specific questions, a relatively high agreement was found between the witnesses' perception of the voice in the scale-rating interview (Study IV). This finding is rather encouraging. That is, if voice perceptions do not vary too much between witnesses, this means that the use of scale based descriptions may play an important role for speaker profiling and lineup construction. However, the same high agreement was not found for the checklist used by the Swedish Security Service (Study III). The exact same set of questions was not used, which is one obvious possible explanation for the discrepancy. Another potential explanation may be the different response formats. The specific questions in the checklist were asked in a yes/no-format instead of ratings on a scale. As some of the attributes changed over the phone call (e.g., talking loudly in the beginning but not in the end), participants expressed some difficulty in giving a simple yes/no answer. Thus, on a more fine-grained scale (as a 6-point scale compared to yes or no), it is possible to take that into consideration and choose an alternative that best represents the overall perception. Hence, it seems as if the response format for specific questions is of importance, and at least for high agreement, the use of a scale seems to be preferable.

However, to decide whether to use a free description or specific questions is not that easy. Free recall is often accurate because witnesses only report features that they have a rather clear memory for, though, resulting in rather few descriptions. Further, verbally describing a stimuli has been found to impair subsequent recognition performance, an effect termed "verbal overshadowing" (Schooler & Engstler-Schooler, 1990). In line with eye-witnesses, earwitness research using a free recall has found a similar effect (Perfect et al., 2002; Vanags et al., 2005). The use of a checklist counteracts the problem with the lack of a proper vocabulary and therefore offers more complete descriptions. Though, on the other hand, a negative effect is that it might result in listeners answering questions which they did not attend to at encoding and therefore also result in more incorrect answers (Meissner et al., 2007). Hence, it is evident that memory enhancing techniques for eliciting more detailed and accurate voice descriptions are needed.

The Beliefs Held by Swedish Judges and Lay Judges

As legal practitioners make important assessments concerning the reliability of earwitness testimony, it is important to know how well their beliefs correspond with research findings. Therefore, a survey examining the beliefs held by Swedish judges and lay judges about a number of variables tested in the present thesis was conducted (Öhman, Granhag, & Eriksson, 2012). General questions showed that an earwitness' verbal statement (memory for conversation and sounds) was believed to be more commonly referred to in court than voice identification. This is interesting since most research within this domain has focused on earwitness identifications. The legal practitioners were further asked how much weight they would give a voice identification. Although on the whole very low, lay judges were found to give significantly more weight to a voice identification compared to professional judges. As for the variables examined in the present thesis, the survey showed that the expressed beliefs were in line with research findings concerning the negative effect of time delay and change of tone (although we did not find an effect of tone of voice, previous research indicates a negative effect), and that it is more difficult to correctly identify a voice compared to a face. Though, a majority incorrectly believed that hearing a voice through a mobile phone would have a negative effect on performance. Further, a "no opinion" alternative was the most frequently indicated for the correspondence between accuracy and confidence, how good people are at describing voices, and how well children (11–13-year-olds) perform in relation to adults for both voice identification and memory for content. That professional judges tend to be careful in taking a stand is in line with previous studies examining Swedish judges' beliefs about eyewitness testimony (Granhag, Strömwall, & Hartwig, 2005). In the district court, professional judges and lay judges decide together on the outcome, for the question of guilt as well as for sanctions (Stridbeck & Granhag, 2010). This might be an explanation for their hesitation to take a stand, as in their profession it is crucial that information is carefully evaluated. The consequences of wrongful decisions can be serious. It would be of interest to examine whether the beliefs of other legal professionals would be similar to that of judges, such as police officers, and especially prosecutors as they (in Sweden) have comparatively great power to decide when and when not to prosecute.

In sum, this rather small survey indicates that Swedish judges and lay judges seem to have rather limited knowledge about earwitness testimony and the factors that might moderate its accuracy. This is of course unfortunate and a problem for the legal system. Hence, it is important to try to educate legal professionals about the reliability of earwitness testimony.

Limitations

The current thesis aimed to test earwitnesses under ecologically valid conditions. However, all features of a real crime are not possible to simulate in an experimental setting. Examples of such features are the effect of stress and arousal that real witnesses might feel. Further, participants in an experiment are probably not as engaged in the task as a real witness would be, because of the lack of personal importance. In addition, the experimental nature of the study might lower the decision criterion for making an identification as there are no consequences for a wrongful answer. In fact, real eyewitnesses are found to be more reluctant to make an identification (Valentine et al., 2003; Wright & McDaid, 1996). Hence, the results of the present thesis do not exactly mirror what might be expected in real life cases.

One might argue that it was an advantage that the same general setup and method was used in all studies, as it opened up for a number of interesting comparisons. However, as only one voice was used for the perpetrator the findings cannot be generalized to voices that vary in distinctiveness. A further possible limitation is that the most selected foil in Study I was also included in the lineups in Study II & IV. Since the same pattern was found in those studies, it might be the case that the inclusion of this particular foil, to some extent, masked any “true” effects of the manipulated variables.

The aim of Study II was to test if mobile phone sound quality caused by technical factors had an impact on voice identification accuracy, therefore other factors connected with mobile phone speech, like speaking style, were not included. This might be seen as a further limitation. The direct and mobile speech samples were recorded simultaneously which most likely had the result that the speakers did not feel as though they were speaking in a mobile phone and therefore did not adjust their speech in the way that they might have done otherwise. One factor that often influences mobile phone speech is background noise, which in the normal case causes the speaker to talk more loudly (Byrne & Foulkes, 2004; Foulkes & Barron, 2000). Though, there was no background noise since the recordings were made in a quiet room.

In Study III all interviews started with questions about what the perpetrator said. It might be possible that focusing on the content worsened the memory for the voice. Therefore, it would have been interesting to include a control group that received no questions about the perpetrator's account. However the rationale for systematically including the question on what the perpetrator said is that in real-life investigations such a question is highly relevant to the police. Though, the order of the questions (content vs. voice) could have been manipulated. Further, as we did not include a control group in Study IV (that would receiving no questions about the voice), we cannot rule out the possibility of a verbal-overshadowing effect. We can only speculate whether stronger effects of tone of voice and time delay would have emerged if examined without having had to describe the voice. It was decided to use only description conditions because a lineup is unlikely to be conducted unless the witness has provided some information about the voice and further, it is suggested that all witnesses should be interviewed about the voice to enable a voice profile of the offender (Broeders & Rietveld, 1995).

The present thesis only included a TA condition in Study I. As pointed out from eyewitness research, it is not possible to generalize findings from TP to TA outcomes, and vice versa (Clark, Howell, & Davey, 2008). Therefore, we cannot draw any conclusions about whether the examined variables in Study II, III, & IV affect earwitnesses' decisions when the perpetrator is not present in the lineup.

Legal Implications and Future Directions

After testing a total of 949 witnesses under a number of different conditions, the message is clear; voice identification under reasonably realistic conditions is a highly difficult task.

As one aim of the present thesis is to inform practitioners, some practical recommendations should be discussed. First of all, one has to consider that, as mentioned, it is a very challenging task. Secondly, it is important to remember that an earwitness's subjective confidence, before as well as after an identification, is not a good predictor of accuracy. Although the perpetrator was always present in the studies in the present thesis, many witnesses identified a foil or falsely rejected the lineup. This implies that a foil selection or rejection is not a strong indication that the suspect is in fact innocent. The finding that children aged 11–13 performed at the same level, or even better than the adults at voice identification has the implication that *if* we are prepared to accept adults as earwitnesses, then children in this age-group should also be accepted.

As for memory of content, adults were found to have a relatively good memory for the perpetrator's account when interviewed shortly after the event. However, this positive result might have been at the expense of the voice memory. That attention is primarily focused on what is being *said*, instead of the voice as such, is probably true for most real criminal cases. Therefore, it might be useful to ask earwitnesses what aspect they focused on during their observations. The older children (11–13-years) reported fewer details than adults, but they should not as a result be interpreted as being unreliable witnesses. That is, they reported less information, but they did not confabulate more often than adults. Memory for content was found to be greatly negatively affected by a two week delay. Hence, it is important that witnesses' statements are obtained as soon as possible after the event, and this seems to be even more important for child witnesses.

Although speaker identification is becoming somewhat more common in legal settings, it is still not used very often. Hence, there is a need for establishing best practice standards for conducting voice lineups (Hollien, 2012). Furthermore, it is important to identify factors that may distort a lineup (like articulation rate or the perception of specific voices) to ensure a fair composition of foils.

Although the outcomes for many of the examined variables are found to be mixed, I hesitate to recommend that future research should carry on examining the effect of different variables. As earwitnesses are found to perform so poorly, focus should instead be on developing interview techniques that could enhance earwitnesses' ability to make correct identifications and reduce the high number of false identifications. The Cognitive Interview, which stems from eyewitness research, is based on well-established knowledge of human memory and has shown impressive results for recall (e.g., Memon et al., 2010). The present thesis advances previous knowledge as the CI was found to enhance recall for a brief conversation. The next step should be to try to find an interview technique or procedure that is beneficial for *recognition*. Such an interview technique should be based on established principles of memory. One suggestion for the future is to focus on how we encode voices. Such knowledge could be useful for developing better lineup procedures and methods for conducting interviews with earwitnesses. The interview method currently used by the Swedish Security Service was not found to have very positive effects on either recognition or recall. Hence, there is a strong need for developing memory enhancing methods.

To conclude, in line with previous research (e.g., Read & Craik, 1995; Yarmey, 2007), the results of the present thesis indicate that earwitness evidence should be considered as quite weak evidence. As earwitnesses seem to perform more poorly than eyewitnesses in all three domains (descriptions, memory for content, and identification), the current findings highlight the importance of prioritizing other types of evidence when possible. Actors in the legal system need to treat voice identification evidence with great caution. Although research tends to show low accuracy scores, voice identification is possible. However, for earwitnesses to be really useful we must find ways of improving their performance.

References

Armstrong, H. A., & McKelvie, S. J. (1996). Effect of face context on recognition memory for voices. *The Journal of General Psychology, 123,* 259–270. doi:10.1080/00221309.1996.9921278

Baddeley, A. (1990). *Human memory: Theory and practice*. Hove, UK: Lawrence Erlbaum.

Baddeley, A. (2000). Short-term and working memory. In E. Tulving & F. I. M. Craik (Eds.), *Oxford handbook of memory* (pp. 77–92). Oxford, UK: Oxford University Press.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology, 63,* 1–29. doi:10.1146/annurev-psych-120710-100422

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158–173. doi:10.1037/0033-295X.105.1.158

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27,* 586–595. doi:10.1016/0749-596X(88)90028-9

Bahr, R., & Pass, K. (1996). The influence of style-shifting on voice identification. *Forensic Linguistics, 3,* 24–38.

Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychology and Aging, 12,* 12–21. doi:10.1037/0882-7974.12.1.12

Barnecutt, P., Pfeffer, K., & Creswell, L. (1999). ‘Earwitness’: A comparison of auditory, visual and audio-visual judgements of vehicle speed. *Psychology, Crime & Law, 5,* 319–329. doi:10.1080/10683169908401775

Bartholomeus, B. (1973). Voice identification by nursery school children. *Canadian Journal of Psychology, 27*, 464–472. doi:10.1037/h0082498

Bates, E., Masling, M., & Kintsch, W. (1978). Recognition memory for aspects of dialogue. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 187–197. doi:10.1037/0278-7393.4.3.187

Belin, P., & Zatorre, R. J. (2003). Adaption to speaker's voice in the right anterior temporal lobe. *NeuroReport, 14,* 2105–2109. doi:10.1097/00001756-200311140-00019

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403,* 309–312. doi:10.1038/35002078

Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, 47*, 296–312. doi:10.1111/j.1469-7610.2006.01611.x

Bower, G. H. (1967). A multicomponent view of a memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation, Vol 1* (pp. 230–325). New York: Academic Press.

Boydell, C. A., & Read, J. D. (2011). Accuracy of and confidence in mock jailhouse informants' recall of criminal accounts. *Applied Cognitive Psychology, 25*, 255–264. doi:10.1002/acp.1672

Brainerd, C. J., & Reyna, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review, 100,* 42–67. doi:10.1037/0033-295X.100.1.42

Brickman, A. M., & Stern, Y. (2009). Aging and memory in humans. In L. R. Squire (Ed.), *Encyclopedia of Neuroscience, Vol 1* (pp. 175–180). Oxford: Academic Press.

Broeders, A. P. A., & Rietveld, T. (1995). Speaker identification by earwitnesses. In A. Braun & J.-P. Köster (Eds.), *Studies in forensic phonetics* (pp. 1–11). Trier: Wissenschaftlicher Verlag.

Broeders, A. P. A., & van Amelsvoort, A. G. (1999). Lineup construction for forensic earwitness identification: A practical approach. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1373–1376). San Francisco.

Brümmer, N., & Strasheim, A. (2009). *AGNITIO's Speaker recognition system for EVALITA 2009*. Paper presented at The 11th Conference of the Italian Association for Artificial Intelligence.

Bull, R., & Clifford, B. R. (1984). Earwitness voice recognition accuracy. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness Testimony: Psychological Perspectives* (pp. 92–123). Cambridge: Cambridge University Press.

Byrne, C., & Foulkes, P. (2004). The 'mobile phone effect' on vowel formants. *The International Journal of Speech, Language and the Law*, *11*, 83–102.

Campos, L., & Alonso-Quecuty, M. L. (2006). Remembering a criminal conversation: Beyond eyewitness testimony. *Memory, 14*, 27–36. doi:10.1080/09658210444000476

Campos, L., & Alonso-Quecuty, M. L. (2008). Language crimes and the cognitive interview: Testing its efficacy in retrieving a conversational event. *Applied Cognitive Psychology, 22*, 1211–1227. doi:10.1002/acp.1430

Cerrato, L., Falcone, M., & Paoloni, A. (2000). Subjective age estimation of telephonic voices. *Speech Communication, 31,* 107–112. doi:10.1016/S0167-6393(99)00071-0

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior, 32,* 187–218. doi: 10.1007/s10979-006-9082-4

Clifford, B. R. (1980). Voice identification by human listeners: On earwitness reliability. *Law and Human Behavior, 4,* 373–394. doi:10.1007/BF01040628

Clifford, B. R., Rathborn, H., & Bull, R. (1981). The effects of delay on voice recognition accuracy. *Law and Human Behavior, 5,* 201–208. doi:10.1007/BF01044763

Clifford, B. R., & Toplis, R. (1996). A comparison of adults' and children's witnessing abilities. In N. K. Clark & G. M. Stephenson (Eds.), *Investigative and forensic decision-making: selected papers from the Division of Criminological and Legal Psychology Annual Conference 1995* (pp. 76–83). Leicester: Division of Criminological and Legal Psychology, British Psychological Society.

Cole, C. B., & Loftus, E. F. (1987). The memory of children. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds.), *Children's eyewitness memory* (pp. 178–208). New York: Springer-Verlag.

Compton, A. J. (1963). Effects of filtering and vocal duration upon the identification of speakers aurally. *Journal of the Acoustical Society of America, 35*, 1748–1752. doi:10.1121/1.1918810

Cook, S., & Wilding, J. (1997a). Earwitness testimony: Never mind the variety, hear the length. *Applied Cognitive Psychology, 11*, 95–111. doi:10.1002/(SICI)1099-0720(199704)11:2<95::AID-ACP429>3.0.CO;2-O

Cook, S., & Wilding, J. (1997b). Earwitness testimony 2: Voices, faces and context. *Applied Cognitive Psychology, 11*, 527–541. doi:10.1002/(SICI)1099-0720(199712)11:6<527::AID-ACP483>3.0.CO; 2-B

Cook, S., & Wilding, J. (2001). Earwitness testimony: Effects of exposure and attention on the Face Overshadowing Effect. *British Journal of Psychology, 92,* 617–629. doi:10.1348/000712601162374

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294. doi:10.1037/0096-3445.104.3.268

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decisionmaking in eyewitness identification cases. *Law and Human Behavior, 12*, 41–55. doi:10.1007/BF01064273

Davies, G. M. (1996). Children's identification evidence. In S. L. Sporer, R. S. Malpass & G. Köhnken (Eds.), *Psychological issues in eyewitness identification* (pp. 233–258). Mahwah, NJ: Erlbaum.

Davis, D., & Friedman, R. D. (2006). Memory for conversation: The orphan child of witness memory researchers. In M. P. Toglia, J. D. Read, D. R. Ross, & R. C. L. Lindsay (Eds.), *Handbook of Eyewitness Memory* (Vol. 1): *Memory for Events,* (pp. 3–52). Mahwah, NJ: Erlbaum.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voice. *Science, 208,* 1174–1176. doi:10.1126/science.7375928

Deffenbacher, K. A., Cross, J. F., Handkins, R. E., Chance, J. E., Goldstein, A. G., Hammersley, R., et al. (1989). Relevance of voice identification research criteria for evaluating reliability of an identification. *Journal of Psychology, 123*, 109–119. doi:10.1080/00223980.1989.10542967

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84,* 795–805. doi:10.1037/0021-9010.84.5.795

Elaad, E., Segev, S., & Tobin, Y. (1998). Long-term working memory in voice identification. *Psychology, Crime & Law, 4,* 73–88. doi:10.1080/10683169808401750

Eriksson, A. (2008). Rättsfonetik. In P. A. Granhag & S-Å. Christianson (Eds.), *Handbok i rättspsykologi* (pp. 325–339). Stockholm: Liber AB.

Fenn, K. M., Shintel, H., Atkins, A. S., Skipper, J. I., Bond, V. C., & Nusbaum, H. C. (2011). When less is heard than meets the ear: Change deafness in a telephone conversation. *The Quarterly Journal of Experimental Psychology, 64,* 1442–1456. doi:10.1080/17470218.2011.570353

Finger, K., & Pezdek, K. (1999). The effect of cognitive interview on face identification accuracy: Release from verbal overshadowing. *Journal of Applied Psychology, 84*, 340–348. doi:10.1037/0021-9010.84.3.340

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL, England: Charles C Thomas.

Fisher, R. P., Ross, S. J., & Cahill, B. S. (2010). Interviewing witnesses and victims. In P. A. Granhag (Ed.), *Forensic psychology in context: Nordic and international approaches* (pp. 56–74). Cullompton, Devon, UK: Willian Publishing.

Foulkes, P., & Barron, A. (2000). Telephone speaker recognition amongst members of a close social network. *Forensic Linguistics, 7,* 180–198. doi:10.1558/sll.2000.7.2.180

Fowler, A. E. (1991). How early phonological development might set the stage for phoneme awareness. *Haskins Laboratories Status Report on Speech Research,* SR-105/l06, 53–64.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology, 70*, 401–412. doi:10.1037/0021-9010.70.2.401

Gibbons, J., Anderson, D. R., Smith, R., Field, D. E., & Fischer, C. (1986). Young children's recall and reconstruction of audio and audiovisual narratives. *Child Development, 57,* 1014–1023. doi:10.2307/1130375

Gonzalez, J. (2003). Estimation of speakers' weight and height from speech: A re-analysis of data from multiple studies by Lass and colleagues. *Perceptual and Motor Skills, 96,* 297–304.
doi:10.2466/pms.2003.96.1.297

Gordon, B. N., Baker-Ward, L., & Ornstein, P. A. (2001). Children's testimony: A review of research on memory for past experiences. *Clinical Child and Family Psychology Review, 4*, 157–181.
doi:10.1023/A:1011333231621

Granhag, P. A., Strömwall, L. A., & Hartwig, M. (2005). Eyewitness testimony: Tracing the beliefs of Swedish legal professionals. *Behavioral Sciences and the Law, 23,* 709–727. doi:10.1002/bsl.670

Gwyer, P., & Clifford, B. R. (1997). The effects of the cognitive interview on recall, identification, confidence and the confidence/accuracy relationship. *Applied Cognitive Psychology, 11*, 121–145.
doi:10.1002/(SICI)1099-0720(199704)11:2<121::AID-ACP443>3.0.CO;2-L

Hammersley, R., & Read, J. D. (1985). The effect of participation in a conversation on recognition and identification of the speakers' voices. *Law and Human Behavior, 9,* 71–81. doi:10.1007/BF01044290

Hammersley, R., & Read, J. D. (1996). Voice identification by humans and computers. In S. L. Sporer, R. S. Malpass, & G. Köhnken (Eds.), *Psychological issues in eyewitness identification* (pp. 117–152). Mahwah, NJ: Erlbaum.

Hanley, J. R., Smith, S. T., & Hadfield, J. (1998). I recognize you but I can't place you: An investigation of familiar-only experiences during tests of voice and face recognition. *Quarterly Journal of Experimental Psychology, 51A,* 179–195. doi:10.1080/713755751

Hatano, H., Kitamur, T., Takemoto, H., Mokhtari, P., Honda, K., & Masaki, S. (2012). Correlation between vocal tract length, body height, formant frequencies, and pitch frequency for the five Japanese vowels uttered by fifteen male speakers. Interspeech 2012. Portland, Oregon.

Hollien, H. (1996). Consideration of guidelines for earwitness lineups. *Forensic Linguistics, 3*, 14–23.

Hollien, H. (2002). *Forensic voice identification*. San Diego, CA: Academic Press.

Hollien, H. (2012). On earwitness lineups. *Investigative Sciences Journal, 4*, 1–17.

Hollien, H., Bennett, G., & Gelfer, M. P. (1983). Criminal identification comparison: Aural versus visual identifications resulting from a simulated crime. *Journal of Forensic Sciences, 28,* 208–221.

Hollien, H., Huntley, R., Künzel, H., & Hollien, P. A. (1995). Criteria for earwitness lineups. *Forensic Linguistics, 2*, 143–153.

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.

Huss, M. T., & Weaver, K. A. (1996). Effect of modality in earwitness identification: Memory for verbal and nonverbal auditory stimuli presented in two contexts. *The Journal of General Psychology, 123,* 277–287. doi:10.1080/00221309.1996.9921280

Ibabe, I., & Sporer, S. L. (2004). How you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology, 18,* 711–726. doi:10.1002/acp.1025

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review, 4,* 577–581. doi:10.3758/BF03214352

Ito, T., Takeda, K., & Itakura, F. (2005). Analysis and recognition of whispered speech. *Speech Communication*, 45, 139–152. doi:10.1016/j.specom.2003.10.005

Kebbell, M. R., & Milne, R. (1998). Police officers' perceptions of eyewitness performance in forensic investigations. *Journal of Social Psychology, 138*, 323–330. doi:10.1080/00224549809600384

Kerstholt, J. H., Jansen, N. J. M., van Amelsvoort, A. G., & Broeders, A. P. A. (2004). Earwitnesses: Effects of speech duration, retention interval and acoustic environment. *Applied Cognitive Psychology, 18*, 327–336. doi:10.1002/acp.974

Kerstholt, J. H., Jansen, N. J. M., van Amelsvoort, A. G., & Broeders, A. P. A. (2006). Earwitnesses: Effects of accent, retention and telephone. *Applied Cognitive Psychology, 20*, 187–197. doi:10.1002/acp.1175

King, M. A., & Yuille, J. C. (1987). Suggestibility and the child witness. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds.), *Children's eyewitness memory* (pp. 24–35). New York: Springer-Verlag.

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology, 49*, 58–69. doi:10.1037/0022-3514.49.1.58

Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology, 38,* 618–625. doi:10.1016/S0022-1031(02)00510-3

Kreiman, J., & Papcun, G. (1991). Comparing discrimination and recognition of unfamiliar voices. *Speech Communication, 10,* 265–275. doi:10.1016/0167-6393(91)90016-M

Künzel, H. J. (2001). Beware of the 'telephone effect': The influence of telephone transmission on the measurement of formant frequencies. *Forensic Linguistic*s*, 8,* 80–99. doi:10.1558/sll.2001.8.1.80

Larsson, A. S., Granhag, P. A., & Spjut, E. (2003). Children's recall and the cognitive interview: Do the positive effects hold over time? *Applied Cognitive Psychology, 17*, 203–214. doi:10.1002/acp.863

Lass, N. J., Barry, P. J., Reed, R. A., Walsh, J. M., & Amuso, T. A. (1979). The effect of temporal speech alternations on speaker height and weight identifications. *Language and Speech, 22,* 163–171. doi:10.1177/002383097902200206

Lavner, Y., Rosenhouse, J., & Gath, I. (2001). The prototype model in speaker identification by human listeners. *International Journal of Speech Technology, 4,* 63–74. doi:10.1023/A:1009656816383

Leander, L., Granhag, P. A., & Christianson, S.-Å. (2005). Children exposed to obscene phone calls: What they remember and tell. *Child Abuse & Neglect, 29,* 871–888. doi:10.1016/j.chiabu.2004.12.012

Legge, G. E., Grosmann, C., & Pieper, C. M. (1984). Learning unfamiliar voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 298–303. doi:10.1037/0278-7393.10.2.298

Lindh, J., & Eriksson, A. (2007). Robustness of long time measures of fundamental frequency. In *Proceedings of Interspeech 2007* (pp. 2025 –2028). Antwerp, Belgium.

Lindsay, R. C. L., Pozzulo, J. D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. *Law and Human Behavior, 21*, 391–404. doi:10.1023/A:1024807202926

Ling, J., & Coombe, A. (2005). Age effects in earwitness recall of a novel conversation. *Perceptual and Motor Skills, 100,* 774–776. doi:10.2466/PMS.100.3.774-776

MacWhinney, B., Keenan, J. M., & Reinke, P. (1982). The role of arousal in memory for conversation. *Memory and Cognition, 10,* 308–317. doi:10.3758/BF03202422

Malpass, R. S., & Devine, P. G. (1981). Guided memory in eyewitness identification. *Journal of Applied Psychology, 66*, 343–350. doi:10.1037/0021-9010.66.3.343

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology, 27*, 153–165. doi:10.1016/0022-0965(79)90067-5

McAllister, H. A., Dale, R. H. I., Bregman, N. J., McCabe, A., & Cotton, C. R. (1993). When eyewitnesses are also earwitnesses: Effects on visual and voice identifications. *Basic and Applied Social Psychology, 14,* 161–170. doi:10.1207/s15324834basp1402_3

McCauley, M. R., & Fisher, R. P. (1995). Facilitating children’s eyewitness recall with the revised cognitive interview. *Journal of Applied Psychology, 8*, 510–516. doi:10.1037/0021-9010.80.4.510

McClelland, E. (2008). Voice recognition within a closed set of family members. Paper presented at the IAFPA 2008 conference, Lausanne, Switzerland.

McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology, 133,* 249–271. doi:10.1080/00221309.1937.9917999

McGehee, F. (1944). An experimental study of voice recognition. *Journal of General Psychology, 31*, 53–65. doi:10.1080/00221309.1944.10545219

McGivern, R. F., Andersen, J., Byrd, D., Mutter, K. L., & Reilly, J. (2002). Cognitive efficiency on a match to sample task decreases at the onset of puberty in children. *Brain and Cognition, 50*, 73–89. doi:10.1016/S0278-2626(02)00012-X

Mehler, J., Bertoncini, J., Barriere, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother’s voice. *Perception, 7,* 491–497. doi:10.1068/p070491

Meissner, C. A., Sporer, S. L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read &

M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol II: Memory for people* (pp. 3–34). Mahwah, NJ: Erlbaum.

Melara, R. D., DeWitt-Rickards, T. S., & O'Brien, T. P. (1989). Enhancing lineup identification accuracy: Two codes are better than one. *Journal of Applied Psychology, 74,* 706–713. doi:10.1037/0021-9010.74.5.706

Memon, A., Meissner, C. A., & Fraser, J. (2010). The cognitive interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law, 16*, 340–372. doi:10.1037/a0020518

Memon, A., & Yarmey, A. D. (1999). Earwitness recall and identification: Comparison of the cognitive interview and the structured interview. *Perceptual and Motor Skills, 88*, 797–807. doi:10.2466/pms.1999.88.3.797

Miller, J. B., deWinstanley, P. A., & Carey, P. (1996). Memory for conversation. *Memory, 4*, 615–631. doi:10.1080/741940999

Mullennix, J. W., Ross, A., Smith, C., Kuykendall, K., Conard, J., & Barb, S. (2011). Typicality effects on memory for voice: Implications for earwitness testimony. *Journal of Applied Cognitive Psychology, 25,* 29–34. doi:10.1002/acp.1635

Mullennix, J. W., Stern, S. E., Grounds, B., Kalas, R., Flaherty, M., Kowalok, S., et al. (2010). Earwitness memory: Distortions for voice pitch and speaking rate. *Journal of Applied Cognitive Psychology, 24,* 513–526. doi:10.1002/acp.1566

Mulligan, N. W., & Brown, A. S. (2003). Attention and implicit memory. In L. Jimenez (Ed.), *Attention and implicit learning* (pp. 297–333). Philadelphia, PA, USA: John Benjamins Publishing Company.

Nass, C., & Gong, L. (2000). Speech interfaces from an evolutionary perspective. *Communications of the ACM, 43,* 36–43. doi:10.1145/348941.348976

Neisser, U. (1981). John Dean's memory: A case study. *Cognition, 9,* 1–22. doi:10.1016/0010-0277(81)90011-1

Nelson, D. L., Reed, V. S., & McEvoy, C. L. (1977). Learning to order pictures and words: A model of sensory and semantic encoding. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 485–497. doi:10.1037/0278-7393.3.5.485

Nevid, J. S. (2003). *Psychology: Concepts and Applications*. Boston: Houghton Mifflin Company.

Nittrouer, S. (2006). Children hear the forest. *Journal of the Acoustical Society of America*, *120*, 1799–1802. doi:10.1121/1.2335273

Nittrouer, S., Crowther, C. S., & Miller, M. E. (1998). The relative weighting of acoustic properties in the perception of [s]+stop clusters by children and adults. *Perception & Psychophysics, 60*, 51–64. doi:10.3758/BF03211917

Nittrouer, S., Manning, C., & Meyer, G. (1993). The perceptual weighting of acoustic cues changes with linguistic experience. *Journal of the Acoustical Society of America, 94*, 1865 (1 page). doi:10.1121/1.407649

Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America, 101,* 2253–2266. doi:10.1121/1.418207

Nolan, F., & Grabe, E. (1996). Preparing a voice lineup. *Forensic Linguistics, 3,* 79–94.

Nolan, F., McDougall, K., & Hudson, T. (2009). Voice similarity and the effect of the telephone: a study of the implications for earwitness evidence: Full Research Report, ESRC End of Award report, RES-000-22-2582. Swindon: ESRC.

Olsson, N., Juslin, P., & Winman, A. (1998). Realism of confidence in earwitness versus eyewitness identification. *Journal of Experimental Psychology: Applied, 4*, 101–118. doi:10.1037/1076-898X.4.2.101

Orchard, T. L., & Yarmey, A. D. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology, 9*, 249–260. doi:10.1002/acp.2350090306

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.

Papcun, G., Kreiman, J., & Davis, A. (1989). Long-term memory for unfamiliar voices. *Journal of the Acoustical Society of America, 85,* 913–925. doi:10.1121/1.397564

Parker, J. F., & Carranza, L. E. (1989). Eyewitness testimony of children in target-present and target-absent lineups. *Law and Human Behavior, 13*, 133–149. doi:10.1007/BF01055920

Perfect, T. J., Hunt, L. J., & Harris, C. M. (2002). Verbal overshadowing in voice recognition. *Applied Cognitive Psychology, 16,* 973–980. doi:10.1002/acp.920

Peters, D. P. (1987). The impact of naturally occurring stress on children's memory. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds.), *Children's eyewitness memory* (pp. 122–141). New York: Springer-Verlag.

Petrini, K., & Tagliapietra, S. (2008). Cognitive maturation and the use of pitch and rate information in making similarity judgments of a single talker. *Journal of Speech, Language, and Hearing Research, 51*, 485–501. doi:10.1044/1092-4388(2008/035)

Petrushin, V. A., Tsirulnik, L. I., & Makarova, V. (2010). Whispered speech prosody modeling for TTS synthesis. *Proc. Speech Prosody 2010*, 1151–1154.

Pezdek, K., & Prull, M. (1993). Fallacies in memory for conversation: Reflections in Clarence Thomas, Anita Hill, and the like. *Applied Cognitive Psychology, 7,* 299–310. doi:10.1002/acp.2350070404

Pfeffer, K., & Barnecutt, P. (1996). Children's auditory perception of movement of traffic sounds. *Child: Care, Health and Development, 22,* 129–137. doi:10.1111/j.1365-2214.1996.tb00780.x

Philippon, A. C., Cherryman, J., Bull, R., & Vrij, A. (2007). Earwitness identification performance: the effect of language, target, deliberate strategies and indirect measures. *Applied Cognitive Psychology, 21,* 539–550. doi:10.1002/acp.1296

Pickel, K. L., French, T. A., & Betts, J. M. (2003). A cross-modal weapon focus effect: The influence of a weapon's presence on memory for auditory information. *Memory, 11,* 277–292.
doi:10.1080/09658210244000036

Pickel, K. L., & Staller, J. B. (Published online: 21 January 2011). A perpetrator's accent impairs witnesses' memory for physical appearance. *Law and Human Behavior.*
doi: 10.1007/s10979-011-9263-7

Pollack, J., Pickett, J. M., & Sumby, W. H. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America, 26,* 403–406. doi:10.1121/1.1907349

Pozzulo, J. D. (2007). Person description and identification by child witnesses. In R. C. L. Lindsay, D. F. Ross, J. D. Read & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol II: Memory for people* (pp. 283–307). Mahwah, NJ: Erlbaum.

Procter, E. E., & Yarmey, A. D. (2003). The effect of distributed learning on the identification of normal-tone and whispered voices. *The Korean Journal of Thinking & Problem Solving, 13,* 17–29.

Quinlivan, D. S., Neuschatz, J. S., Jimenez, A., Cling, A. D., Douglass, A. B., & Goodsell, C. A. (2009). Do prophylactics prevent inflation? Post-identification feedback and the effectiveness of procedures to protect against confidence-inflation in earwitnesses. *Law and Human Behavior, 33,* 111–121. doi:10.1007/s10979-008-9132-1

Rathborn, H. A., Bull, R. H., & Clifford, B. R. (1981). Voice recognition over the telephone. *Journal of Police Science and Administration, 9,* 280–284.

Read, D., & Craik, F. I. M. (1995). Earwitness identification: Some influences on voice recognition. *Journal of Experimental Psychology: Applied, 1*, 6–18. doi:10.1037/1076-898X.1.1.6

Reisberg, D. (2010). *Cognition: Exploring the science of the mind, 4th edition.* New York: Norton.

Relander, K., & Rämä, P. (2009). Separate neutral processes for retrieval of voice identity and word content in working memory. *Brain Research, 1252,* 143–151. doi:10.1016/j.brainres.2008.11.050

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1–75. doi:10.1016/1041-6080(95)90031-4

Ricci, C. M., & Beal, C. R. (1998). Child witnesses: Effect of event knowledge on memory and suggestibility. *Journal of Applied Developmental Psychology, 19,* 305–317. doi:10.1016/S0193-3973(99)80042-2

Ricci, C. M., & Beal, C. R. (2002). The effect of interactive media on children's story memory. *Journal of Educational Psychology, 94,* 138–144. doi:10.1037/0022-0663.94.1.138

Rikspolisstyrelsen. (2005). *Vittneskonfrontation*. RPS Rapport 2005:2

Roebuck, R., & Wilding, J. (1993). Effects of vowel variety and sample length on identification of a speaker in a lineup. *Applied Cognitive Psychology, 7,* 475–481. doi:10.1002/acp.2350070603

Rothman, H. B. (1977). A perceptual (aural) and spectrographic identification of talkers with similar sounding voices. In J. S. Jackson (Ed.), *International Conference on Crime Countermeasures – Science and Engineering* (pp. 37–42). Oxford: University of Oxford.

Saslove, H., & Yarmey, A. D. (1980). Long-term auditory memory: speaker identification. *Journal of Applied Psychology, 65*, 111–116. doi:10.1037/0021-9010.65.1.111

Sauerland, M., Sagana, A., & Otgaar, H. (Available online: 16 Apr 2012). Theoretical and legal issues related to choice blindness for voices. *Legal and Criminological Psychology.* doi:10.1111/j.2044-8333.2012.02049.x

Saywitz, K. J. (1987). Children's testimony: Age related patterns of memory errors. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds.), *Children's eyewitness memory* (pp. 36–52). New York: Springer-Verlag.

Schiller, N. O., & Köster, O. (1998). The ability of expert witnesses to identify voices: a comparison between trained and untrained listeners. *Forensic Linguistics, 5,* 1–9. doi:10.1558/sll.1998.5.1.1

Schmidt-Nielsen, A., & Crystal, T. H. (1998). Human vs. machine speaker identification with telephone speech. Paper presented at ICSLP, 1998.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36–71. doi:10.1016/0010-0285(90)90003-M

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100*, 139–156. doi:10.1037/0033-2909.100.2.139

Sidtis, D., & Kreiman, J. (2012). In the beginning was the familiar voice: personally familiar voices in the evolutionary and contemporary biology of communication. *Integrative Psychological & Behavioral Science, 46,* 146–159. doi:10.1007/s12124-011-9177-4

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review, 5,* 710–715. doi:10.3758/BF03208850

Solan, L. M., & Tiersma, P. M. (2003). "Falling on deaf ears: Scientists say that earwitnesses are unreliable. Why aren't courts listening?" *Legal Affairs, 71* (Nov./Dec. 2003). http://www.legalaffairs.org/issues/November-December-2003/story_solan_novdec03.msp

Spence, M. J., Rollins, P. R., & Jerger, S. (2002). Children's recognition of cartoon voices. *Journal of Speech, Language, and Hearing Research, 45*, 214–222. doi:10.1044/1092-4388(2002/016)

Sporer, S. L. (1996). Psychological aspects of person descriptions. In S. L. Sporer, R. S. Malpass, & G. Köhnken (Eds.), *Psychological issues in eyewitness identification* (pp. 53–86). Mahwah, NJ: Erlbaum.

Sroufe, L. A., Cooper, R. G., & DeHart, G. B. (1992). *Child Development: Its Nature and Course*. New York, NY: Mcgraw-Hill Book Company.

Stafford, L., Burggraf, C. S., & Sharkey, W. F. (1987). Conversational memory: The effects of time, recall mode, and memory expectancies on remembrances of natural conversations. *Human Communication Research, 14,* 203–229. doi:10.1111/j.1468-2958.1987.tb00127.x

Stafford, L., & Daly, J. A. (1984). Conversational memory: The effects of recall mode and memory expectancies on remembrances of natural conversations. *Human Communication Research, 10,* 379–402. doi:10.1111/j.1468-2958.1984.tb00024.x

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25,* 207–222. doi:10.1080/14640747308400340

Steblay, N. K. (2007). *2001+6: An updated meta-analysis of eyewitness lineup performance under sequential versus simultaneous formats.* Paper presented at the conference Off the witness stand: Using psychology in the practice of justice. New York, March 1–3, 2007.

Stern, S. E., Mullennix, J. W., Corneille, O., & Huart, J. (2007). Distortions in the memory of the pitch of speech. *Experimental Psychology, 54,* 148–160. doi:10.1027/1618-3169.54.2.148

Stevenage, V. S., Howland, A., & Tippelt, A. (2011). Interference in eyewitness and earwitness recognition. *Applied Cognitive Psychology, 25,* 112–118. doi: 10.1002/acp.1649

Stridbeck, U., & Granhag, P. A. (2010). Legal procedures in the Nordic countries and in the USA: a comparative overview. In P. A. Granhag (Ed.), *Forensic psychology in context: Nordic and international approaches* (pp. 14–32). Cullompton, Devon, UK: Willian Publishing.

Toglia, M. P., Shlechter, T. M., & Chevalier, D. S. (1992). Memory for directly and indirectly experienced events. *Applied Cognitive Psychology, 6,* 293–306. doi:10.1002/acp.2350060403

Traunmüller, H., & Eriksson, A. (1995). The perceptual evaluation of F0-excursions in speech as evidenced in liveliness estimations. *Journal of the Acoustical Society of America, 97,* 1905–1915. doi:10.1121/1.412942

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373. doi:10.1037/h0020071

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17,* 969–993. doi:10.1002/acp.939

Vanags, T., Carroll, M., & Perfect, T. J. (2005). Verbal overshadowing: a sound theory in voice recognition? *Applied Cognitive Psychology, 19*, 1127–1144. doi:10.1002/acp.1160

van Dommelen, W. A., & Moxness, B. H. (1995). Acoustic parameters in speaker height and weight identification: Sex-specific behaviour. *Language and Speech, 38,* 267–287. doi:10.1177/002383099503800304

van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior, 21,* 661–685. doi:10.1023/A:1024812831576

van Lancker, D., & Kreiman, J. (1985). Unfamiliar voice discrimination and familiar voice recognition are independent and unordered abilities. *UCLA Working Papers in Phonetics, 62*, 50–60.

van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologica, 25,* 829–834. doi:10.1016/0028-3932(87)90120-5 van Wallendael, L. R., Surace, A., Hall-Parsons, D., & Brown, M. (1994). "Earwitness" voice recognition: Factors affecting accuracy and impact on jurors. *Applied Cognitive Psychology, 8*, 661–677. doi:10.1002/acp.2350080705

Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 333–342. doi:10.1037/0096-1523.29.2.333

von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A.L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research, 17,* 48–55. doi:10.1016/S0926-6410(03)00079-X

von Kriegstein, K., & Giraud, A.L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage, 22,* 948–955. doi:10.1016/j.neuroimage.2004.02.020

Wells, G. L., Malpass, R. S., Lindsay, R. C. L., Fisher, R. P., Turtle, J. W., & Fulero, S. M. (2000). From the lab to the police station. A successful application of eyewitness research. *American Psychologist, 55,* 581–598. doi:10.1037/0003-066X.55.6.581

Wells, G. L, & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54,* 277–295. doi:10.1146/annurev.psych.54.101601.145028

Wilding, J., Cook, S., & Davis, J. (2000). Sound familiar? *The Psychologist, 13,* 558–562.

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real lineups. *Applied Cognitive Psychology, 10,* 75–84. doi:10.1002/(SICI)1099-0720(199602)10:1<75::AIDACP364>3.0. CO; 2-E

Yarmey, A. D. (1986). Verbal, visual and voice identification of a rape suspect under different levels of illumination. *Journal of Applied Psychology, 71,* 363–370. doi:10.1037/0021-9010.71.3.363

Yarmey, A. D. (1991a). Descriptions of distinctive and non-distinctive voices over time. *Journal of the Forensic Science Society, 31,* 421–428. doi:10.1016/S0015-7368(91)73183-6

Yarmey, A. D. (1991b). Voice identification over the telephone. *Journal of Applied Social Psychology, 21*, 1868–1876. doi:10.1111/j.1559-1816.1991.tb00510.x

Yarmey, A. D. (1992). The effects of dyadic discussion on earwitness recall. *Basic and Applied Social Psychology, 13,* 251–263. doi:10.1207/s15324834basp1302_8

Yarmey, A. D. (1993). Stereotypes and recognition memory for faces and voices of good guys and bad guys. *Applied Cognitive Psychology, 7,* 419–431. doi:10.1002/acp.2350070505

Yarmey, A. D. (1995). Earwitness speaker identification. *Psychology, Public Policy, and Law, 1*, 792–818. doi:10.1037/1076-8971.1.4.792

Yarmey, A. D. (2001). Earwitness descriptions and speaker identification. *Forensic Linguistics, 8*, 113–122. doi:10.1558/sll.2001.8.1.113

Yarmey, A. D. (2003). Earwitness identification over the telephone and in field settings. *Forensic Linguistics, 10,* 65–77. doi:10.1558/sll.2003.10.1.62

Yarmey, A. D. (2007). The psychology of speaker identification and earwitness memory. In R. C. L. Lindsay, D. F. Ross, J. D. Read & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol II: Memory for people* (pp. 101–136). Mahwah, NJ: Erlbaum.

Yarmey, A. D., & Matthys, E. (1992). Voice identification of an abductor. *Applied Cognitive Psychology, 6*, 367–377. doi:10.1002/acp.2350060502

Yarmey, A. D., Yarmey, A. L., & Yarmey, M. J. (1994). Face and voice identifications in showups and lineups. *Applied Cognitive Psychology, 8,* 453–464. doi:10.1002/acp.2350080504

Yarmey, A. D., Yarmey, A. L., Yarmey, M. J., & Parliament, L. (2001). Commonsense beliefs and the identification of familiar voices. *Applied Cognitive Psychology, 15*, 283–299. doi:10.1002/acp.702

Zetterholm, E., Sarwar, F., & Allwood, C. M. (2009). Earwitnesses: The effect of voice differences in identification accuracy and the realism in confidence judgments. In P. Branderud & H. Traunmüller (Eds.), *Proceedings of FONETIK 2009, TheXXIIth Swedish Phonetics Conference* (pp. 180–185). Stockholm: Stockholm University.

Öhman, L., Eriksson, A., & Granhag, P. A. (2012). What children and adults remember from a perpetrator's verbal account. Manuscript under preparation.

Öhman, L., Granhag, P. A., & Eriksson, A. (2012). Evaluating earwitness evidence: Does Swedish judges and lay-judges hold correct beliefs? Manuscript under preparation.

Appendix

I. Öhman, L., Eriksson, A., & Granhag, P.A. (2011). Overhearing the planning of a crime: Do adults outperform children as earwitnesses? *Journal of Police and Criminal Psychology, 26,* 118–127. doi: 10.1007/s11896-010-9076-5

II. Öhman, L., Eriksson, A., & Granhag, P.A. (2010). Mobile phone quality vs. direct quality: How the presentation format affects earwitness identification accuracy. *The European Journal of Psychology Applied to Legal Context, 2,* 161–182.

III. Öhman, L., Eriksson, A., & Granhag, P.A. (Available online: 27 Feb 2012). Enhancing adults' and children's earwitness memory: Examining three types of interviews. *Psychiatry, Psychology and Law*. doi: 10.1080/13218719.2012.658205

IV. Öhman, L., Eriksson, A., & Granhag, P.A. (2013). Angry voices from the past and present: Effects on adults' and children's earwitness memory. *Journal of Investigative Psychology and Offender Profiling, 10,* 57–70. doi:10.1002/jip.1381